Efficient modelling of signalling networks using Integer Linear Programming

Enio Gjerga

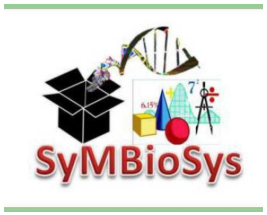
Paris, France - July 2017

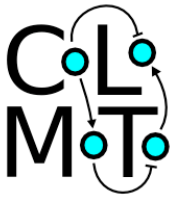
www.saezlab.org

@sysbiomed

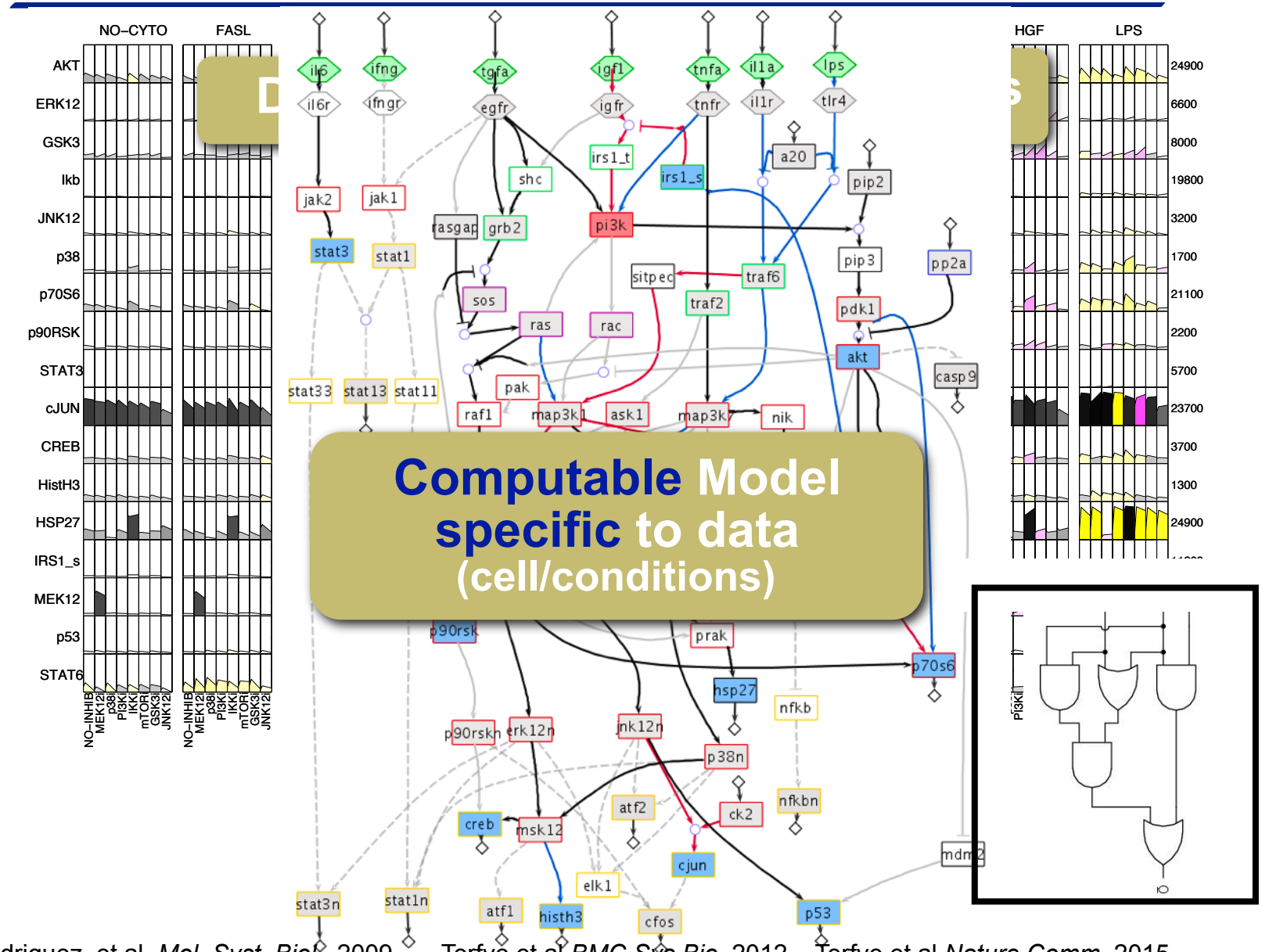
RWTH AACHEN UNIVERSITY

Joint Research Centre for Computational Biomedicine (JRC-COMBINE)



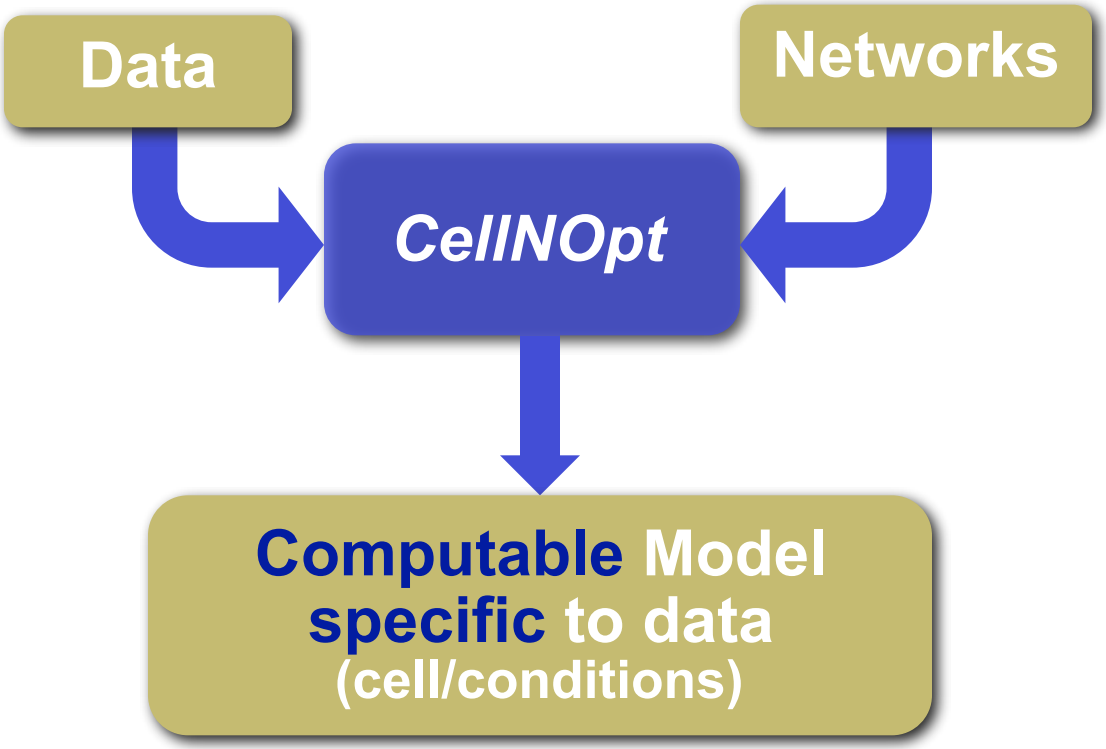


Logic modeling to link signaling networks with perturbation data



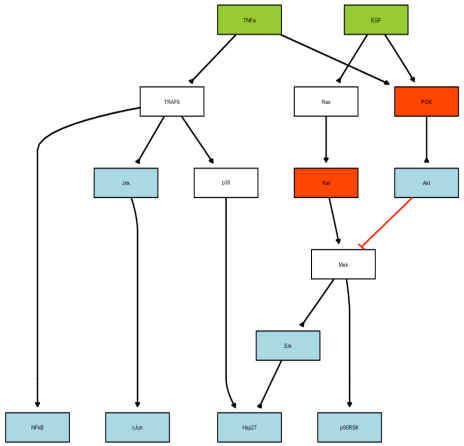
Saez-Rodriguez, et al. *Mol. Syst. Biol.*, 2009
Eduati et al. J *Bioinformatics*, 2012
Terfve et al *BMC Sys Bio*, 2012
MacNamara et al. *Phys Biol*, 2012
Terfve et al *Nature Comm*, 2015
Traynard et al., *CPT:PSP, in press*

Core task: Optimisation of model against data

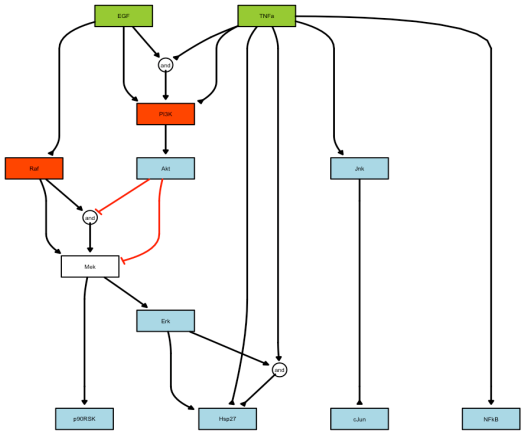


CellNOpt pipeline

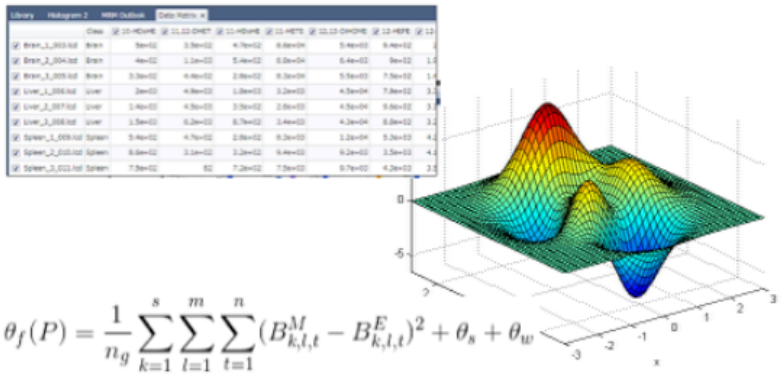
Prior Knowledge Network



Preprocessing

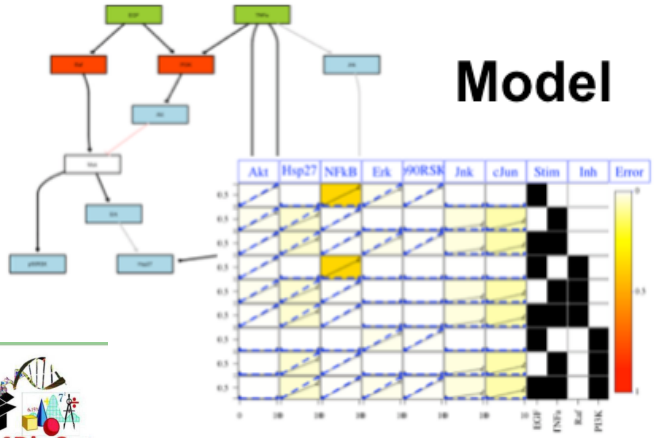


Training



$$\theta_f(P) = \frac{1}{n_g}\sum_{k=1}^{s}\sum_{l=1}^{m}\sum_{t=1}^{n}(B^M_{k,l,t} - B^E_{k,l,t})^2 + \theta_s + \theta_w$$

Model



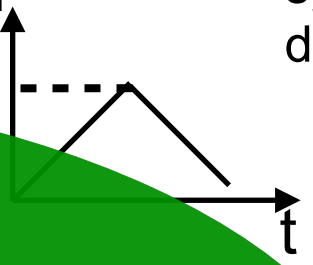
Saez-Rodriguez J, *Mol Sys Bio* 5:331,2009

From Boolean to continuous and dynamic models within CellNOpt

RWTH AACHEN

Boolean (bina
logic steady

Boolean multi
time-scale

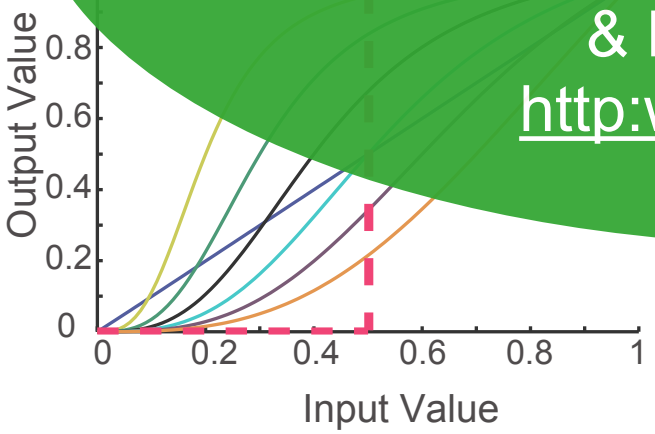


sync.
dynamics

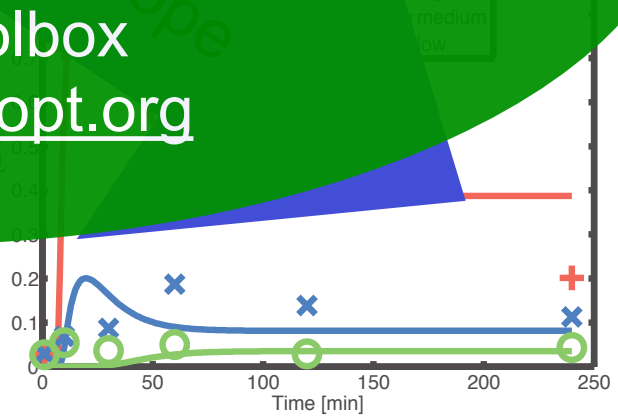
Aidan MacNamara

CellNOpt

R/Bioconductor packages (+Python-wrapper)
& Cytoscape plug-in (CytoCopter)
& Matlab toolbox
http://www.cellnopt.org



rris et al., PloS Comp Bio 2011



Wittmann et al. BMC Sys Bio 2009

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

David Henriques

CellNOpt pipeline - Inputs

Data

Code	Description	handled in CellNOptR
ID	identifiers	
TR	treatment	yes
DA	Data aquistion	yes
DV	Data value	yes

MIDAS
Minimum Information for Data Analysis in Systems Biology

TR:mock:CellLine	TR:EGF	TR:TNFa	TR:PI3Ki	DA:Akt	DA:Hsp27	DV:Akt	DV:Hsp27
1	1	0	0	0	0	0	0
1	0	1	0	0	0	0	0
1	1	0	0	10	10	1	0.2
1	0	1	0	10	10	1	0.5

CSV format file containing the cues, signals and readouts

Inhibitors are coded by adding the letter *i* after the name

Saez-Rodriguez et.al. *Bioinformatics* 2007

Prior Knowledge Network

```
nodeA relationship nodeB
nodeC relationship nodeA
nodeD relationship nodeE nodeF nodeB
```

SBMLqual data format can also be used to load a model

Chaouiya et.al., *BMC Sys Biol,* 2013

SIF - Simple Interaction Format

Cytoscape compatible

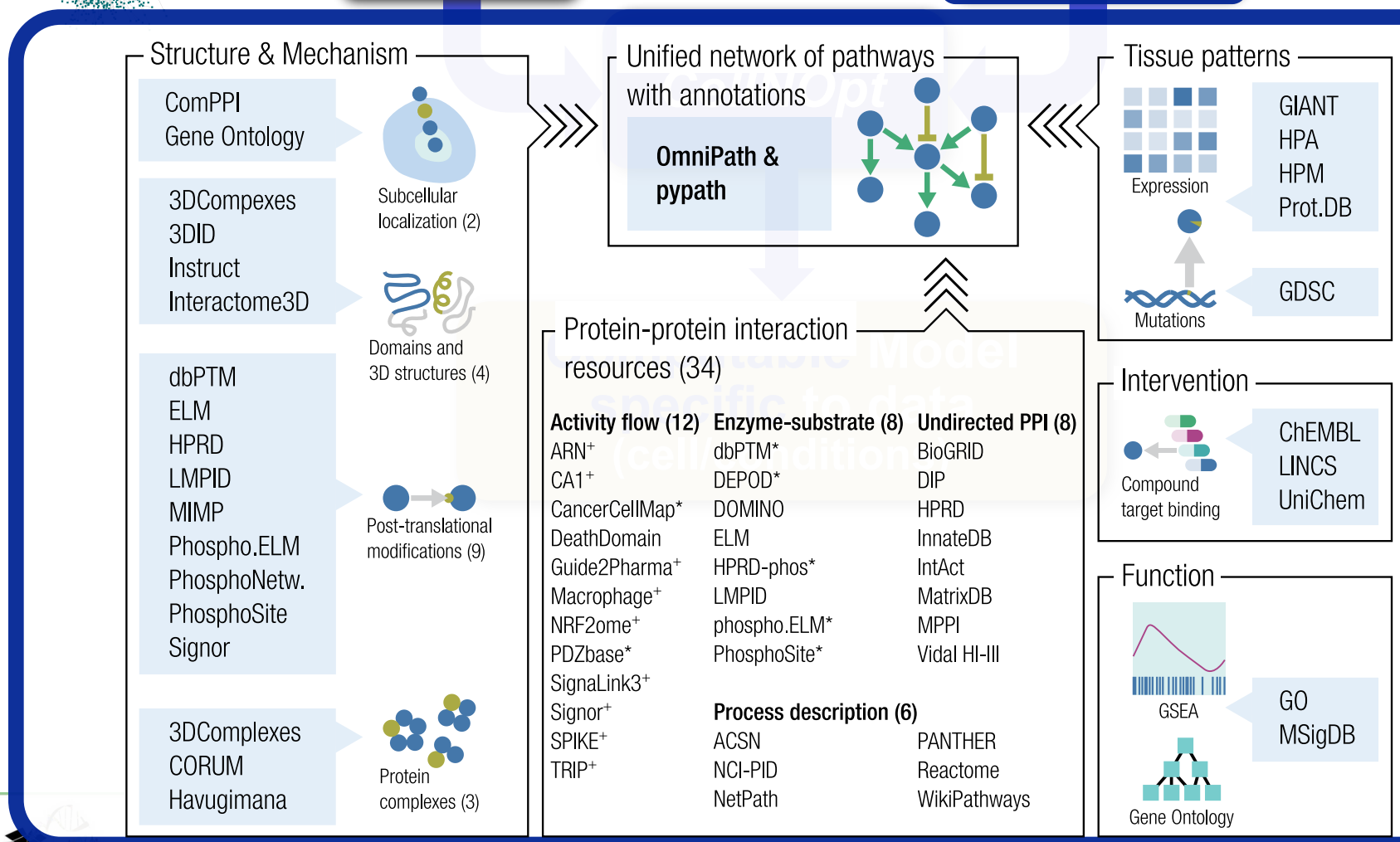
relationship can be 1 or -1 (for inhibition)

Simple, but no layout information included

Omnipath: Integration of existing pathway resources to improve modelling

www.omnipathdb.org

Networks

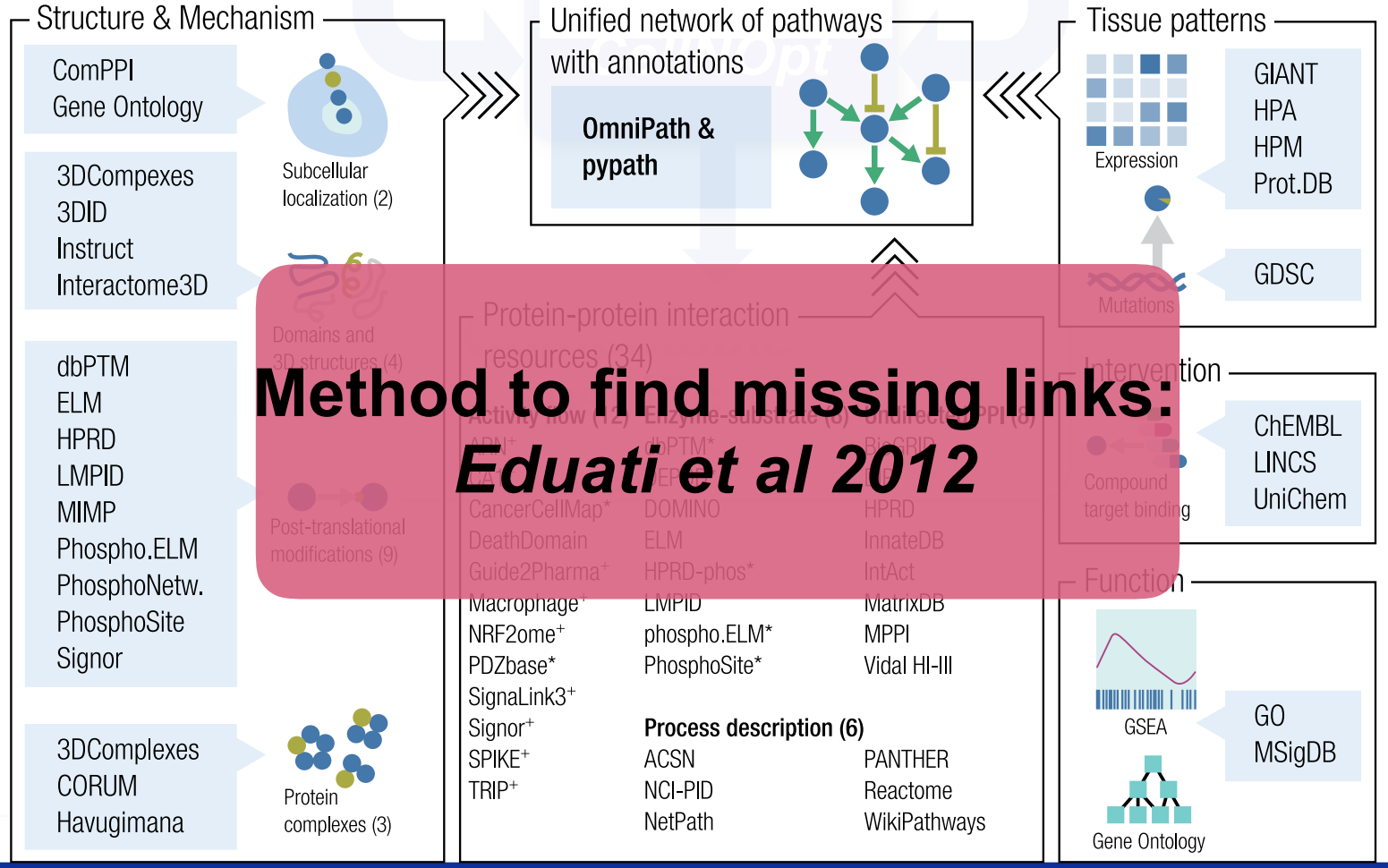


Turei, Korcsmaros, Saez-Rodriguez, *Nature Methods,* 2016

Omnipath: Integration of existing pathway resources to improve modelling

www.omnipathdb.org

Networks



Method to find missing links:
Eduati et al 2012

Turei, Korcsmaros, Saez-Rodriguez, *Nature Methods,* 2016

Modelling combinations of data-types

Data | **Networks**

- Antibody-based population data (protein arrays, luminex, …)

- **Single cell**

- **Mass spectrometry** phospho-proteomics
  - Limited conditions & replicates

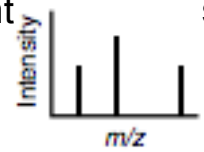
- **Metabolic regulation**

Using incomplete & noisy mass spectrometry data to build logic models of signal transduction

RWTH AACHEN

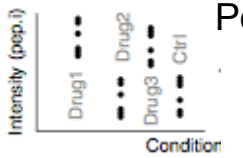
Protein extraction
phospho-enrichment

Cells treated with kinase inhibitors

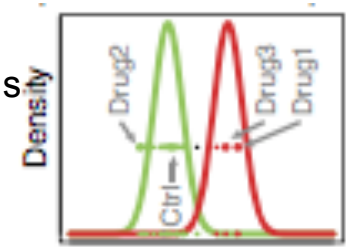


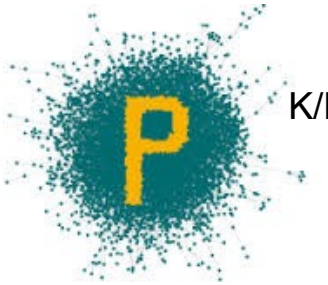
MS analysis

Normalisation
stat. significance



GMM
Perturbation status



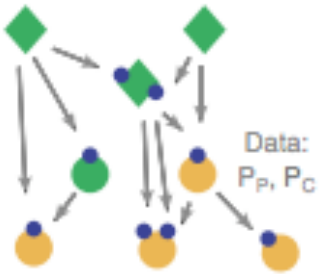


Omnipath

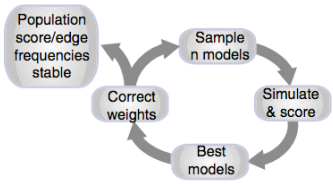
K/P-S interactions

Bipartite Network

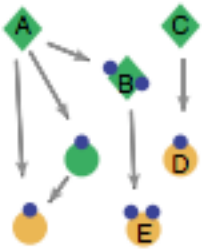
Building background
network



Training BN to
data



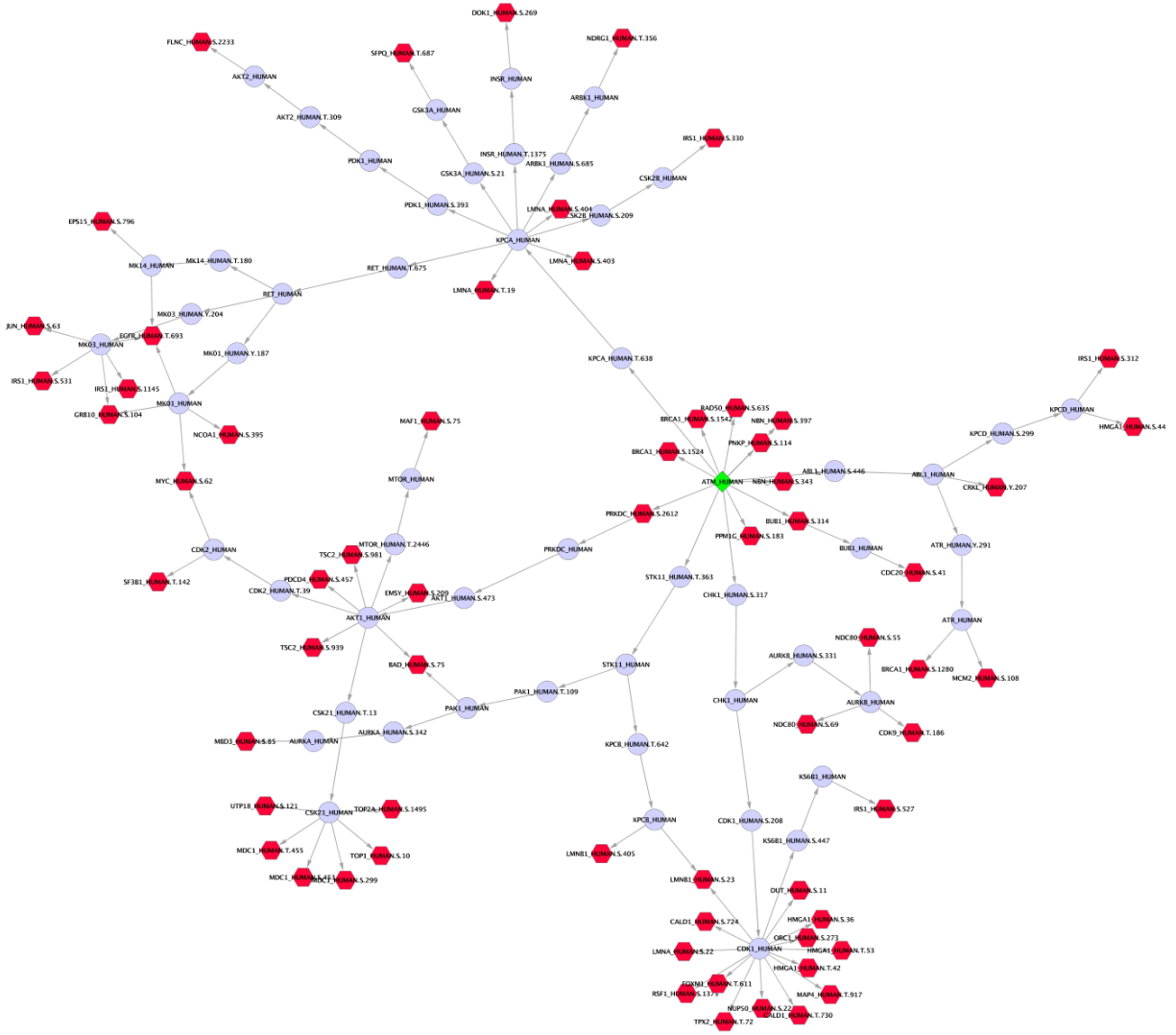
Finding logical
perturbation flow



Terfve et.al., *Nature Comms*, 2015

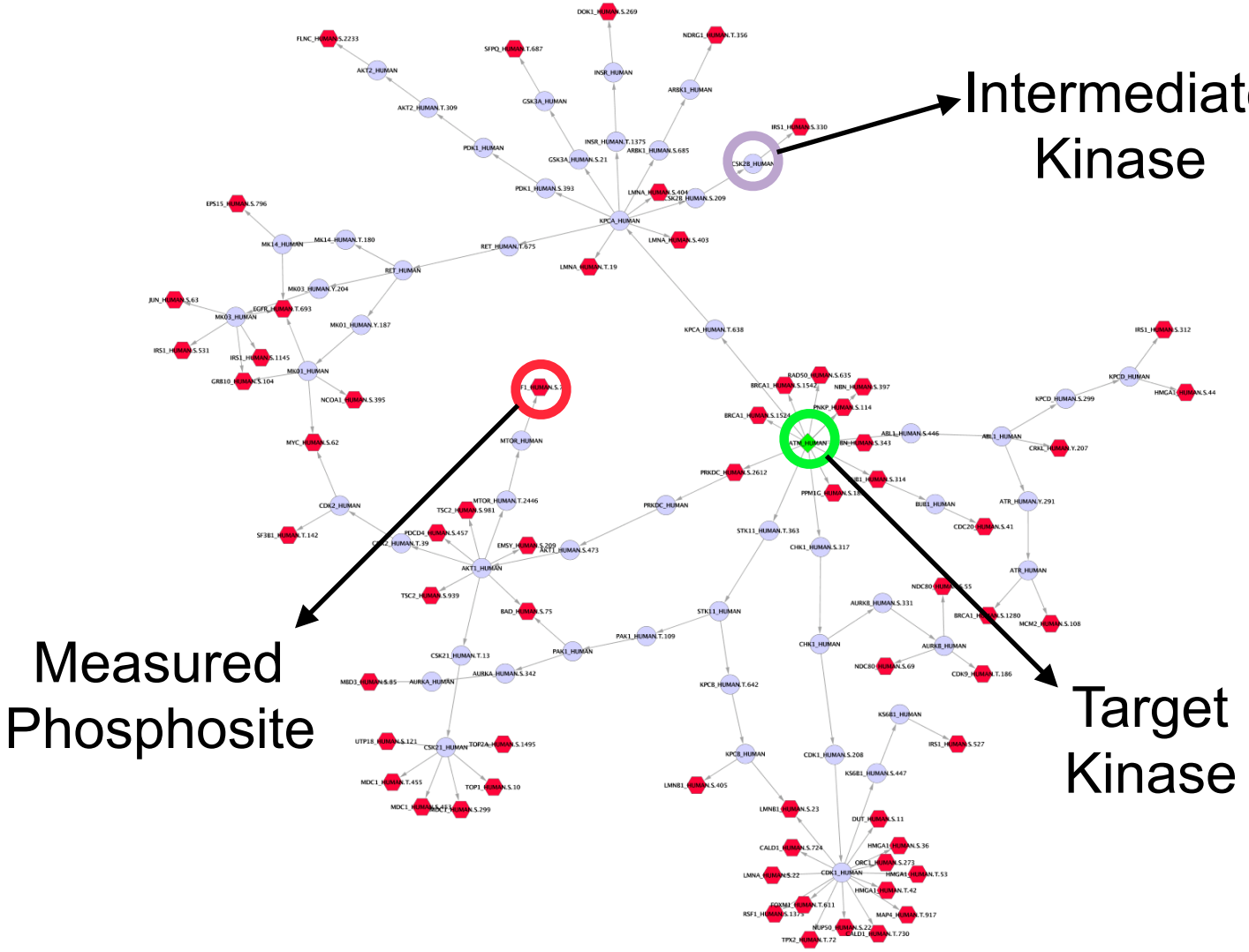
SyMBioSys

PHONEMeS



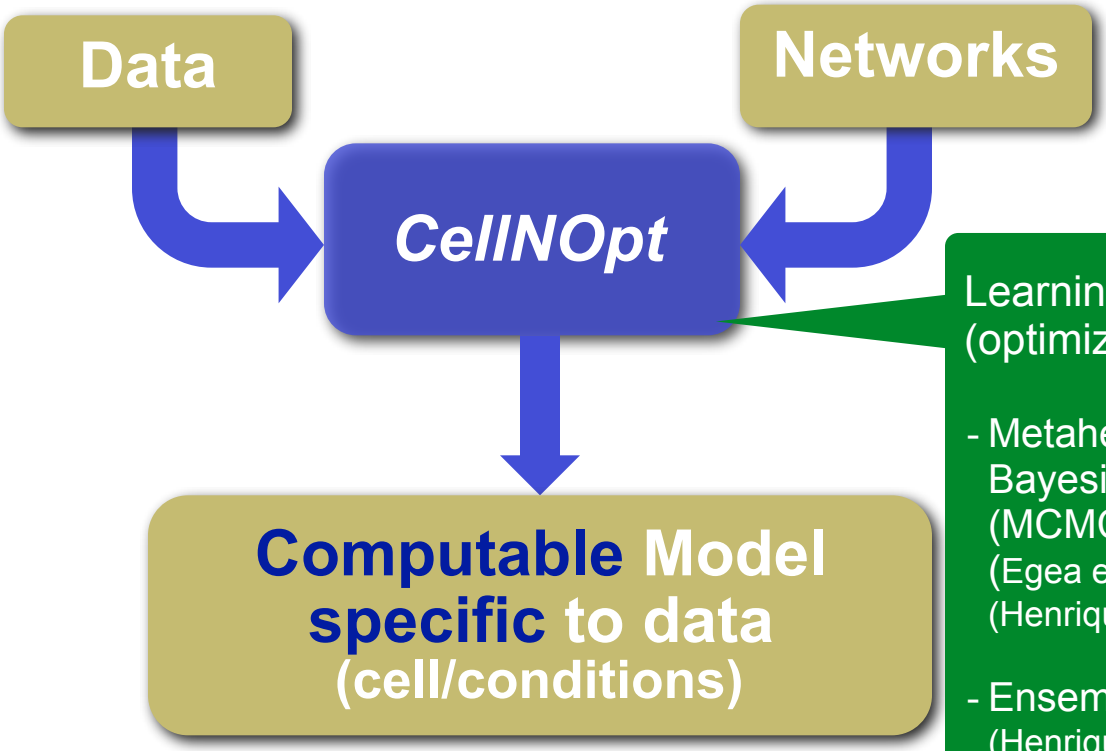
E. Gjerga

PHONEMeS



E. Gjerga
J. Wirbel

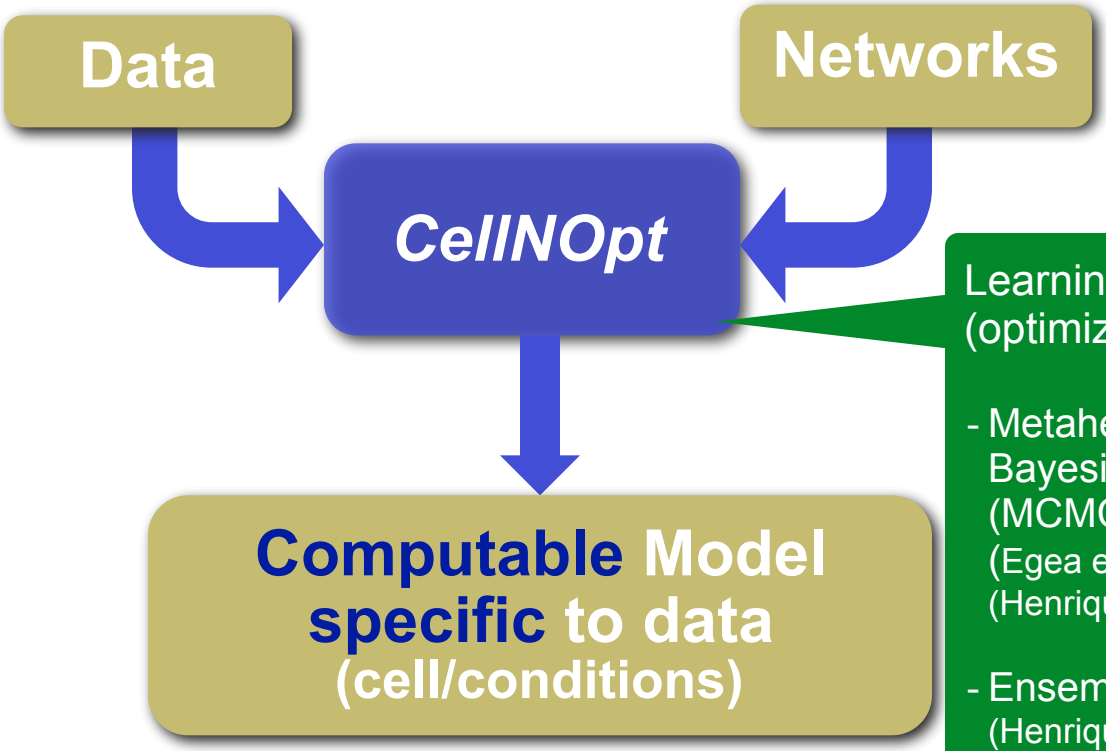
Computational & algorithmically developments



Learning algorithms (optimization):

- Metaheuristics & Bayesian Inference (MCMC) (Egea et al. *BMC Bioinf* 2014; (Henriques et al. *Bioinf* 2015)
- Ensembles of models (Henriques et al. *PLoS CB*, 2017)
- Use of Answer Set Programming (Guziolowski et al. *Bioinf* 2013, Videla et al. *Bioinf* 2017) and Integer Linear Programming (Mitsos et al *PLoS CB* 2009)

Computational & algorithmically developments



Learning algorithms (optimization):

- Metaheuristics & Bayesian Inference (MCMC) (Egea et al. *BMC Bioinf* 2014; (Henriques et al. *Bioinf* 2015)
- Ensembles of models (Henriques et al. *PLoS CB*, 2017)
- Use of Answer Set Programming (Guziolowski et al. *Bioinf* 2013, Videla et al. *Bioinf* 2017) and **Integer Linear Programming** (Mitsos et al *PLoS CB* 2009)

PHONEMeS ILP Implementation

ILP is a mathematical optimisation problem in which the objective function and constraints are linear while the variables are integers

General ILP statement: $\min_{x}\left(c^T x\right) \qquad s.t.: Ax \le b, \;\; x \in \mathrm{Z}^n$

Problems with binary decision variables occur often in many signalling model formulations

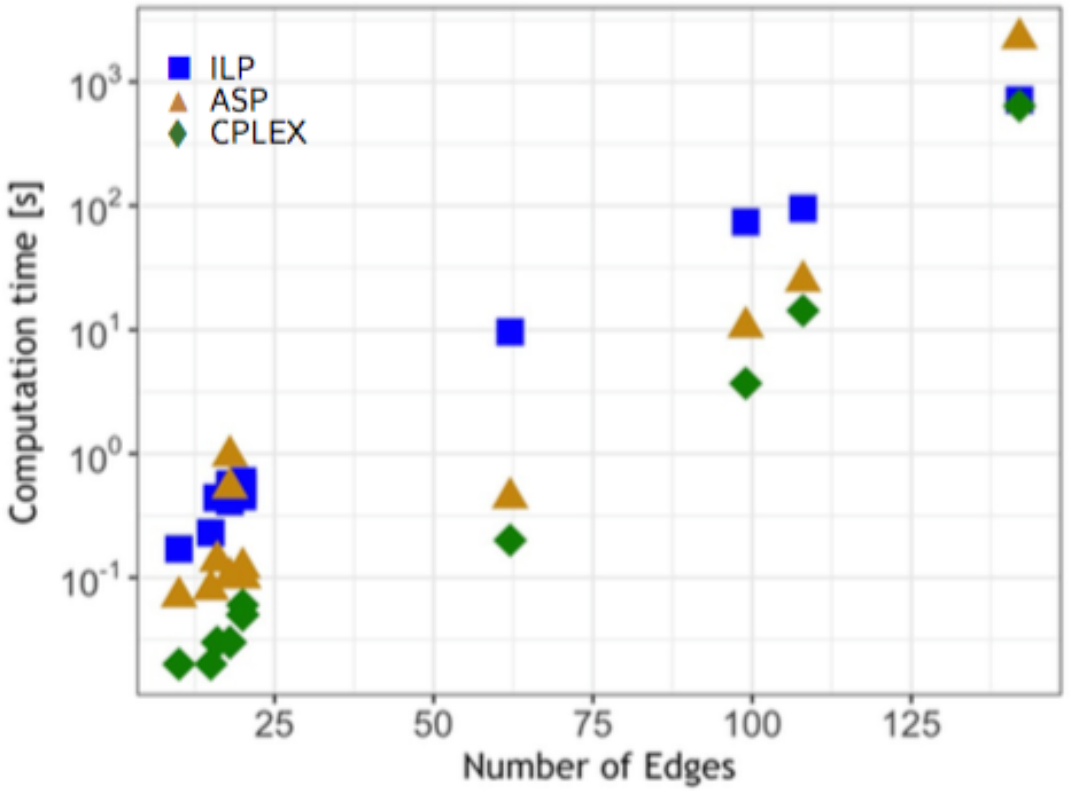
ILP implementations for **CellNOpt** and **PHONEMeS**

CPLEX solver

ILP implementation of CellNOpt

ILP implementation for CellNOpt (*Mitsos et.al.*)

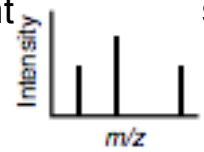
Comparisons between different algorithms applied



Accelerating model building from mass spectrometry data with ILP

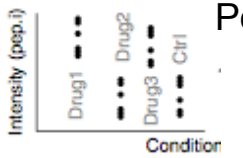
Protein extraction
phospho-enrichment

Cells treated with kinase inhibitors

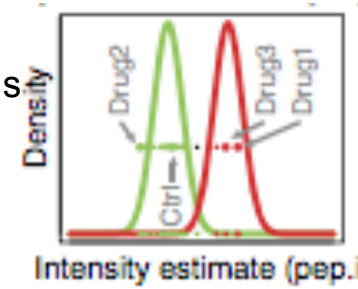


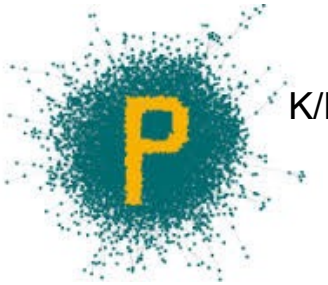
MS analysis

Normalisation
stat. significance



GMM
Perturbation status



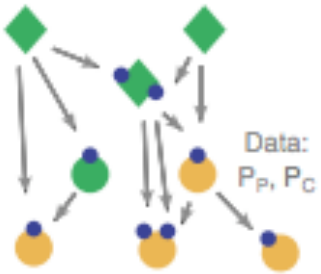


Omnipath

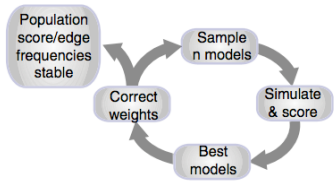
K/P-S interactions

Bipartite Network

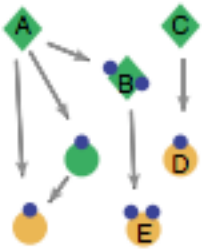
Building background
network



Training BN to
data



Finding logical
perturbation flow



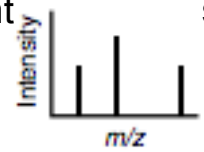
Terfve et.al., *Nature Comms*, 2015

SyMBioSys

Accelerating model building from mass spectrometry data with ILP

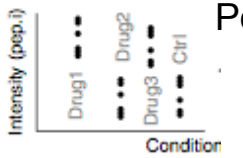
Protein extraction
phospho-enrichment

Cells treated with kinase inhibitors

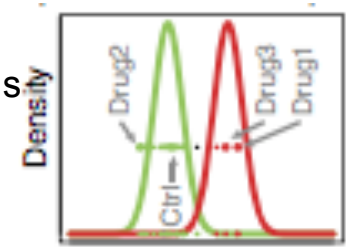


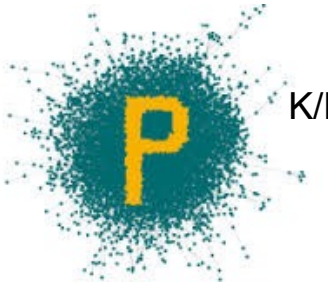
MS analysis

Normalisation
stat. significance



GMM
Perturbation status



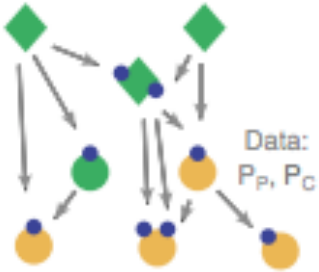


Omnipath

K/P-S interactions

Bipartite Network

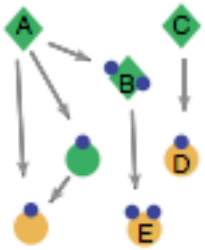
Building background
network



Training BN to
data

IBM CPLEX

Finding logical
perturbation flow



ILP for PHONEMeS

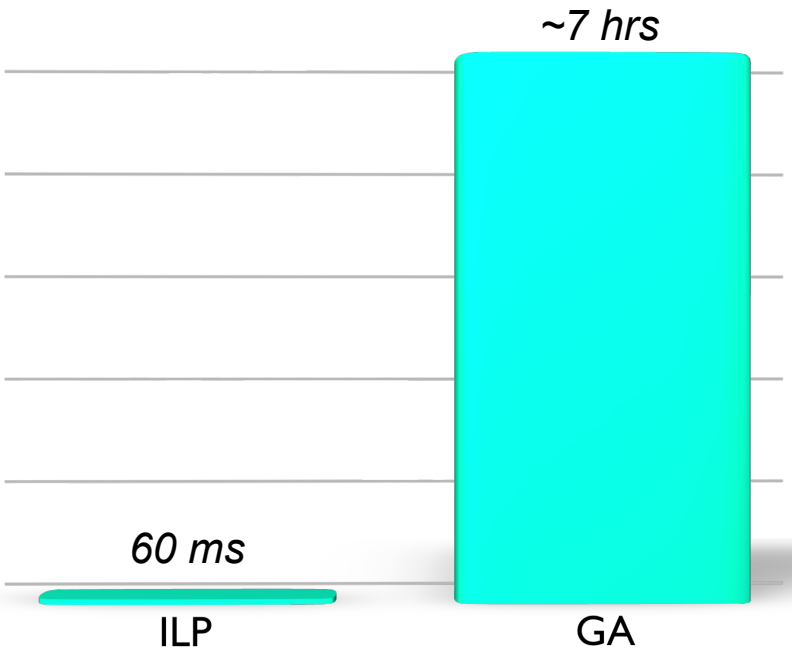
Can we make it faster??

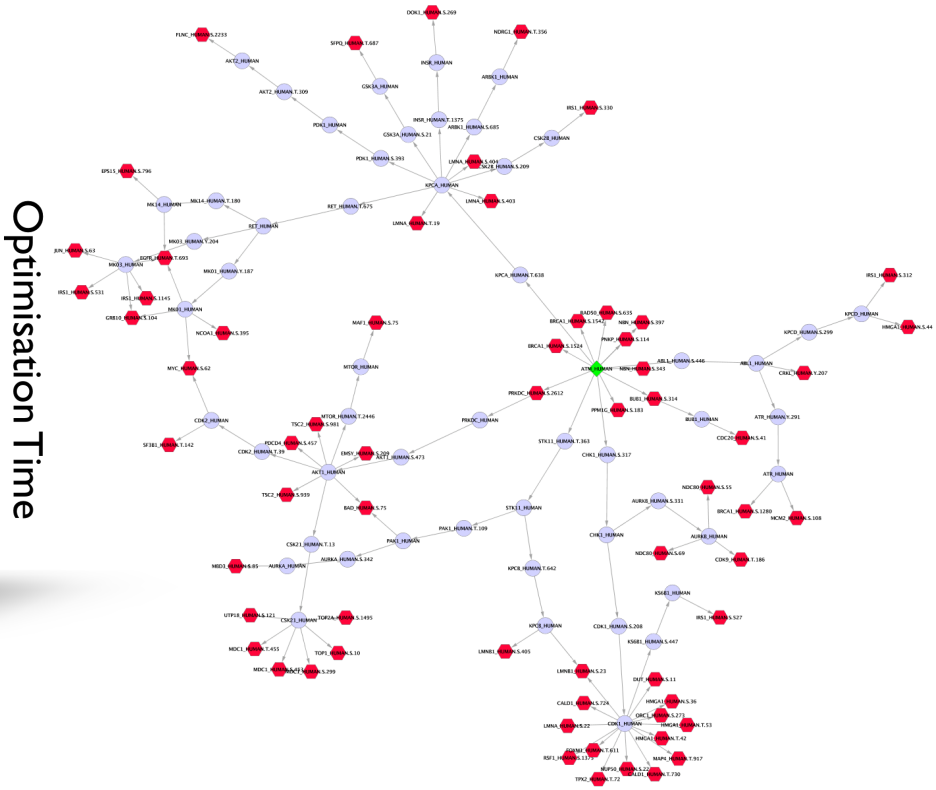
E. Gjerga

SyMBioSys

Massive speed up with ILP

PHONEMeS ILP is orders of magnitude more efficient than the previous implementation - **Time Efficient !!!**





Ongoing work with ILP and Phonemes

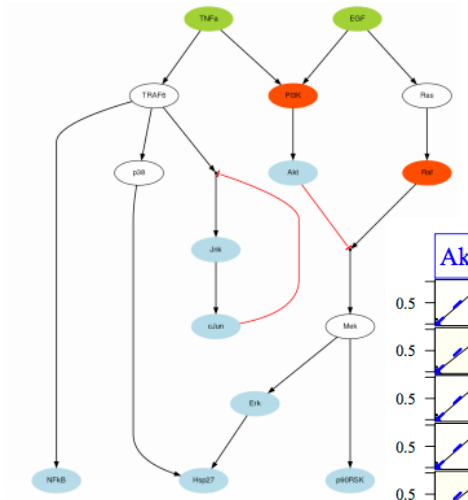
Modelling through multiple targets

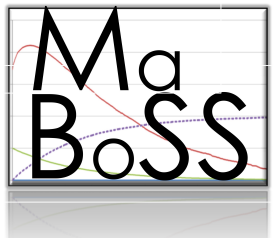
Modelling of perturbation propagation from a receptor level through hybrid interactions

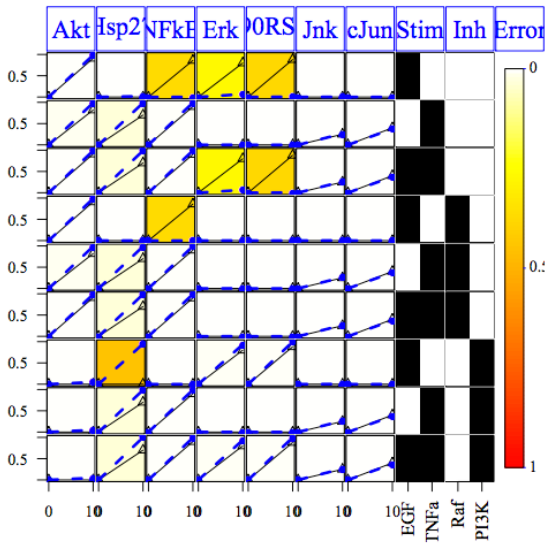
Time course PHONEMeS

CellNOpt new features - MaBoSS

MaBoSS implementation in CellNOpt







MaBoSS, an algorithm for modelling biological networks in discrete framework with continuous time

Simulates the probabilities for each species to be active at each condition

MaBoSS simulator for CellNOpt

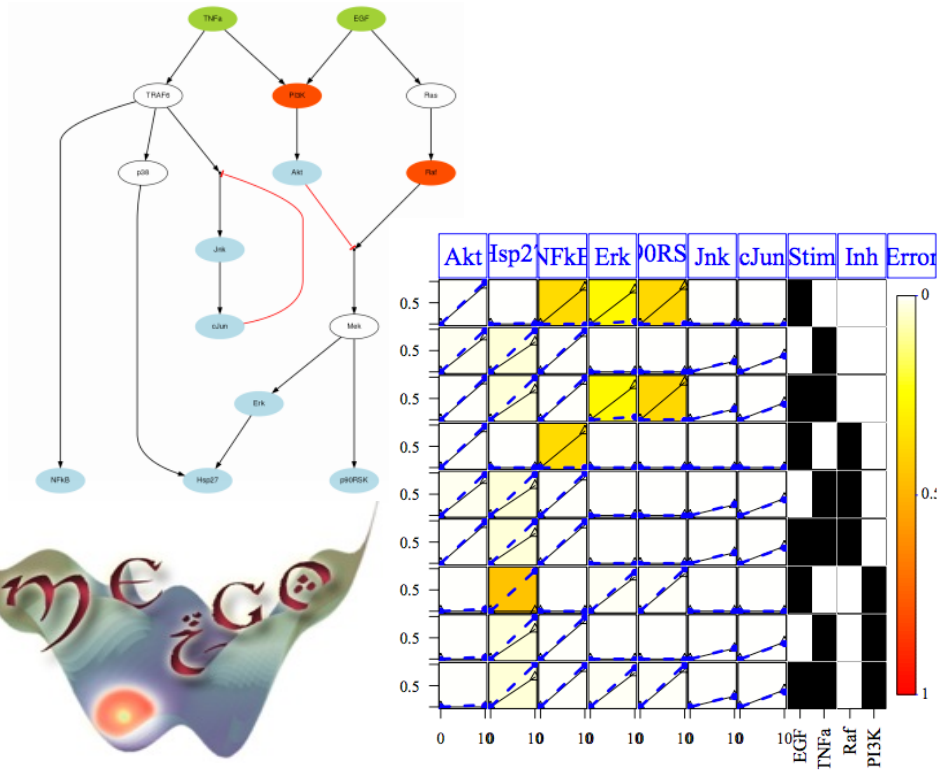
Stoll et.al., *BMC Sys Biol,* 2012

C. Chevalier

SyMBioSys

CellNOpt new features - CNORode2017

New **CNORode** version released in *Github*



CNORode allows a quantitative description of a boolean model through the identification of logic ODE parameters

New features:

L1 regularisation penalty term is added to the objective function

An additional **steady state penalty** priorities this parameter sets in which the steady state is reached

Bootstrap or optimisation with random resampling allows to determine the confidence intervals for each parameter

Download package in: https://github.com/saezlab/CNORode2017

see Eduati et al. Cancer Research 2017

Acknowledgements

Saez-Rodriguez group, specially:

Luis Tobalina

Aurelien Dugourd

Jakob Wirbel

Attila Gabor

Celine Chevalier

Federica Eduati

Panuwat Trairatphisan

Collaborators: Julio Banga
Bioprocess Engineering Group
Vigo, Spain

Funding: H2020 SymbioSys - Marie Skłodowska-Curie Action

Thank You

www.saezlab.org

@sysbiomed