# Analyzing Multi-Dimensional Biological Model

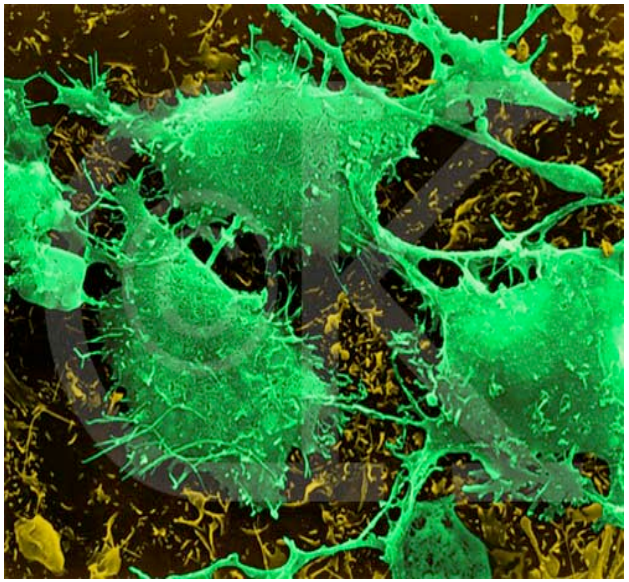
Matthieu Pichené

# Biological problem

- Design efficient cancerous tumor treatments.

- Efficient protocol = Optimize drug quantity :
  - frequency of treatment
  - choice of concentration

- Testing many treatments *in vivo* is long/ costly.



Goal : Propose *in silico* method to sort candidate protocols

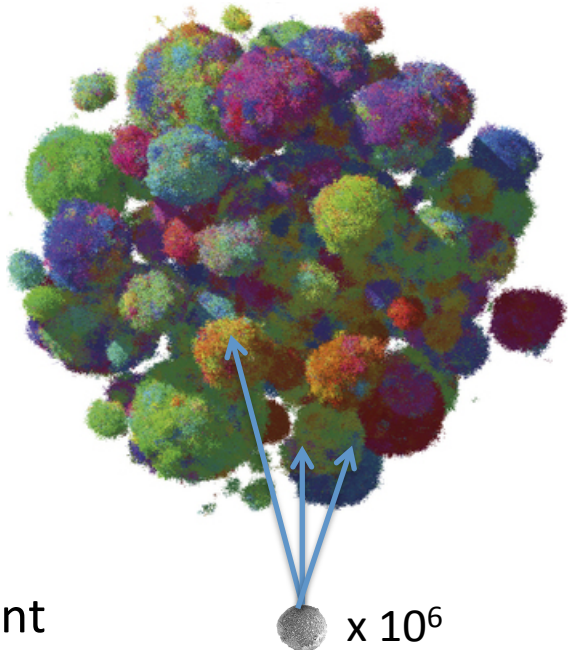
Study case : HeLa cells (cervical cancer). TRAIL protein triggering the apoptosis (programmed cell death) process.

# Challenge

Modeling treatment of non-vascularized tumor:

- Tumor up to $10^6$ cells.
- Survivors often possess a temporary resistance to treatment (depend on proteic concentrations)

=> decrease efficiency of repeated TRAIL treatment




Consider two scales:

- Tissue : Tumor evolution, treatment diffusion
- Cell : Effect of the treatment, Transient treatment resistance

Issue: High complexity model (combinatory explosion) => Abstractions

# Overview

# Tissular level: Abstraction

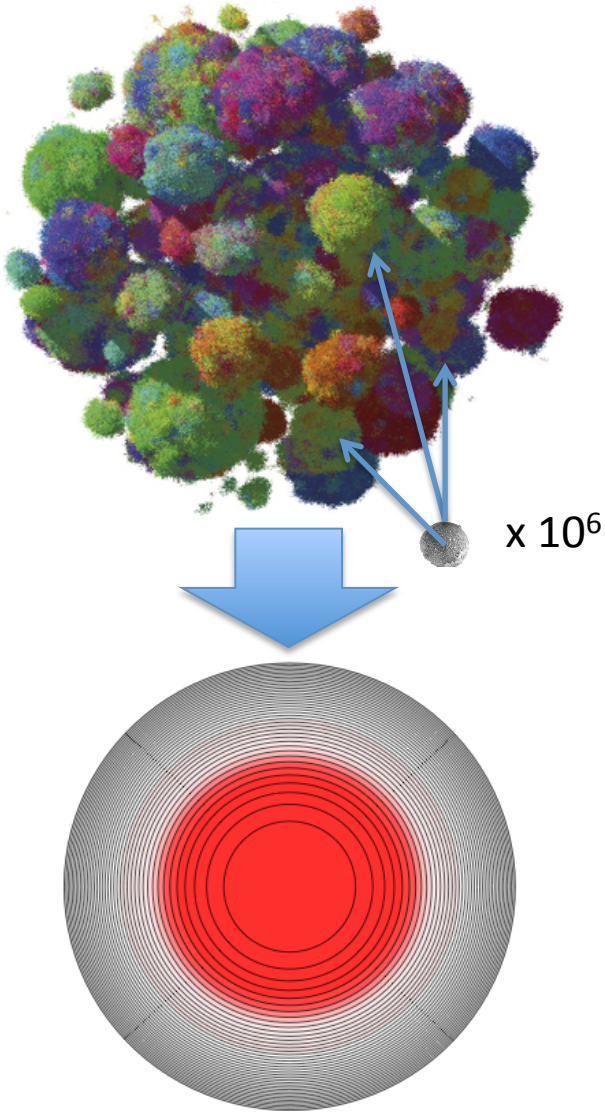
- Not counting all cells individually

  Populations instead of agents (=cells)

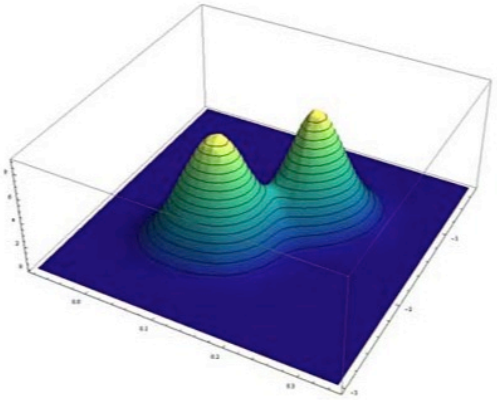
- Different conditions at different depths:
  Separate in several subpopulations (layers).

- To handle subpopulations resistance:

  use distribution of proteic concentration over the subpopulation



# Handling the distribution



How to represent the distribution of many protein concentrations ?

Idea: Discretize concentrations + Approximation.

How to evaluate the behavior of a cell from a distribution of discretized concentrations ?

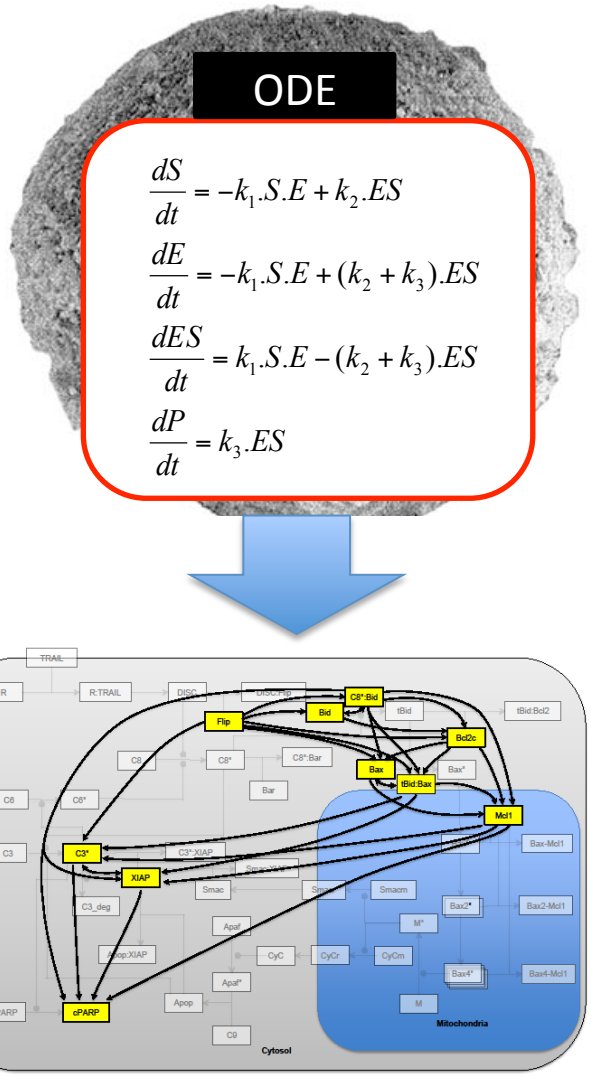
Low level model with continuous variables (e.g. ODEs)
ODEs behavior from a discrete configuration ?
⇒ One way: Thousands of simulations and average out
Quite inefficient

# Cellular level

Thousands of ODEs simulations and average out
Precise but time consuming

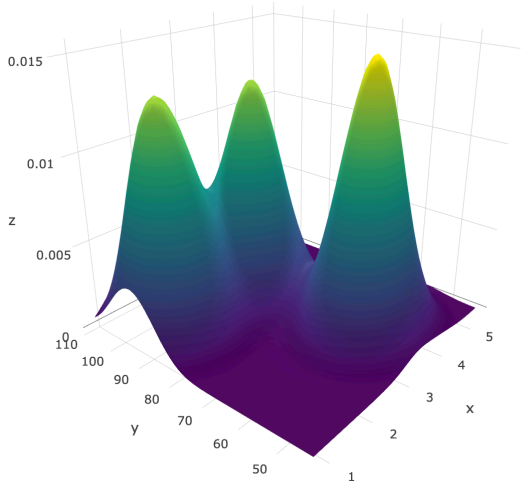
Abstraction:
⇒ Stochastic discrete abstraction
starts from a discrete configuration

More time efficient model:
1 simulation represents thousands of ODE simulations

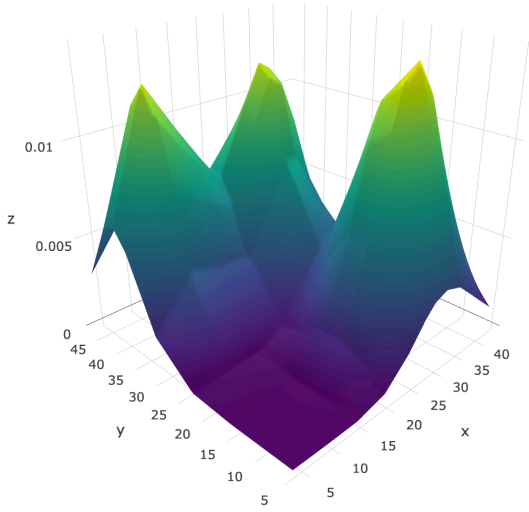


# Contributions

# Representing Distributions

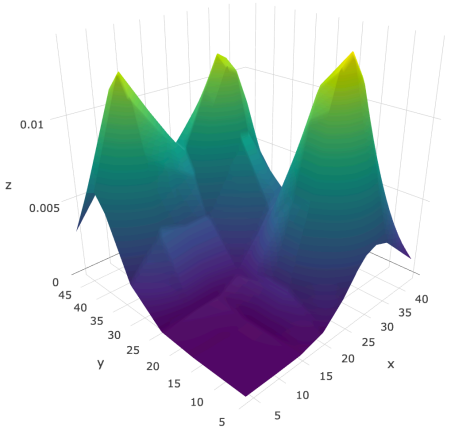


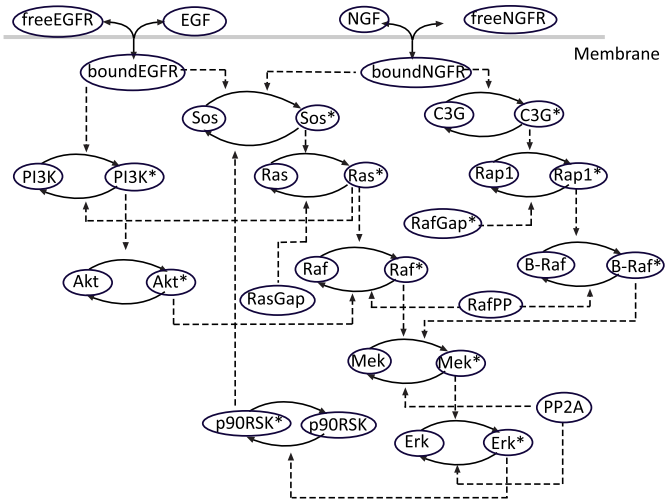

Discretize concentrations + Approximation.




Matthieu Pichené, Sucheendra Palaniappan,Eric Fabre, Blaise Genest. Non-Disjoint Clustered Representation for Distributions over a Population of Cells. *Submitted*.

# Distribution Representation




n = 2

Represent distribution explicitly:
P(X1=x1,X2=x2) for all x1,x2

n = 36

Distribution representation ?

Exact representation : $S^n$ values

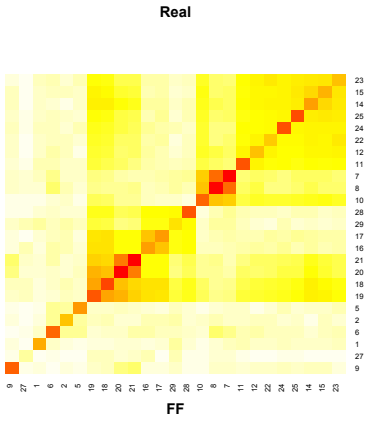
S = number of discrete values, n = number of species

# Distribution Representation

P(X1=x1,…,Xn=xn)
$S^n$ values

Explicit representation
Can't be used realistically.
Need for approximation:

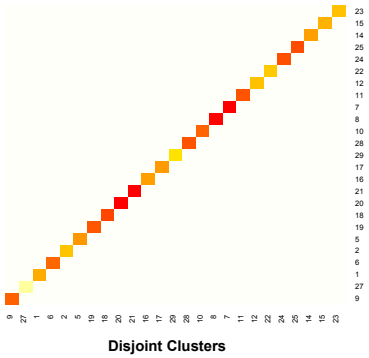
Mutual Information




P(X1=1), …
P(X1=S), …
P(Xn=S)
n S values

Naïve option: consider species as independent variables (fully factored).

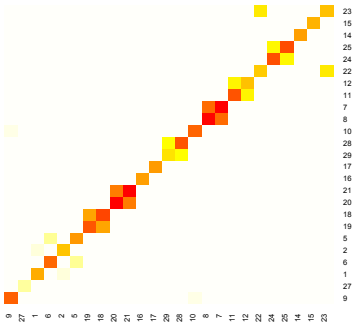
$$P_{FF}(X_1 = x_1, ..., X_n = x_n) = \prod_i P(X_i = x_i)$$




P(X1=1,X2=1),
P(X1=S,X2=1)…
P(Xn=S,Xk=S)
c $S^d$ values

More precise: Disjoints clusters $K_1$.. $K_c$.

$$P_C(X_1 = x_1, \cdots, X_n = x_n) = \prod_{j \leq c} P(X_i = x_i, i \in K_j)$$

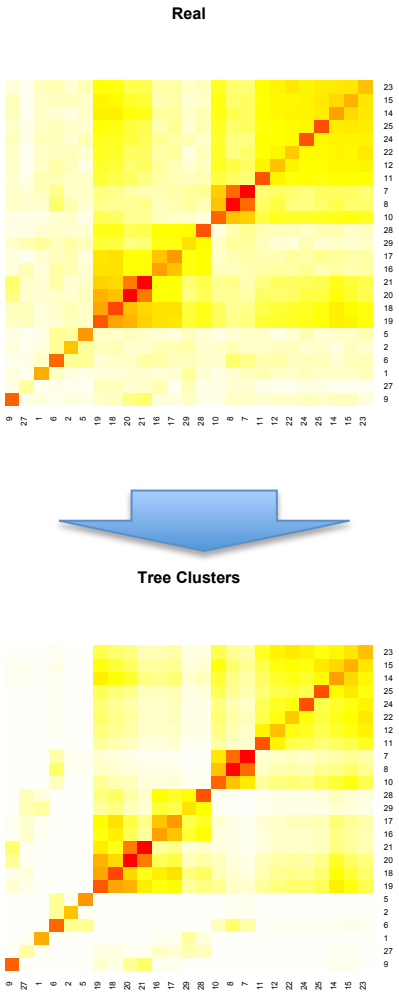


# Distribution Representation

Idea: non disjoint clusters

P(X1=1,X2=1),
P(X1=S,X2=1)…
P(Xn=S,Xk=S)
c $S^d$ values

$$P_{NDC}(X_1 = x_1, \cdots, X_n = x_n) = \prod_{j \leq c} \frac{P(X_i = x_i, i \in K_j)}{P(X_i = x_i, i \in \bigcup_{\ell<j} K_\ell \cap K_j)}$$

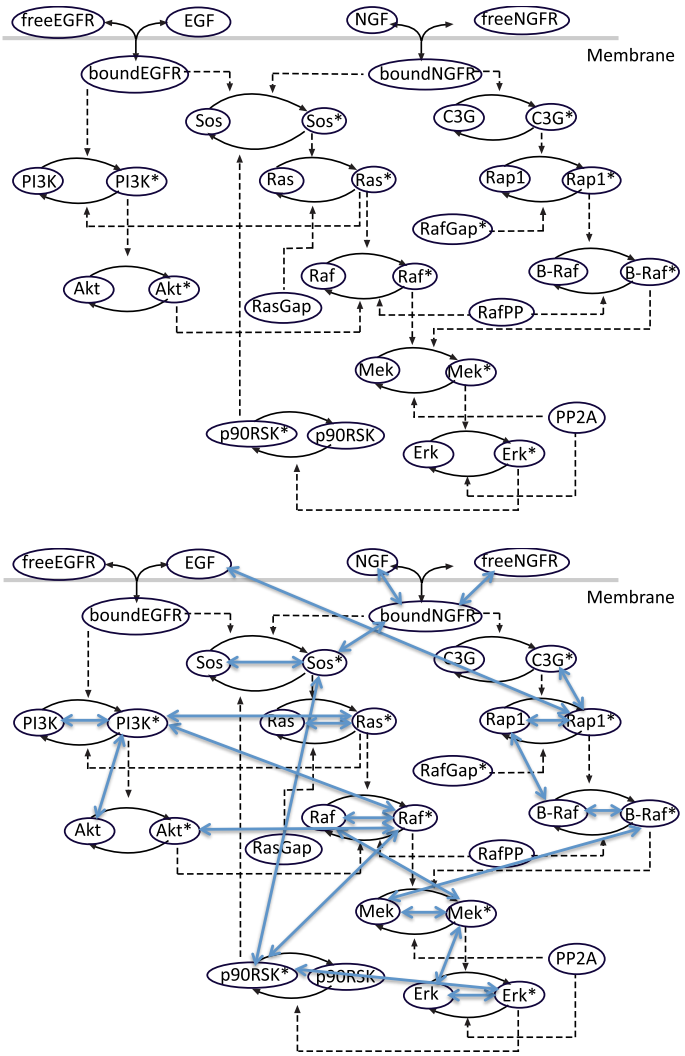



Correlations are quite preserved

# Distribution Representation

How to choose optimal clusters ?

Use of the [Chow-Liu 1968] algorithm in polynomial time to find optimal clusters making a tree / trees



# Experimental results

Enzyme catalytic reaction, probability distribution at 2 minutes:

| Representation | Mean MI | max MI Error | Max P error | KL diverg. |
|---|---|---|---|---|
| FF | 0.22 | 0.27 | 0.22 | 0.31 |
| Disjoint Cluster | 0.26 | 0.11 | 0.05 | 0.12 |
| Tree Cluster | 0.277 | 0.04 | 0.005 | 0.001 |
| Exact | 0.278 | 0 | 0 | 0 |

Apoptosis pathway, probability distribution at 105 minutes:

| Representation | mean MI | max MI Error | Size of representation |
|---|---|---|---|
| FF | 0.06 | 0.32 | 50 |
| Tree Cluster | 0.1 | 0.12 | 225 |
| Exact | 0.12 | 0 | $10^7$ |

EGF-NGF pathway, probability distribution at 5 minutes:

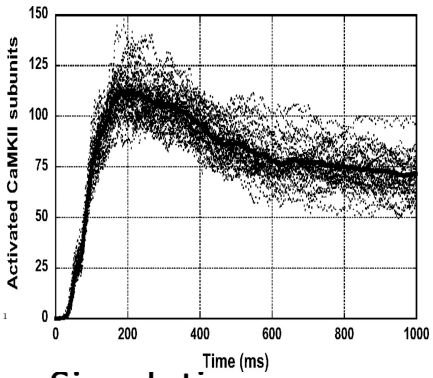
| Representation | mean MI | max MI Error | Size of representation |
|---|---|---|---|
| FF | 0.016 | 0.6 | 160 |
| Disjoint Cluster | 0.019 | 0.2 | 775 |
| Tree Cluster | 0.023 | 0.07 | 775 |
| Exact | 0.026 | 0 | $10^{22}$ |

# Cellular Level

Sucheendra Palaniappan, François Bertaux, Matthieu Pichené, Eric Fabre, Gregory Batt, Blaise Genest. Discrete Stochastic Abstraction of Biological Pathway Dynamics: A case study of the Apoptosis Pathway. *Bioinformatics*, Oxford University Press, to appear.

# Cellular level

$$\frac{dc_1}{dt} = -k_1.c_1.c_2 + k_2.c_3$$
$$\frac{dc_2}{dt} = -k_1.c_1.c_2 + k_2.c_3 + k_{17}.c_{18} + k_{11}.c_{11}$$
$$\frac{dc_3}{dt} = k_1.c_1.c_2 - k_2.c_3 - k_3.c_3.c_4 + k_4c_5$$
$$\frac{dc_4}{dt} = -k_3.c_3.c_4 + k_4c_5 + k_{11}.c_{11} + k_{20}.c_{21}$$
$$\frac{dc_5}{dt} = k_3.c_3.c_4 - k_4.c_5 - k_5.c_5.c_6 + k_6.c_7$$
$$\frac{dc_6}{dt} = -k_5.c_5.c_6 + k_6.c_7 + k_{11}.c_{11} + k_{20}.c_{21}$$
$$\frac{dc_7}{dt} = k_5.c_5.c_6 - k_6.c_7 - k_7.c_7.c_8 + k_8.c_9$$
$$\frac{dc_8}{dt} = -k_7.c_7.c_8 + k_8.c_9 + k_{11}.c_{11} + k_{20}.c_{21}$$
$$\frac{dc_9}{dt} = k_7.c_7.c_8 - k_8.c_9 - k_9.c_9.c_{10} + k_{10}.c_{11} - k_{15}.c_9.c_{17} + k_{16}.c_{18}$$




HSD + Simulations

Mathematical model of apoptosis :

HSD=ODE + Stochastic [Bertaux et al. 2014]

Simulating HSD from discrete configuration:

- Lots of simulations

Stochastic discrete abstraction

Bertaux, F., Stoma, S., Drasdo, D., and Batt, G. (2014). Modeling Dynamics of Cell-to-Cell Variability in TRAIL-Induced Apoptosis Explains Fractional Killing and Predicts Reversible Resistance. PLoS Comput Biol ,10(10), 14

# Cellular level : Abstraction

Stochastic discrete abstraction

Compact Markov chain (CMC):

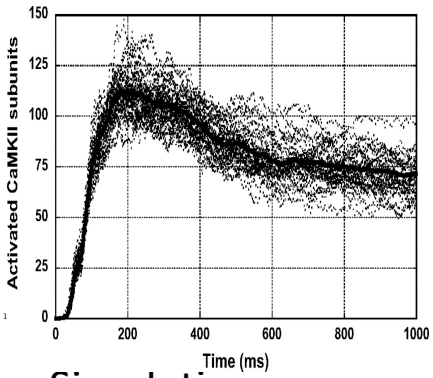
- Discretized concentrations
- Stochastic transitions
- Variables reduced to interest proteins
- Less time steps required (min vs. sec)

Time efficient:

1 simu CMC 20x faster than 1 simu HSD

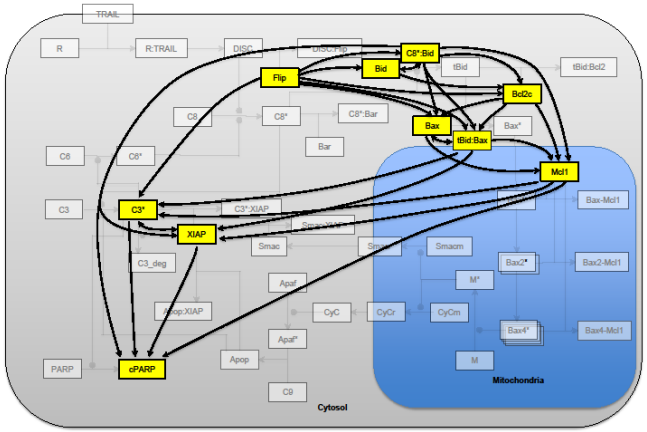
1 simulation of CMC = many simulations of HSD

$$\frac{dc_1}{dt} = -k_1.c_1.c_2 + k_2.c_3$$
$$\frac{dc_2}{dt} = -k_1.c_1.c_2 + k_2.c_3 + k_{17}.c_{18} + k_{11}.c_{11}$$
$$\frac{dc_3}{dt} = k_1.c_1.c_2 - k_2.c_3 - k_3.c_3.c_4 + k_4c_5$$
$$\frac{dc_4}{dt} = -k_3.c_3.c_4 + k_4c_5 + k_{11}.c_{11} + k_{20}.c_{21}$$
$$\frac{dc_5}{dt} = k_3.c_3.c_4 - k_4.c_5 - k_5.c_5.c_6 + k_6.c_7$$
$$\frac{dc_6}{dt} = -k_5.c_5.c_6 + k_6.c_7 + k_{11}.c_{11} + k_{20}.c_{21}$$
$$\frac{dc_7}{dt} = k_5.c_5.c_6 - k_6.c_7 - k_7.c_7.c_8 + k_8.c_9$$
$$\frac{dc_8}{dt} = -k_7.c_7.c_8 + k_8.c_9 + k_{11}.c_{11} + k_{20}.c_{21}$$
$$\frac{dc_9}{dt} = k_7.c_7.c_8 - k_8.c_9 - k_9.c_9.c_{10} + k_{10}.c_{11} - k_{15}.c_9.c_{17} + k_{16}.c_{18}$$




HSD + Simulations

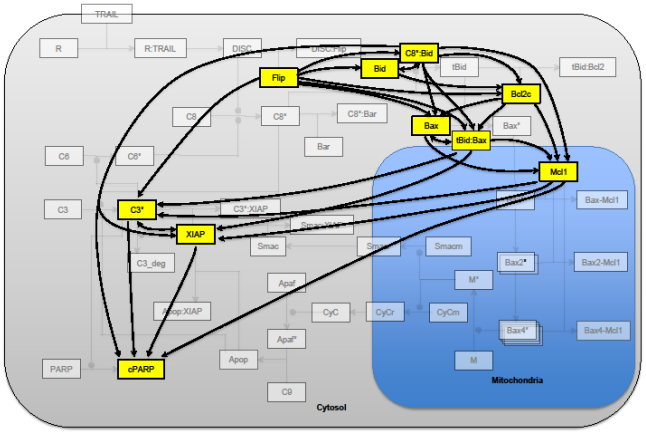
DBNizer

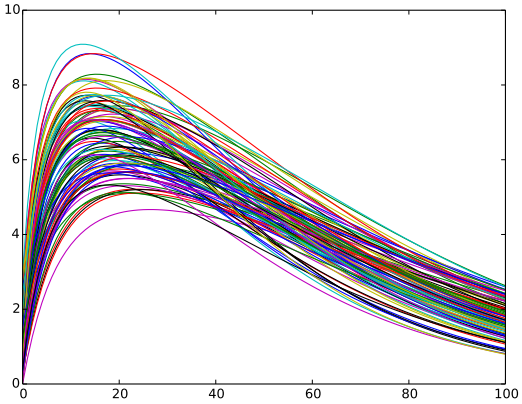



| Model | cell death (**HSD: 69.9%**) | discerning power (**HSD: 100%**) | Time / 1000 simulations (**HSD: 56s**) |
|---|---|---|---|
| $MIDBN_7$ | 70.43% | 96.14% | **2.13s (26.3X)** |
| $MIDBN_8$ | **69.57%** | 96.31% | 2.64s (21.21X) |
| $MIDBN_9$ | 69.33% | 96.37% | 2.98s (18.8X) |
| $MIDBN_{10}$ | 69.03% | **96.84%** | 3.30s (17X) |
| $MIDBN_{58}$ | 66.85% | 94.12% | 73.05s |
| $RNDBN$ | 92.29% | 85.53% | 299s |

# Simulations vs Inferences

To obtain the probability distribution produced by the CMC




Lots of simulations

$$P^t(\boldsymbol{X}=\boldsymbol{x}) = \sum_{\boldsymbol{u}\in V^X} P^{t-1}(\boldsymbol{X}=\boldsymbol{u}) \prod_{i=1}^{n} CPT_{t,i}(\boldsymbol{x}_i \mid \boldsymbol{u}_{\hat{i}})$$

Inference (1 computation)

Can't represent explicitely P(**X=u**) => approximation


Matthieu Pichené, Sucheendra Palaniappan,Eric Fabre, Blaise Genest.
Non-Disjoint Clustered Representation for Distributions over a Population of Cells. *Submitted*.

# Inference

Test of different approximate distributions for inference in compact Markov chains.

Program : ClusterAlgo (based on different distribution approximations)

Enzyme catalytic reaction:

| Method | Max. Error | Mean Error (normalized) | Nb. Error > 0.1 | Comput. Time |
|---|---|---|---|---|
| FF | 0.17 | 100 | 49 | 0.2s |
| Disj. Cluster | 0.12 | 65 | 16 | 0.5s |
| Tree Cluster | 0.004 | 3 | 0 | 0.6s |

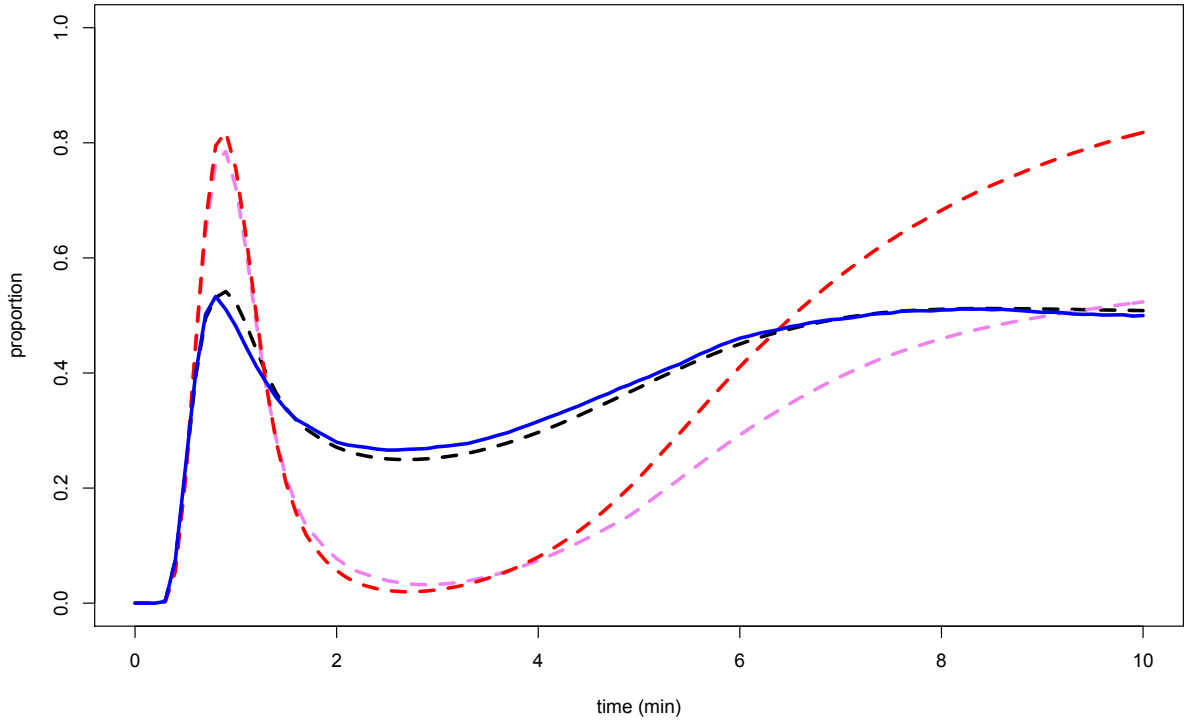
Apoptosis pathway:

| Method | Max. Error | Mean Error (normalized) | Nb. Error > 0.1 | Comput. Time |
|---|---|---|---|---|
| FF | 0.44 | 100 | 124 | 2.2s |
| Tree Cluster | 0.06 | 13.7 | 0 | 96s |

EGF-NGF pathway (normalized wrt FF for comparison with HFF):

| Method | Max. Error | Mean Error | Nb. Error > 0.1 | Comput. Time |
|---|---|---|---|---|
| FF | 100 | 100 | 100 | 1 |
| Disjoint Cluster | 84 | 79 | 84 | 1.9 |
| Tree Cluster | 32 | 14 | 16 | 12 |
| HFF (3k) | 62 | 60 | 50 | 10 |
| HFF (32k) | 49 | 38 | 35 | 1100 |

# Inference with approximate distribution




EGF-NGF pathway Proba(ErkAct = 2)

FF (factored Frontier) :
No correlations between var.

Disjoint clusters

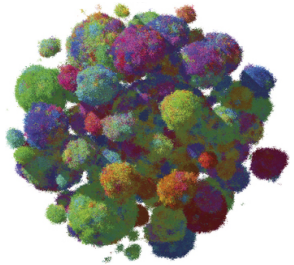
Tree (or forest) clusters

Simulations

# Tissular level

# Tissular level : Abstraction

- Obtaining tumor simulations using (modified) *TumorSimulator* (agent-based) [Waclaw et al. 2015]

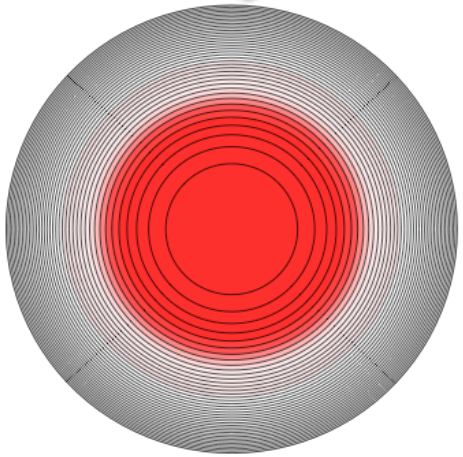


Simulations of TumourSimulator

*LayPopGenerator*

- Abstraction : Compact Markov chain

Several layers, each representing subpopulation with similar conditions (same depth).

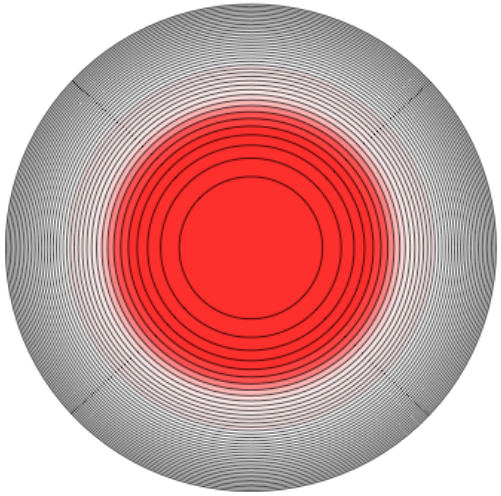


Waclaw, B., Bozic, I., Pittman, M., Hruban, M., Vogelstein, B., Nowak, M. (2015).
A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. **Nature** 525, 261-264.

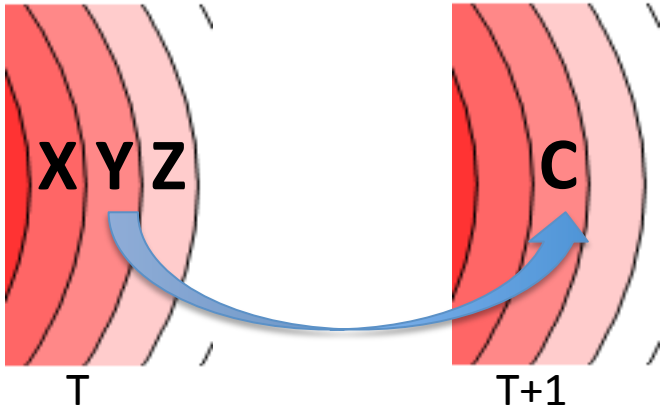
# Compact Markov Chain

*Work in progress.*

- Variables : concentrations of cells in layers



How concentration C relates to concentrations X, Y, Z ?

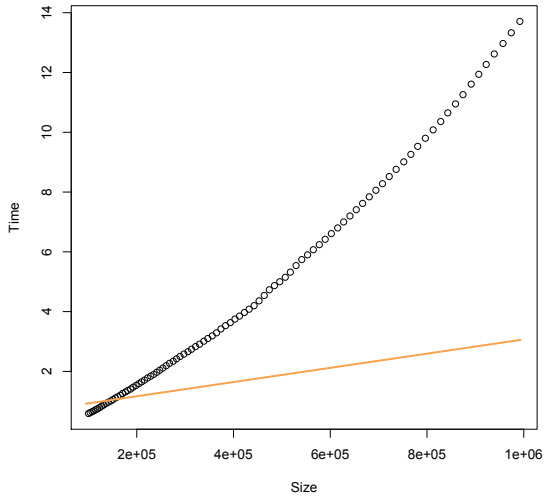



~5.000 simulations to learn the « rules »

# Running Time

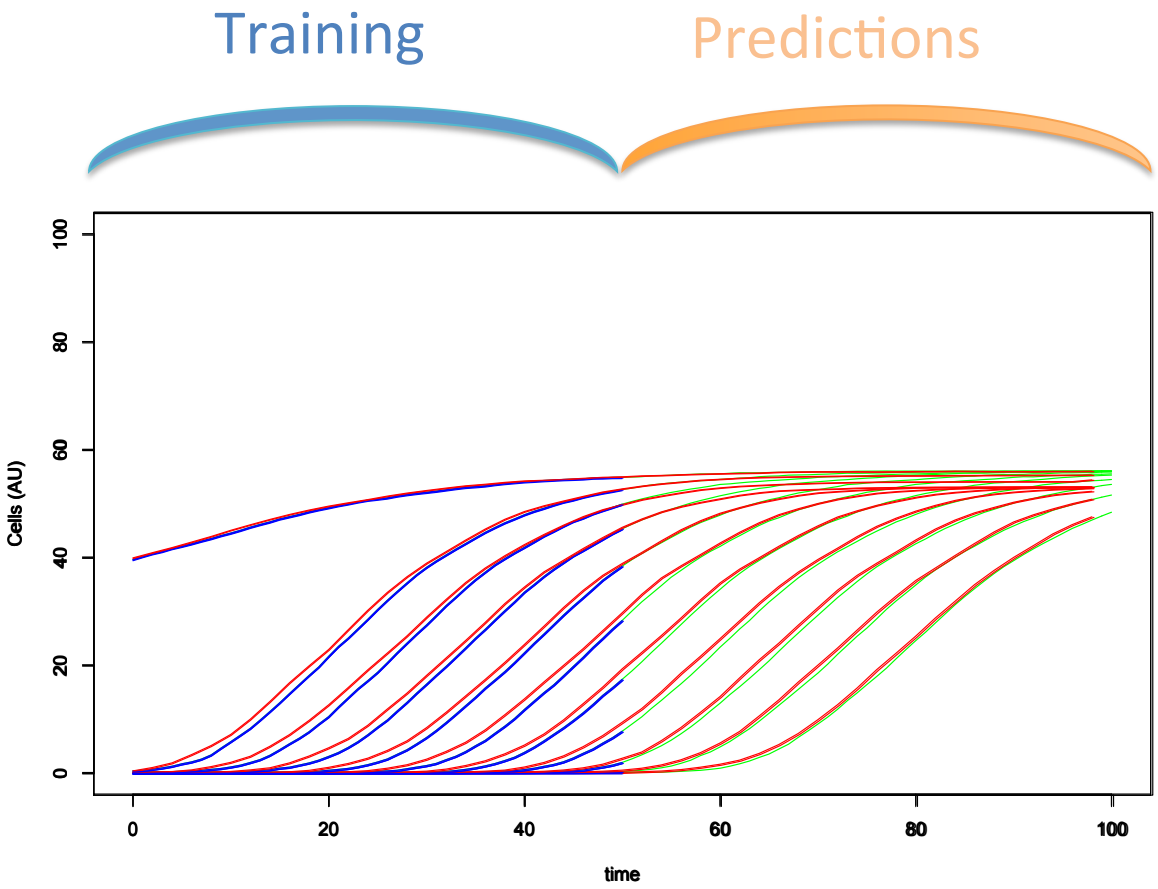
Growth (without TRAIL treatment) till one million cells

- Original : 14.3 seconds
- Our program : 2.46 seconds
  (6X faster)



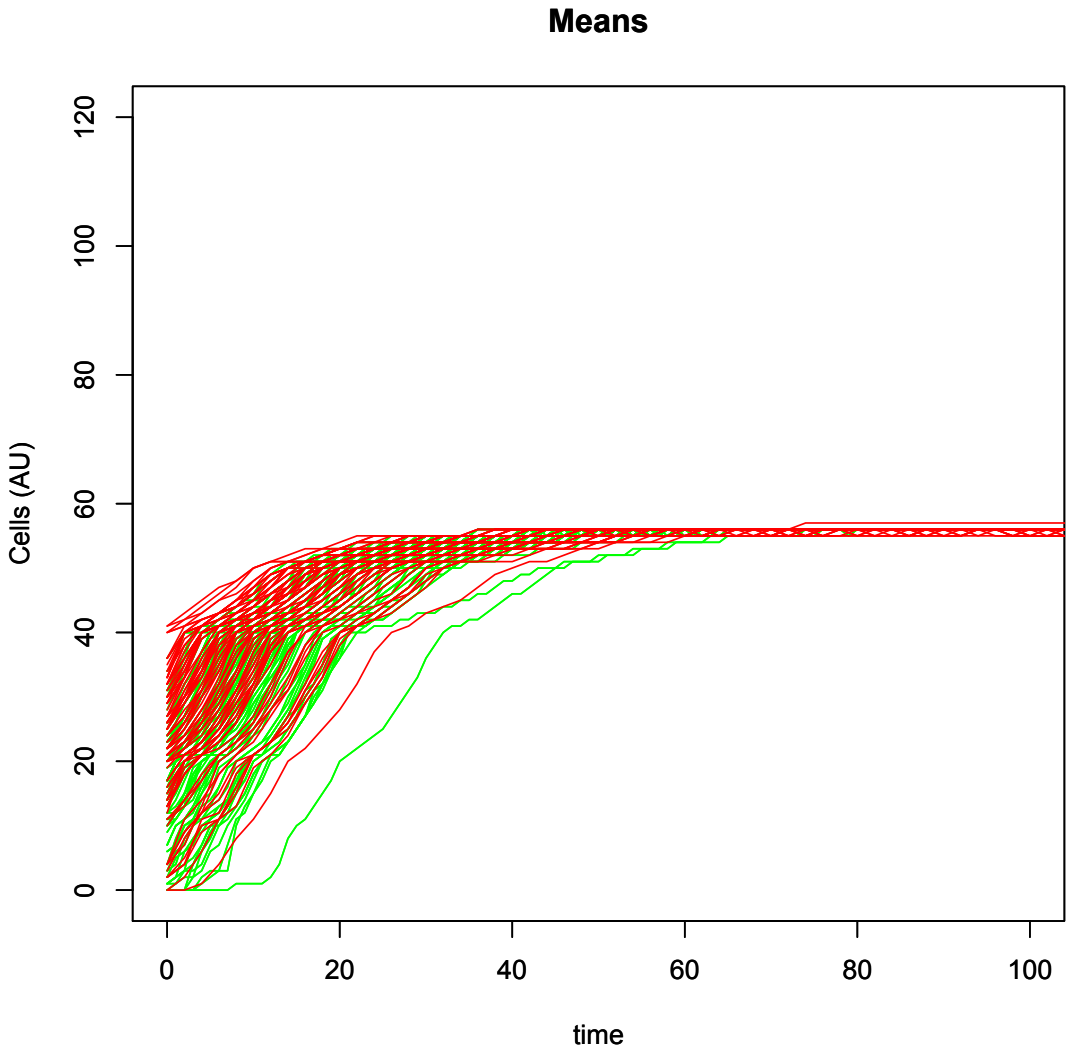

AND can use distributions instead of millions of cells for TRAIL treatment (later)

# Tissular level : Results



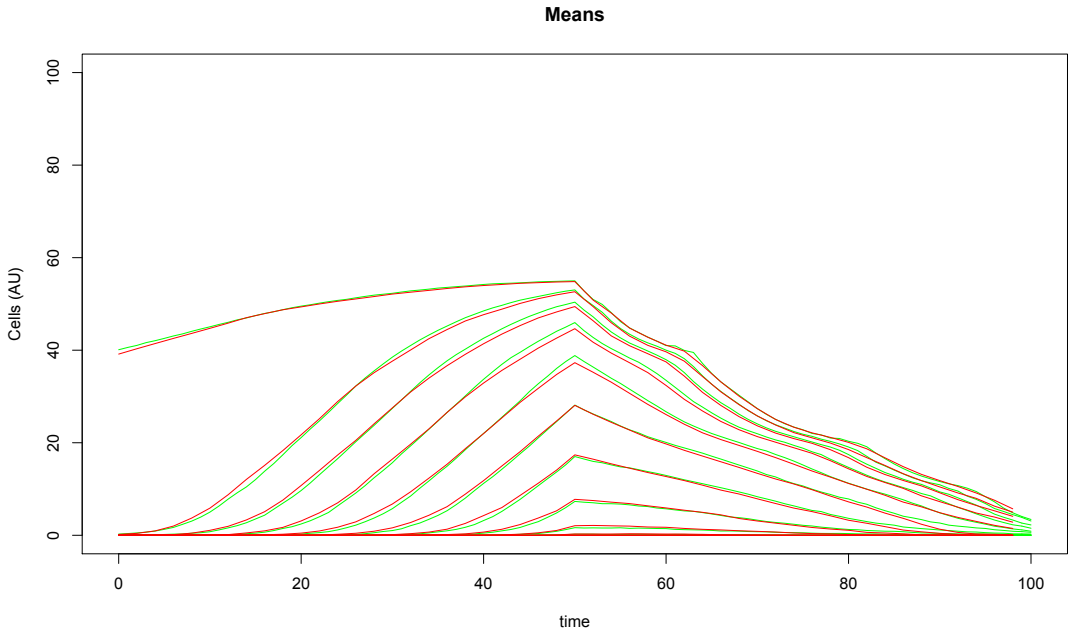

Blue/Green : Initial model
Blue : Used for training
Red : Model prediction

# Tissular level : Results




Individual runs comparison (layer 2)

# Tissular level : Results




Homogeneous treatment

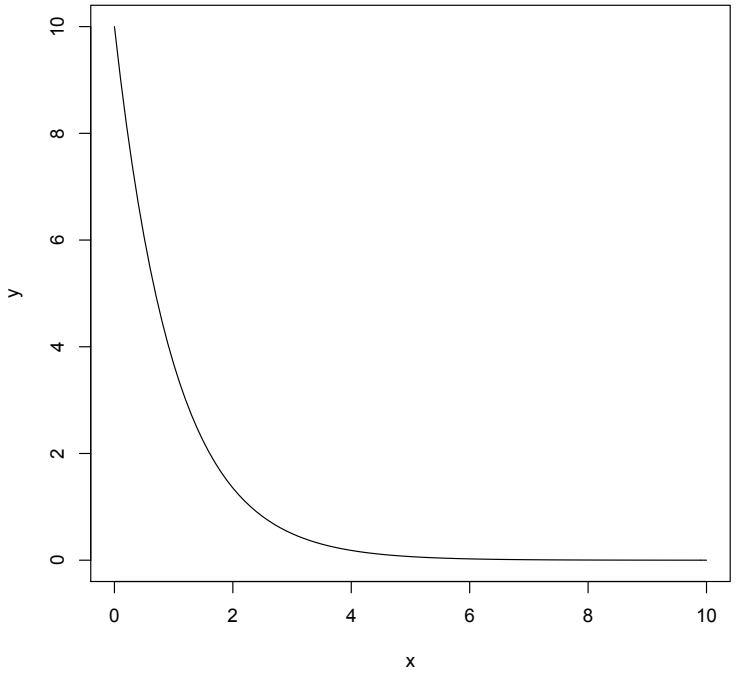
+ Resimulation accurate

- Problem to infere

# Conclusion and future work

- Fast cellular abstraction that can be used in each sub-population of the tumor
- Tumor abstraction in progress
- Link between the two levels to do.

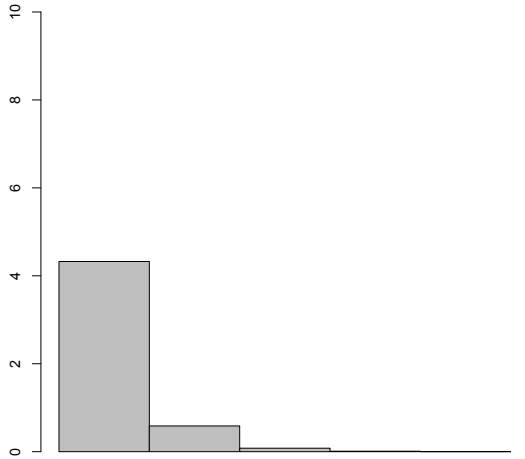
# Discretization Method

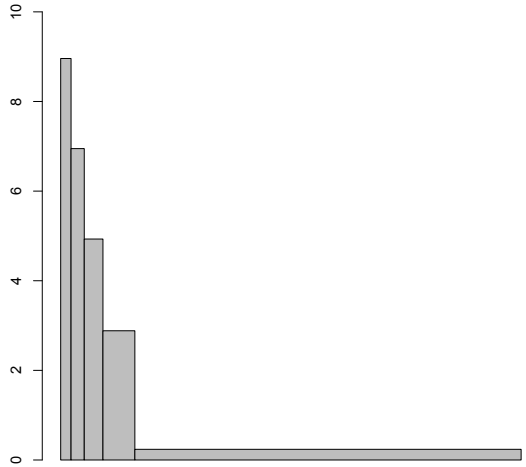



Naïve

Max entropy

…
(Lloyd-Max reducing distorsion)




5 discretized values