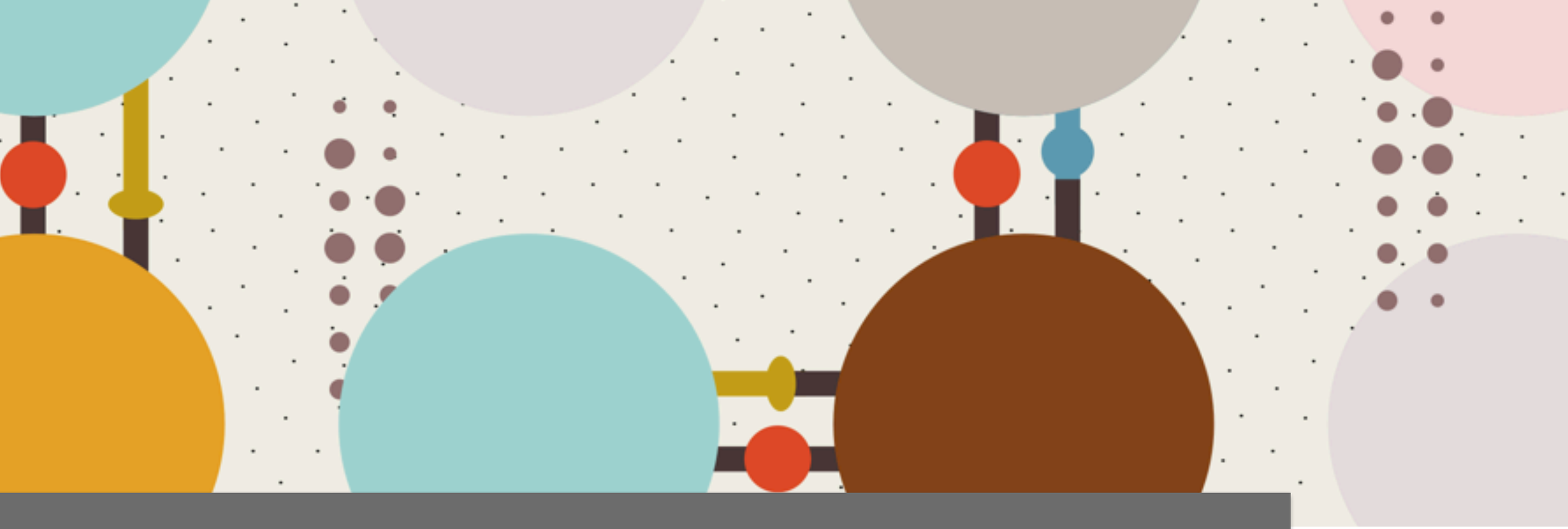


IO17 | Large Scale Bioinformatics for Immuno-Oncology

Modeling framework and Boolean logic models

Francesca Finotello, Federica Eduati, and Pedro L. Fernandes
Part of the slides are curtesy of Julio Saez-Rodriguez

GTPB | The Gulbenkian Training Programme in Bioinformatics
Instituto Gulbenkian de Ciência, Oeiras, Portugal | Sept 19th-22nd, 2017

INSTITUTO GULBENKIAN DE CIÊNCIA

Hallmarks of mathematical modeling

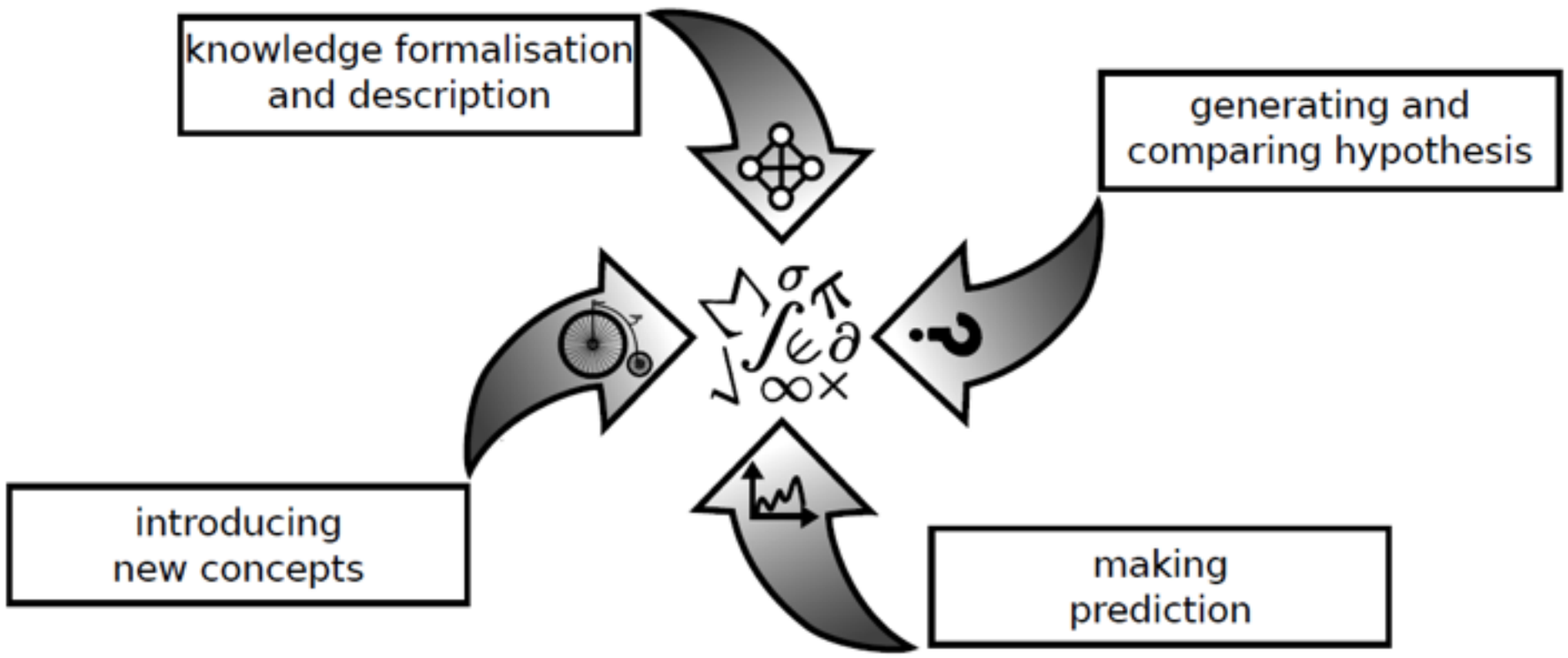
Mathematical models in **systems biology** are abstract representation of biological systems, aimed at at mimicking the reality with a certain degree of approximation.

"All models are wrong but some are useful"

George Box, 1976

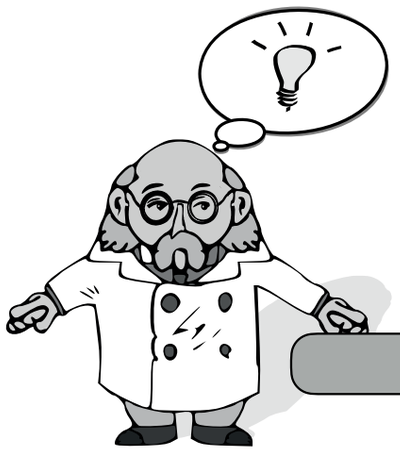
"truth … is much too complicated to allow anything but approximations"

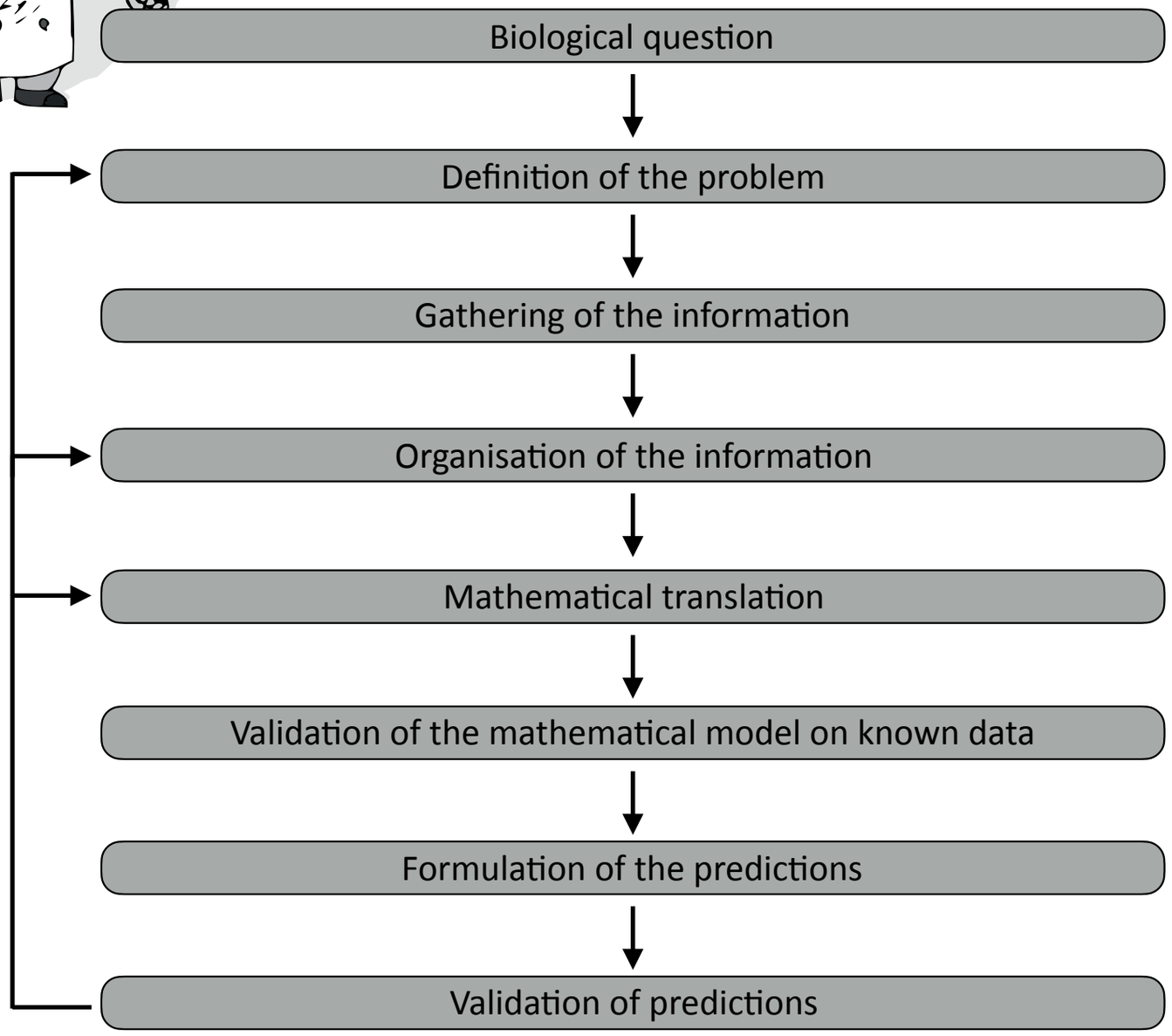
John von Neumann, 1947



from: E. Barillot, et al. Computational Systems Biology of Cancer, Chapman & Hall, 2012

Mathematical modeling flowchart

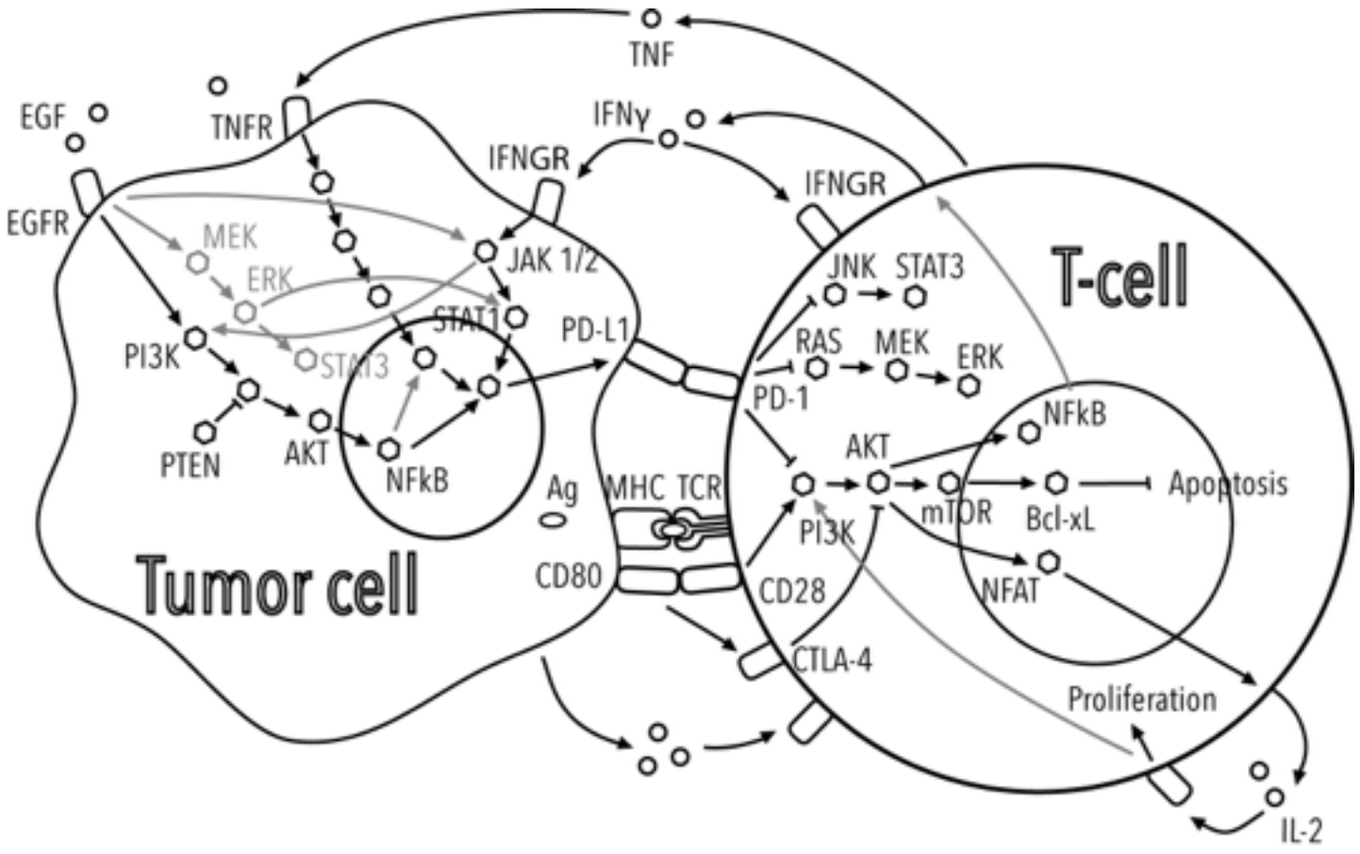




from: E. Barillot, et al. Computational Systems Biology of Cancer, Chapman & Hall, 2012

Why mathematical models in immnuno-oncology?

Interactions between tumor cells and T-cells is mediated by signaling pathways, which are complex networks that can be described by mathematical models.

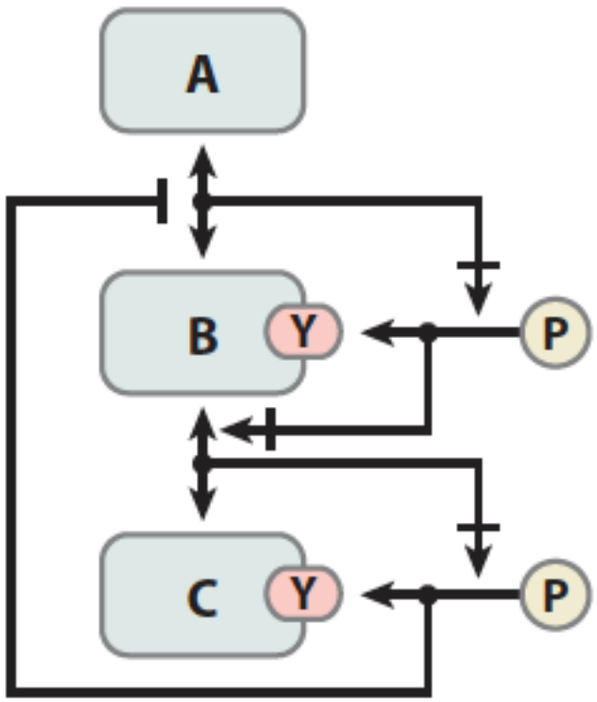


Immune-oncology questions that can be addressed with mathematical models are:

- How are signaling pathways deregulated in cancer?
- How can we target these pathways to restore normal behaviour?
- What is the effect of perturbations (targeted agents/checkpoint inhibitors) on the pathways?
- ….

Choice of modelling formalism

Physicochemical modeling



i $A + B \underset{k_{1r}}{\overset{k_{1f}}{\rightleftharpoons}} AB$

ii $AB \xrightarrow{k_2} A + B^*$

iii $B^* + C \underset{k_{3r}}{\overset{k_{3f}}{\rightleftharpoons}} B^*C$

iv $B^*C \xrightarrow{k_4} B^* + C^*$

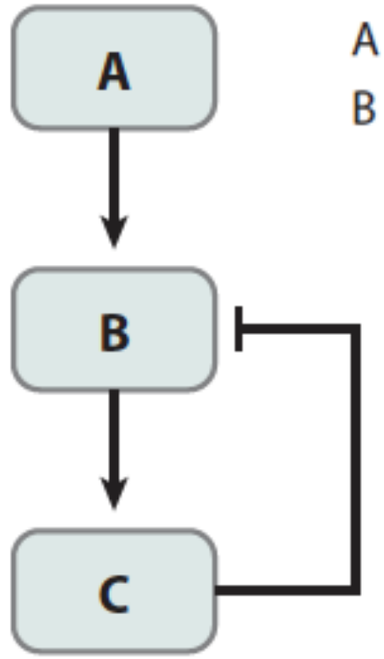
v $A + C^* \underset{k_{5r}}{\overset{k_{5f}}{\rightleftharpoons}} AC^*$

vi $\frac{dA}{dt} = k_2.[AB] - k_{1f}.[A].[B] + k_{1r}.[AB] - k_{5f}.[A].[C^*] + k_{5r}.[AC^*]$

vii $\frac{dB^*}{dt} = k_2.[AB] - k_{3f}.[B^*].[C] + k_{3r}.[B^*C] + k_4.[B^*C]$

viii $\frac{dC^*}{dt} = k_4.[B^*C] - k_{5f}.[A].[C^*] + k_{5r}.[AC^*]$

Causal (logic) modelling



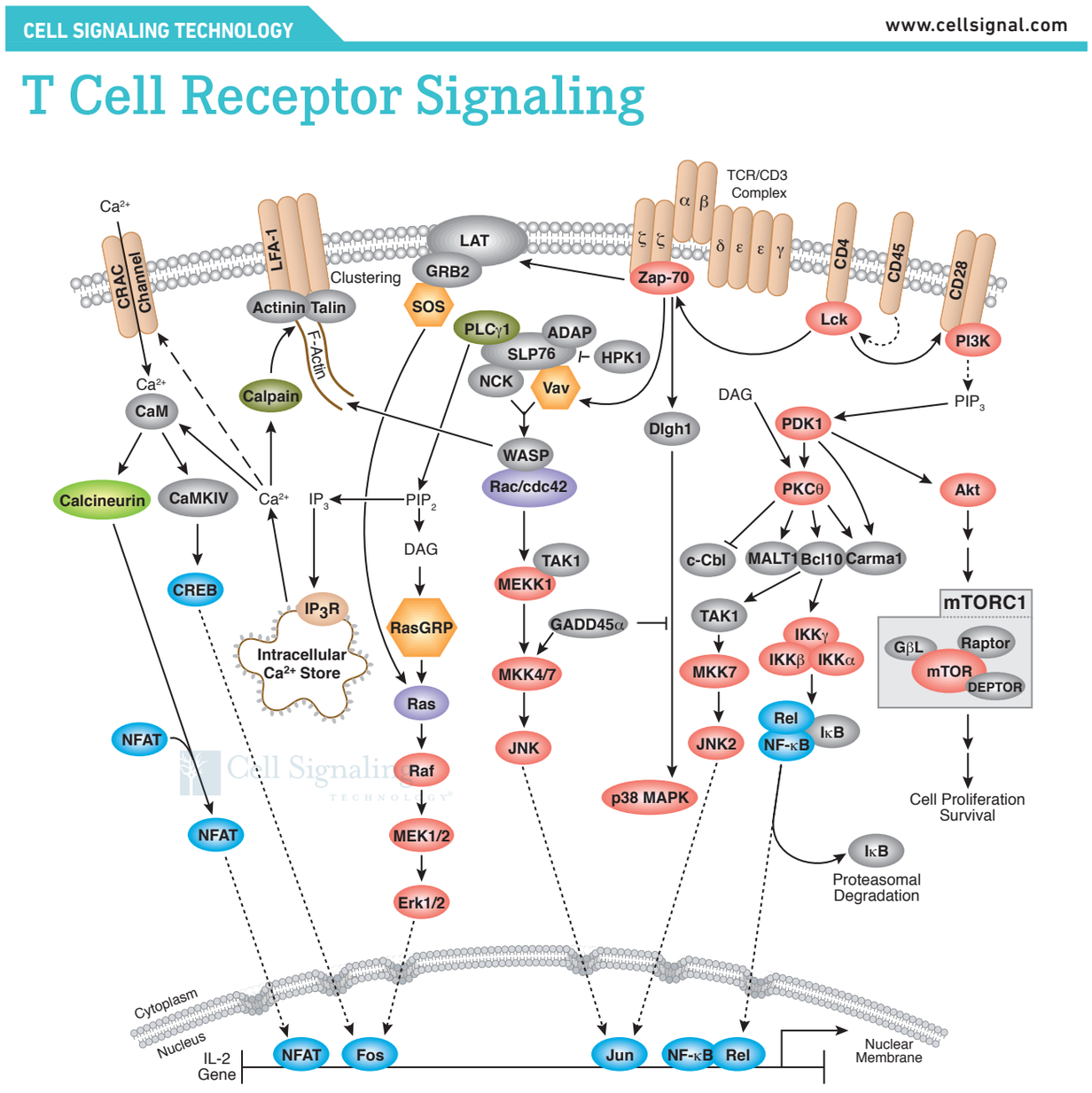
A and (not C) → B

B → C

The amount of details to include in the model and the **mathematical formalism** used to describe the process should be **lead by the biological question** (and by available data).

Figures from: Saez-Rodriguez J, et al. *Annual Rev Biomed Eng*, 2015

Modeling signaling pathways with logic models



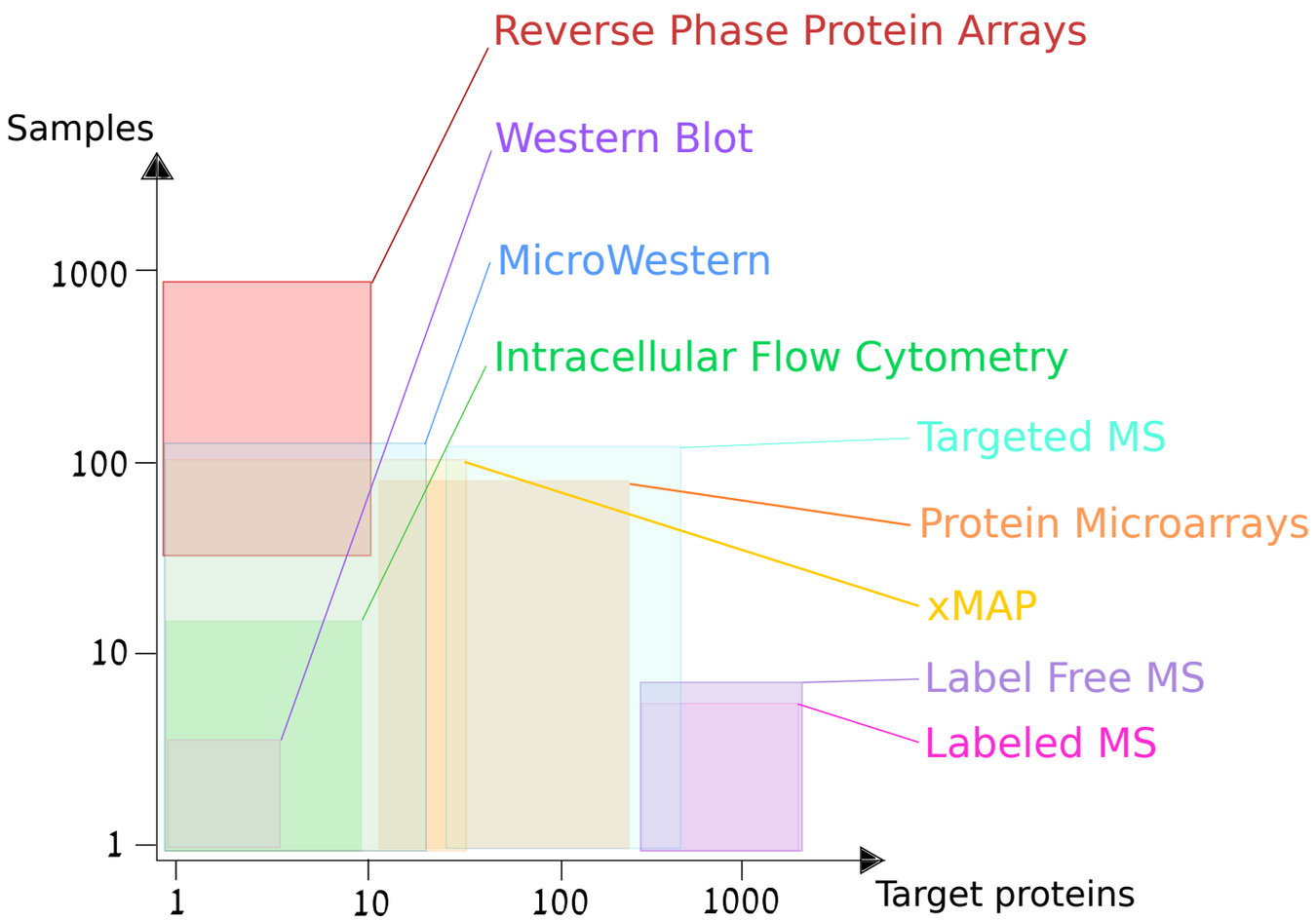
- complex network with many species and many interactions
- different post-translational modifications convey the signal through the network
- lack of molecular information available in pathway maps

can be modelled with **logic models** using **phosphorylation events** as markers of proteins activation and deactivation

Logic models have the advantages of being:

- straightforward
- robust
- compatible with quantitative data

Proteomics to look at signal transduction



Antibody-based methods:

low coverage

many conditions

Mass-spectrometry methods:

high coverage

few conditions

Terfve C, Saez-Rodriguez J, *Adv. Syst. Biol.*, 2012

Saez-Rodriguez J, et al. *Annual Rev Biomed Eng*, 2015

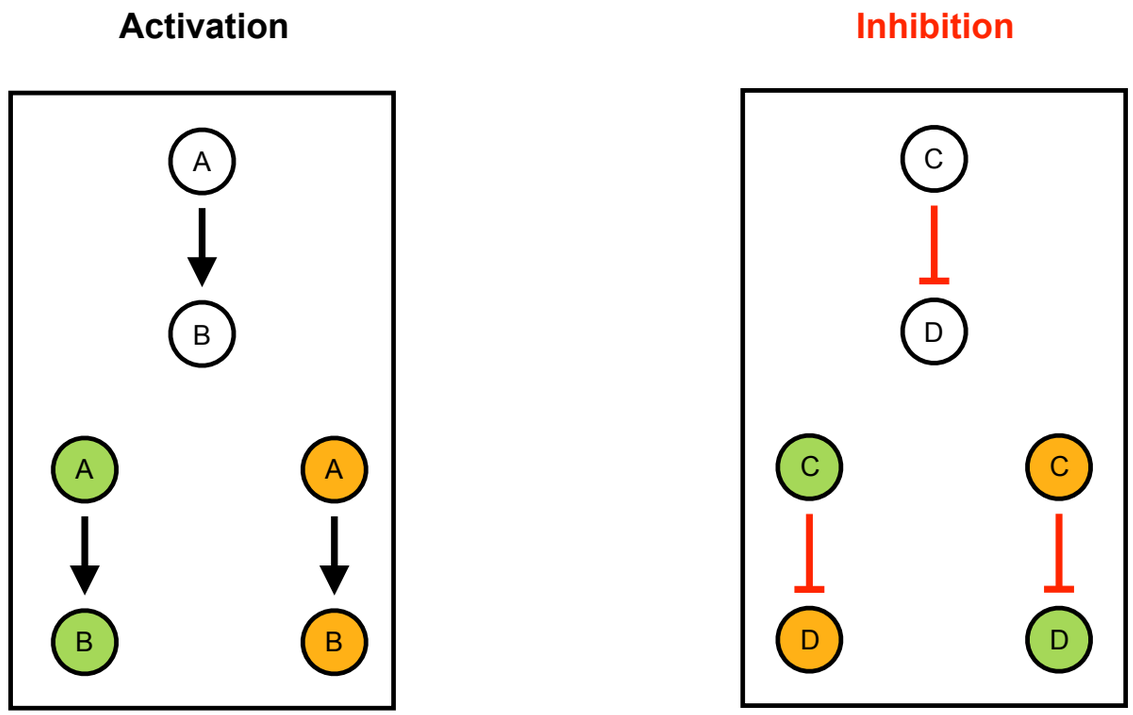
Main principles of Boolean logic models

Simplest type of logic models are Boolean models

- 2 signaling states:
  - ON (= 1)
  - OFF (= 0)

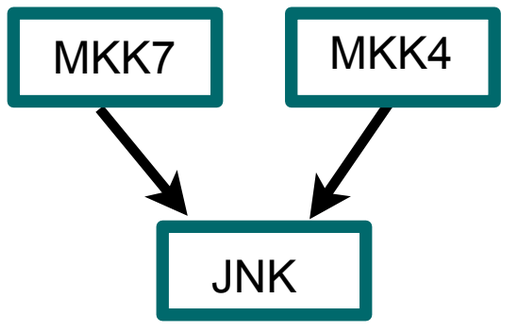
- 2 reaction types between
  - 1 regulator (e.g. kinase)

  and
  - 1 regulated node (e.g. substrate)

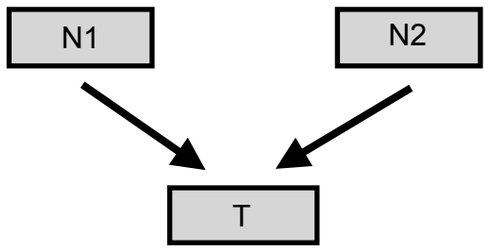


Main principles of Boolean logic models

OR gate

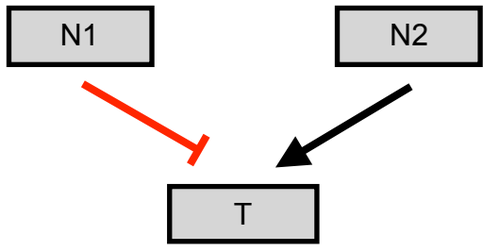


Two activating input nodes



N1	N2	N1 OR N2
0	0	0
0	1	1
1	0	1
1	1	1

One activating and one inhibitory input node

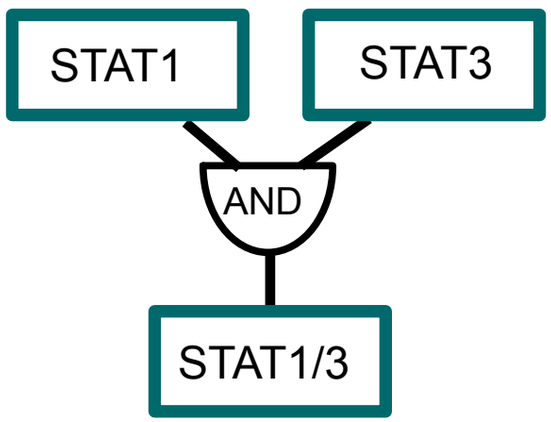


Activator dominant

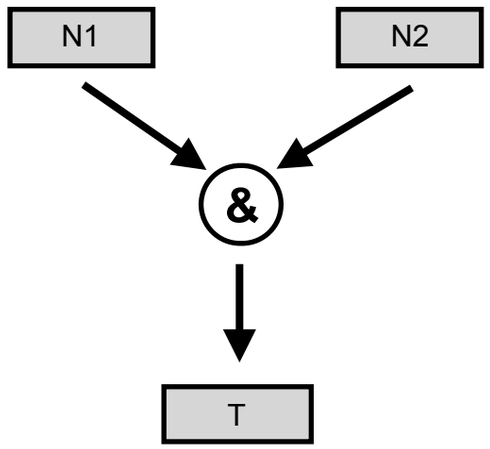
N1	N2	iN1 OR N2
0	0	1
0	1	1
1	0	0
1	1	1

iN1 = ¬ N1
1
1
0
0

AND gate

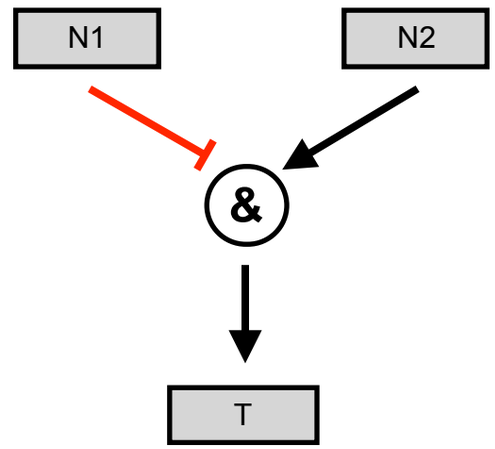


Two activating input nodes



N1	N2	N1 AND N2
0	0	0
0	1	0
1	0	0
1	1	1

One activating and one inhibitory input node

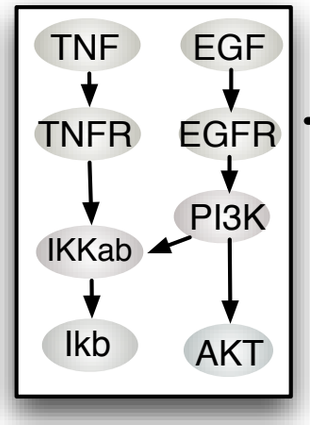


Inhibitor dominant

N1	N2	iN1 AND N2
0	0	0
0	1	1
1	0	0
1	1	0

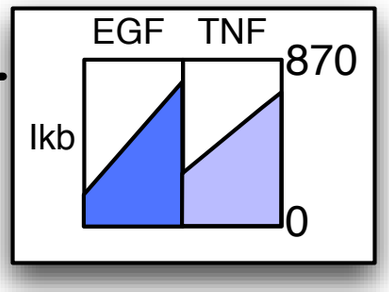
iN1 = ¬ N1
1
1
0
0

Linking pathway maps to data of signal transduction

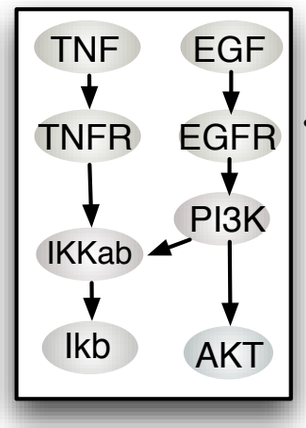


Define Pathway Map
from literature/Database

Perform Experiments



Linking pathway maps to data of signal transduction



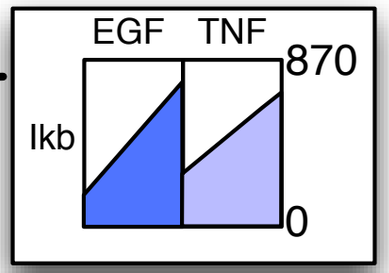
Define Pathway Map
from literature/Database

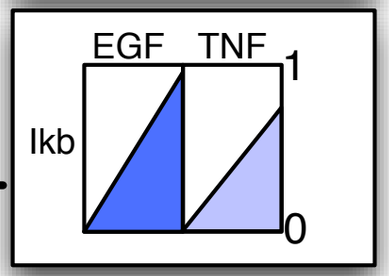
Perform Experiments

Import Data

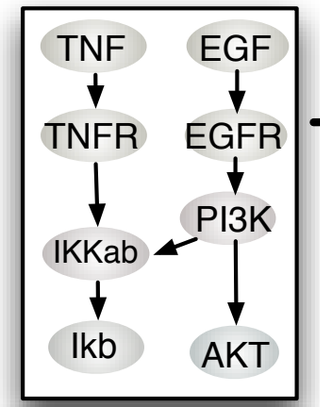
Process Data
Normalize between 0,1
Filter noise, saturation

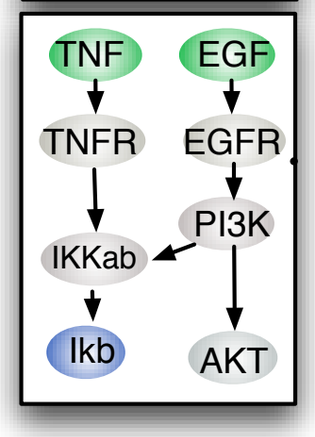
DataRail





Linking pathway maps to data of signal transduction





Define Pathway Map
from literature/Database

Import Map

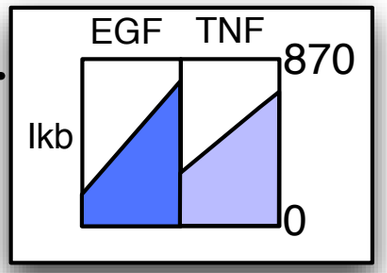
Process Map
Compress & Remove non-observables

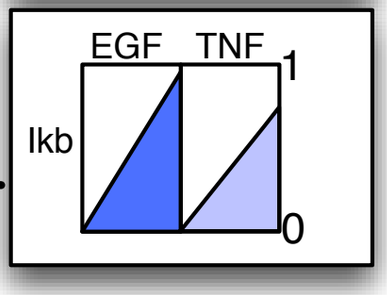
Perform Experiments

Import Data

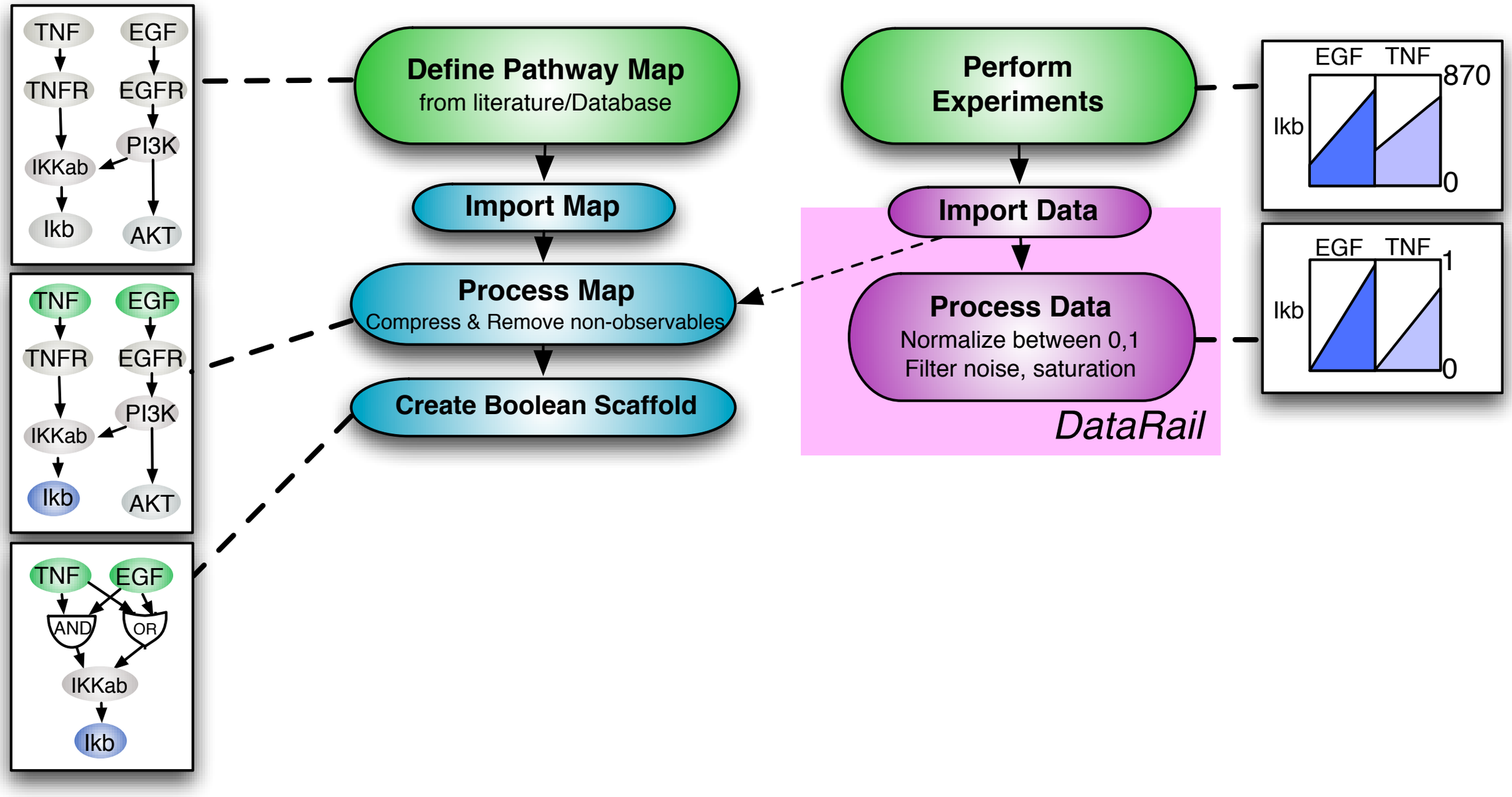
Process Data
Normalize between 0,1
Filter noise, saturation

DataRail

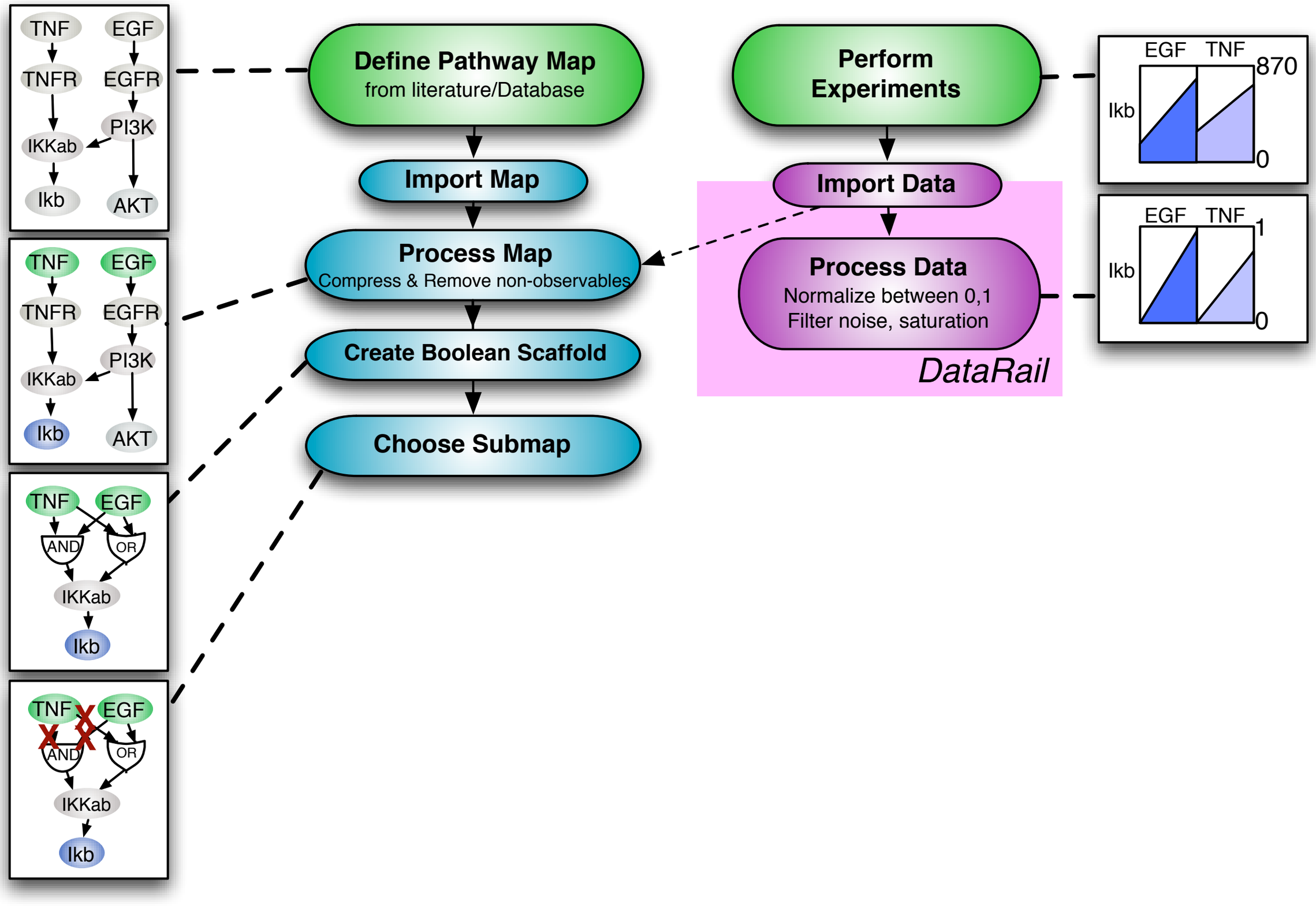




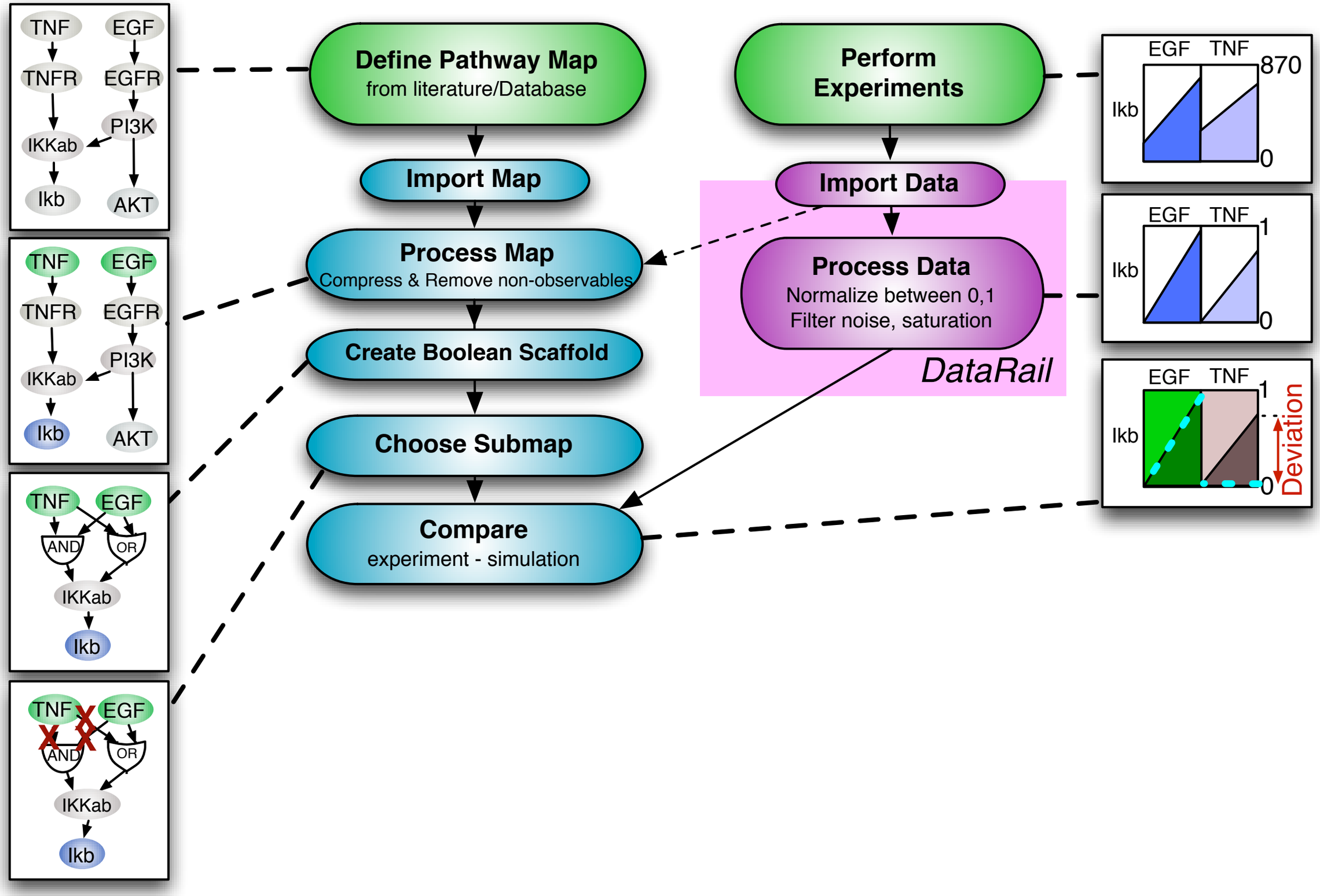
Linking pathway maps to data of signal transduction



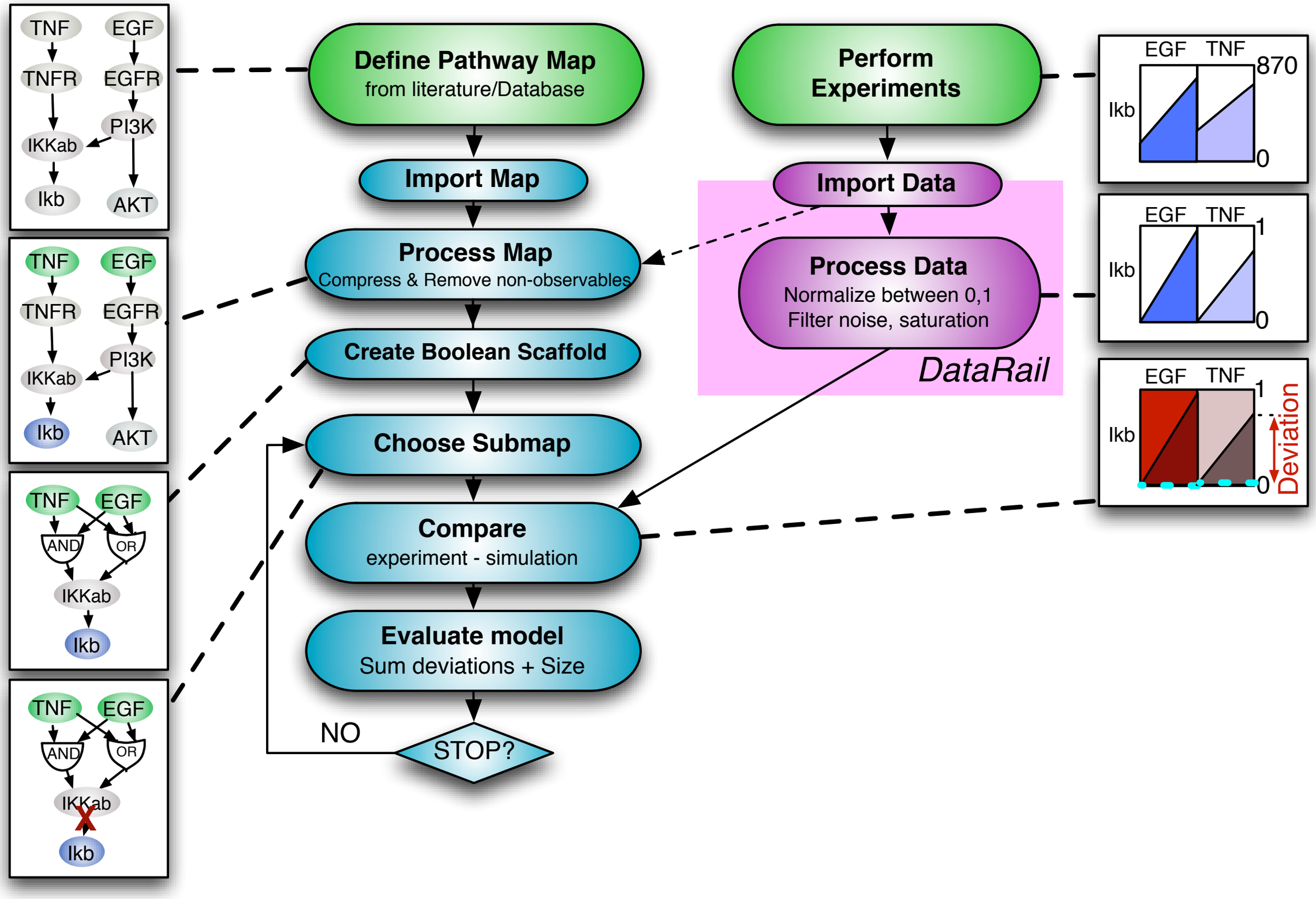
Linking pathway maps to data of signal transduction



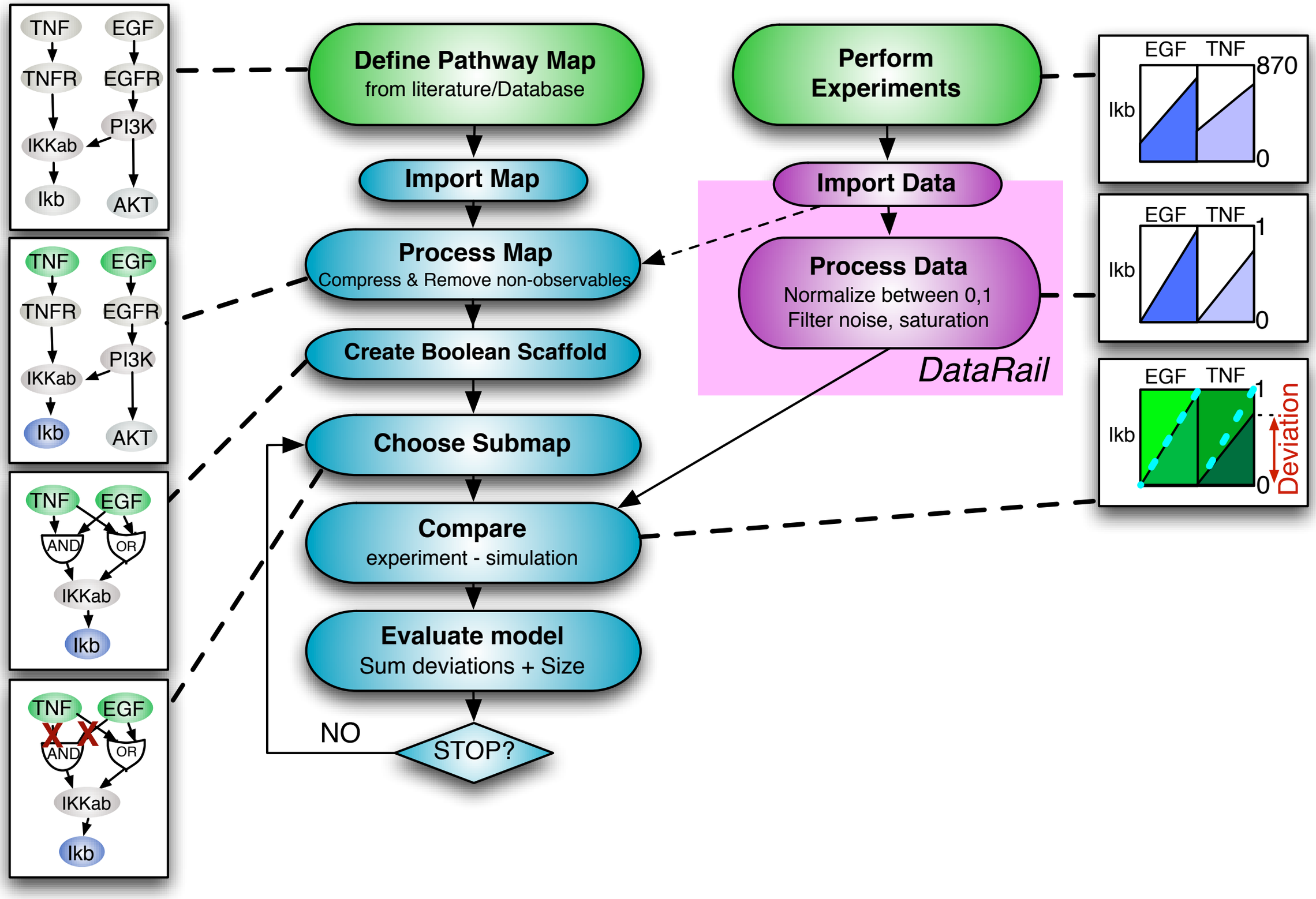
Linking pathway maps to data of signal transduction



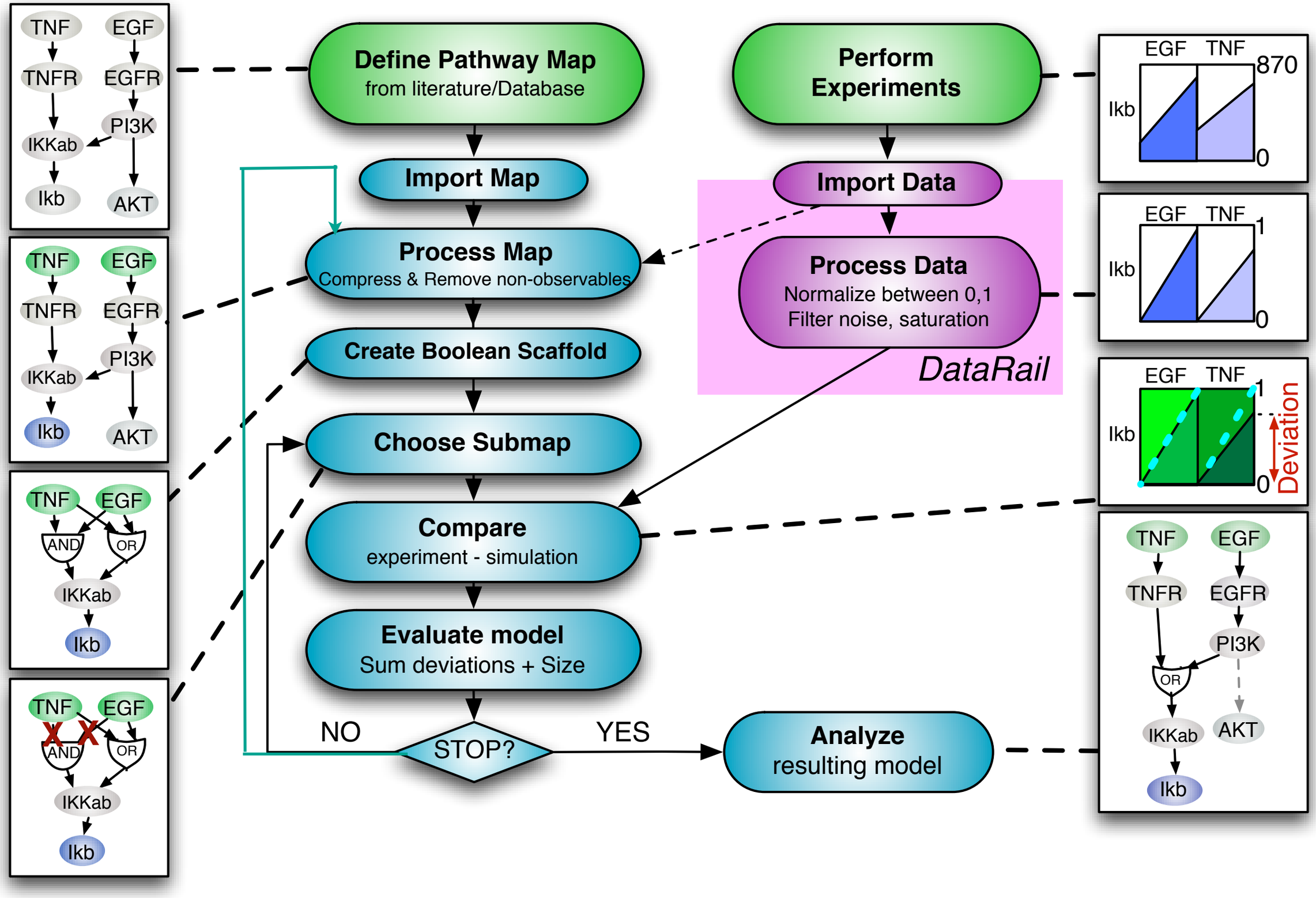
Linking pathway maps to data of signal transduction



Linking pathway maps to data of signal transduction



Linking pathway maps to data of signal transduction



Linking pathway maps to data of signal transduction

Pipeline implemented in

CellNOpt *(for CellNetOptimizer), aka CNO*

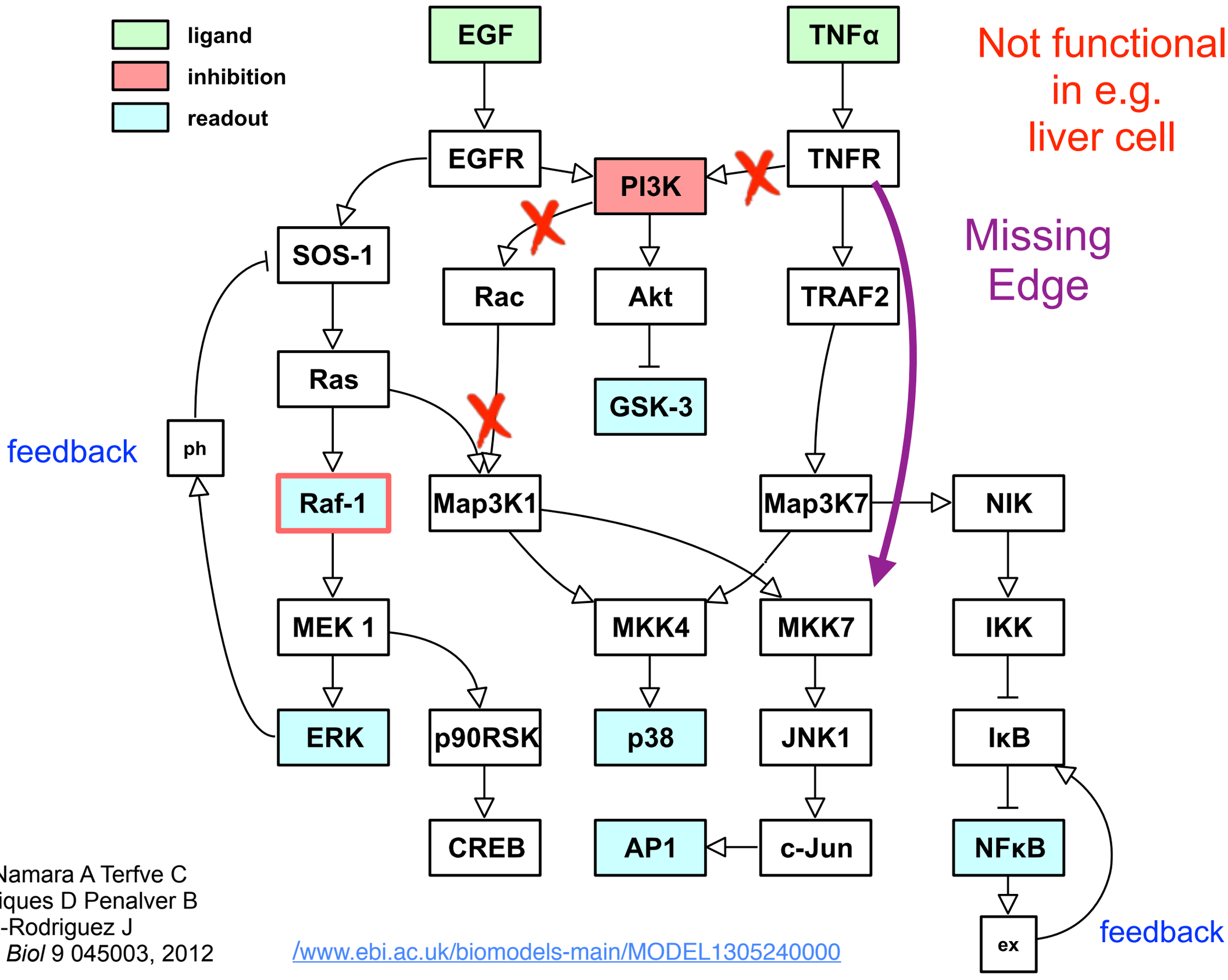
a Bioconductor, Python (and Matlab) toolbox

freely available at http:www.cellnopt.org

Bioconductor:
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

Matlab:
Morris MK, Melas I, Saez-Rodriguez J, *Methods Mol. Biol,* 930:179-214, 2013

A Toy model



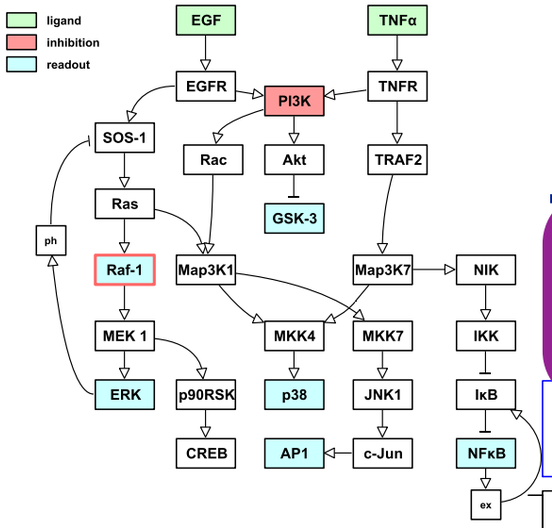
Not functional in e.g. liver cell

Missing Edge

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

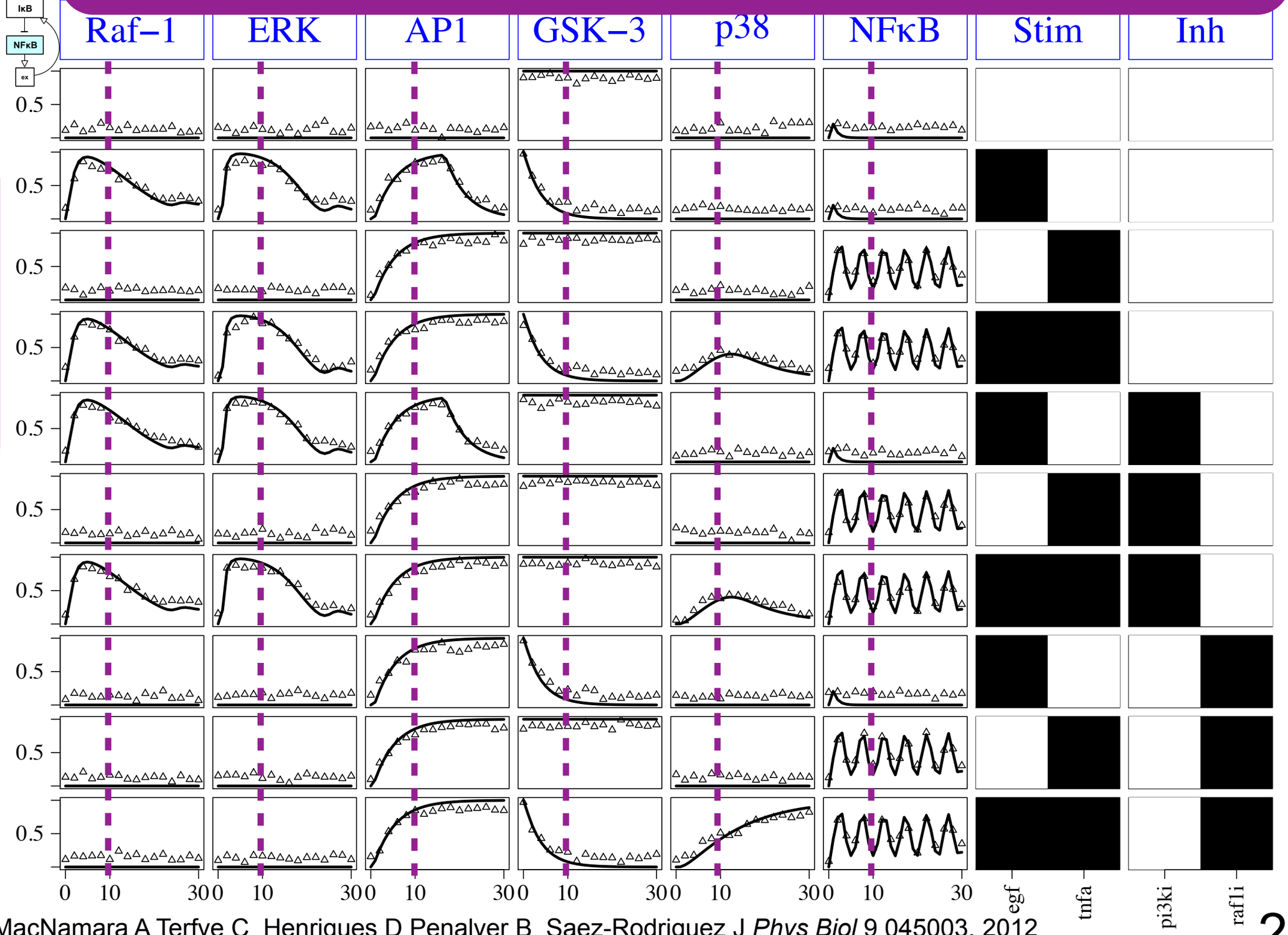
/www.ebi.ac.uk/biomodels-main/MODEL1305240000

The ‘real’ data



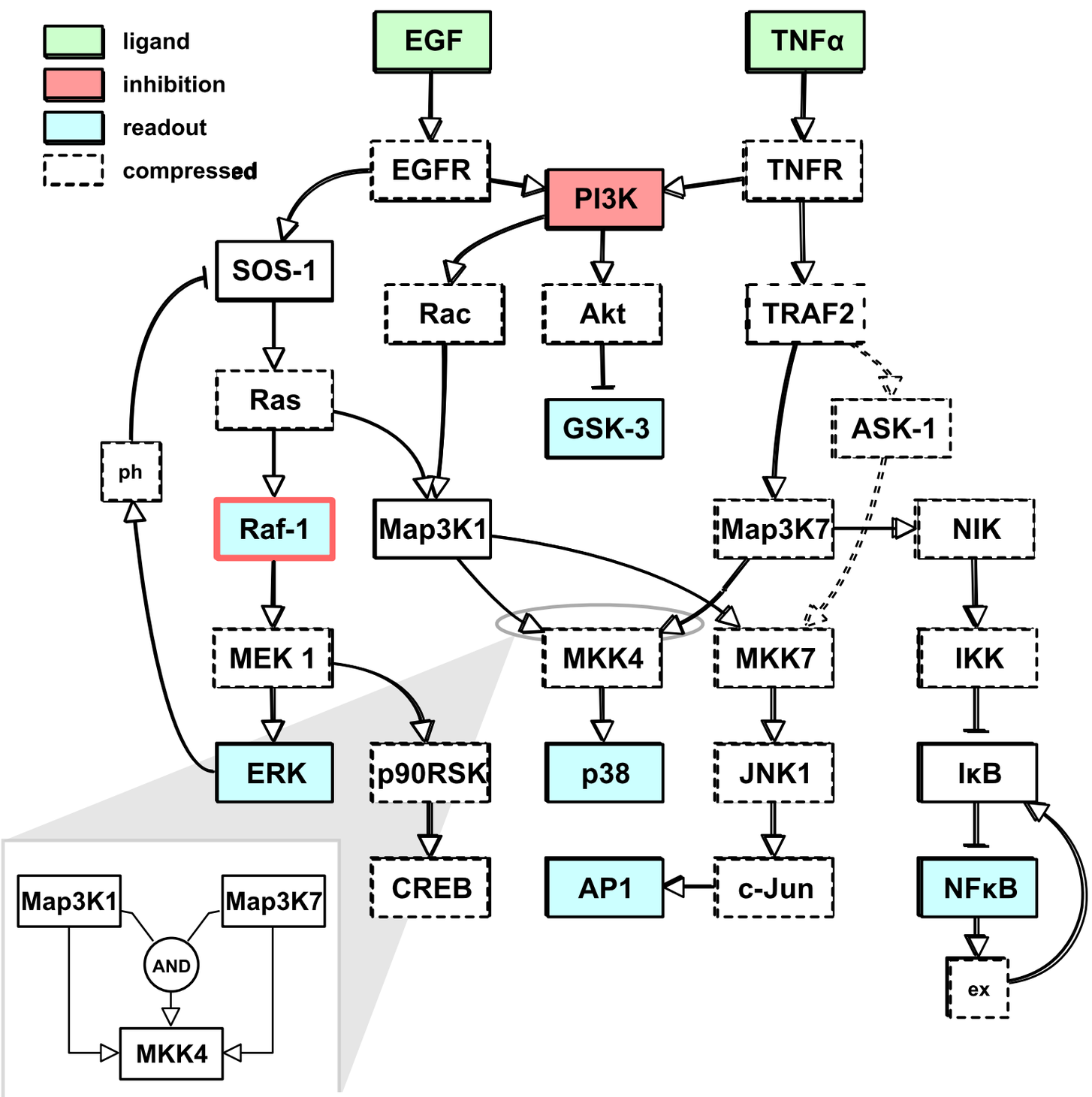
If you can only pick one ($$), choose one representatitive of a ‘time scale’

How to pick right time to measure?

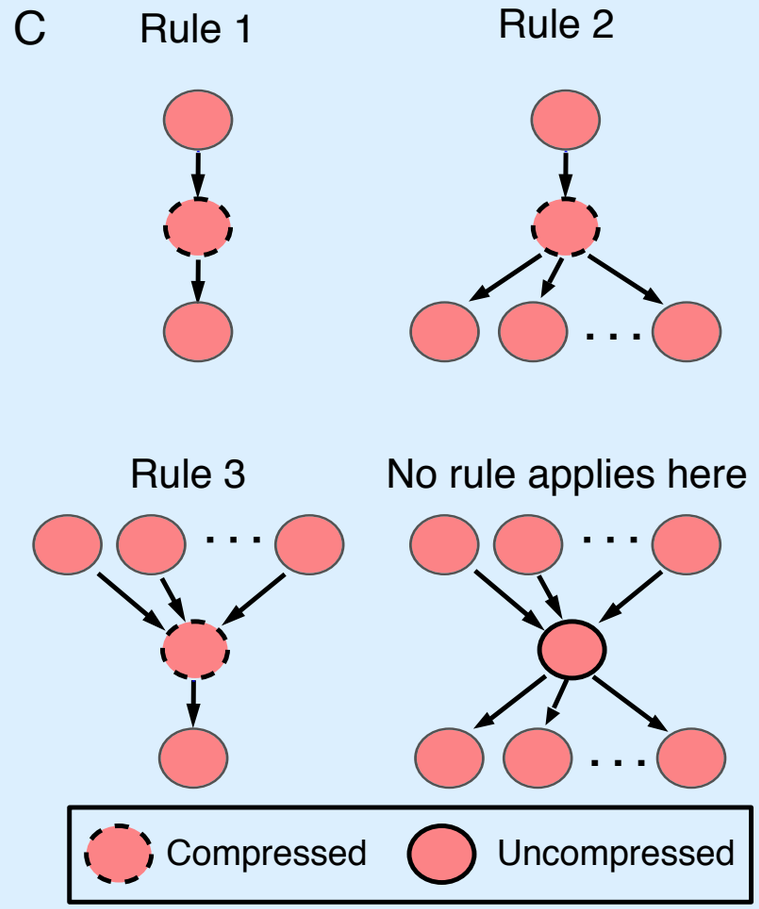


MacNamara A Terfve C Henriques D Penalver B Saez-Rodriguez J *Phys Biol* 9 045003, 2012

Model preprocessing: compression and expansion of gates



Model compression& removal of non-controlable & non-observable branches



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

How to choose model: balance of fit of data and size of model

A good model should describe (and predict) data well and be as simple as possible

Metric

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

Fit to data

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

$Bi_{kl}^{M} \in \{0,1\}$ $Bi_{kl}^{E} \in [0,1)$

Data is normalized between 0 and 1

Relative importance Fit vs. Size

Size of model

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

Best model ~ minimum metric
(optimization problem) - can be solved algorithmically