Efficient algorithms for the contextualization of molecular network models

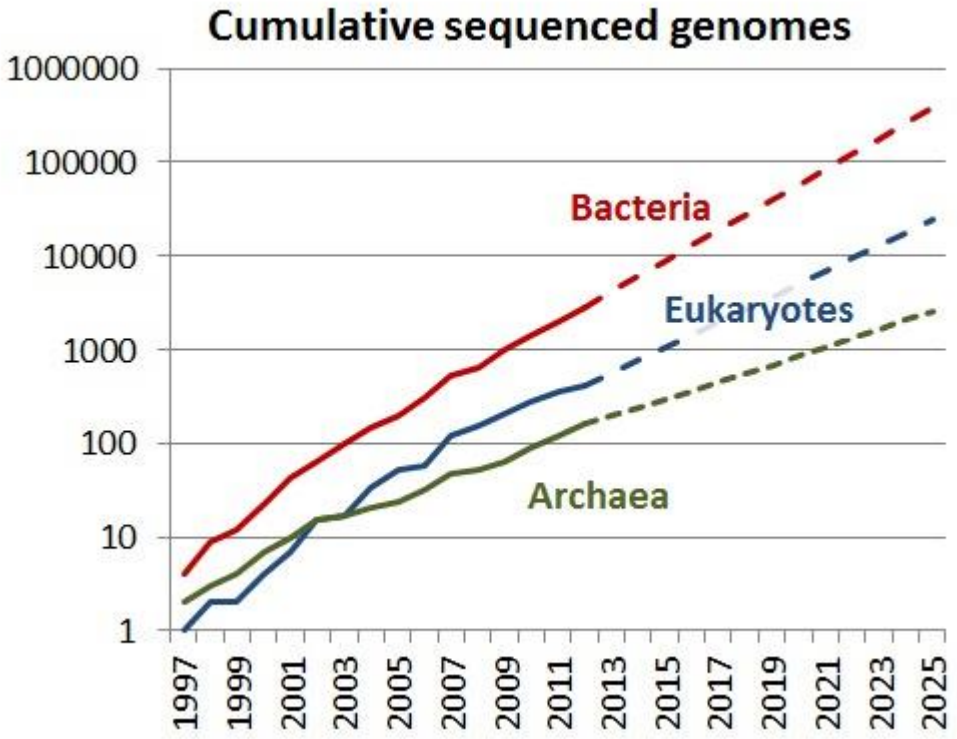
1. **Introduction: Network modeling / Systems Biology**
2. **Specific metabolic network models with fastcore / fastcormics**
3. **Specific signalling network models with optPBN / Falcon**

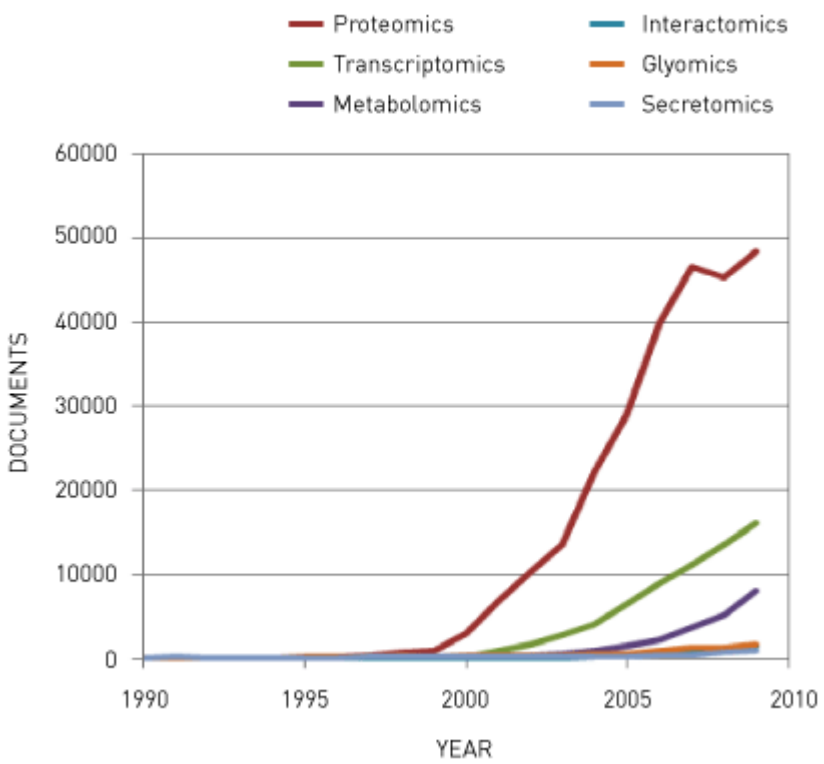
Thomas Sauter

System Biology group

University of Luxembourg

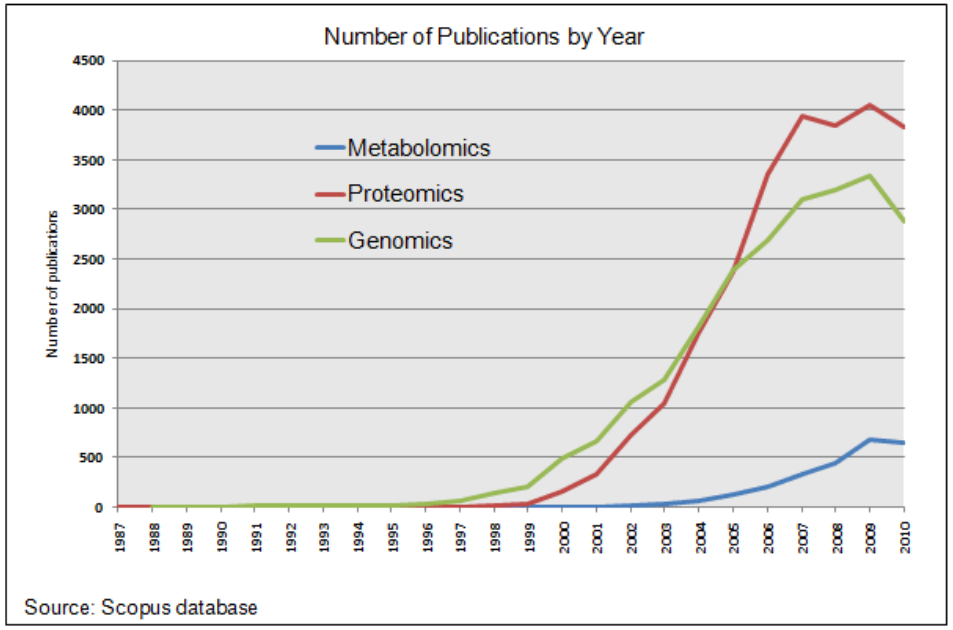
“We are witnessing the emergence of the "data rich” era in biology…”





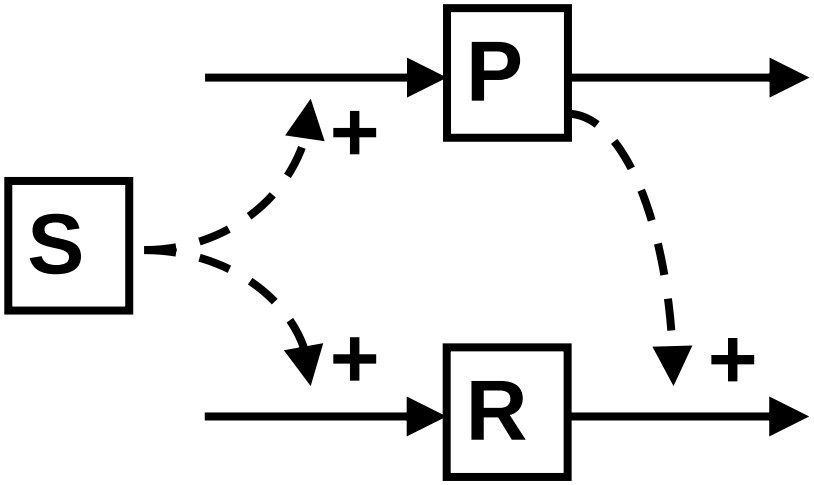
[S. Subramaniam, 2004, Jones, Research Trends, 2010
Su, sulab.org, 2013;
Gilbert et al, Stand. Genomic Sci., 2012]

“The bottleneck … has shifted from generating the data to interpreting results so as to derive insights into biological mechanisms.”



“We cannot understand complex systems without modeling.”

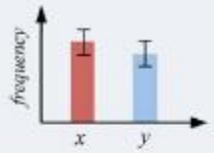
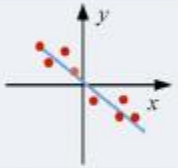
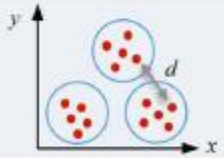
“The role of the model is to make something complex intelligible or understandable.”



[O. Wolkenhauer, Front Physiol, 2014, “Why model?”]

“Why model?

The role of the model is to make something complex intelligible or understandable.”

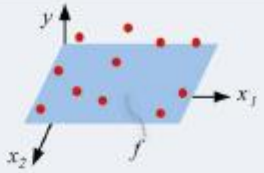
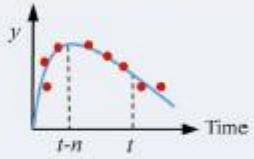
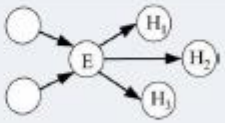
Testing for differences	**Significance Analysis**	Is there a difference between the two groups?
“before/after” “with/without”	$95\%\ \mathrm{CI} = \overline{X} \pm \left(t_{(n-1)}(5\%) \times \overline{\mathrm{SE}}\right)$	
Analysing covariaton	**Correlation Analysis**	Does a small/large value of x, coincide with a small/large value of y?
“coincides with”	$r_{x,y} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$	
Identifying groups	**Cluster/Discriminant Analysis**	Can the data be grouped on the basis of features x and y?
“is similar to”	$D = \min_i \left\{ \min_j \left\{ \frac{d(i,j)}{\max_k d'(k)} \right\} \right\}$	
Condensing data	**Component/Factor Analysis**	Explain the variability through a weighted linear combination of principal components.
“reduces to”	$X = U \Sigma W^T$	

Ease of experiments

[O. Wolkenhauer, Front Physiol, 2014, “Why model?”]

“Why model?

The role of the model is to make something complex intelligible or understandable.”

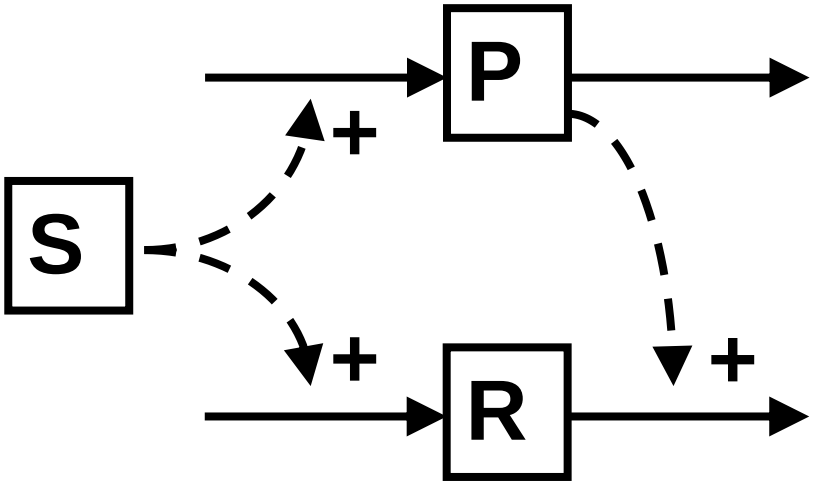
Fitting data	Regression Analysis	What is the predicted value of y, given x_1 and x_2?
“relates to”	$y = f(x_1,\ldots,x_n)$	
Numerical predictions	**Time Series Analysis**	What is the predicted value of y, knowing past values of y?
“follows”	$y(t) = f(y(t-1), y(t-2),\ldots)$	
Analyzing influences	**Bayesian (network) Analysis**	From the observed evidence, H_2 is most probable.
“Given E, the probability of H is”	$P(H_j\|E) = \frac{\Pr(E\|H_j)\Pr(H_j)}{\sum \Pr(E\|H_j)\Pr(H_j)}$	
Investigating mechanisms	**Dynamical Systems Theory**	The response can be explained by a (bistable) switching mechanism.
“causally entails” “If ... Then ...”	$\frac{d}{dt}x_i = f_i(x_1(t),\ldots,x_n(t),u(t))$	

Power of explanation

[O. Wolkenhauer, Front Physiol, 2014, “Why model?”]

“Systems biology is the science that studies how biological function emerges from the interactions between the components of living systems and how these emergent properties enable and constrain the behavior of those components.”

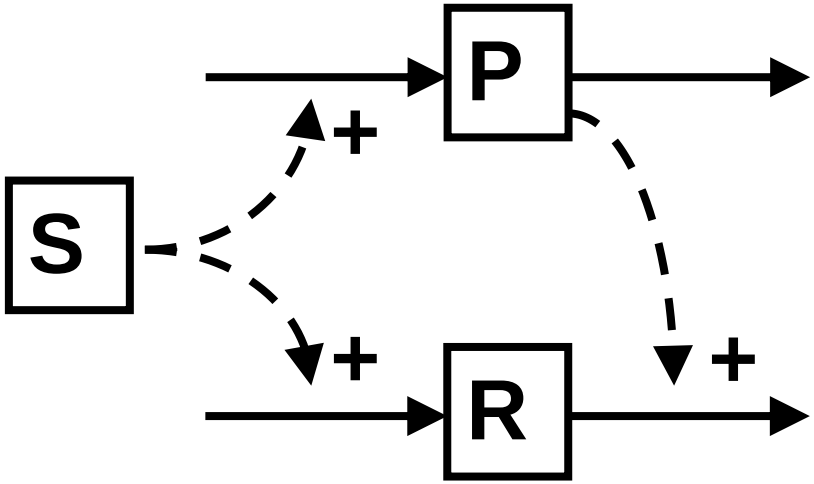
Example Network Motif: Function?



[O. Wolkenhauer, Front Physiol, 2014, “Why model?”]

Systems approach / Function:

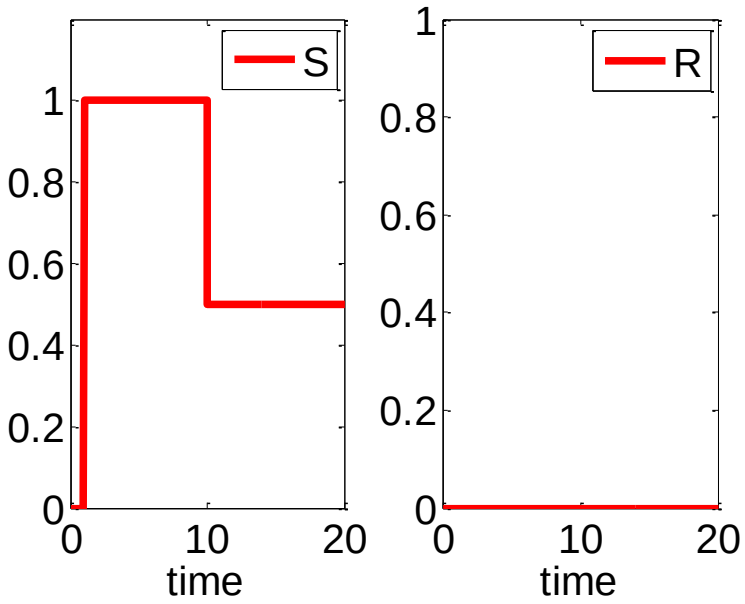
Output behavior as a function of input and time



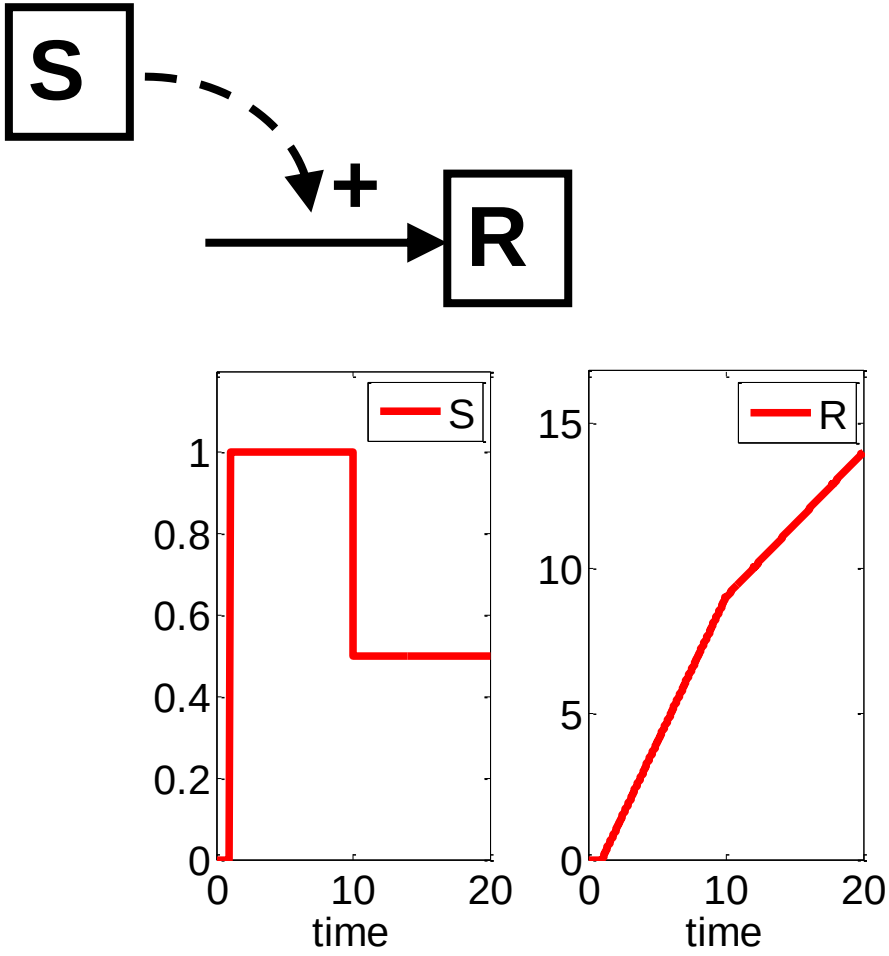
Output as a function of input and time

S

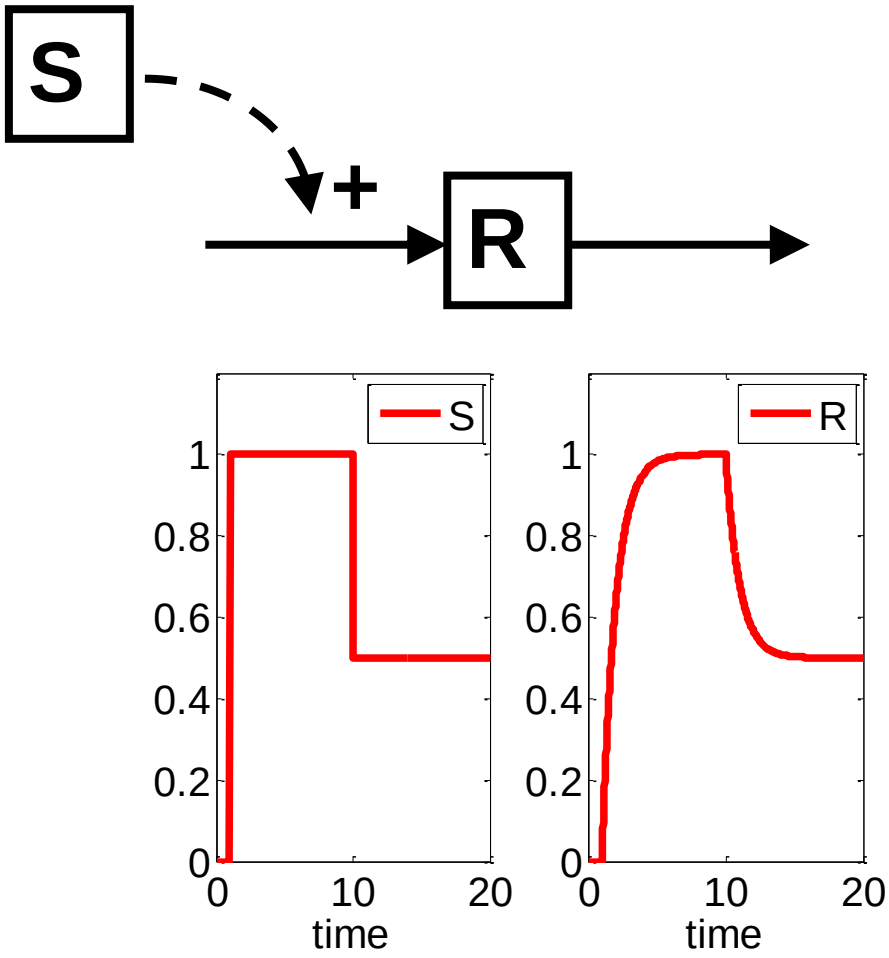
R



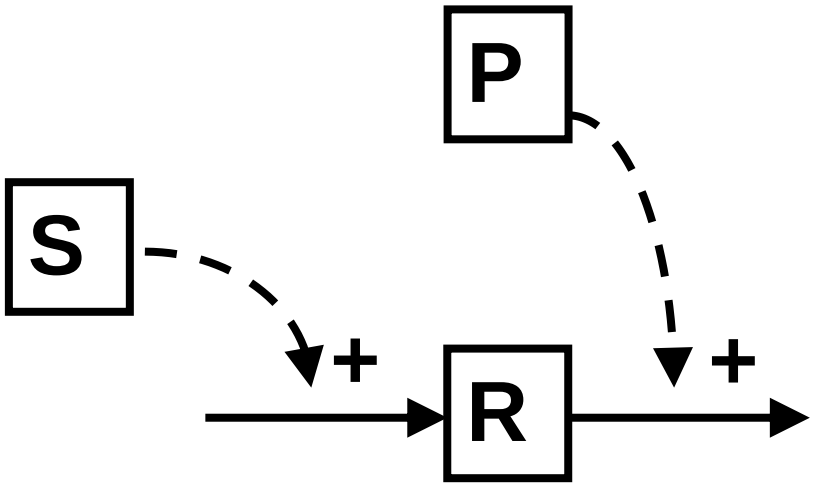
Output as a function of input and time

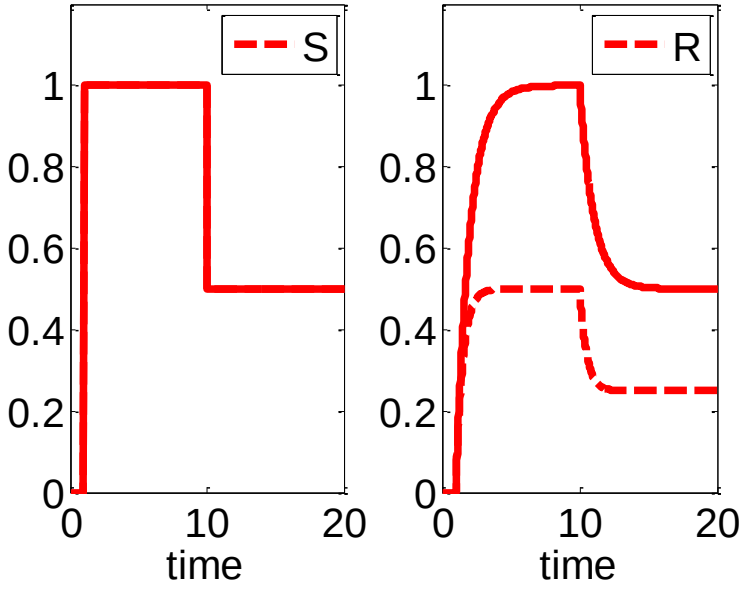


Output as a function of input and time

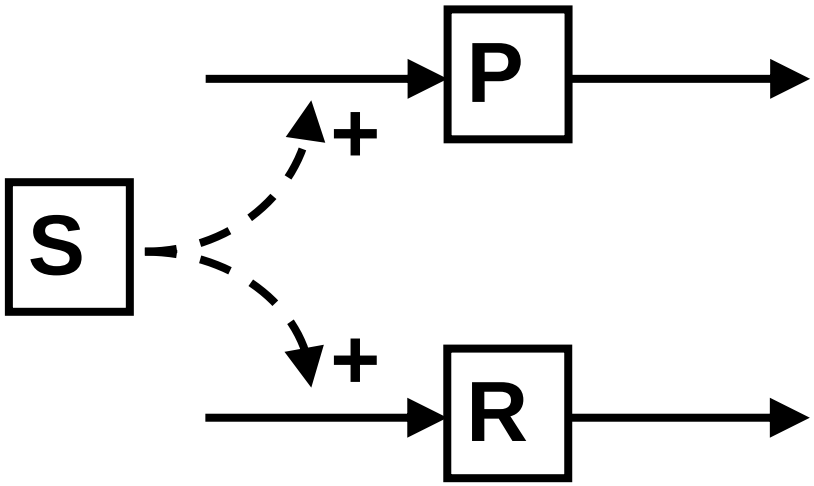


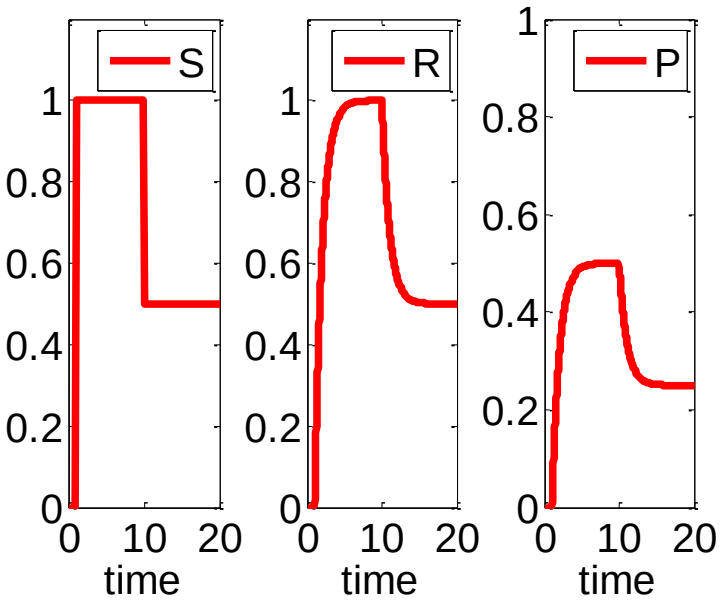
Output as a function of input and time



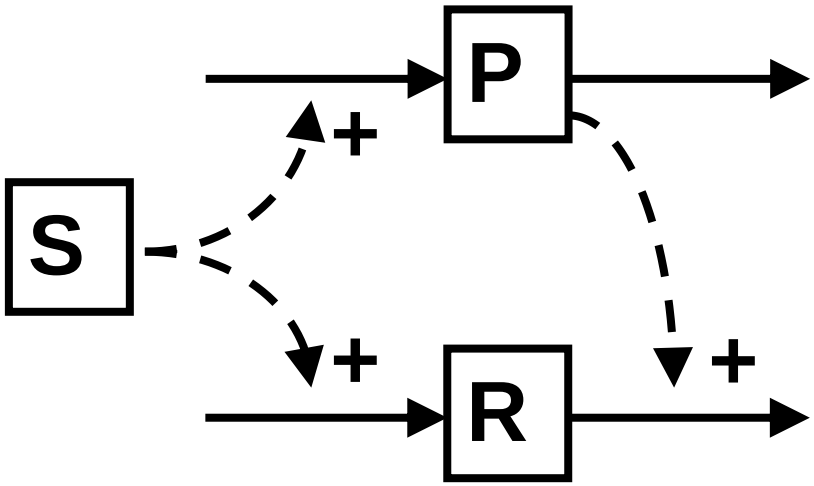


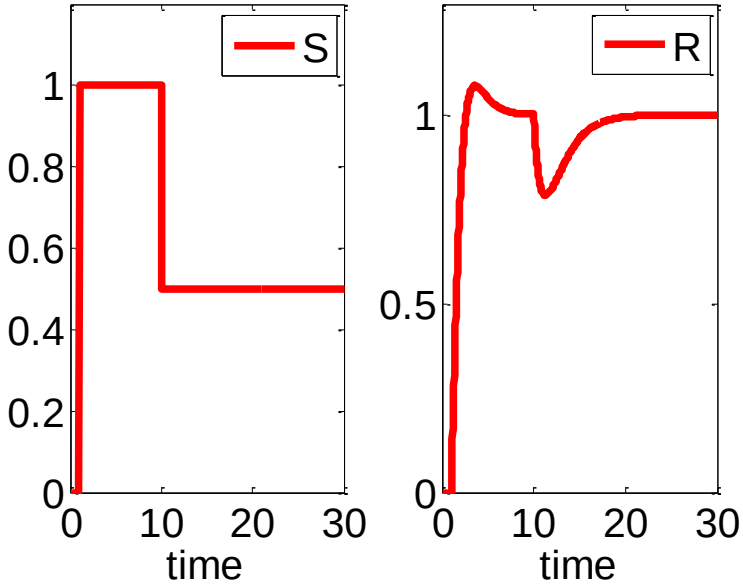
Output as a function of input and time





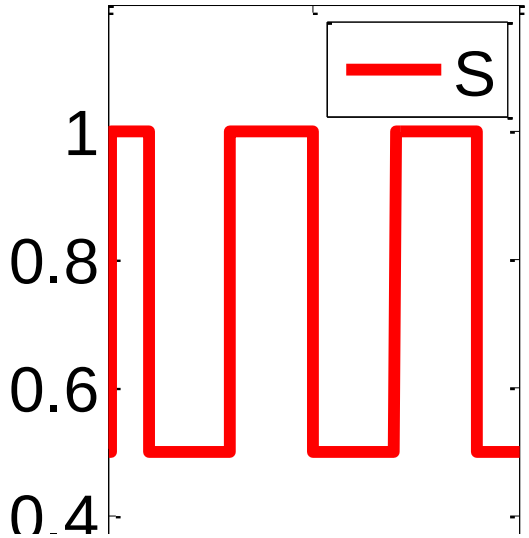
Output as a function of input and time

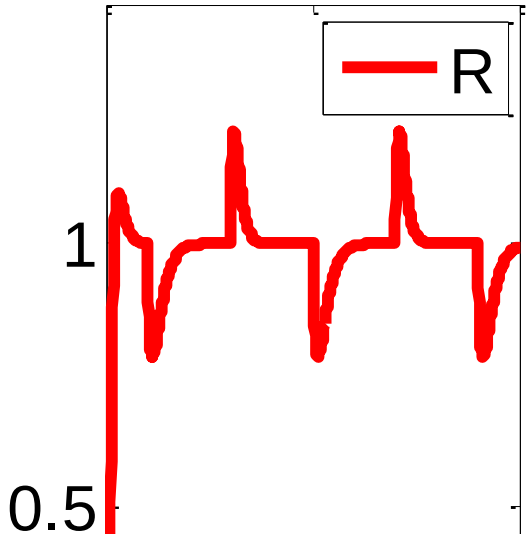




Function:

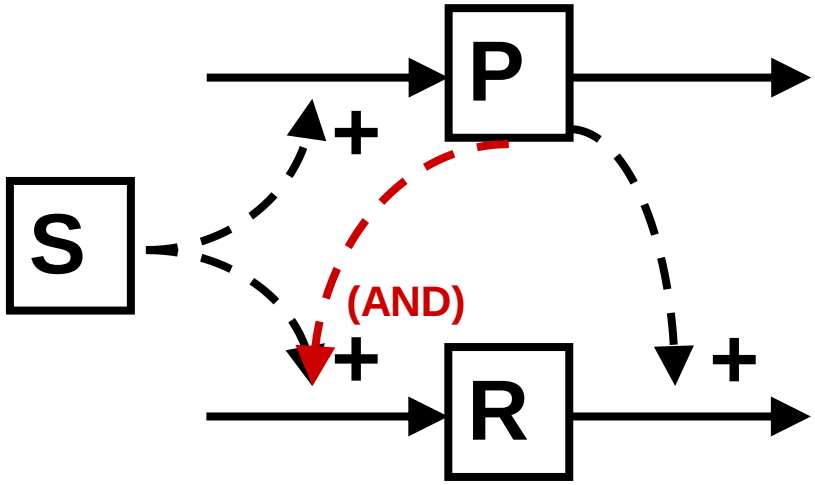
Change detection / Adaptation

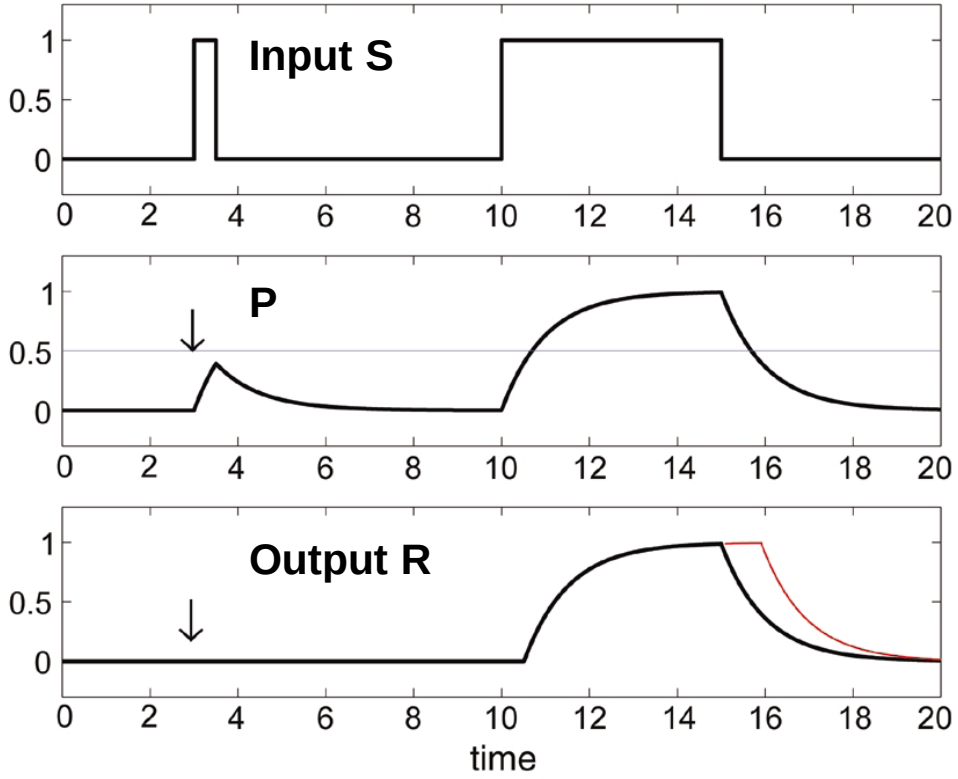




Systems approach:

Second example: Different wiring

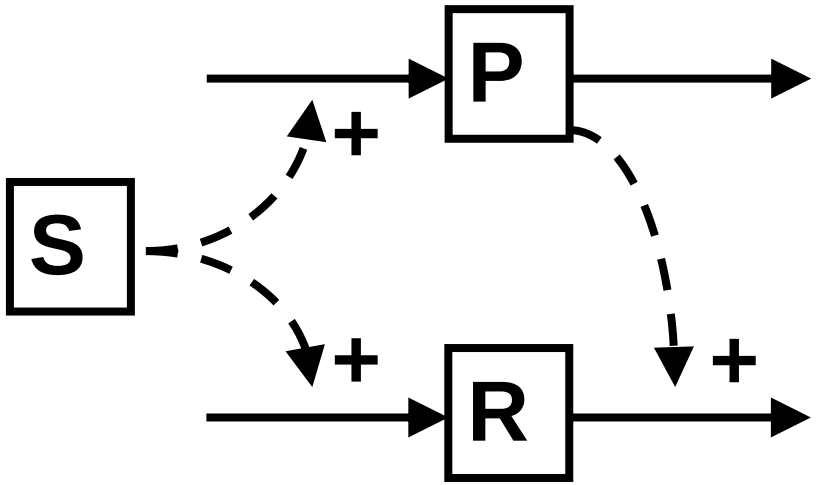




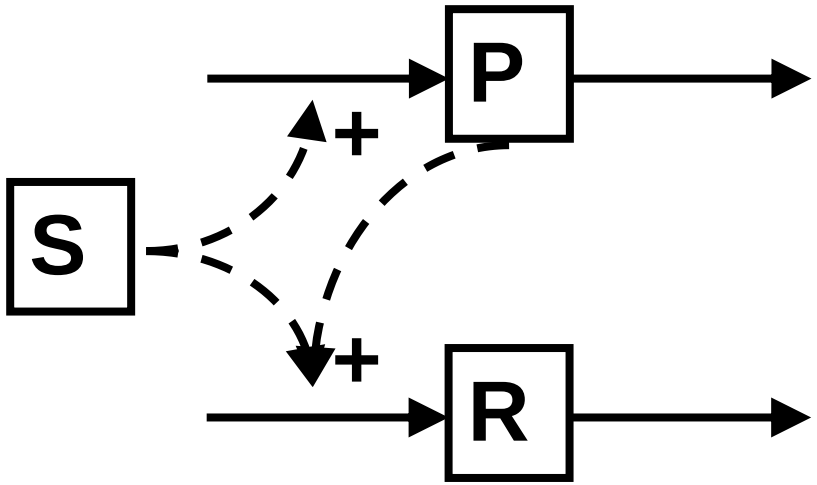
Function: Filter – removes short inputs & act upon long inputs

Structure: Feedforward loops

incoherent:



coherent:



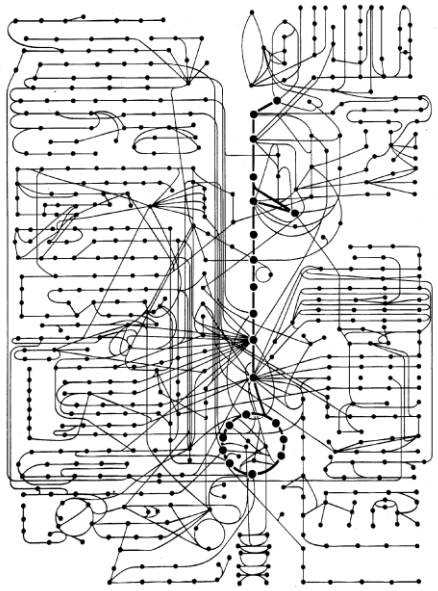
Functions:

Change detection

Short input filter

Model based Data Integration and Analysis of Disease specific Networks

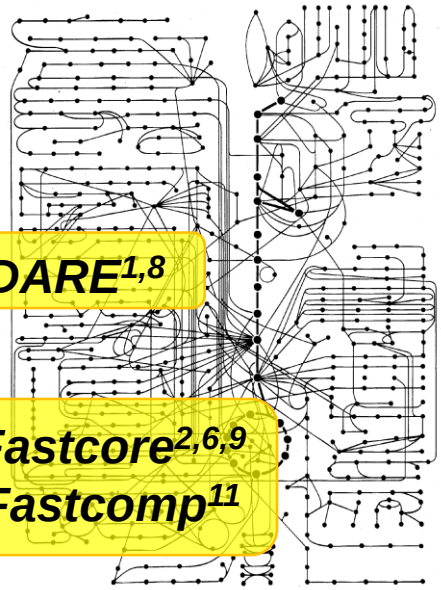
Data integration



Network analysis

Model based Data Integration and Analysis of Disease specific Networks

Data integration → **Network analysis**

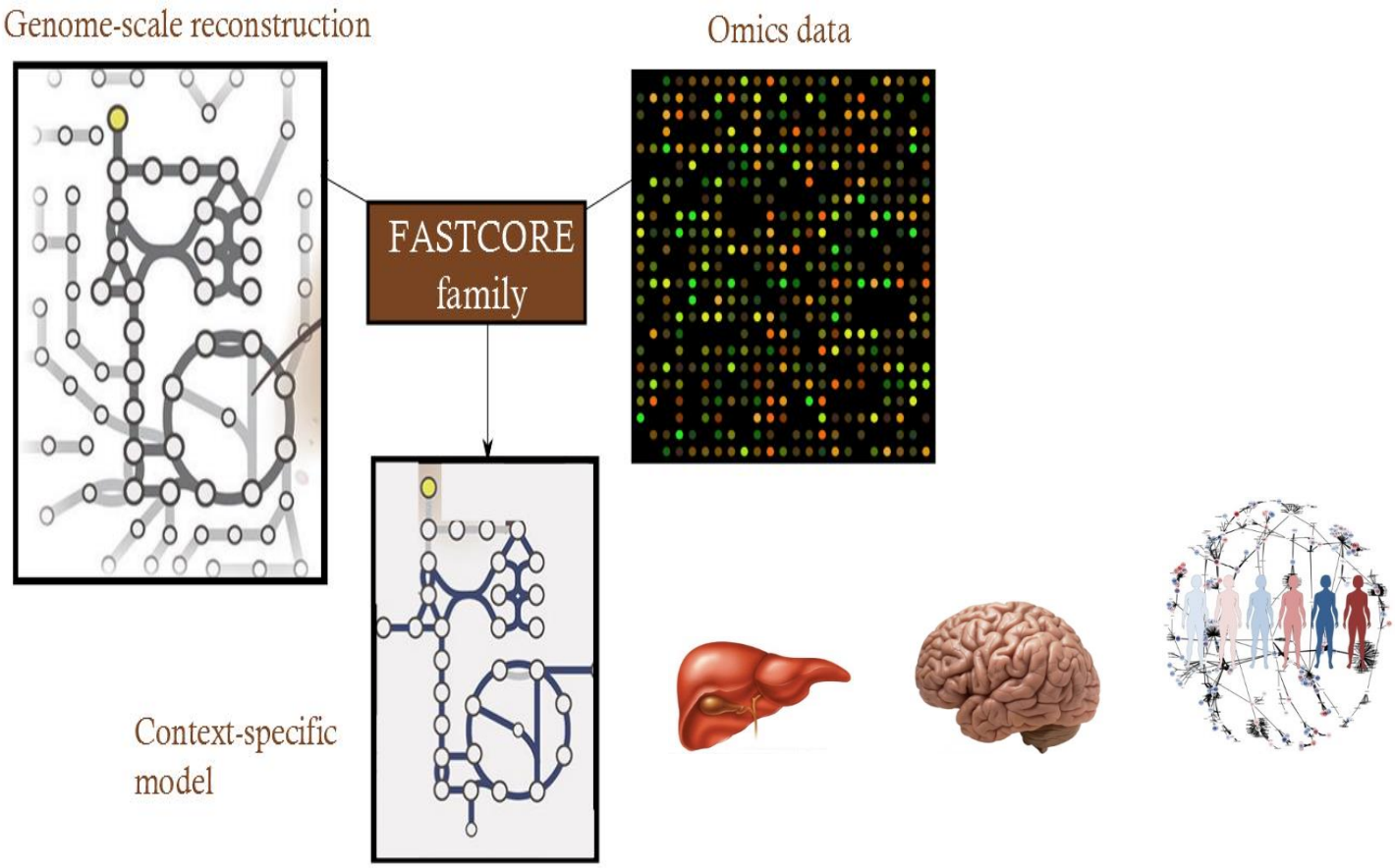


- Omics data **visualization**
  - *IDARE*[1,8]
- Metabolic network **reconstruction**
  - *Fastcore*[2,6,9]
  - *Fastcomp*[11]
- Signalling network **curation**
  - *optPBN*[3,4]
  - *FALCON*[5,7]
- Gene regulatory networks from **epigenetic** data
  - *EPIC-DREM*[10]
- **Data mining / machine learning**

1 Nucleic Acids Research 2013, 1;42(3):1474-96
2 PLoS Comput Biol. 2014, Jan;10(1)
3 Cell Commun Signal. 2013, Jul 1;11:46
4 PLoS One. 2014 Jul 1;9(7):e98001
5 Bioinformatics 2017 1;33(21):3431-3436
6 Nature Protocols, accepted
7 Frontiers Physiol 2018, 9:550
8-11 in prep, subm, under rev

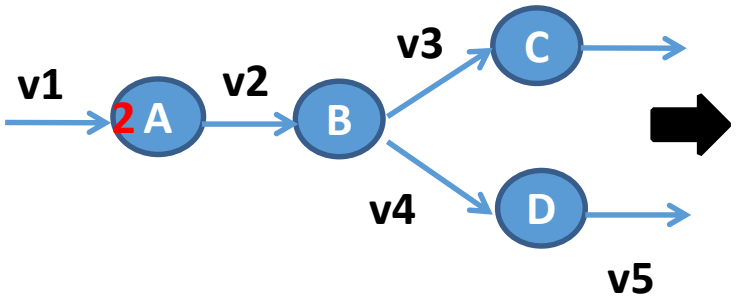
Metabolic models are

scaffolds for the integration of omics data



FASTCORE: Vlassis, Pires Pacheco & Sauter, 2014, Plos Comp Biol
FASTCORMICS: Pires Pacheco et *al*, 2015, BMC genomics
Benchmarking: Pires Pacheco *et al*, 2016, Front Physiol
FASTCORMICS RNAseq: Pires Pacheco et *al*, in preparation
In COBRA toolbox: Nature Protocols, accepted

Fast Reconstruction of Compact Context-Specific Metabolic Network Models: Constraint Based Modelling



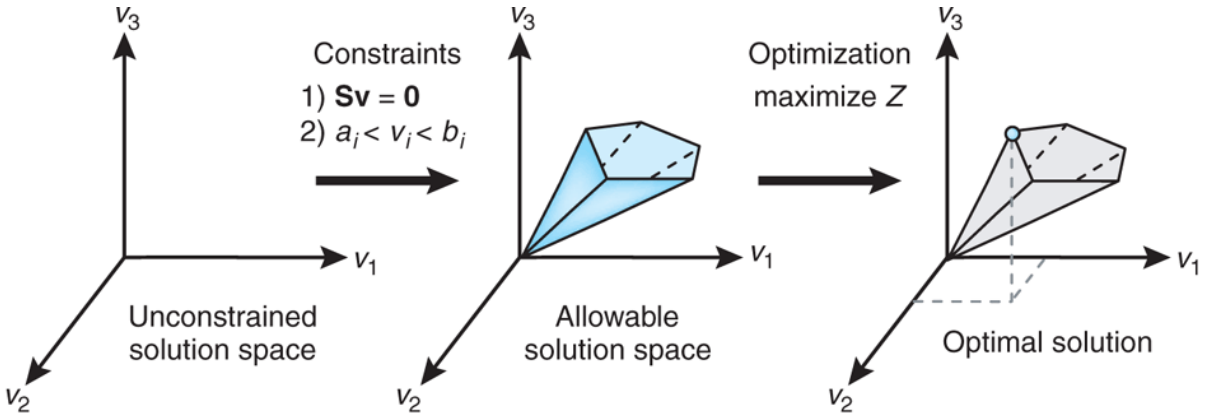
Metabolic network

	v1	v2	v3	v4	v5	v6
A	2	-2	0	0	0	0
B	0	1	-1	-1	0	0
C	0	0	1	0	0	-1
D	0	0	0	1	-1	0

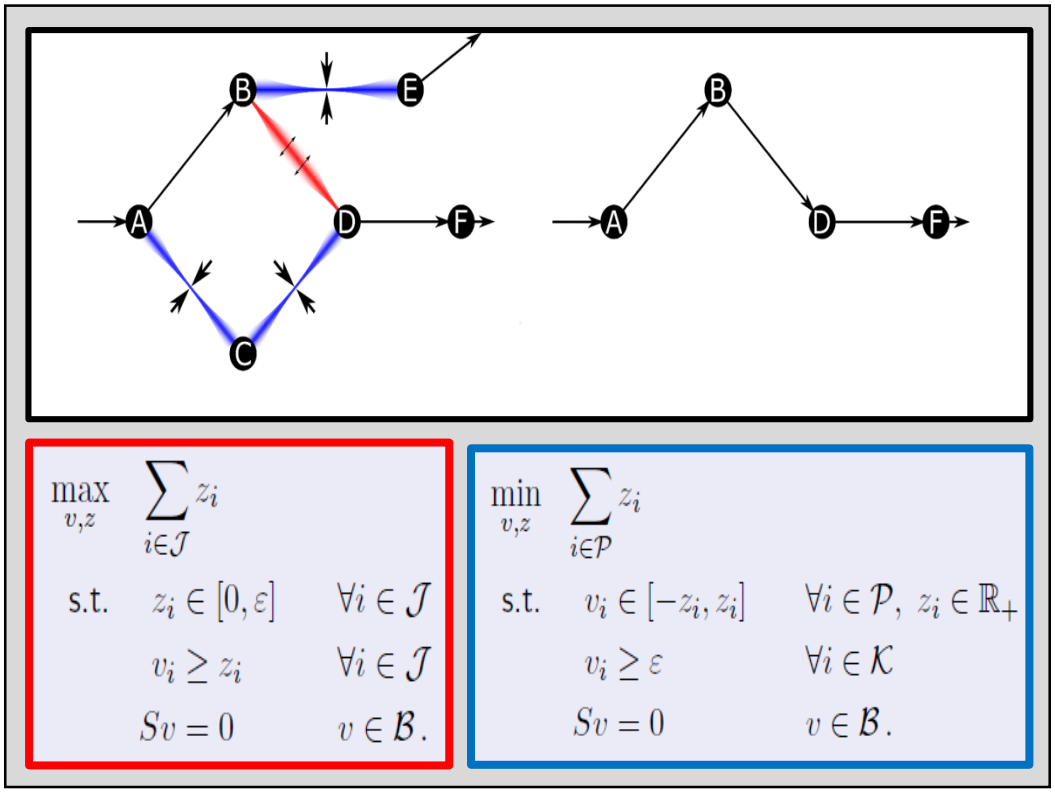
S

Mathematical representation: S-matrix

Constraints
Bounds: vmax, v>0...
Balances : S*v =0



The FASTCORE family* principle



$$\max_{v,z} \sum_{i \in \mathcal{J}} z_i$$

$$\text{s.t.} \quad z_i \in [0, \varepsilon] \qquad \forall i \in \mathcal{J}$$

$$v_i \geq z_i \qquad \forall i \in \mathcal{J}$$

$$Sv = 0 \qquad v \in \mathcal{B}.$$

$$\min_{v,z} \sum_{i \in \mathcal{P}} z_i$$

$$\text{s.t.} \quad v_i \in [-z_i, z_i] \qquad \forall i \in \mathcal{P},\ z_i \in \mathbb{R}_+$$

$$v_i \geq \varepsilon \qquad \forall i \in \mathcal{K}$$

$$Sv = 0 \qquad v \in \mathcal{B}.$$

1) **Core reactions** are known to be **expressed** (red) in the context of interest.
2) Core reactions are forced to carry a non-zero flux by the **maximization of an approximation of the cardinality function**
3) The inclusion of non-core (blue) reactions is **minimized by a L1-regularization**

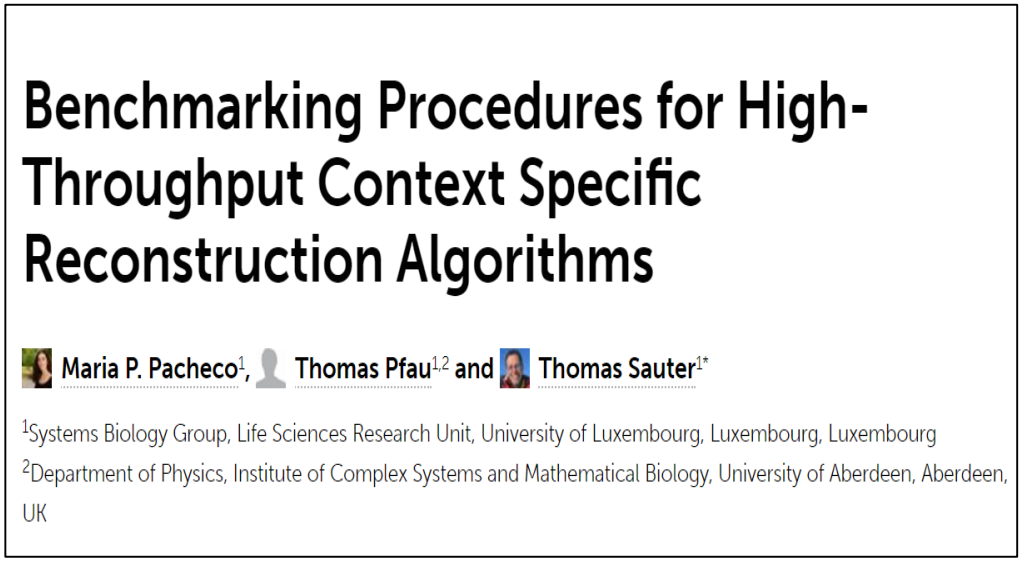
FASTCORE: Vlassis, Pires Pacheco & Sauter, 2014, Plos Comp Biol
FASTCORMICS: Pires Pacheco et *al*, 2015, BMC genomics
Benchmarking: Pires Pacheco *et al*, 2016, Front Physiol
FASTCORMICS RNAseq: Pires Pacheco et *al*, in preparation
In COBRA toolbox: Nature Protocols, accepted

Benchmarking of the FASTCORE family

Criteria:

- **computational demands**
- **accuracy**
- **Resolution power**

(ability to capture metabolic variations between two context i.e. cell types)



Benchmarking Procedures for High-Throughput Context Specific Reconstruction Algorithms

Maria P. Pacheco[1], Thomas Pfau[1,2] and Thomas Sauter[1*]

[1]Systems Biology Group, Life Sciences Research Unit, University of Luxembourg, Luxembourg, Luxembourg
[2]Department of Physics, Institute of Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen, UK

Data used:

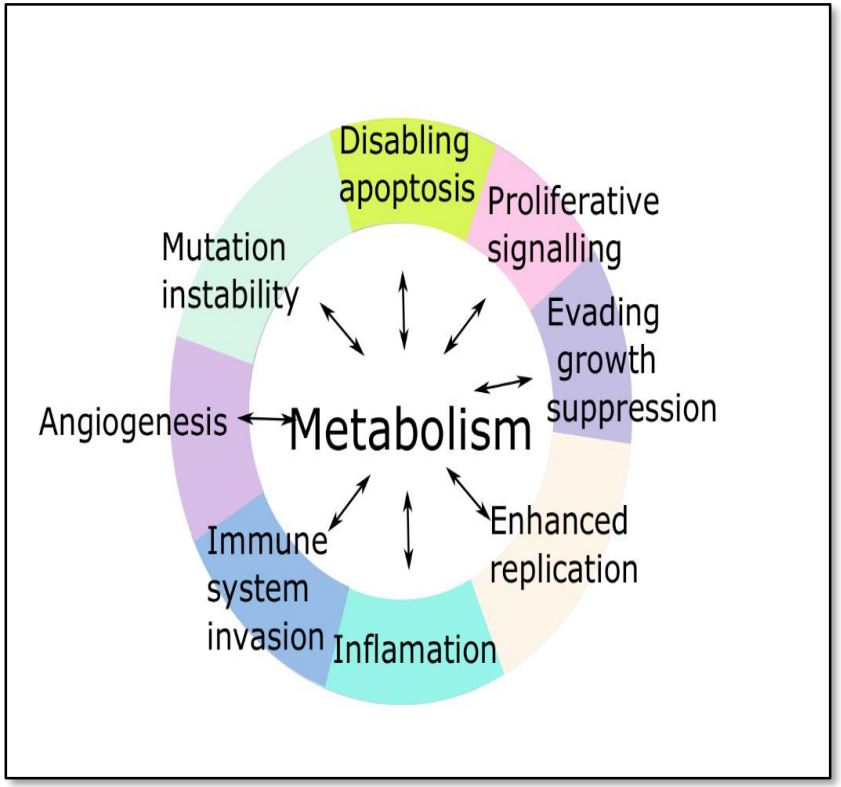
- **Artificial data**
- **Real data**

Strategies used:

- **Comparison to independent references (Human Protein atlas, HepatoNet, etc.)**
- **Cross-validations**
- **Functionalities**

Pires Pacheco, Pfau & Sauter, 2016, Front Physiol.

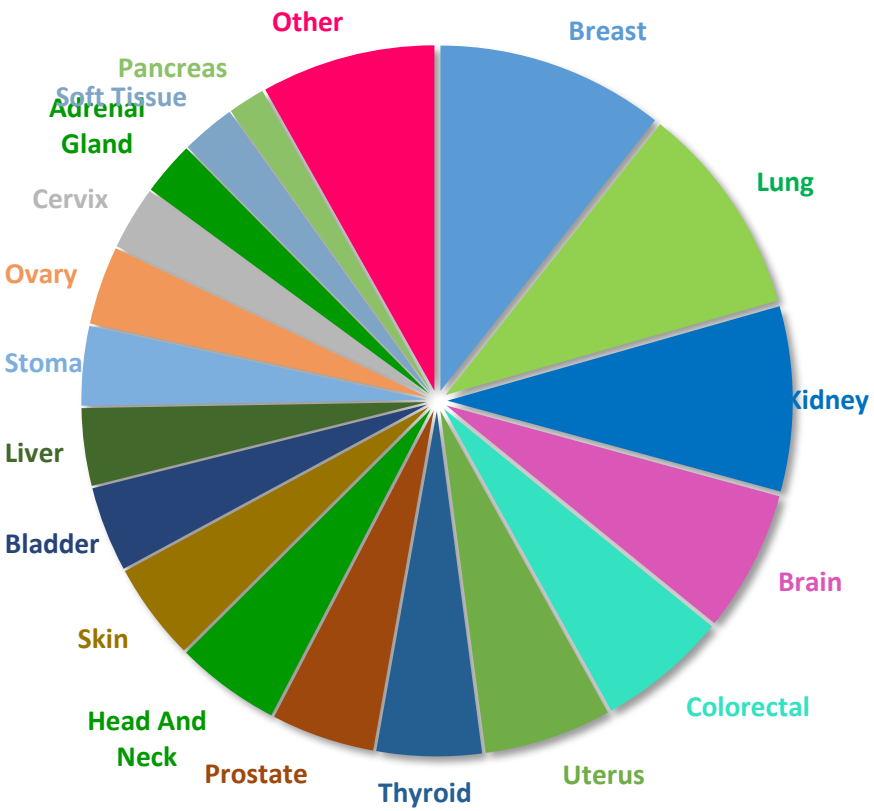
Cancer cells have an altered metabolism that can be targeted



- **Metabolic rewiring affects others hallmarks of cancer**
- **Specifically targeting cancer metabolism allows to reduce toxicity (on other cell types)**

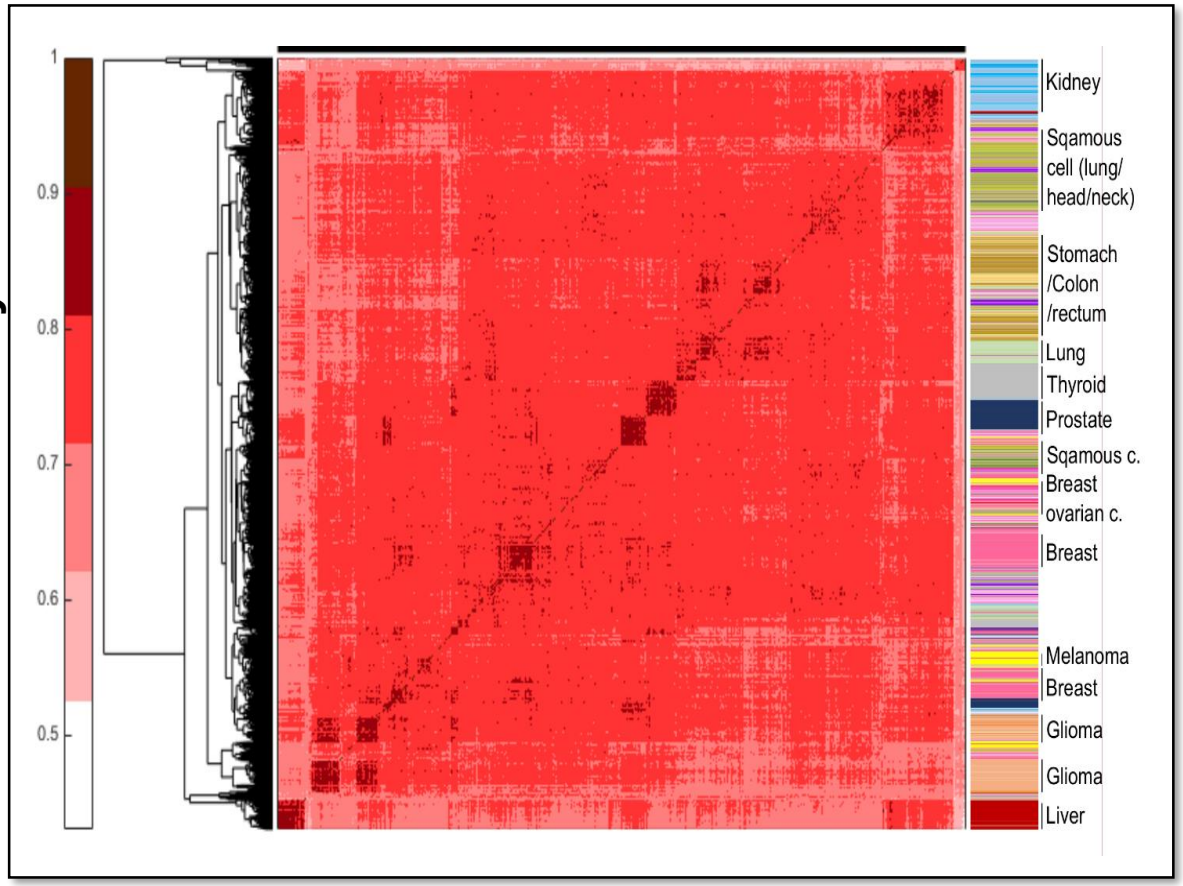
10005 models were built using the FASTCORMICS RNA-seq workflow

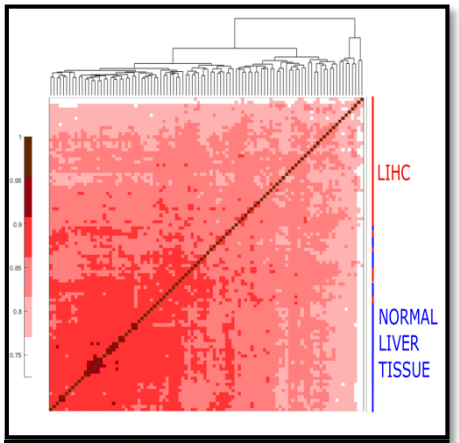
- **TCGA dataset contains 10005 RNA-seq samples across 24 cancer types**
- **All 10005 models were built in less than 5 days**

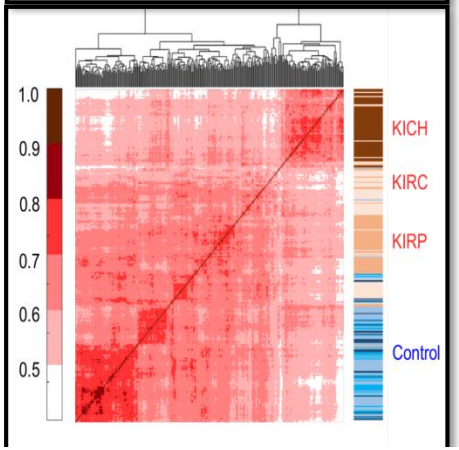


FASTCORMICS RNAseq: Pires Pacheco et *al*, in preparation

FASTCORMICS RNA-seq captures metabolic variation between different cancer types



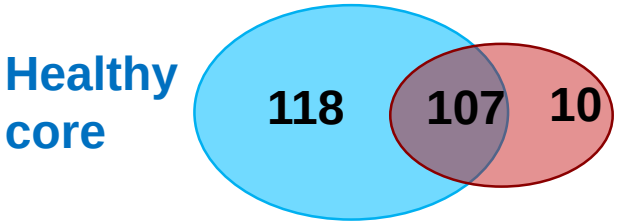




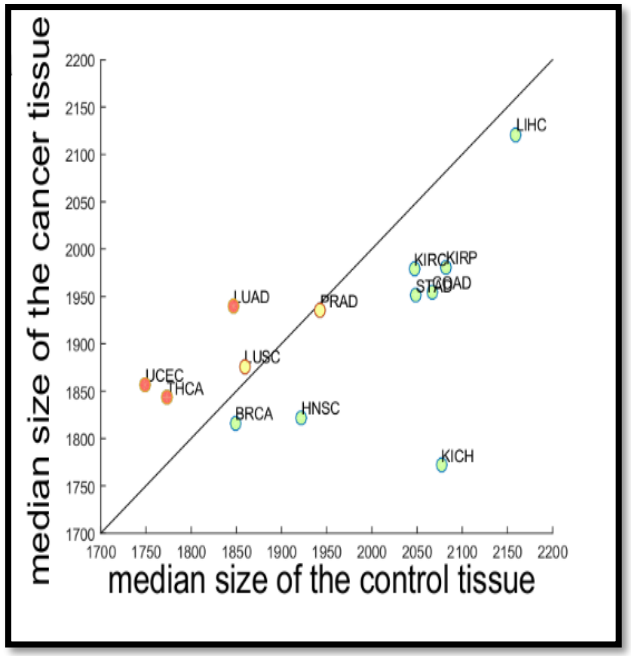
KICH=Kidney Chromophobe
KIRC= Kidney renal clear cell carcinoma
KIRP=Kidney renal papillary cell carcinoma

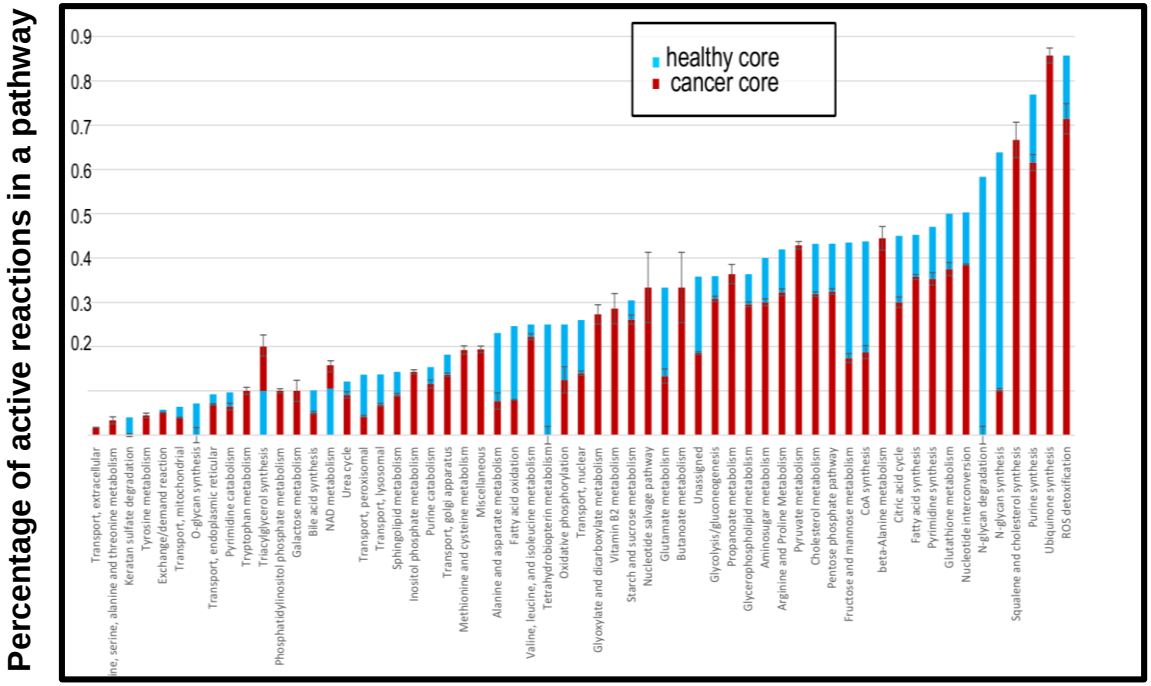
FASTCORMICS RNAseq: Pires Pacheco et *al*, in preparation

Cancer-cells have a more compact metabolism



- Cancer cells contain more essential genes
  = more potential drug targets



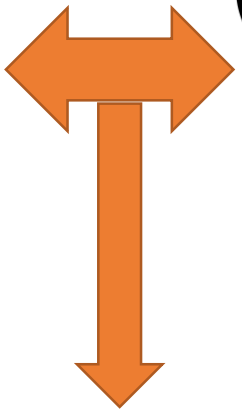


FASTCORMICS RNAseq: Pires Pacheco et *al*, in preparation

Application 2:

Identification of potential drug targets for drug repositioning

Gene essentiality analysis and datamining to find potential drug targets



DRUG Databases

- 17 drug targets
- 43 unique drugs for 17 targets

in silico gene essentiality analysis

- 107 essential genes

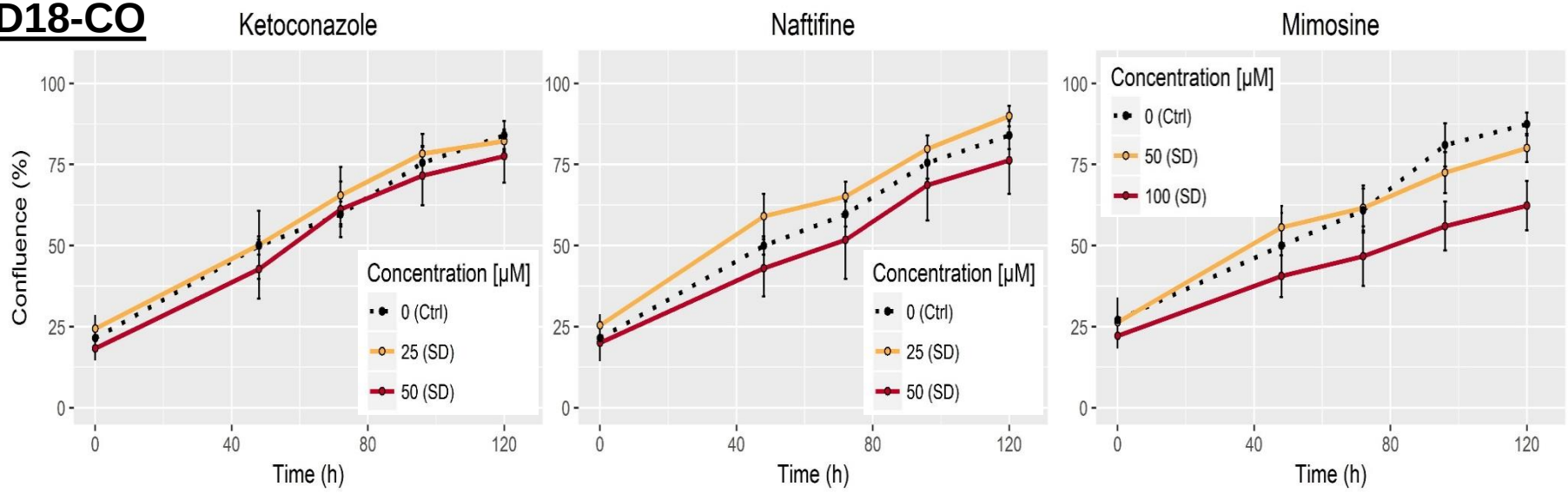
Selection of potential non-cancer drugs to treat colorectal cancer

- 17 drugs without anticancer reference

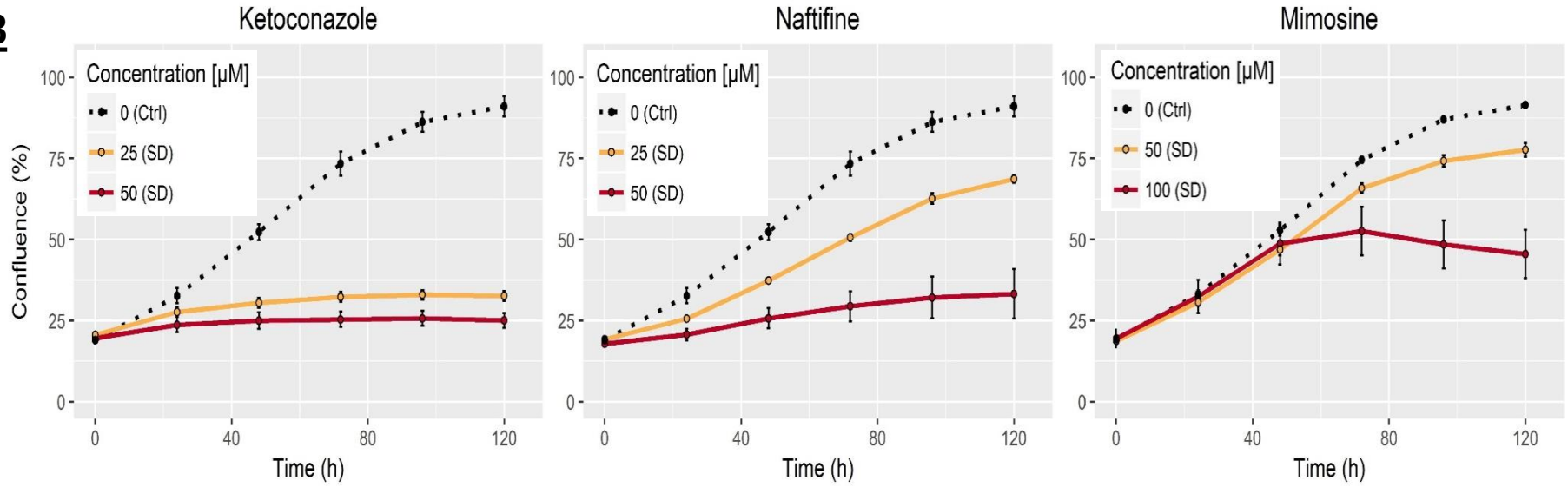
FASTCORMICS RNAseq: Pires Pacheco et *al*, in preparation

Healthy (CCD18-CO – colonic fibroblast) vs. Cancer (T18 – primary) cells

CCD18-CO



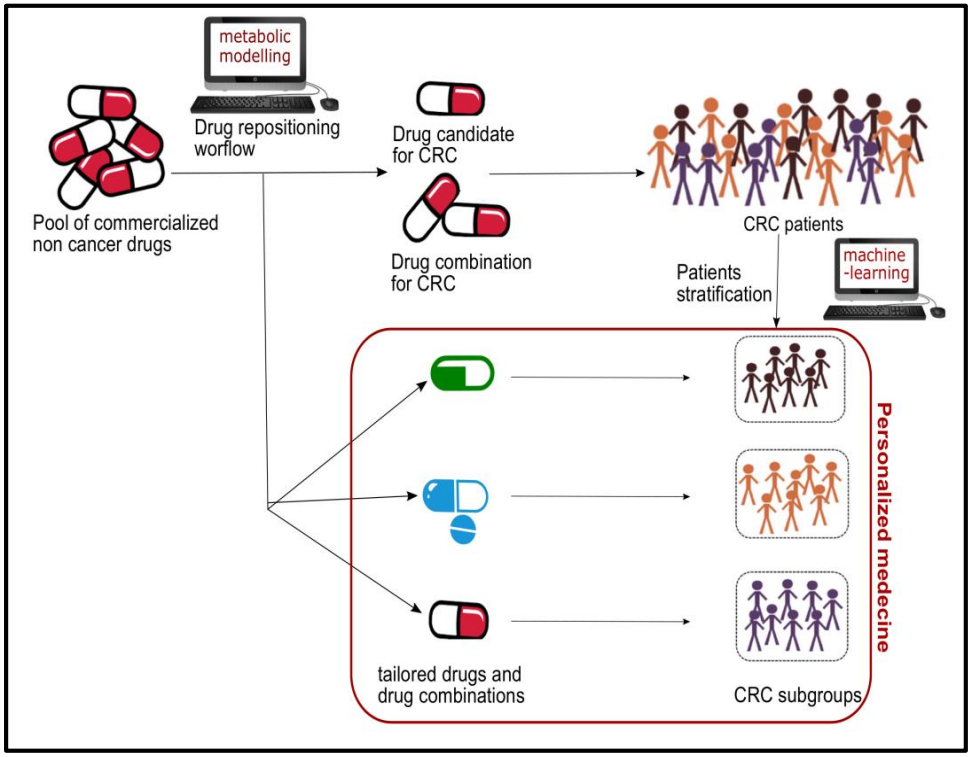
T18



- **24 hour time interval because of missing time points for CCD18-CO cells**
- **Graphs show representative experiment in techn. triplicates +/- SD, experiments were repeated 3x (CCD18) and 10x (T18)**

Outlook

- **Refine the workflow to capture metabolic variation between patients sub-groups and to predict tailored treatments**

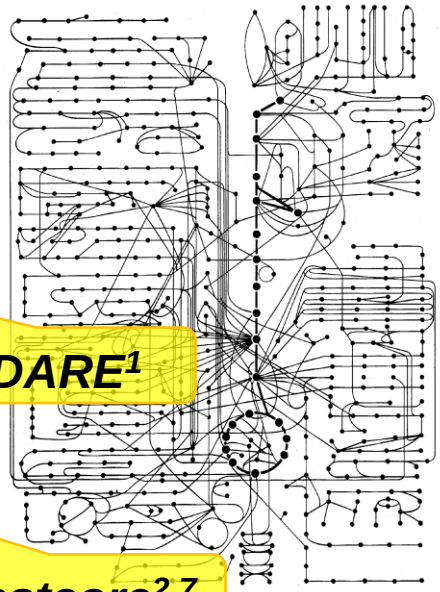


FASTCORE: Vlassis, Pires Pacheco & Sauter, 2014, Plos Comp Biol
FASTCORMICS: Pires Pacheco et *al*, 2015, BMC genomics
FASTCORMICS RNAseq: Pires Pacheco et *al*, in preparation
In COBRA toolbox: Nature Protocols, accepted

Computational model based data integration and network analysis

Data integration

- Omics data **integration** & Visualization
- Compact network **reconstruction**
- Network **curation** using (probabilistic) Boolean modelling



IDARE[1]

Fastcore[2,7]

optPBN[3,5]
FALCON[6]

Network analysis

- **Metabolic** networks
- **Signaling** networks
- **Sensitivity** Analysis / Early **warning**

DyGloSA[4]

1 Nucleic Acids Research 2013, 1;42(3):1474-96
2 PLoS Comput Biol. 2014, Jan;10(1)
3 Cell Commun Signal. 2013, Jul 1;11:46
4 PLoS One. 2013, Dec 18;8(12):e82973
5 PLoS One. 2014 Jul 1;9(7):e98001
6 Bioinformatics 2017 1;33(21):3431-3436
7 Nature Protoc., accepted

Systems Biology

FALCON: A Toolbox for the Fast Contextualization of Logical Networks

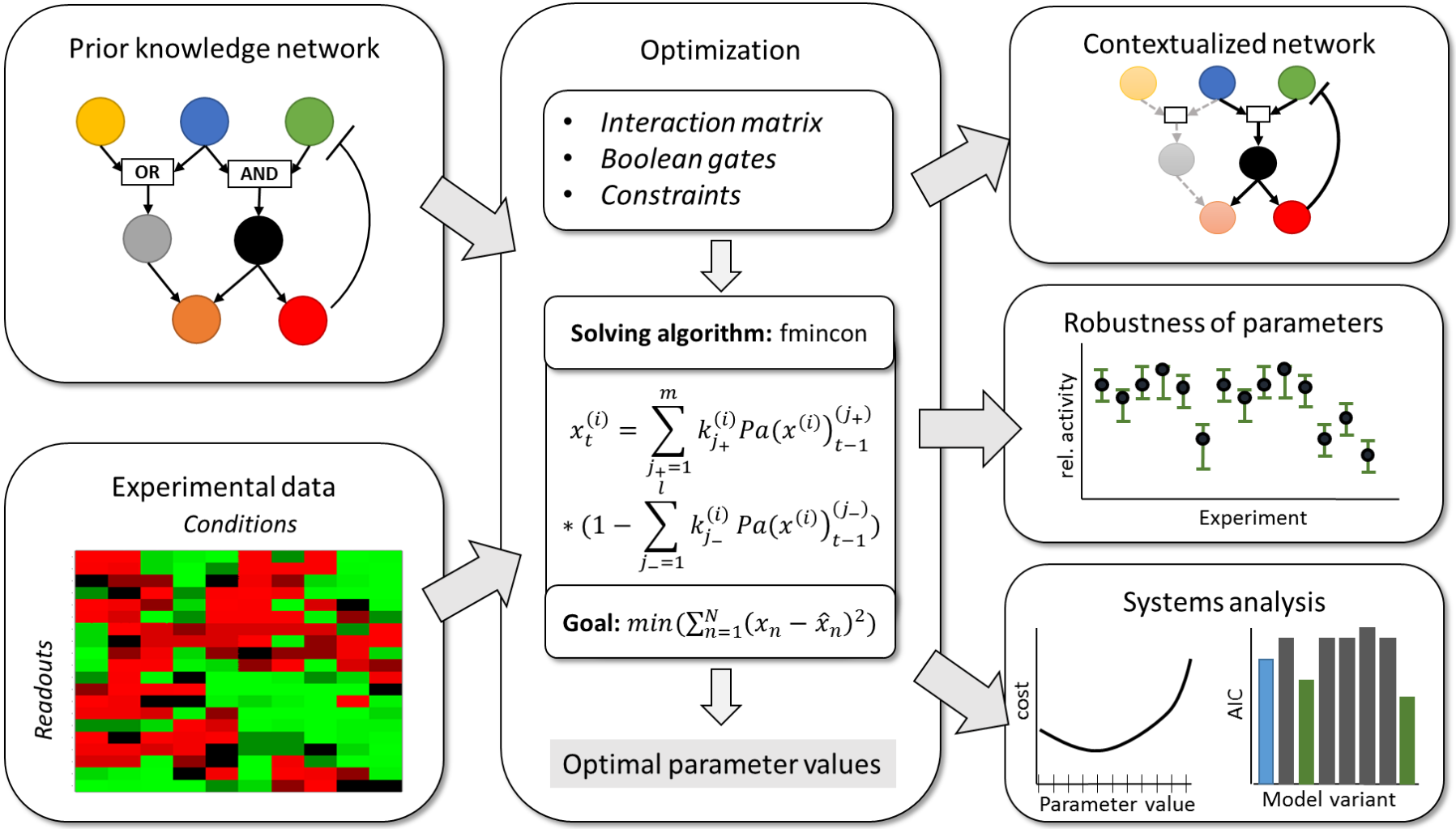
Sebastien De Landtsheer [1,†], Panuwat Trairatphisan [1,†], Philippe Lucarelli [1] and Thomas Sauter [1,*]

[1] Systems Biology Group, Life Sciences Research Unit, University of Luxembourg, Luxembourg, Luxembourg

GitHub

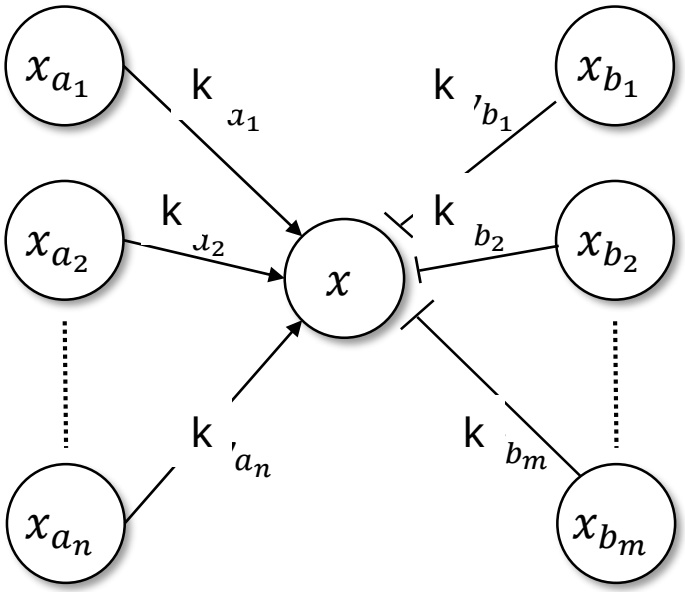
DeLandtsheer *et al.* (2017) FALCON: A Toolbox for the Fast Contextualisation of Logical Networks. Bioinformatics. 2017 Jun 29. doi: 10.1093/bioinformatics/btx380.

Pipeline Structure: Optimization



DeLandtsheer *et al.* (2017) FALCON: A Toolbox for the Fast Contextualisation of Logical Networks. Bioinformatics. 2017 Jun 29. doi: 10.1093/bioinformatics/btx380.

Concept



Biological equivalent	Graphical form	Algebraic computation
Activation	A → Z (k)	$Z_{t+1} = A_t * k$
Inhibition	A -| Z (k)	$Z_{t+1} = 1 - (A_t * k)$
Complex formation	A AND B → Z (k)	$Z_{t+1} = A_t * B_t * k$
Competitive interaction	A OR B → Z (k)	$Z_{t+1} = 1 - [(1-A_t) * (1-B_t) * k]$
Non-competitive interaction	A → Z (k_1) B → Z (k_2)	$Z_{t+1} = A_t * k_1 + B_t * k_2$ (with $k_1 + k_2 = 1$)

$$\boldsymbol{X_t^{(i)}} = \sum_{\boldsymbol{j_+=1}}^{\boldsymbol{m}} \boldsymbol{k_{j_+}^{(i)} Pa(X^{(i)})_{t-1}^{(j_+)}} * \left(\mathbf{1} - \sum_{\boldsymbol{j_-=1}}^{\boldsymbol{l}} \boldsymbol{k_{j_-}^{(i)} Pa(X^{(i)})_{t-1}^{j_-}} \right)$$

Sums of activating signals

$$\sum_{j_+=1}^{m} k_{j_+}^{(i)} = 1$$

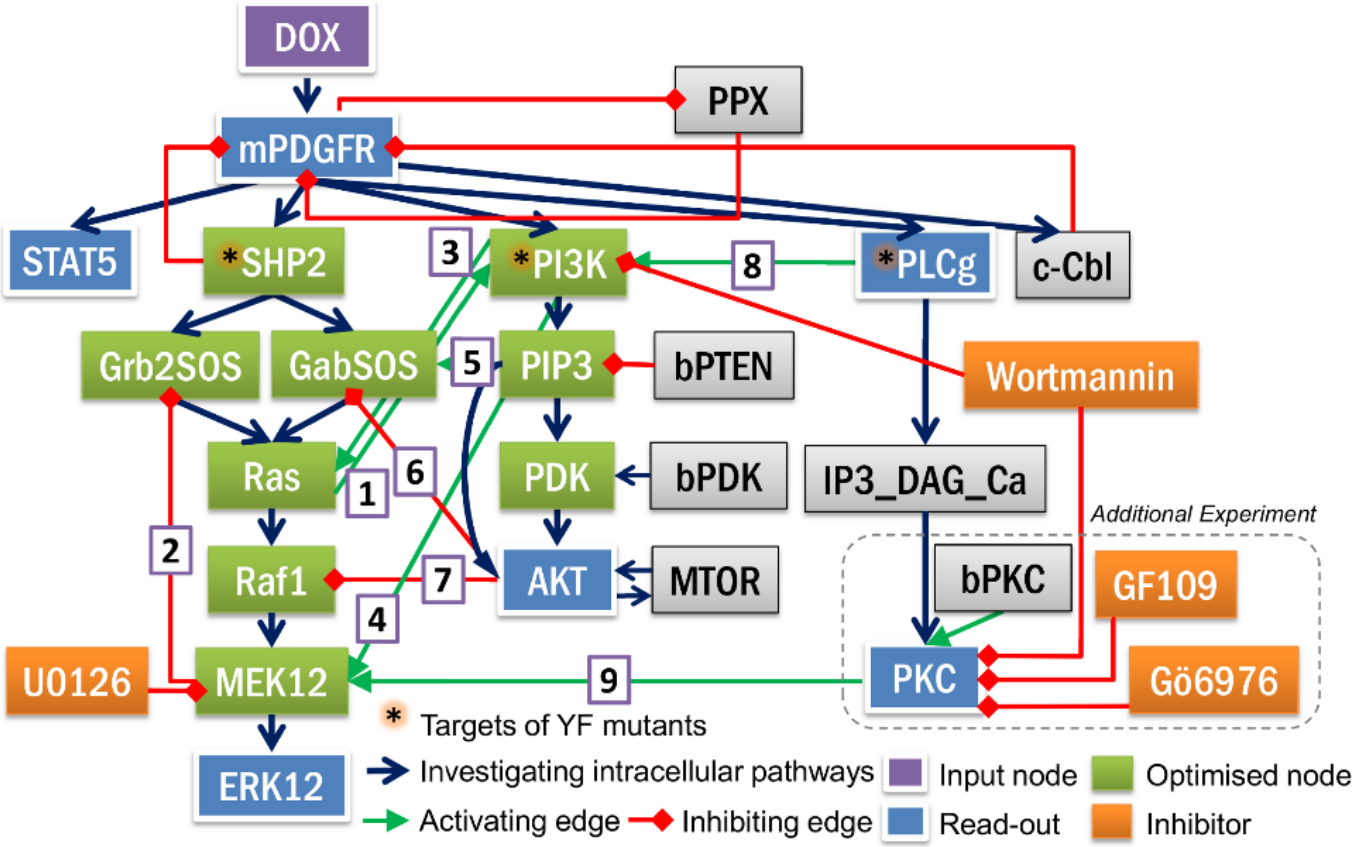
Sums of inhibiting signals

$$0 \leq \sum_{j_-=1}^{l} k_{j_-}^{(i)} \leq 1 .$$

DeLandtsheer *et al.* (2017) FALCON: A Toolbox for the Fast Contextualisation of Logical Networks. Bioinformatics. 2017 Jun 29. doi: 10.1093/bioinformatics/btx380.

FALCON: cancer specific PDGF signalling

With: P. Trairatphisan, M. Wiesinger, S. Haan, J. Pang, A. Mizera



Typical study:

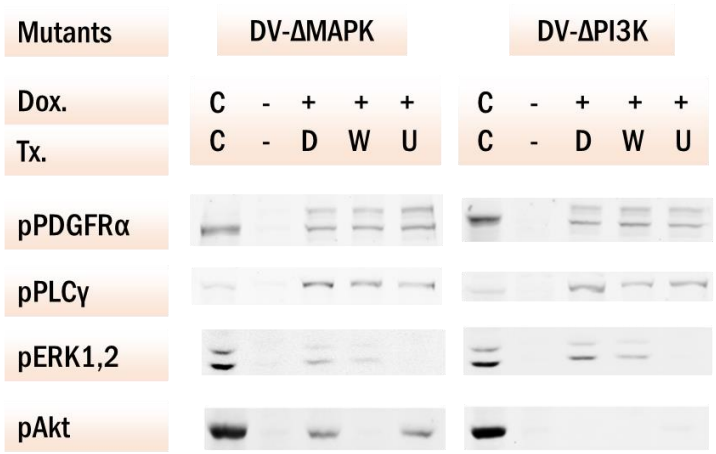
- **WB dataset**
- **Different cell lines**
- **Different drugs**
- **Putative cross-talks**

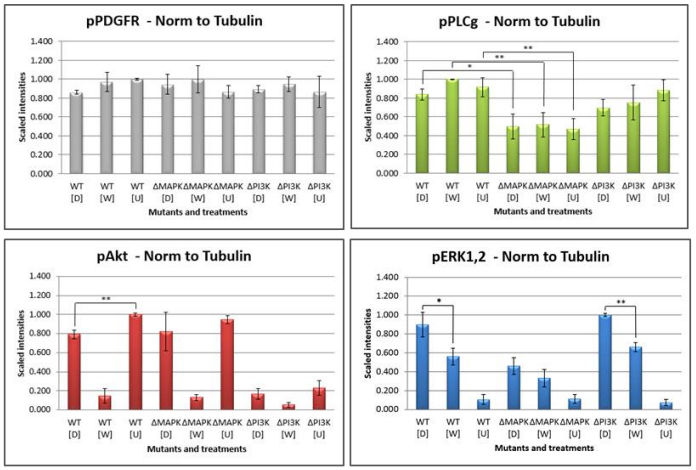
- **Training on single perturbations**
- **Predicting combined perturbations**

Trairatphisan et al, *PLoS One*, 2014

FALCON: cancer specific PDGF signalling

With: P. Trairatphisan, M. Wiesinger, S. Haan, J. Pang, A. Mizera





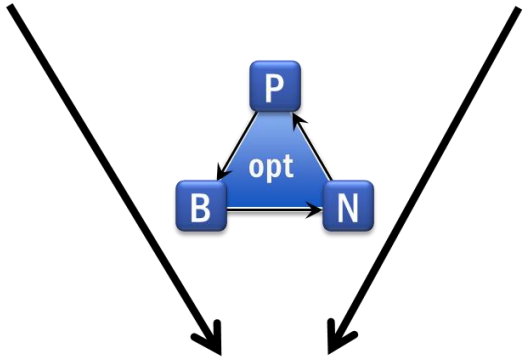
* PDGFRα mutants provided by Prof. Serge Haan

PDGFRα mutants* + inhibitors

Literature review**

Western Blots investigation

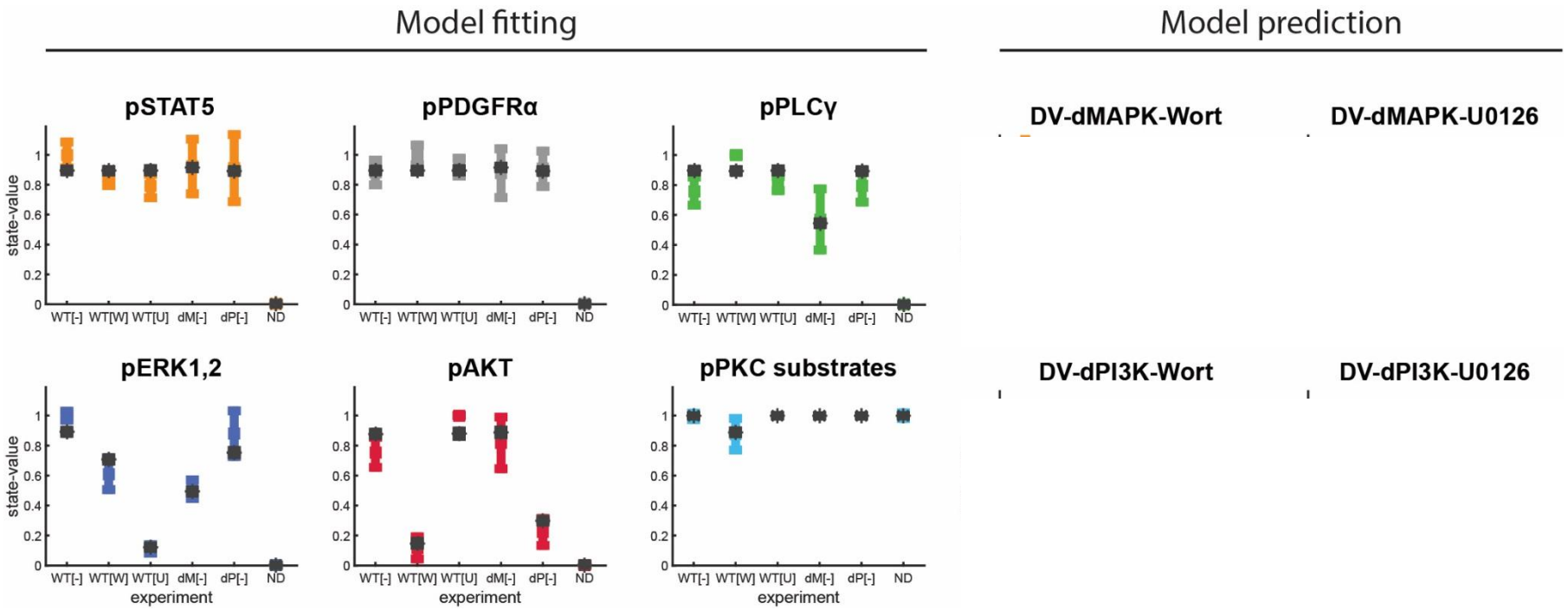
PDGF model topology



PDGF model in GIST context

FALCON: cancer specific PDGF signalling

With: P. Trairatphisan, M. Wiesinger, S. Haan, J. Pang, A. Mizera



Trairatphisan et al, *PLoS One*, 2014

Discussion & Conclusion FALCON

- **Our modelling framework is capable of contextualizing logical networks with phosphoproteomics data**
  - **Lower fitting costs**
  - **Very fast**

- **The low computational costs allow for:**
  - **Larger networks and big data**
  - **Systems Analysis:**
    - **Parameter localization**
    - **Identifiability**
    - **KO simulations**
    - **Cell line comparison**

Acknowledgements

L. Sinkkonen, P. Nguyen, T. Pfau, M. Galhardo, Y. Gui, P. Lucarelli,
P. Trairatphisan, M. Pires Pacheco, D. Gerard, S. Nickels, P. Martins Conde,
S. De Landtsheer, M. Albrecht, M. Wiesinger, T. Bintener, J. Becker, M. Gavriil, J. Ohnmacht

U Luxembourg: E. Lettelier, S. Haan, J.-L. Bueb, A. Mizera, J. Pang, E. Reckinger-Schaffner
LIH Luxembourg: P. Nazarov, L. Vallar, F. He, F. Azuaje
LCSB Luxembourg: N. Vlassis, J. Schwamborn, J. Schneider, R. Halder, R. Fleming, A. del Sol

U Dresden: D. Kulms
MPI Saarbrücken: M. Schulz
U Kuopio: M. Heinäniemi, M. Liivrand
U Lorraine: T. Bastogne, S. Dobra
Institute Curie Paris: A. Zinovyev
CNRS Paris: L. Pauleve
King's College London: M. Tavassoli
U Bordeaux: G. Sciume
U Stuttgart: J. Witt, S. Barisic, R. Schlatter, S. Henkel, R. Feuer, M. Ederer, O. Sawodny
U Heidelberg / Synlab Holding GmbH: I. Gergei, W. März

Merrimack Pharmaceuticals: (B. Schöberl), A. Raue
Bayer Technology Services: T. Eissing, L. Küpfer
optimata Ltd: Y. Kogan, Z. Agur
ProtATonce: L. Alexopoulos

contact: thomas.sauter@uni.lu