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

Pandu Raharja [1,2] Rene Schoeffel [1,2] Michael Strasser [3]

[1]Technische Universität München

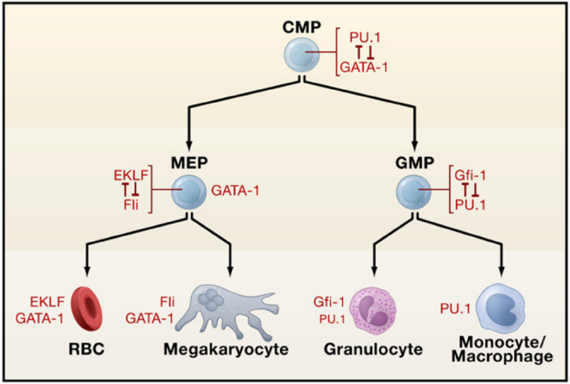
[2]Ludwig-Maximilians-Universität München

[2]Institute of Computational Biology (ICB), Helmholtz Zentrum München

October 19, 2016

# Introduction

- Dynamics of hematopoetic stem cell maturation cell from Common Myeloid Progenitor (CMP) to Megakaryocyte-Erythroid Progenitor (MEP) and Granulocyte-Macrophage Progenitor (GMP)

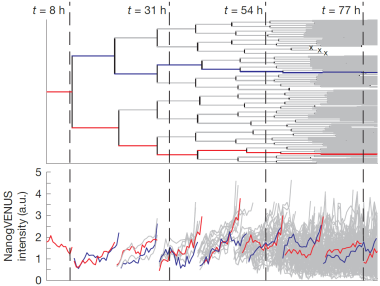


[1]

[1] Graf & Enver, 2009, *Nature*

# Introduction (cont'd)

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



[2]

[2] Feigelman, 2016, Ph.D. Thesis

# Problems

- Stochaticity in single cell resolution is more punctuated
- Tree structure of the data add ore complexity: inheritance of information during inference process is not trivial

# Ideas

- Sequential Monte Carlo simulations along the time-apsed data to infer "good" parameters

# Ideas

- Sequential Monte Carlo simulations along the time-apsed data to infer "good" parameters
- **problem:** Overfitting due to single-cell biased

# Ideas

- Sequential Monte Carlo simulations along the time-apsed data to infer "good" parameters
- **problem:** Overfitting due to single-cell biased
- **solution:** Inference across cell lineages

# Ideas

- Sequential Monte Carlo simulations along the time-apsed data to infer "good" 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 along the time-apsed data to infer "good" 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,

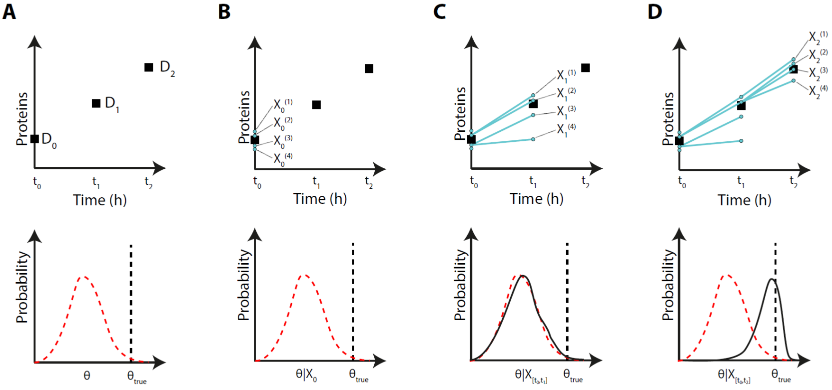
$$P(X, \theta | D) \overset{\text{Bayes}}{=} \frac{P(D|X, \theta)P(X, \theta)}{P(D)} \tag{2}$$

  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
   $$Ki := (K_{i1}, K_{i2}, \ldots, K_{im}) \tag{3}$$
   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.
   $$w_i^k = P(D_i|X_i^k) = \mathcal{N}(\mathcal{D}_i|X_i^k) \tag{4}$$
   3.4 Parameter update for every K,
   $$\theta^k \propto P(\theta|X^k_{[to,ti]}) \tag{5}$$

# Particle Filtering: visualization




[3]

[3] Feigelman, 2016, Ph.D. Thesis