

Department of Information and Computer Science

Bayesian Multi-Way Models for Data Translation in Computational Biology

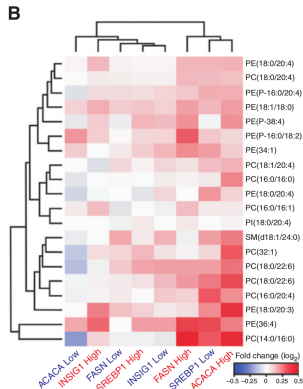
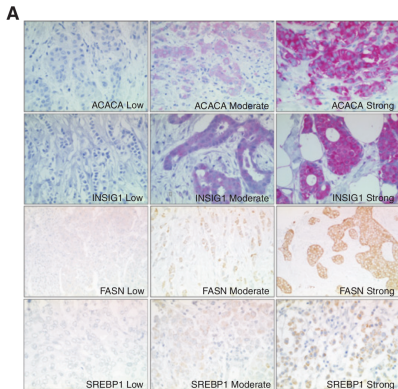
Tommi Suviataival



Introduction

- ▶ *Molecular measurements of biological organisms to study response to:*
 - ▶ disease
 - ▶ medical treatment
 - ▶ environment

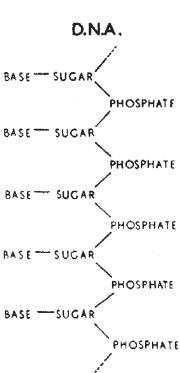
- ▶ Measurements can be made:
 - ▶ *in vivo*: cell extracts from humans or model organisms
 - ▶ *in vitro*: cell lines grown in laboratory



Hilvo *et al.*,
Cancer Res.

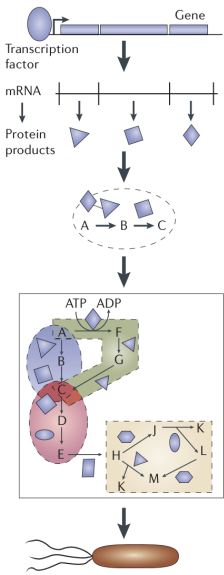
2011

Molecular activity in biological cell



Watson & Crick, *Nature* 1953

- Components**
- Genomics
 - Transcriptomics
 - Proteomics
 - Metabolomics
 - Glycomics
 - Lipidomics
 - Localizomics



- Interactions**
- Protein-DNA interaction
 - Protein-protein interaction

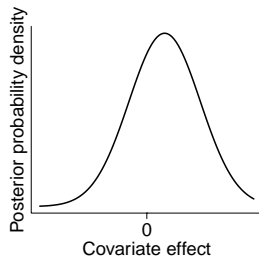
- Functional states**
- Fluxomics
 - Phenomics

Joyce & Palsson, *Nat. Rev. Mol. Cell Biol.* 2006

Machine learning for computational biology

- ▶ Molecular measurements:
 - ▶ Large data sets
 - ▶ Uncertainty/noise
- ⇒ Automated and robust *data-driven* analysis tools needed

- ▶ Bayesian approach to probability:
 - ▶ Take uncertainty into account
 - ▶ Describe the generative process of the data
 - ⇒ Integration of multiple measurement sources
 - ▶ Incorporate existing knowledge by specifying:
 - ▶ the model structure
 - ▶ priors

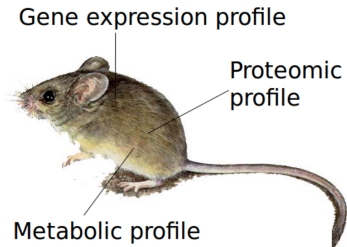


Computational medicine & contributions

- ▶ Model organisms for studying effects of:
 - ▶ genomic mutations
 - ▶ new medical treatments, potentially dangerous

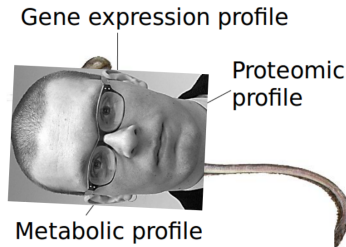
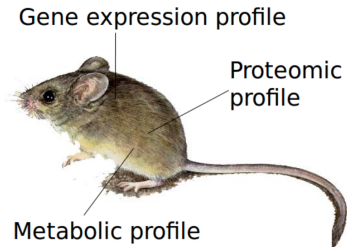
Computational medicine & contributions

- ▶ Model organisms for studying effects of:
 - ▶ genomic mutations
 - ▶ new medical treatments, potentially dangerous
- ▶ Dissertation: statistical modeling of effects in molecular measurement data with
 - ▶ high-dimensional, noisy measurements
 - ▶ multiple measurement types
 - ▶ multiple organisms



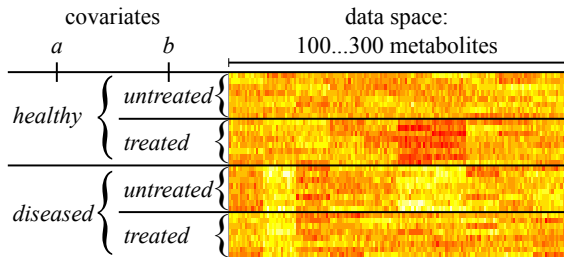
Computational medicine & contributions

- ▶ Model organisms for studying effects of:
 - ▶ genomic mutations
 - ▶ new medical treatments, potentially dangerous
- ▶ Dissertation: statistical modeling of effects in molecular measurement data with
 - ▶ high-dimensional, noisy measurements
 - ▶ multiple measurement types
 - ▶ multiple organisms

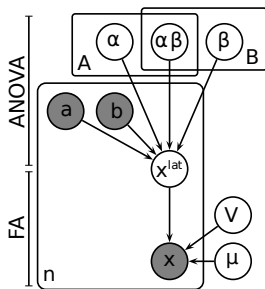


P I: Multi-Way Model for " $n < p$ "

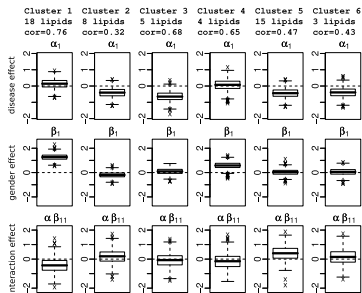
(1) Data:



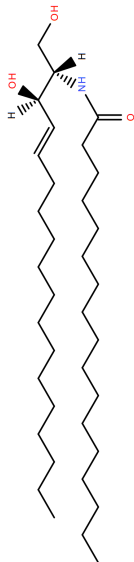
(2) Model:



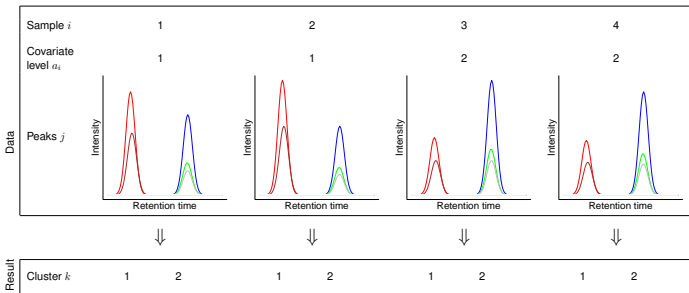
(3) Result:



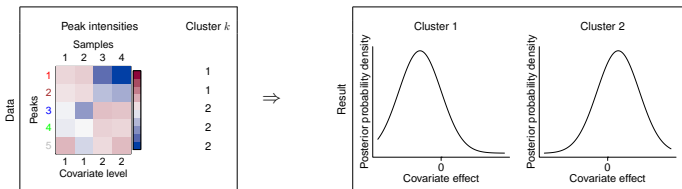
P II-III: Multi-Way Models for Multi-Peak Metabolomics



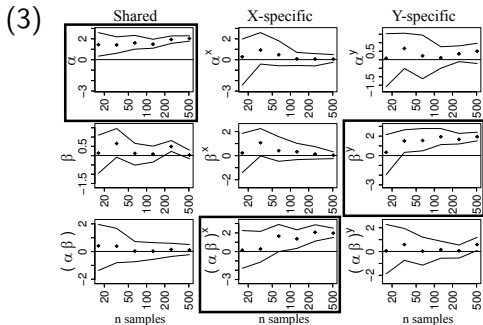
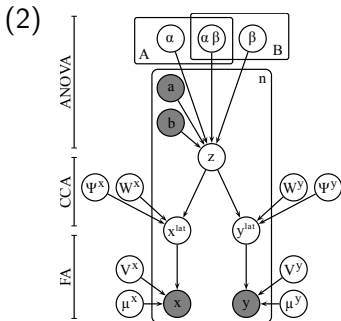
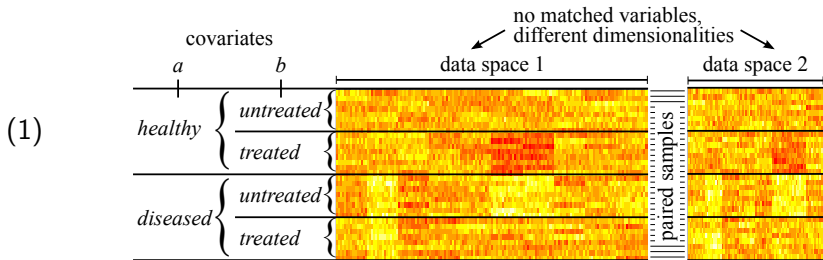
a) Peak clustering based on shapes



b) Inference of covariate effects based on intensity



P IV: Multi-Way Model for Multiple Sources



Summary

New machine learning models for:

- P I Small sample size, high dimensionality ($n < p$)
- P II–III Incorporating prior information about the measurement process
- P IV–V Multiple data sources with co-occurring samples
- P VI–VII Multiple data sources without co-occurring samples

Department of Information and Computer Science

Bayesian Multi-Way Models for Data Translation in Computational Biology

Tommi Suviataival

