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

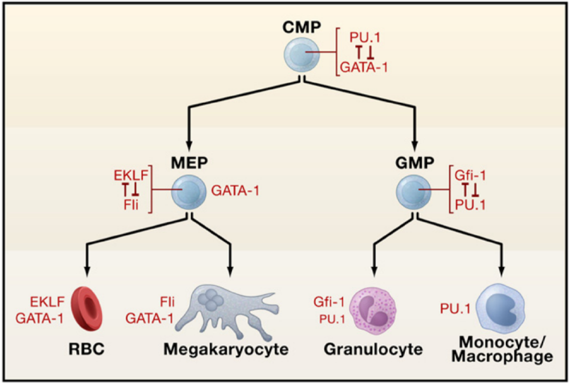
[2]Ludwig-Maximilians-Universität München

[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)

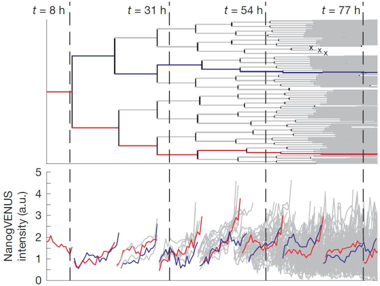


[1]

[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

A particle K is defined as a triple of trajectory X (*time series data*), parameter set θ and assumed model M,

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

- X describes the simulation trajectory
- M describes the model of the simulation
- θ describes the parameters of the simulation's model

Particle Filtering (7)

Particle Filtering: algorithm

1. Initialization of parameters θ.
2. Input of data $\mathcal{D}$.
3. Particle filtering routine:
   3.1 Generation of initial particles for step i

$$Ki := (K_{i1}, K_{i2}, \dots, K_{im}) \tag{2}$$

   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.

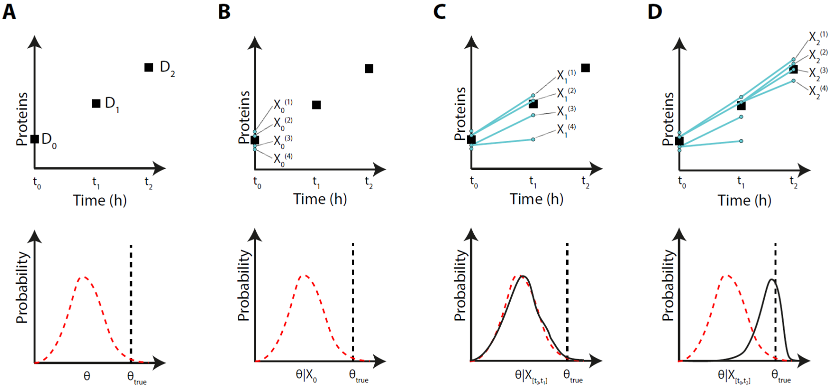
$$w_i^k = P(D_i|X_i^k) = \mathcal{N}(\mathcal{D}_i|X_i^k) \tag{3}$$

   3.4 Parameter update for every K,

$$\theta^k \propto P(\theta|X^k_{[to,ti]}) \tag{4}$$

4. Model comparison.

Particle Filtering: visualization



3

---

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