Systematic evaluation of the consensus across pathway databases

Daniel Domingo-Fernández

Fraunhofer

SCAI

Introduction

- Deconvoluting biology into networks
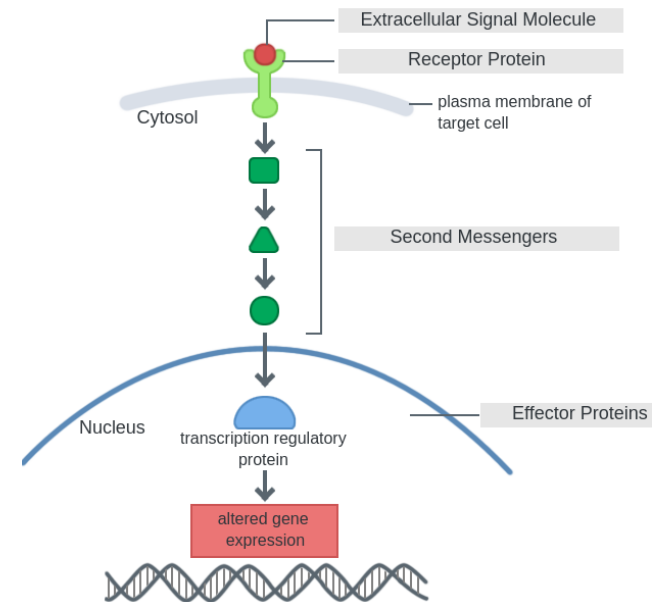
Fraunhofer
SCAI

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.)

Fraunhofer
SCAI

Introduction

- Deconvoluting biology into networks

  - Nodes represent biological entities (e.g, proteins, chemicals, etc.)

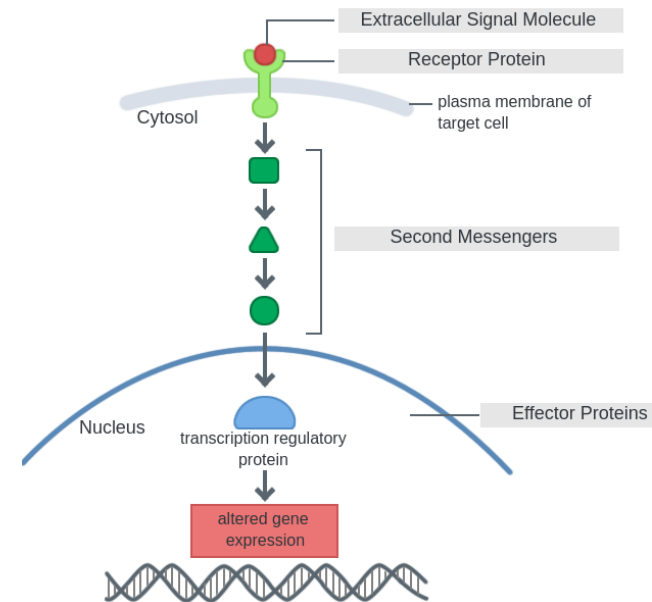 - Edges represent the type of interaction (e.g., activation, inhibition, etc.)

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

Extracellular Signal Molecule
Receptor Protein
Cytosol
plasma membrane of target cell
Second Messengers
Nucleus
transcription regulatory protein
Effector Proteins
altered gene expression

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

Fraunhofer
SCAI

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
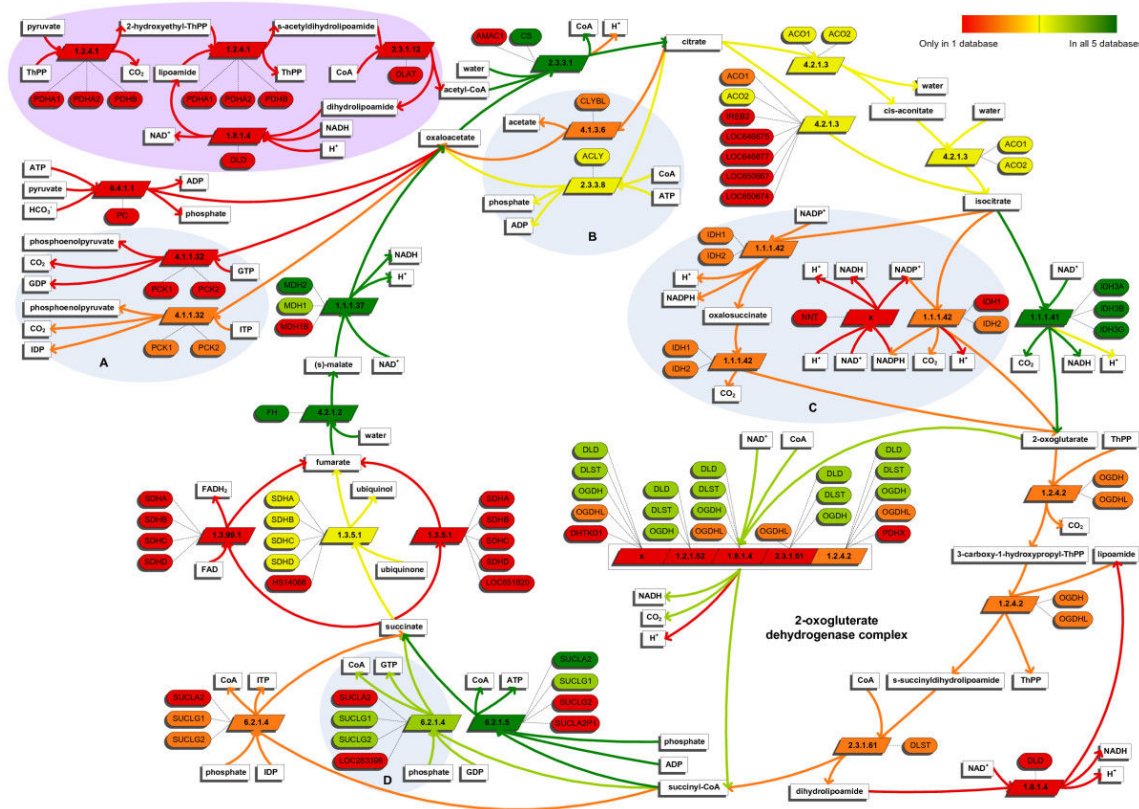
The Problem (2)

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
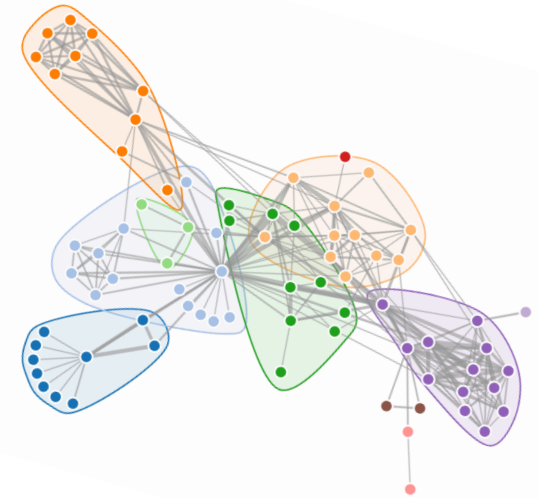*Green* = multiple databases

Integrating pathway databases

Harmonization of resources
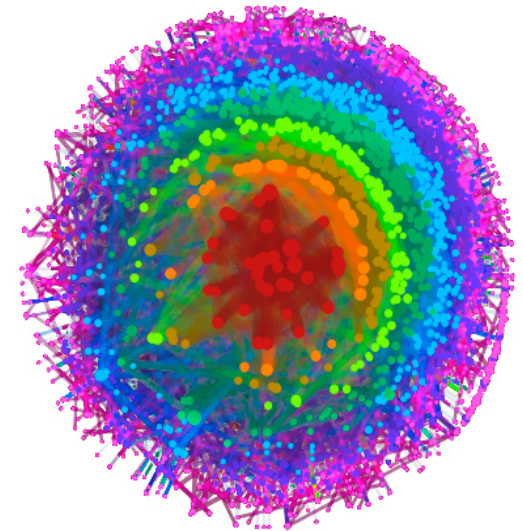
Integration in an unified schema

Integrating pathway databases

Harmonization of resources

Integration in an unified schema

Network representing biological interactions in multiple databases

8
7
6
5
4
3
2
1

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

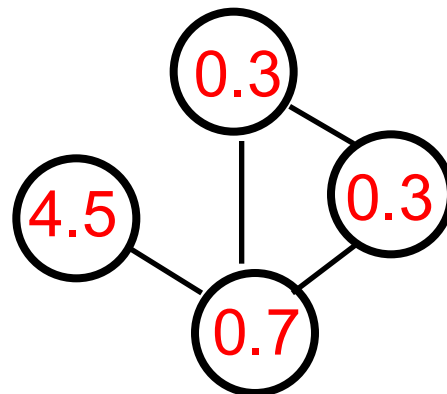**Different network features**

Prediction

Disease vs normal

Use patients' signatures to predict treatments

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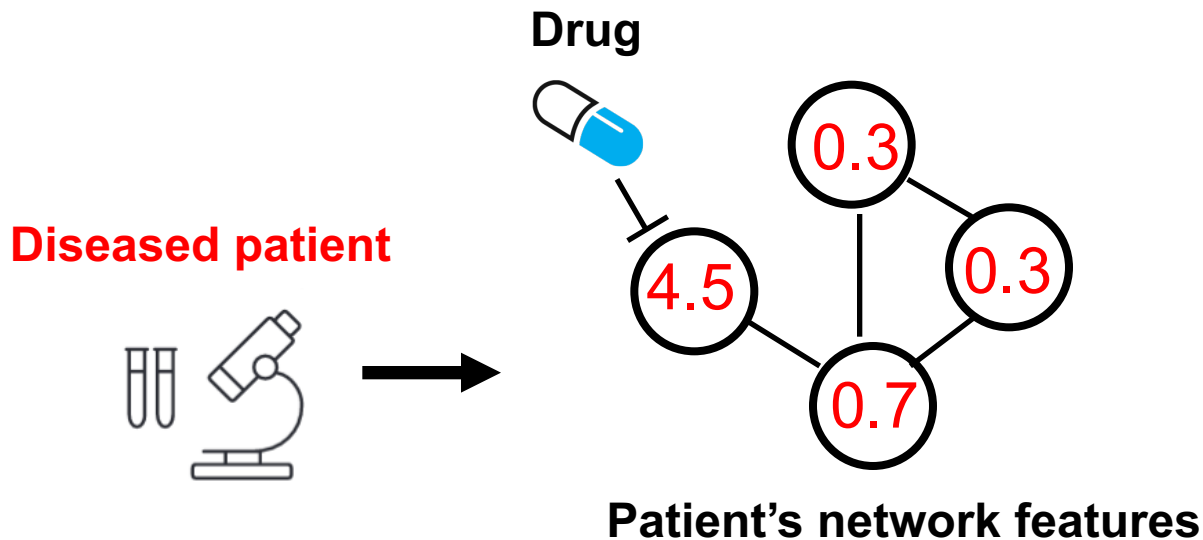- Use the trained classifier looking for a positive prediction

Diseased patient

Patient's network features

Use patients' signatures to predict treatments

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

Diseased patient

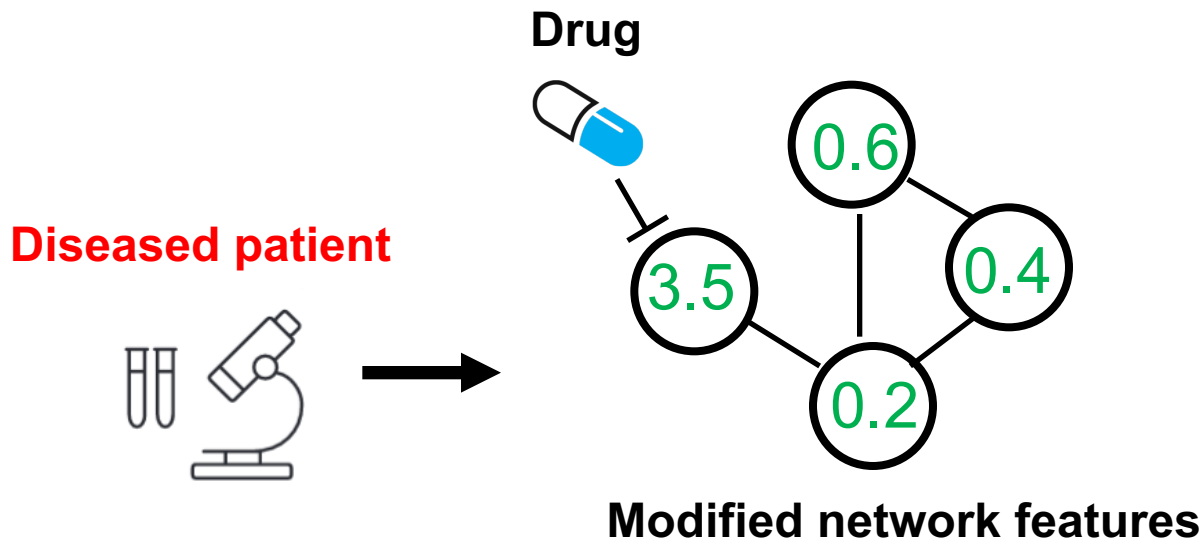**Drug**

Patient's network features

Use patients' signatures to predict treatments

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

Drug

Diseased patient

0.6

0.4

3.5

0.2

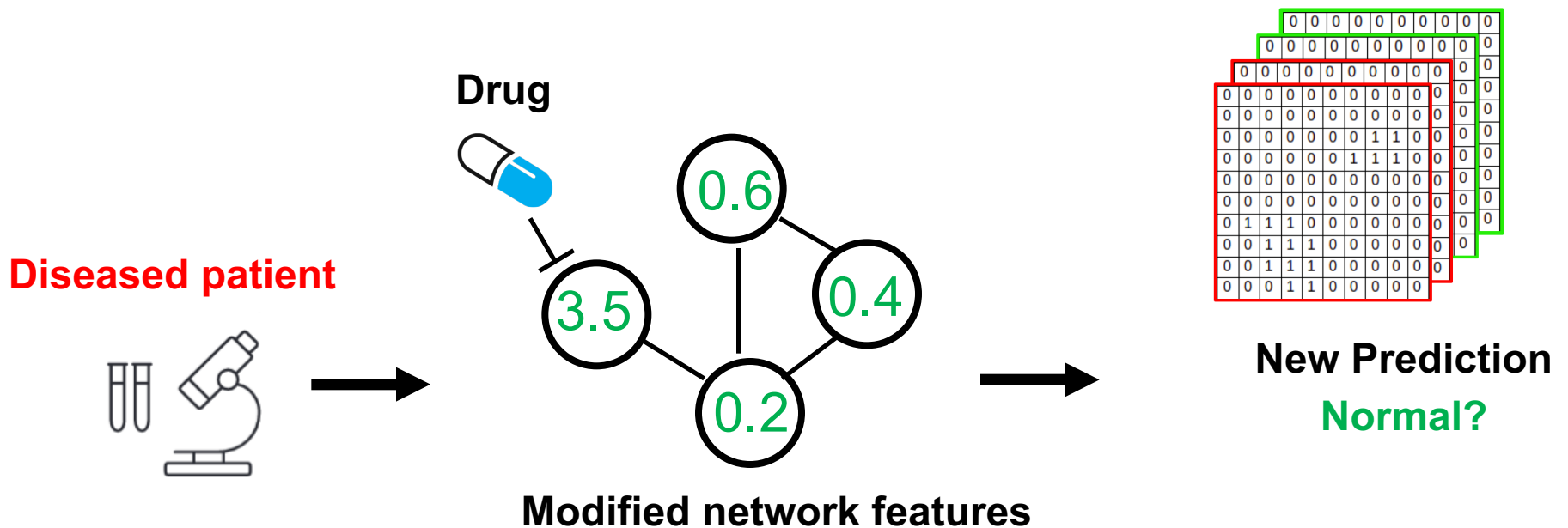**Modified network features**

Fraunhofer

SCAI

Use patients' signatures to predict treatments

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

Diseased patient

Drug

Modified network features

New Prediction
Normal?

Fraunhofer
SCAI

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?)

  - …

Fraunhofer
SCAI

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!

Fraunhofer

SCAI

Acknowledgments

- Prof. Dr. Martin Hofmann-Apitius

- Sarah Mubeen

- Josep Marín Llaó

- Charles Tapley Hoyt

Fraunhofer
SCAI

References

1. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., ... & Milacic, M. (2017). The reactome pathway knowledgebase. *Nucleic acids research*, *46*(D1), D649-D655

2. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, *45*(D1), D353-D361.

3. Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., ... & Ehrhart, F. (2017). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic acids research*, *46*(D1), D661-D667.

4. Stobbe, M. D., Houten, S. M., Jansen, G. A., van Kampen, A. H., & Moerland, P. D. (2011). Critical assessment of humanmetabolic pathway databases: a stepping stone for future integration. BMC systems biology, 5(1), 165.

Fraunhofer

SCAI

Evaluating biological knowledge (ongoing work)