

# Model inference from protein time-course in Hematopoietic Stem Cells (HSC)

Pandu Raharja <sup>1,2</sup>   Rene Schoeffel <sup>1,2</sup>   Michael Strasser <sup>3</sup>

<sup>1</sup>Technische Universität München

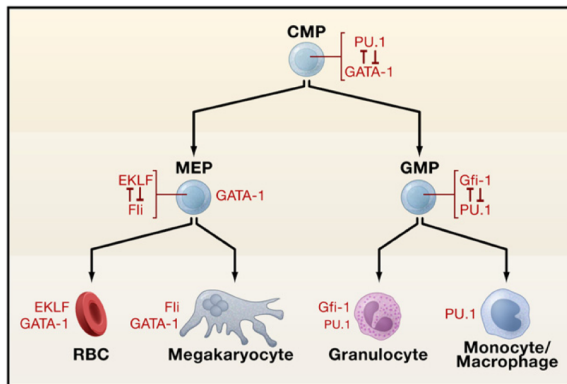
<sup>2</sup>Ludwig-Maximilians-Universität München

<sup>2</sup>Institute of Computational Biology (ICB), Helmholtz Zentrum München

October 20, 2016

# Introduction

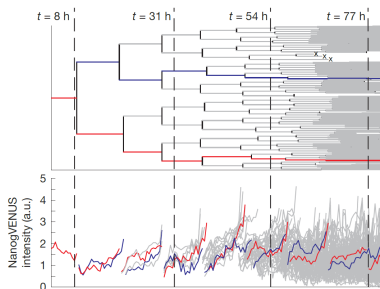
- ▶ Dynamics of hematopoietic stem cell lineage decision from Common Myeloid Progenitor (CMP) to Megakaryocyte-Erythroid Progenitor (MEP) and Granulocyte-Macrophage Progenitor (GMP) is our focus.



1

# Introduction (cont'd)

- ▶ Assumed cross-inhibition dynamics between transcription factors Pu.1 and Gata1 in cell maturation fate:
  - ▶ Dynamics is assumed to be a bistable toggle-switch system
  - ▶ Lineage decision is a stochastic process resulting in uneven yield of MEP and GMP (70% : 30%)
- ▶ Analysis on single-cell time-lapsed data to infer parameters of this dynamics.



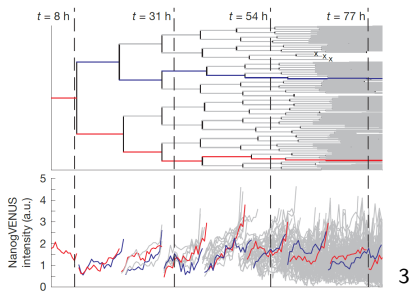
# Problems

## Stochasticity of the system

- ▶ Biophysical reaction is stochastic: reaction mechanism is inherently non-deterministic.
- ▶ Single cell analysis exposes stochasticity of the system: deterministic approaches not suitable in this case

## Branched data set

Tree structure of the data add more complexity: inheritance of information during inference process is not trivial



# Ideas

- ▶ Sequential Monte Carlo simulations using Gillespie algorithm along the time-lapsed data to infer sound parameters.

# Ideas

- ▶ Sequential Monte Carlo simulations using Gillespie algorithm along the time-lapsed data to infer sound parameters.
- ▶ **problem:** Overfitting due to single-cell biased.

# Ideas

- ▶ Sequential Monte Carlo simulations using Gillespie algorithm along the time-lapsed data to infer sound parameters.
- ▶ **problem:** Overfitting due to single-cell biased.
- ▶ **solution:** Inference across cell lineages.

# Ideas

- ▶ Sequential Monte Carlo simulations using Gillespie algorithm along the time-lapsed data to infer sound parameters.
- ▶ **problem:** Overfitting due to single-cell biased.
- ▶ **solution:** Inference across cell lineages.
- ▶ Inferred parameters from all simulated lineages are represented as distribution.



# Ideas

- ▶ Sequential Monte Carlo simulations using Gillespie algorithm along the time-lapsed data to infer sound parameters.
- ▶ **problem:** Overfitting due to single-cell biased.
- ▶ **solution:** Inference across cell lineages.
- ▶ Inferred parameters from all simulated lineages are represented as distribution.
- ▶ Final inferred parameters are expected value  $E$  of the distribution.

# Particle Filtering

- ▶ **Particle**

A particle  $K$  is defined as a triple of previous simulation trajectory  $X$ , parameter set  $\theta$  and assumed model  $M$ ,

$$K := (X, \theta, M) \tag{1}$$

- ▶ **Particle filtering** is an parameters inference method that consists of: (1) *sequentially* performing simulations using *particles*, (2) *updating* the *prior* assumptions of the model using the results of the simulations and (3) rerunning the simulations using updated assumptions (*posterior*).

# Particle Filtering: update rule

## ► Posterior

After each simulation step, a posterior describes the probability of having the trajectory  $X$  and parameter  $\theta$  given the observation  $D$  from real data,

$$P(X, \theta | D) \stackrel{\text{Bayes}}{=} \frac{P(D | X, \theta) P(X, \theta)}{P(D)} \quad (2)$$

This Bayesian update rule is used to update parameters by looking at how well does the simulation follow the real data. I.e. after an iteration we will choose parameters belonging to particles that simulate the trajectory well w.r.t. experimental data.

- **Gamma Distribution** is used as prior since a posterior of a gamma is in turn gamma distributed (*prior conjugate*).

# Particle Filtering: algorithm

1. Initialization of parameters  $\theta$ .
2. Input of data  $\mathcal{D}$ .
3. Particle filtering routine:
  - 3.1 Generation of initial particles for step  $i$  using weighting  $w_{i-1}$  from previous iteration:

$$K_i := (K_{i1}, K_{i2}, \dots, K_{im}) \quad (3)$$

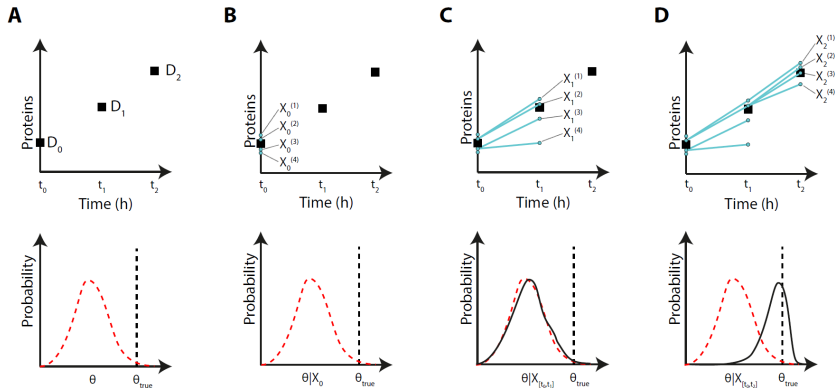
- 3.2 Simulation run of each particle  $K_{ij}$ .
- 3.3 Weighting of each particle. The weight is a function of the probability of observing the data given the simulation result:

$$w_i^k = P(D_i | X_i^k) = \mathcal{N}(\mathcal{D}_i | X_i^k) \quad (4)$$

- 3.4 Parameter update for every  $K$ :

$$\theta^k \propto P(\theta | X_{[t_0, t_i]}^k) \quad (5)$$

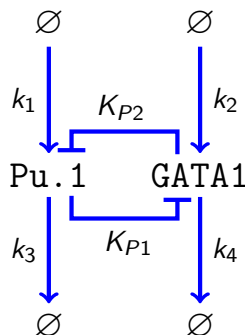
# Particle Filtering: visualization



4

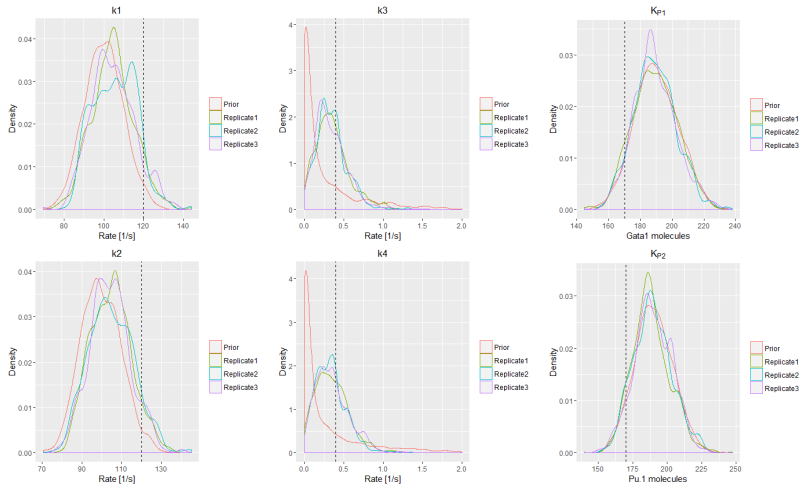
# Reaction Model

- ▶ Simple model in line with cross inhibition assumption.
- ▶ Inhibition modeled through inverse michaelis menten kinetics.
- ▶ Parameterize production and degradation rates  $k_i$  as well as inhibition constant  $K_Y$ .



$$a_i = k_i \cdot \left(1 - \frac{X_i^n}{X_i^n + K_Y^n}\right)$$

# Evaluation on synthetic data with known model



# Model Comparison

- Compare performance of different models via Bayes Factors:

$$B_{M_1, M_2} = \frac{P(M_1|\mathcal{D})}{P(M_2|\mathcal{D})} \stackrel{\text{Bayes}}{=} \frac{P(M_1)P(\mathcal{D}|M_1)}{P(M_2)P(\mathcal{D}|M_2)} \quad (6)$$

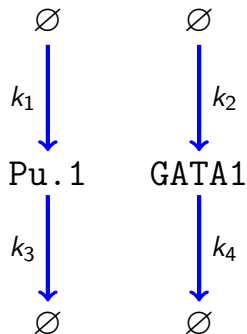
- The average weight of the particle trajectories are used as Bayes Factor:

$$P(\mathcal{D}|M) = P(\mathcal{D}_I) \prod_{l=1}^N P(\mathcal{D}_{i+1}|\mathcal{D}_{0:i}, M) \quad (7)$$



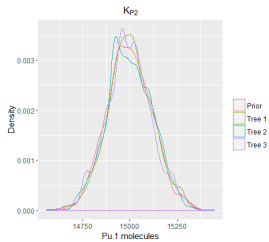
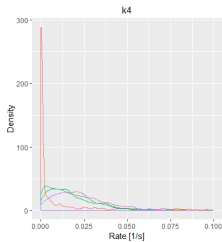
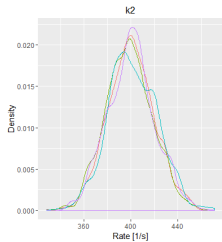
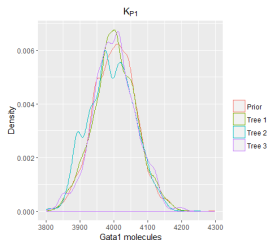
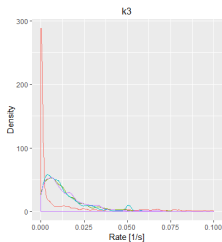
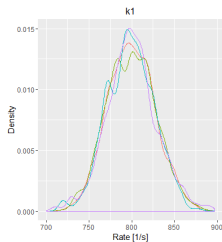
# Model Comparison

- ▶ Independent model used on synthetic data set.
- ▶ Model comparison by Bayes factors.



$$B_{M1,M2} = \frac{P(\mathcal{D}|M_{inhibitory})}{P(\mathcal{D}|M_{independent})} = \frac{10^{-2221.498}}{10^{-2234.828}} \approx 2.14 \cdot 10^{13} \quad (8)$$

# Results on experiment data



# Interpretation

- ▶ Parameters of reaction model have varying susceptibility to fitting through particle filtering.
- ▶ Model cannot accurately describe the dynamic of the transcript factors.
- ▶ Recent finding (Hoppe et al., 2016, Nature) on the same data suggest the lineage choice to be independent from the transcription factor ratios.
- ▶ Particle filtering cannot infer the presence of the external decision factor.

# Conclusion & Outlook

- ▶ Particle filtering enable the parameterization of models using time lapse data of cell lineage trees.
- ▶ Using the weight of particle trajectories models can be compared, allowing quantified comparison of theories.
- ▶ It would be interesting to know whether the decision process is really influenced by several latent factors under different reaction model.

# References

- ▶ Feigelman, J. (2016). "Stochastic and deterministic methods for the analysis of Nanog dynamics in mouse embryonic stem cells." PhD Thesis, Technische Universität München, Munich, Germany.
- ▶ Hoppe, P.S., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K.D., Hilsenbeck, O., Mortz, N., ... & Etzrodt, M. (2016). Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature*, **535(7611)**, 299-302.
- ▶ Graf, T., & Enver, T. (2009). Forcing cells to change lineages. *Nature*, **462(7273)**, 587-594.