Systematic evaluation of the consensus across pathway databases

Daniel Domingo-Fernández
Introduction

- Deconvoluting biology into networks
Introduction

Deconvoluting biology into networks

- Nodes represent biological entities (e.g., proteins, chemicals, etc.)
- Edges represent the type of interaction (e.g., activation, inhibition, etc.)
Introduction

- Deconvoluting biology into networks
  - Nodes represent biological entities (e.g., proteins, chemicals, etc.)
  - Edges represent the type of interaction (e.g., activation, inhibition, etc.)

Taken from: https://theory.labster.com/signaling-cascade/
Introduction

- Deconvoluting biology into networks
  - Nodes represent biological entities (e.g., proteins, chemicals, etc.)
  - Edges represent the type of interaction (e.g., activation, inhibition, etc.)

- Deconvoluting biological networks into pathways
  - Pathways correspond to “small networks” representing a specific molecular process (e.g., inflammation, glucose-uptake, DNA repair, etc.)
  - For researchers, these “small networks” or pathways represent how a specific part of biology works

Taken from: https://theory.labster.com/signaling-cascade/
The Problem (1)

- Numerous pathway databases [1...3]
- Pathways are represented using different schemas and stored in data-silos
- By integrating this information, we can evaluate whether there is consensus among these resources true to the biology
The Problem (2)

*Krebs cycle* pathway (glucose metabolism) network from 5 different databases [4]
**The Problem (2)**

*Krebs cycle* pathway (glucose metabolism) network from 5 different databases [4]

Legend

*Red* = only in one database

*Green* = multiple databases
Integrating pathway databases

Harmonization of resources

Integration in an unified schema
Integrating pathway databases

Harmonization of resources

Integration in an unified schema

Network representing biological interactions in multiple databases
Application example (on going work)

1. Use clinical data as an input into the network
2. Each patient generates a unique signature
3. Machine learning classifies patients and controls

Patient’s Readouts

Different network features

Prediction
Disease vs normal
Use patients’ signatures to predict treatments

- Simulate combination of treatment in specific patients
- Use the trained classifier looking for a positive prediction

Patient’s network features

Diseased patient
Use patients’ signatures to predict treatments

- Simulate combination of treatment in specific patients
- Use the trained classifier looking for a positive prediction

Diseased patient

Patient’s network features

Drug

4.5

0.7

0.3

0.3
Use patients’ signatures to predict treatments

- Simulate combination of treatment in specific patients
- Use the trained classifier looking for a positive prediction

![Diagram showing modified network features and a drug effect on a diseased patient]
Use patients’ signatures to predict treatments

- Simulate combination of treatment in specific patients
- Use the trained classifier looking for a positive prediction

Diseased patient

Modified network features

New Prediction

Normal?
Other applications

- **Machine Learning**
  - Train classifiers for each database and compare performance (which databases/pathways predict better? use pathways as markers?)
  - Remove random "parts" of the network to evaluate changes in performance for prioritization (non relevant in the disease context? noise?)
  - …

- **Graph theory**
  - Biological contradictions (e.g., drug A activates/inhibits protein B)
  - Explore biological landscape (where is science heading to?)
  - …
Summary

- Biology can be deconvoluted as a network

- The entire network is divided into smaller subnetworks that represent specific biological processes -> pathways

- We can interrogate this network with data and in a combination with ML for
  - Prediction
  - Treatment simulation
  - Pathway prioritization
  - …
Thank you!
Acknowledgments

- Prof. Dr. Martin Hofmann-Apitius
- Sarah Mubeen
- Josep Marín Llaó
- Charles Tapley Hoyt
References


Evaluating biological knowledge (ongoing work)