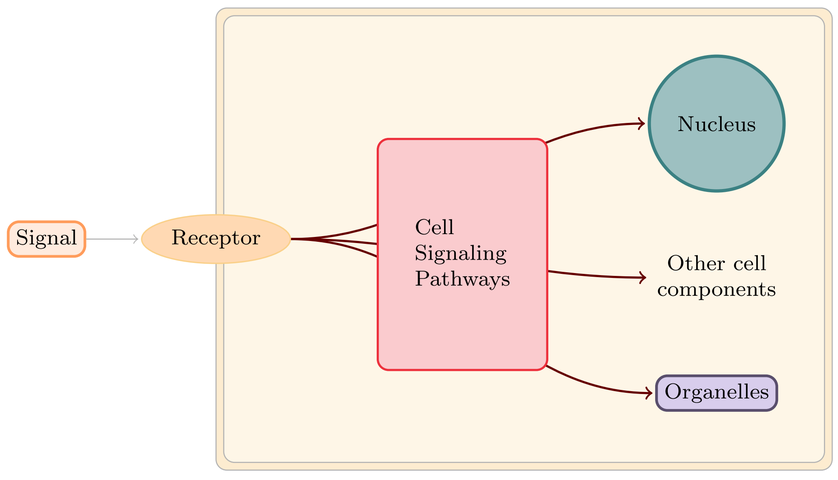
Identification of cell signaling pathways based on biochemical reaction kinetics repositories

Gustavo Estrela de Matos[a, b, c]
Hugo Armelin[b, c], Marcelo S. Reis[b, c]

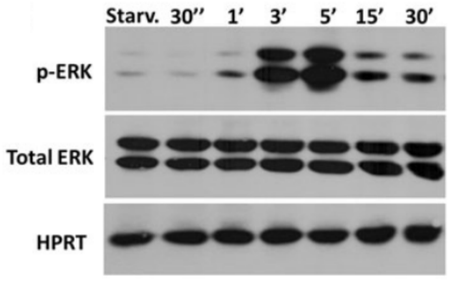
[a]Institute of Mathematics and Statistics, University of São Paulo, Brazil
[b]Center of Toxins, Immune-response and Cell Signaling (CeTICS), Instituto Butantan, Brazil
[c]Special Laboratory of Cell Cycle, Instituto Butantan, Brazil

Cell Signaling Pathways

Cell Signaling is a mechanism that allows the cell to change its behaviour according to the environment.



A signal flows in a cell thorugh a cell signaling pathway, which can be characterized by a sequence of chemical reactions.



We can summarize the state of a cell signaling pathway by measuring the concentration over time of some chemical species that are present on the pathway.

Identification of Signaling Pathways

What is the structure of a cell signaling pathway, given a set of concentration measurements? We answer this question with a computational model, created from a set of chemical reactions, that can reproduce the concentration dynamics observed experimentally. These models are created using the laws of biochemical kinetics, deriving a system of ordinary differential equations.

As an example, we can model the following equation:

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C} \implies \frac{d[C]}{dt} = k[A][B]$$

Where k is a reaction rate constant.

However, to derive the model, we still need to determine what is the set of chemical reactions of the signaling pathway.

Feature Selection

We proposed to solve the identification of cell signaling pathways as a feature selection problem. This problem consists of: given a set of features (reactions) and a score for each subset (the quality of a model), what is the best subset?

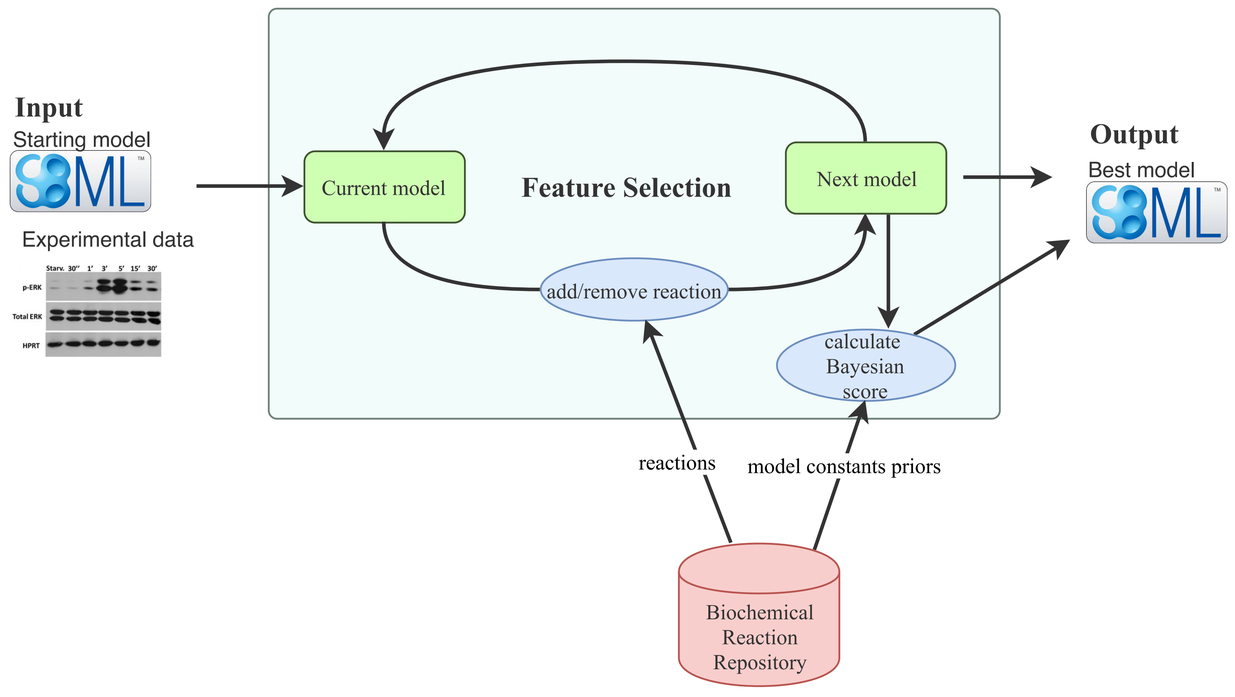
Bayesian Ranking of Models

To determine the quality of a model we implemented a score function that is a estimative of $p(D|M)$, which is the likelihood of observing experiment D under the assumption that model M is correct. To create this estimative we generate samples of the posterior distribution of model parameters (θ).

$$\underbrace{p(\theta|M,D)}_{\text{posterior}} \propto \underbrace{p(D|M,\theta)}_{\text{likelihood}} \underbrace{p(\theta|M)}_{\text{prior}}$$

This ranking score was implemented as a Python package called SigNetMS. It is an open source software and it is available on GitHub.

The Proposed Methodology



Experiments

Conclusão

Acknowledgement

FAPESP CNPq CAPES

{marcelo.reis, gustavo.matos }@butantan.gov.br