Systems Biology Strategies to Study Cancer Metabolism

MSc Thesis presented by *Daniel Ortiz Martínez*
Supervised by *Joaquín Dopazo Blázquez* and *Vicente Arnau Llombart*

Table of contents

Introduction

Introduction

- Systems biology studies the systemic properties and interactions in biological objects
    - relies on mathematical models
    - networks constitute a central concept
- One example of biochemical network is metabolism
- Metabolic diseases are a major source of mortality
- Growth and survival of cancerous cells require alterations in normal metabolism (Cairns et al. 2011)

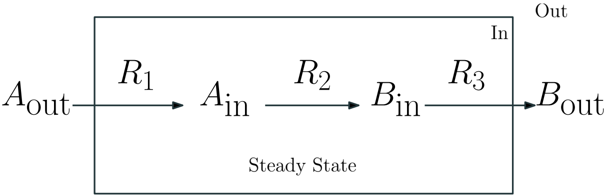
Goals

- **Scientific Goals**:
  - Integration of transcriptomic and metabolic information using flux balance analysis (FBA)
  - Statistical analysis of FBA results
  - Design visualization techniques for biochemical networks
  - Comparison between normal and cancer metabolism
- **Technologic Goals**:
  - Review of available software for FBA
  - Development of open-source software for FBA

Flux Balance Analysis of Metabolism

Flux Balance Analysis (FBA)

- FBA is a widely used approach for studying genome-scale metabolic network reconstructions
- FBA is based on two assumptions:
  - **Steady state**: the fluxes of any compound being produced must be equal to the total amount being consumed
  - **Biological goal optimization**: evolution has operated on metabolism optimizing some biological goal



Flux Balance Analysis (cont.)

- FBA can be stated as a linear programming problem
- Optimal flux values are calculated using solvers
- Biomass production is often used as biological goal
- Matrix notation is typically used:

$$\begin{aligned} \text{Maximize} \quad & Z = \mathbf{c}^{\mathbf{T}} \cdot \mathbf{v} \\ \text{Subject to} \quad & \mathbf{S} \cdot \mathbf{v} = \mathbf{0} \\ & \mathbf{l} \leq \mathbf{v} \leq \mathbf{u} \end{aligned}$$

Flux Variability Analysis (FVA)

- Often the optimal solution of an FBA problem is not unique
- FVA can be used to obtain the range of values that can take the fluxes while producing the optimal objective function value
- FVA solves two optimization problems for each flux v_i:

$$\begin{aligned}
\text{Maximize/Minimize} \quad & Z = v_i \\
\text{Subject to} \quad & \mathbf{S} \cdot \mathbf{v} = \mathbf{0} \\
& \mathbf{c}^{\mathbf{T}} \cdot \mathbf{v} \geq \gamma Z_0 \\
& \mathbf{l} \leq \mathbf{v} \leq \mathbf{u}
\end{aligned}$$

Tissue-Specific FBA

- FBA does not reflect metabolism differences for each tissue
- Shlomi et al. 2008 proposes an FBA-based method integrating a metabolic network reconstruction with gene expression data
- Tissue-specific FBA requires the definition of sets of highly and lowly expressed reactions (R_H and R_L)
- Obtaining R_H and R_L from RNA-Seq data involves:
  1. Identify absent and present genes (Hebenstreit et al. 2011)
  2. Apply gene to protein-reaction logical rules
- Shlomi technique finds the flux distribution that best reflects R_H and R_L while satisfying constraints

Visualization of Results

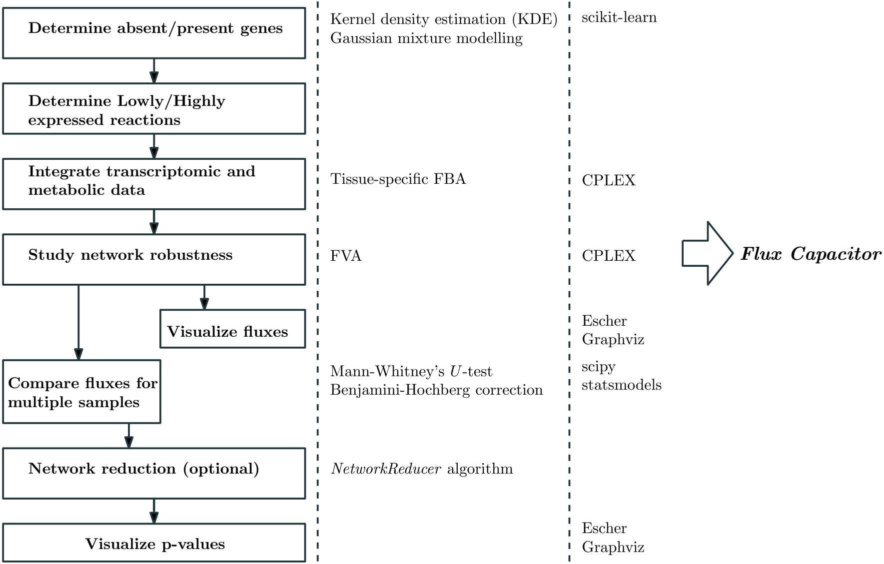
- FBA results should be contextualized to improve understandability
- Biochemical network diagrams are useful for this purpose
- The following strategies were tested:
    - Pre-generated diagrams using Escher
    - Automatically-generated diagrams using Graphviz
    - Network reduction techniques (Erdrich et al. 2015)

Experiments

Datasets

- **Recon 2**: second version of the human metabolic reconstruction provided in the Recon X database
    - 7 440 reactions and 5 063 metabolites
- **KIRC data**: contains diverse information about 537 samples of kidney tissue, including RNA-Seq data
    - 60 cases and 60 controls

Tissue-Specific FBA Results: Overview

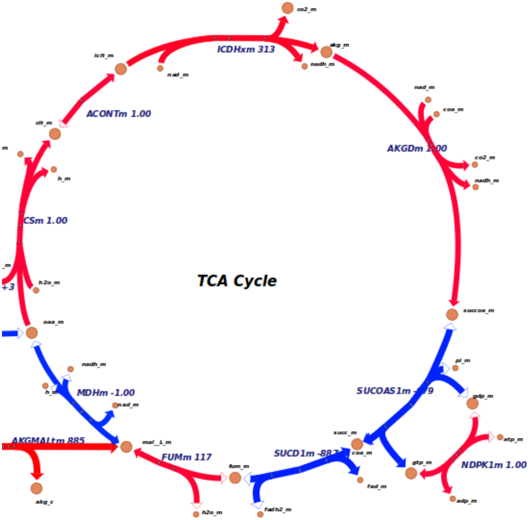


Pipeline	Methods	Software
Determine absent/present genes	Kernel density estimation (KDE) Gaussian mixture modelling	scikit-learn
Determine Lowly/Highly expressed reactions		
Integrate transcriptomic and metabolic data	Tissue-specific FBA	CPLEX
Study network robustness	FVA	CPLEX
Visualize fluxes		Escher Graphviz
Compare fluxes for multiple samples	Mann-Whitney's U-test Benjamini-Hochberg correction	scipy statsmodels
Network reduction (optional)	*NetworkReducer* algorithm	
Visualize p-values		Escher Graphviz

Flux Capacitor

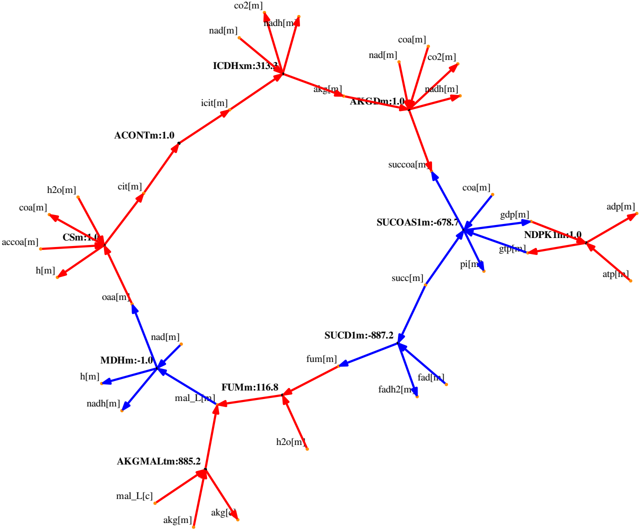
Tissue-Specific FBA Results: Escher Visualization

- TCA cycle for sample `TCGA.A3.3324.01A.02R.1325.07`:



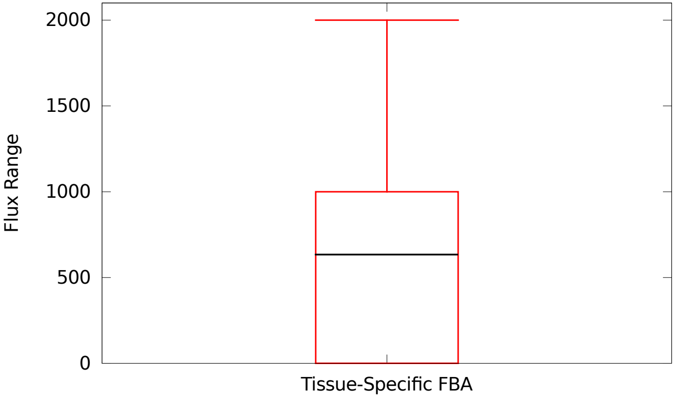
Tissue-Specific FBA Results: Graphviz Visualization

- TCA cycle for sample `TCGA.A3.3324.01A.02R.1325.07`:



Tissue-Specific FBA Results: Network Robustness

- Flux range boxplot for sample `TCGA.A3.3324.01A.02R.1325.07`:



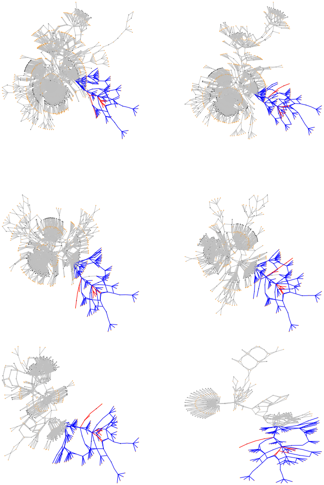
Tissue-Specific FBA Results: Differential Reaction Expression

- Tissue-specific FBA applied to 60 cases and 60 controls
- Fluxes were categorized due to their high variability:
  - inactive reaction (0)
  - active in the direct sense reaction (1)
  - active in the reverse sense reaction (-1)
- Mann-Whitney's U-test + Benjamini-Hochberg correction
  - 428 differentially expressed reactions
- *Biological subsystems* affected:
  - amino-acid metabolism
  - carbohydrate metabolism
  - retinol metabolism[1]

---

[1] already linked to cancer development (R. Blomhoff and H. K. Blomhoff 2006)

Tissue-Specific FBA Results: Network Reducing



- Fast version of *NetworkReducer* was applied
- Algorithm configured to retain all retinol metabolism reactions
- Study of reactions in the vicinity of retinol metabolism was enabled
- Alterations in choline transport were found[2]

---

[2] related to cancer development in (Gillies and Morse 2005)

Open-Source Software

Flux Capacitor

- Flux Capacitor is open-source software distributed under LGPL
- Coded in C++, R, Python and shell scripting
- Hosted on GitHub[3]
- Main Functionality:
  - FBA and FVA
  - Tissue-specific FBA
  - Statistical testing
  - Network visualization and reduction

---

[3] `https://github.com/daormar/flux-capacitor`

Conclusions and Future Work

Conclusions

- Transcriptomic and metabolic data successfully integrated using FBA
- Different mathematical solvers were tested (CPLEX fastest, CBC and CLP acceptable with less restrictive license)
- Escher was useful to contextualize FBA information
- Graphviz-based visualization less intelligible but fully automatic and more versatile
- An open-source toolkit for FBA has been developed

Future Work

- Deeper analysis of tissue-specific FBA results
  - detailed analysis under a biological point of view
- Better network representations using Graphviz
  - fully exploit Graphviz features
- Further experiments with *NetworkReducer*
  - test different criteria to remove reactions
- Improvements and extensions in *Flux Capacitor*
  - reduce dependencies with third-party software

References i

Blomhoff, R. and H. K. Blomhoff (2006). "Overview of retinoid metabolism and function". In: *J. Neurobiol.* 66.7, pp. 606–630.

Cairns, R. A., I. S. Harris, and T. W. Mak (2011). "Regulation of cancer cell metabolism". In: *Nat. Rev. Cancer* 11, pp. 85–95.

Erdrich, P., R. Steuer, and S. Klamt (2015). "An algorithm for the reduction of genome-scale metabolic network models to meaningful core models". In: *BMC Syst Biol* 9, p. 48.

Gillies, R. J. and D. L. Morse (2005). "In vivo magnetic resonance spectroscopy in cancer". In: *Annu Rev Biomed Eng* 7, pp. 287–326.

References ii

- Hebenstreit, D., M. Fang, M. Gu, V. Charoensawan, A. van Oudenaarden, and S. A. Teichmann (2011). “RNA sequencing reveals two major classes of gene expression levels in metazoan cells”. In: *Mol. Syst. Biol.* 7, p. 497.
- Shlomi, T., M. N. Cabili, M. J. Herrgård, B. Ø. Palsson, and Eytan Ruppin (2008). “Network-based prediction of human tissue-specific metabolism”. In: *Nat. Biotechnol.* 26(9), pp. 1003–10.

Questions?

Tissue-Specific FBA Formulation

$$\begin{aligned}
\text{Maximize} \quad & Z = \left(\sum_{i \in R_H} (y_i^+ + y_i^-) + \sum_{i \in R_L} y_i^+ \right) \\
\text{Subject to} \quad & \mathbf{S} \cdot \mathbf{v} = \mathbf{0} \\
& \mathbf{l} \le \mathbf{v} \le \mathbf{u} \\
& \mathbf{v}_i + y_i^+ (\mathbf{l}_i - \epsilon) \ge \mathbf{l}_i, \; i \in R_H \\
& \mathbf{v}_i + y_i^- (\mathbf{u}_i + \epsilon) \le \mathbf{u}_i, \; i \in R_H \\
& \mathbf{l}_i (1 - y_i^+) \le \mathbf{v}_i \le \mathbf{u}_i (1 - y_i^+), \; i \in R_L \\
& y_i^+, y_i^- \in \{0, 1\}
\end{aligned}$$

Methods

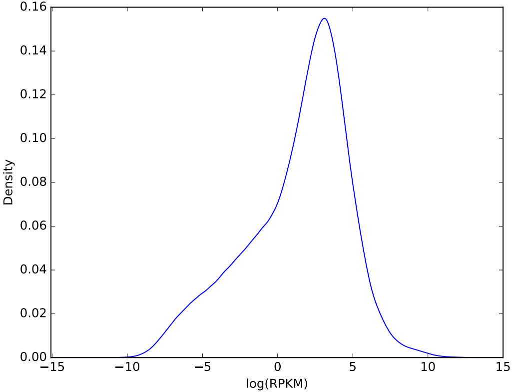
- **Flux balance analysis (FBA)**
- **Flux variability analysis (FVA)**
- **Kernel density estimation (KDE)**
- **Gaussian mixture model estimation**
- **Tissue-specific FBA**
- **U-test + BH procedure**
- **Network reducer algorithm**

Software

- **Solvers**
    - GLPSOL, CLP, CBC, **CPLEX**
- **Visualization tools**
    - Escher, Graphviz
- **Statistical tools**
    - scikit-learn, scipy, statsmodels
- **Software developed for this thesis**

Tissue-Specific FBA Results: Absent/Present Genes

- Two RNA abundance classes were observed:



- Gaussian mixture modelling was applied to classify genes as absent or present

Tissue-Specific FBA Results: Gene and Reaction Statistics

- Statistics for KIRC's sample `TCGA.A3.3324.01A.02R.1325.07`: