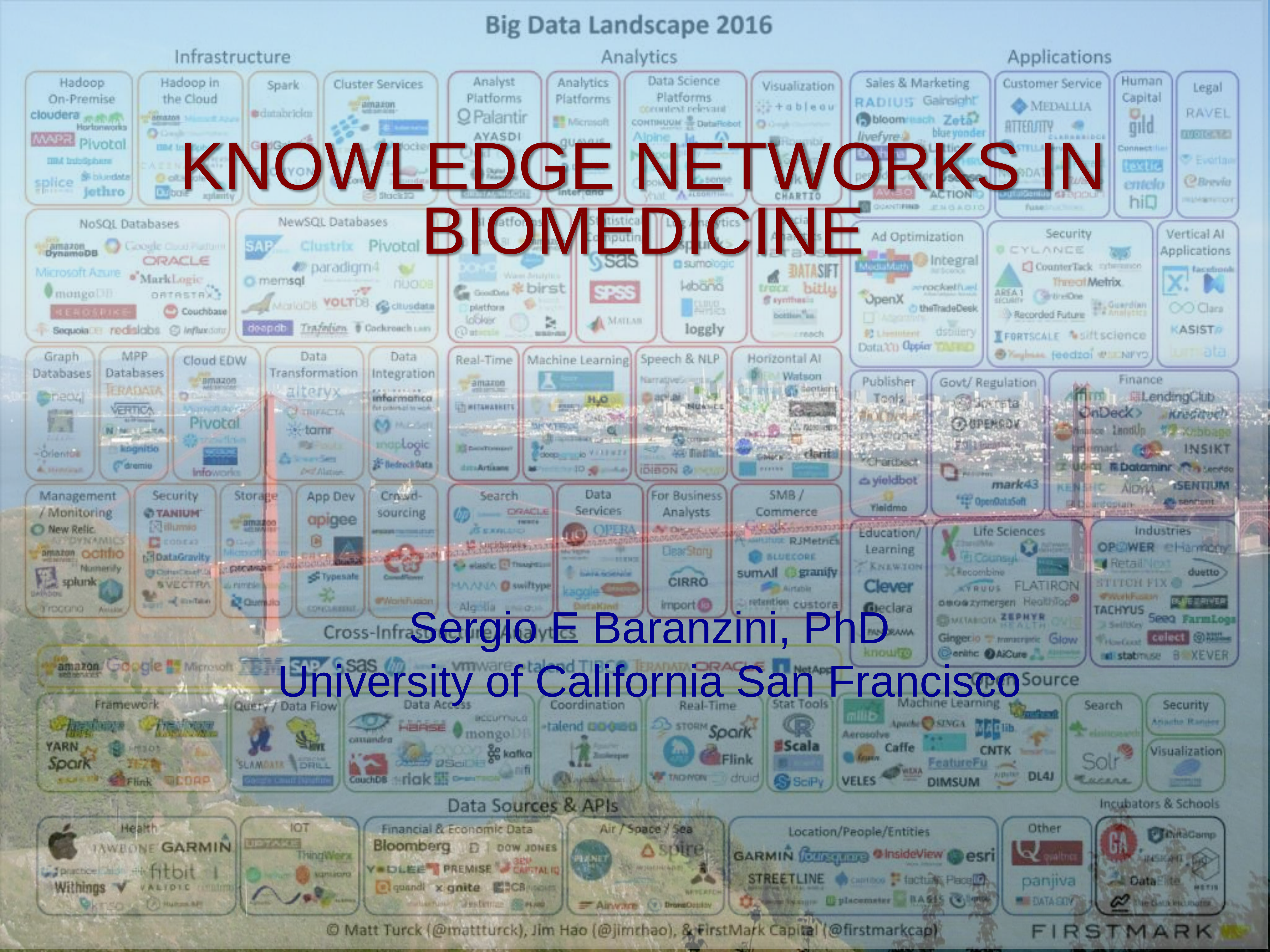
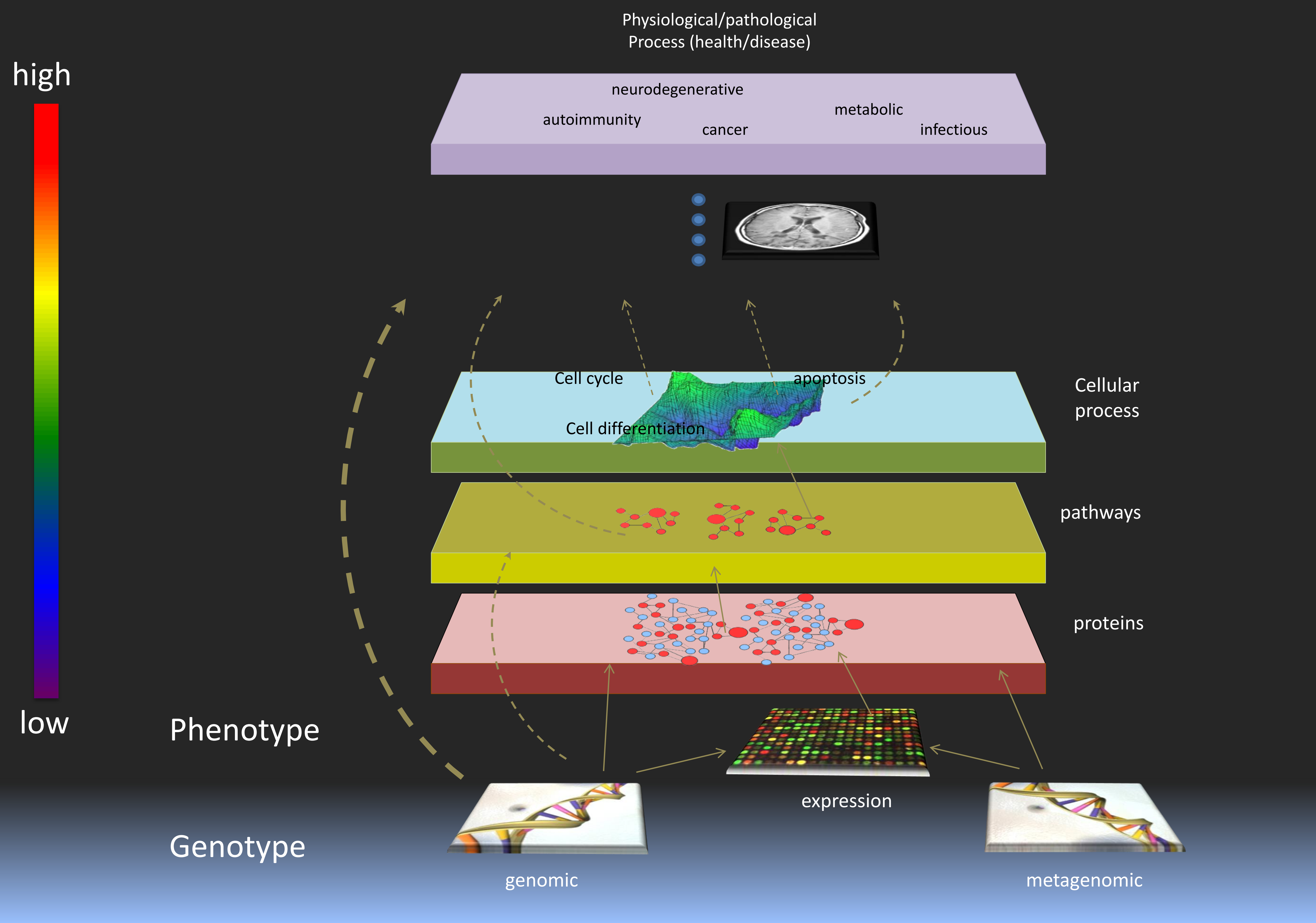
KNOWLEDGE NETWORKS IN BIOMEDICINE

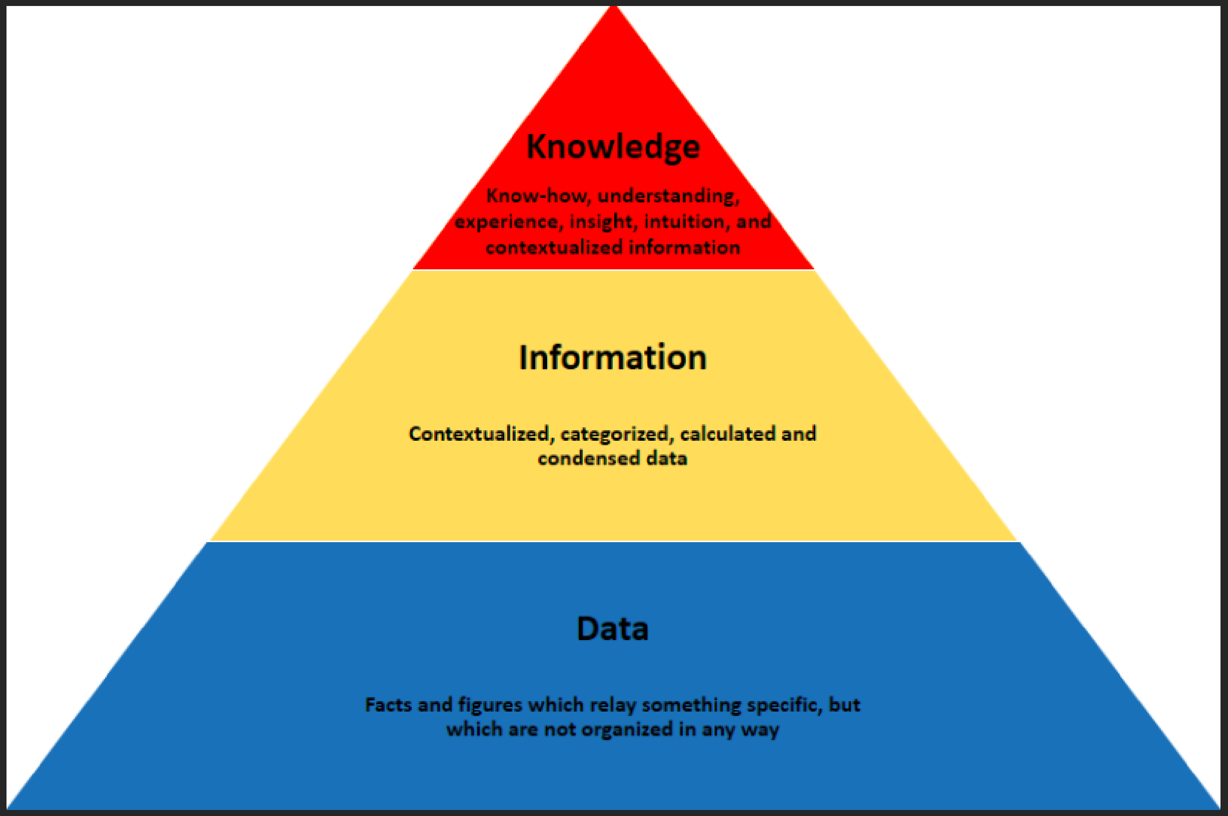
Sergio E Baranzini, PhD
University of California San Francisco

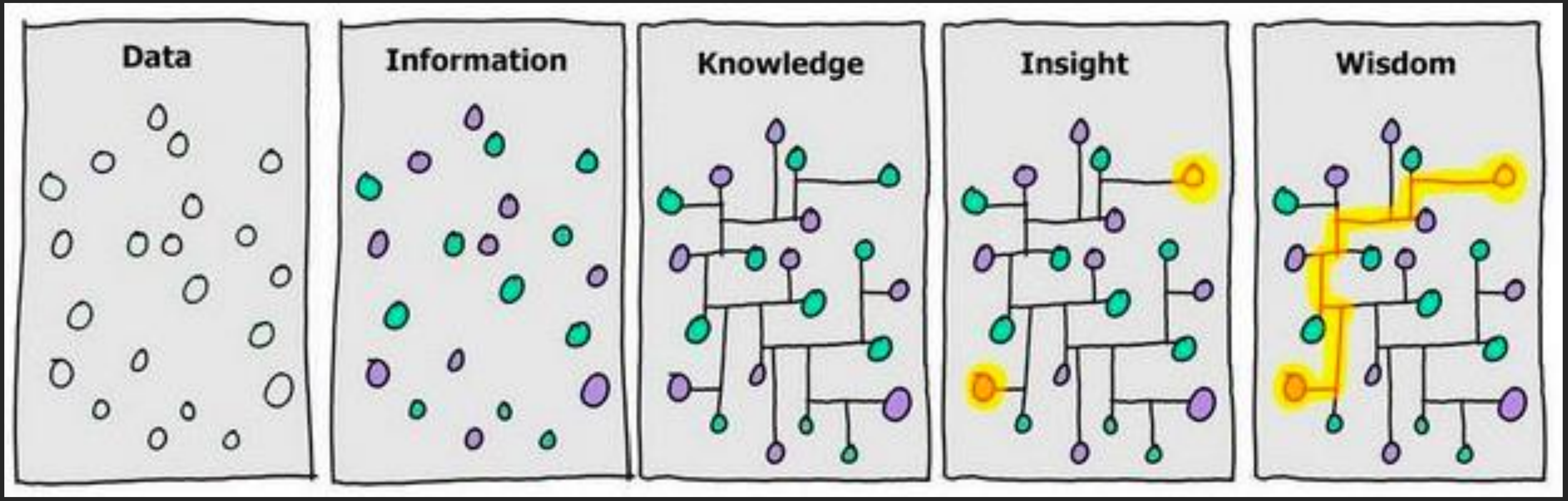


Hierarchical organization of biological complexity



Data -> information -> Knowledge



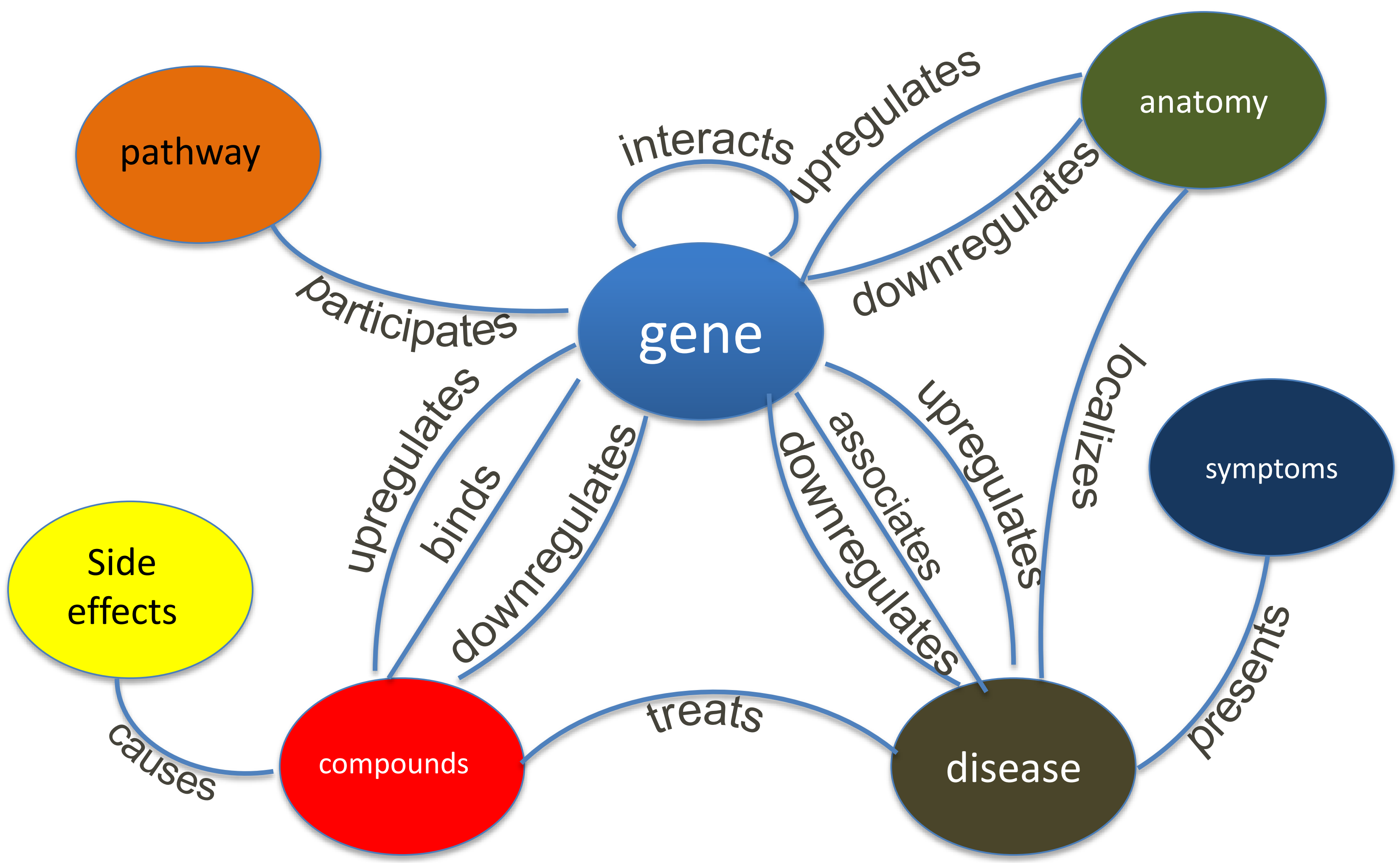


Why is an OKN needed in Biomedicine?

To make use of existing data!

- Reductionism vs. Gestalt
  - Eternal disagreement in “getting details right” vs. “whole picture”?
- Complex vs. complicated
  - Non-linear dynamics
  - Interdependence
  - Emergent properties
- Medicine is behind the curve
  - ML and AI developed decades ago
  - No fully taking advantage of computational power

SPOKE: A database of databases



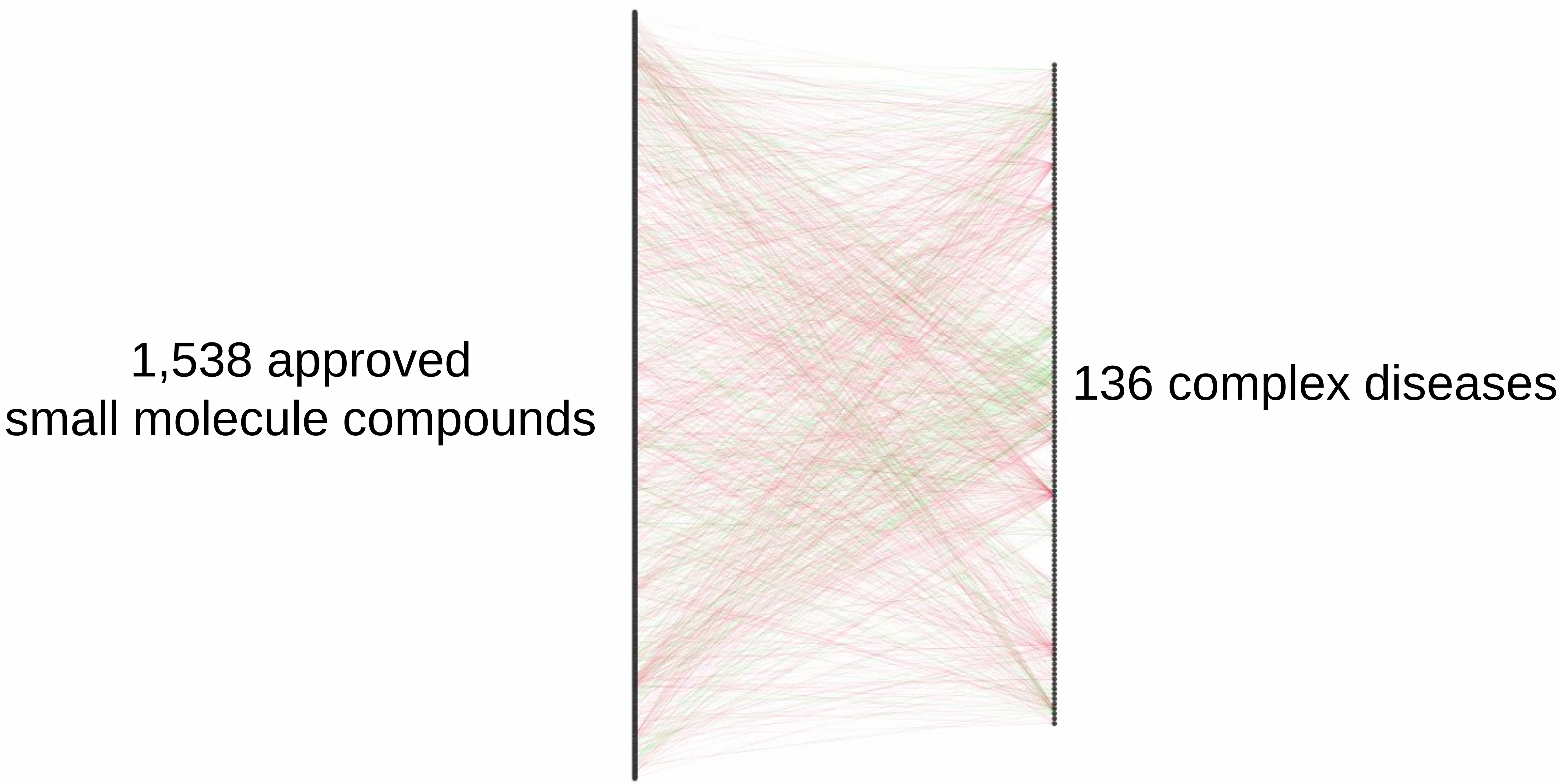
19 databases
50,000 nodes
2M edges

Uses for SPOKE

- Discovery/research
- Drug repurposing
- Context-dependent patient stratification
- Diagnosis/prognosis

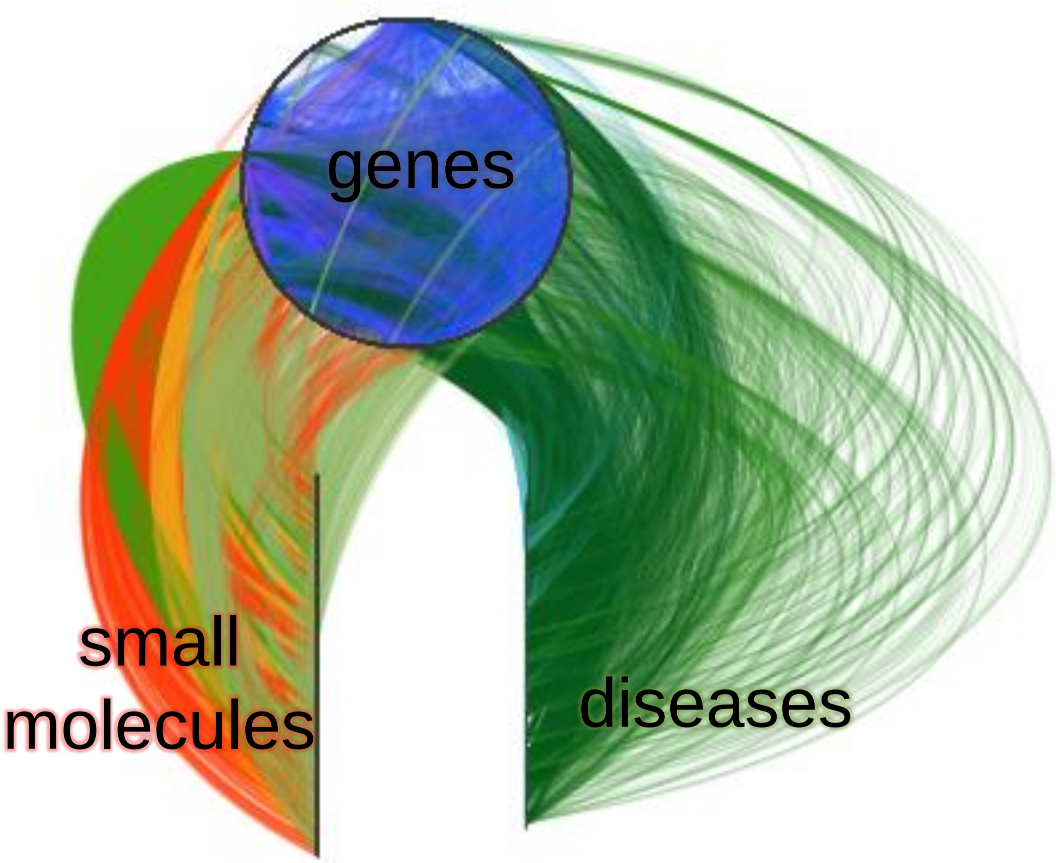
Can every drug be systematically repurposed?

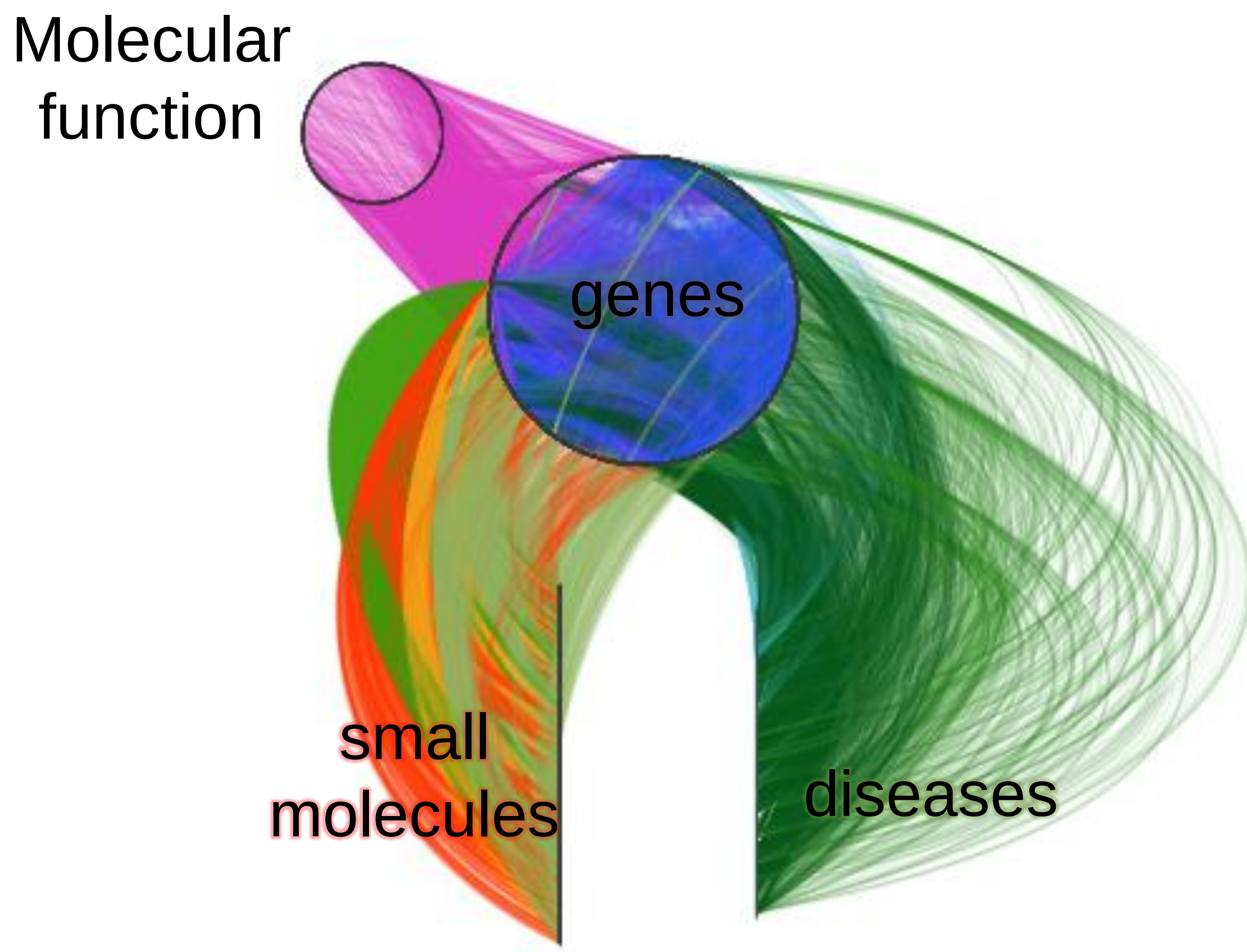
Imagine…


1,538 approved
small molecule compounds
136 complex diseases

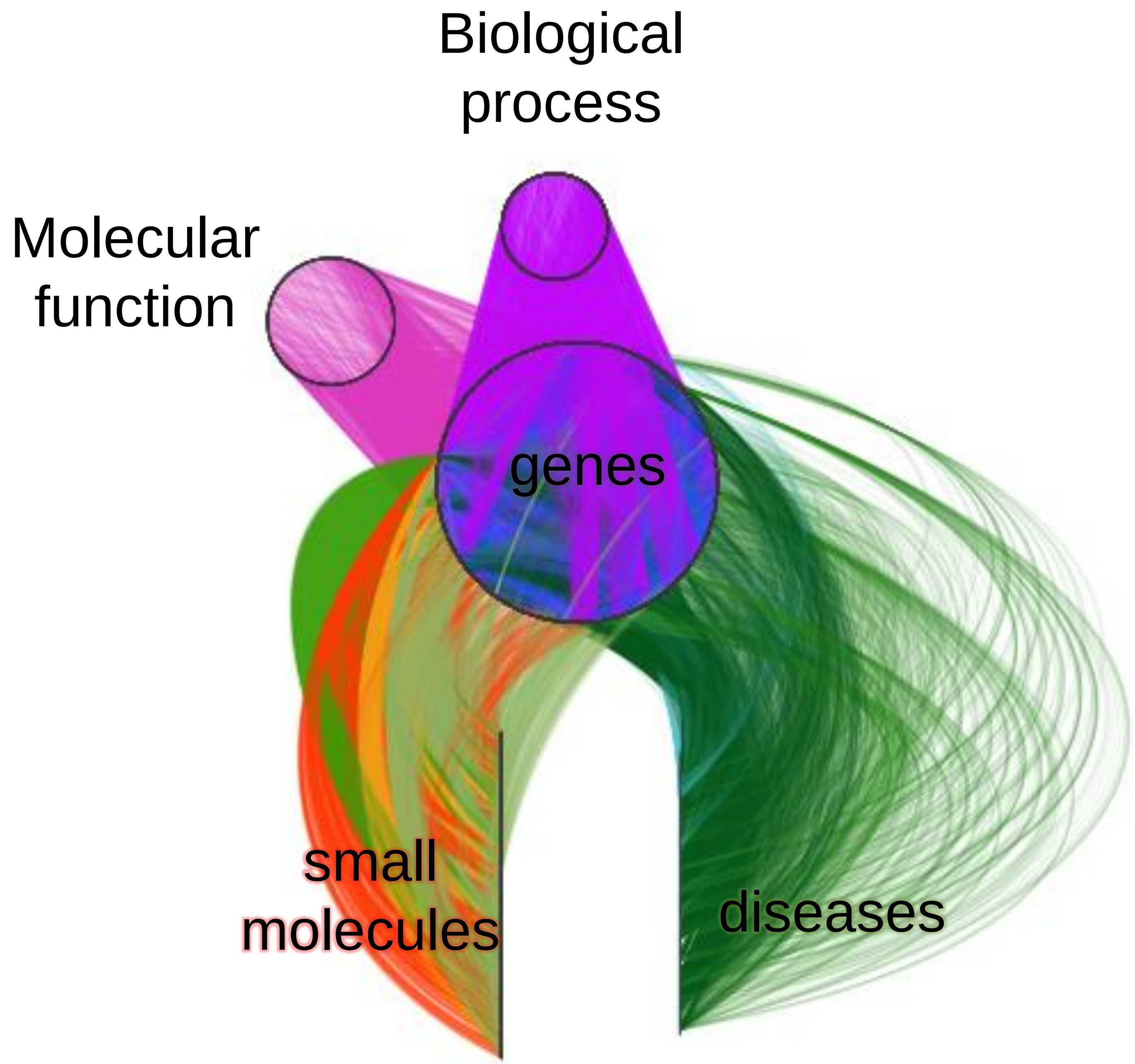
small molecules

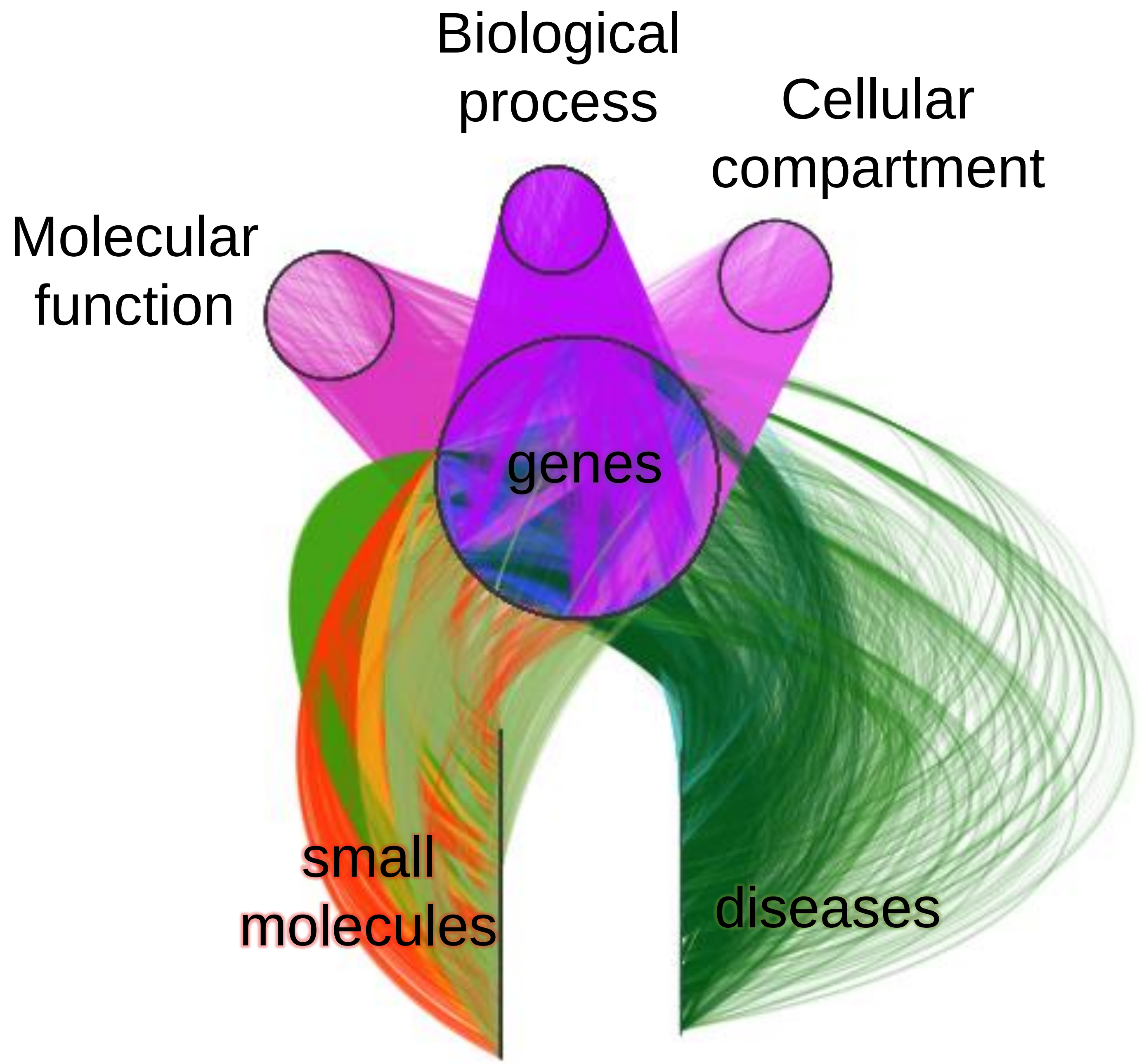
diseases

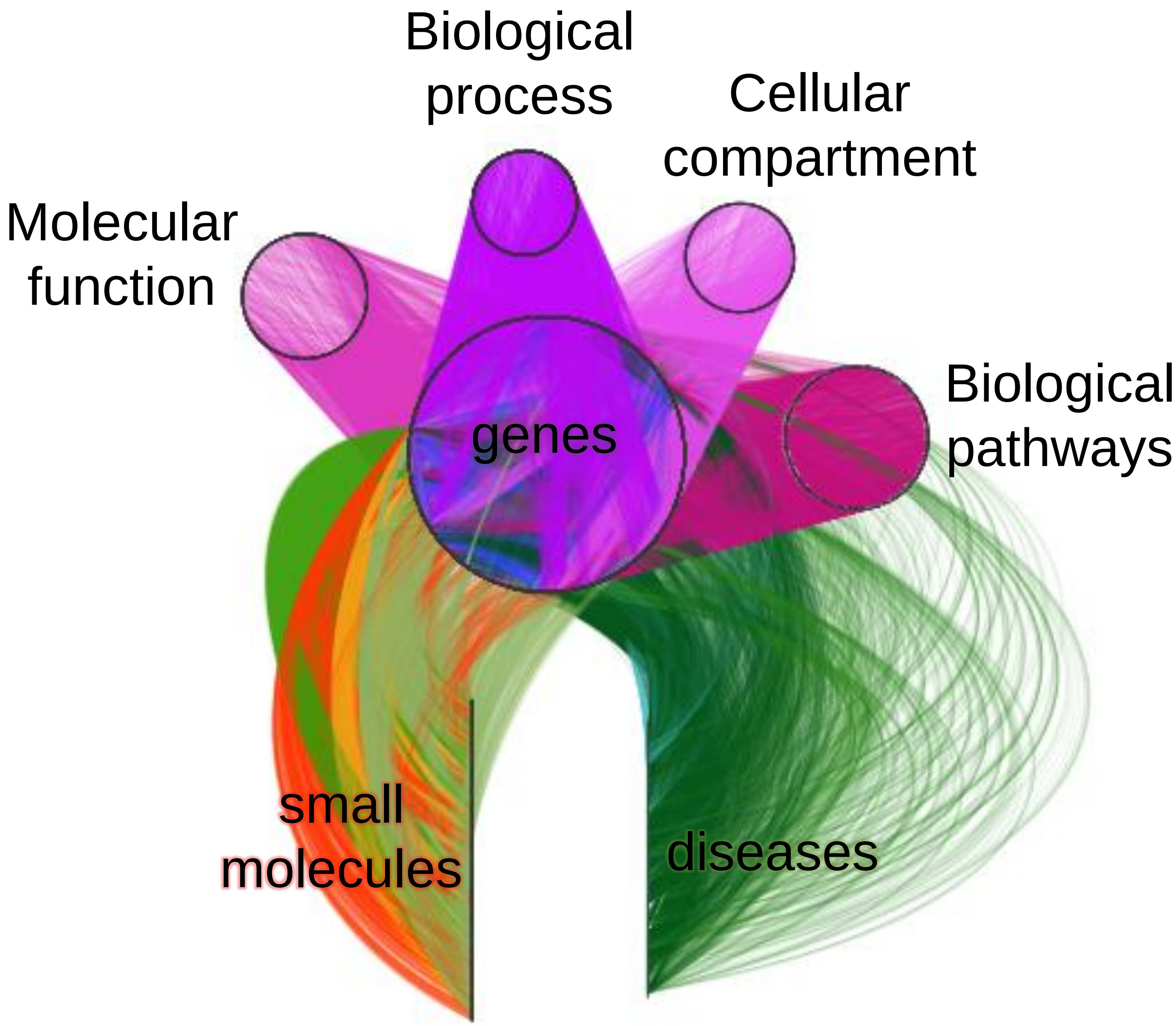

genes
small molecules
diseases



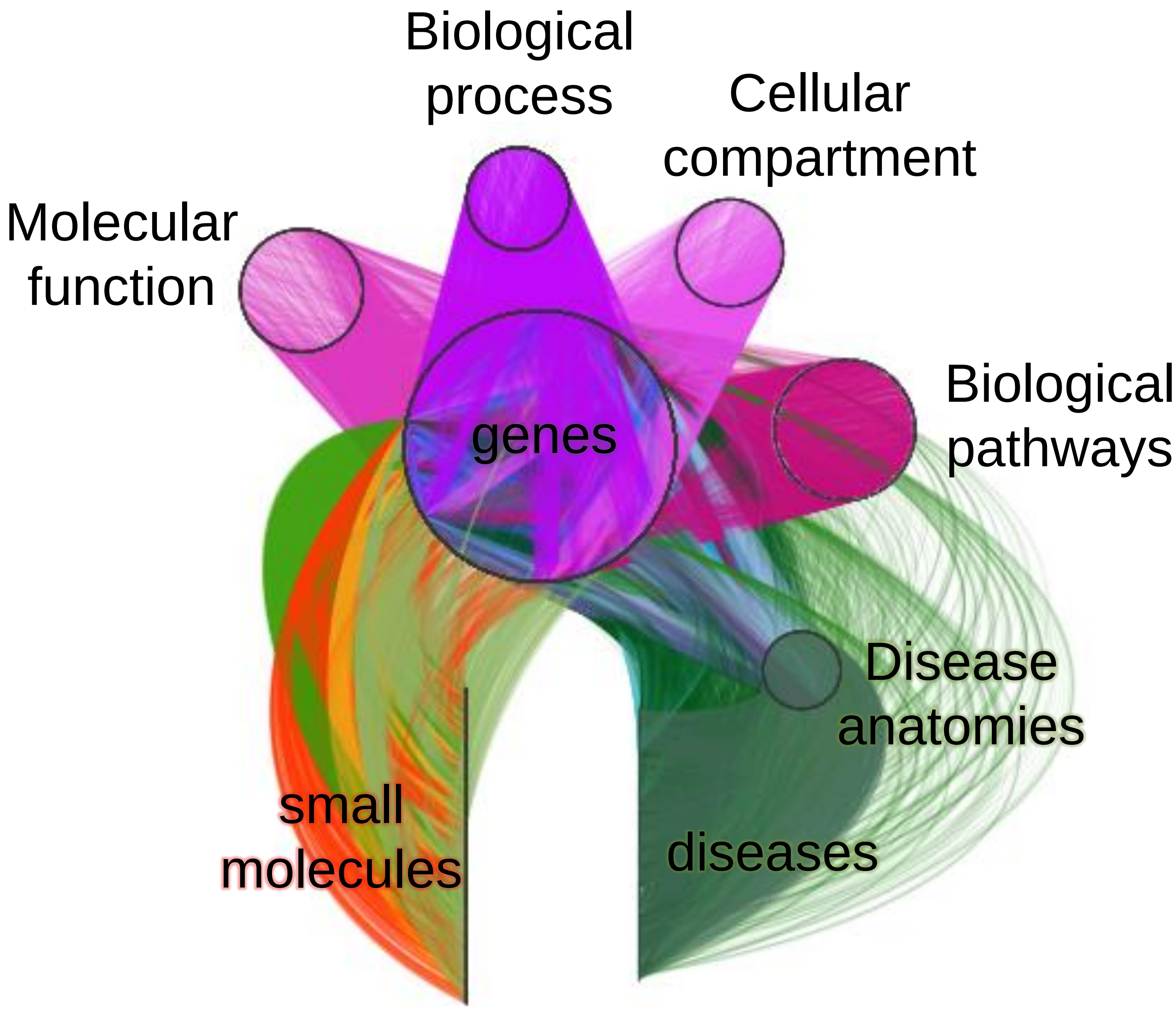
Molecular
function
genes
small
molecules
diseases

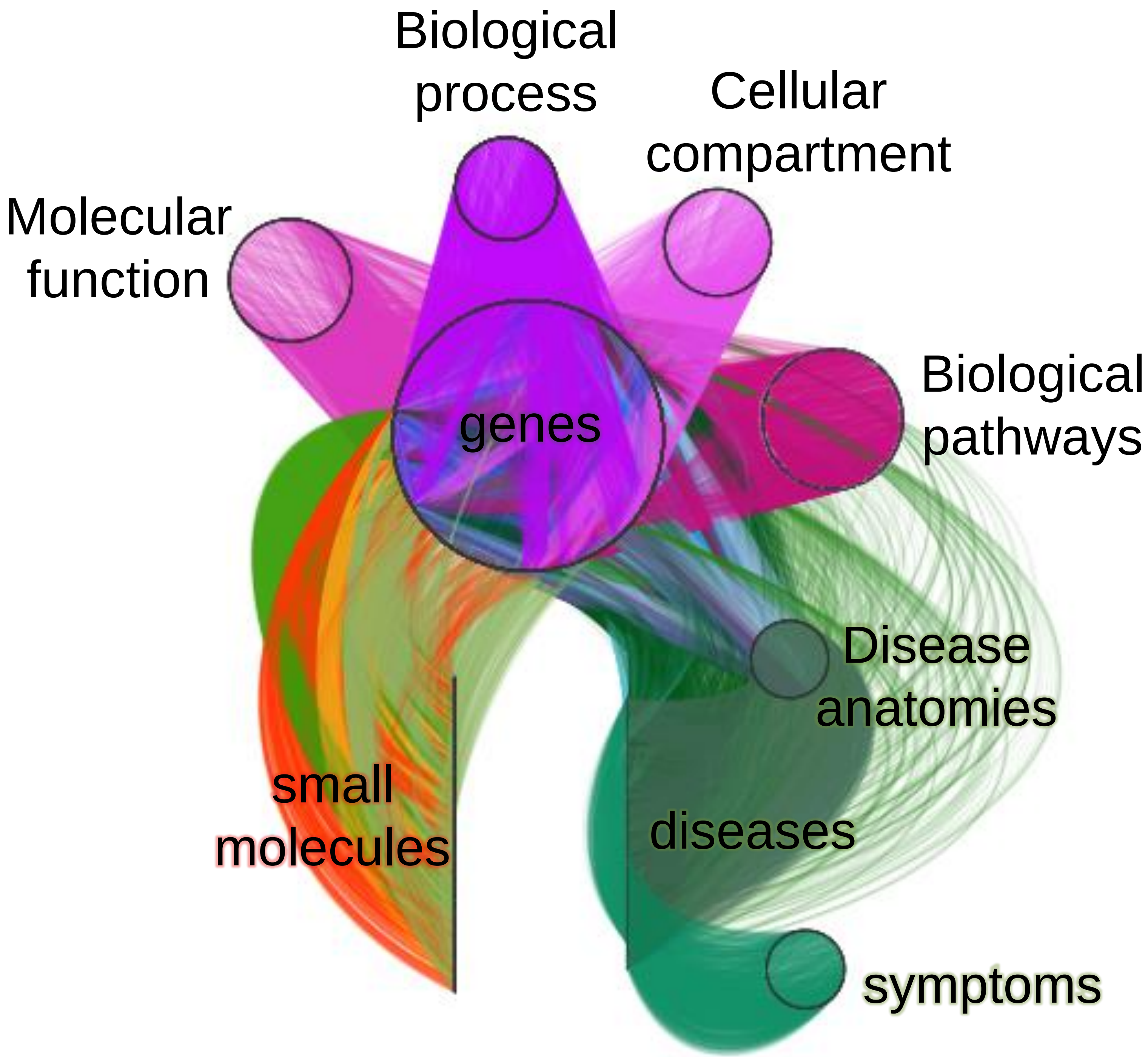

Biological
process
Molecular
function
genes
small
molecules
diseases


Biological process
Cellular compartment
Molecular function
genes
small molecules
diseases

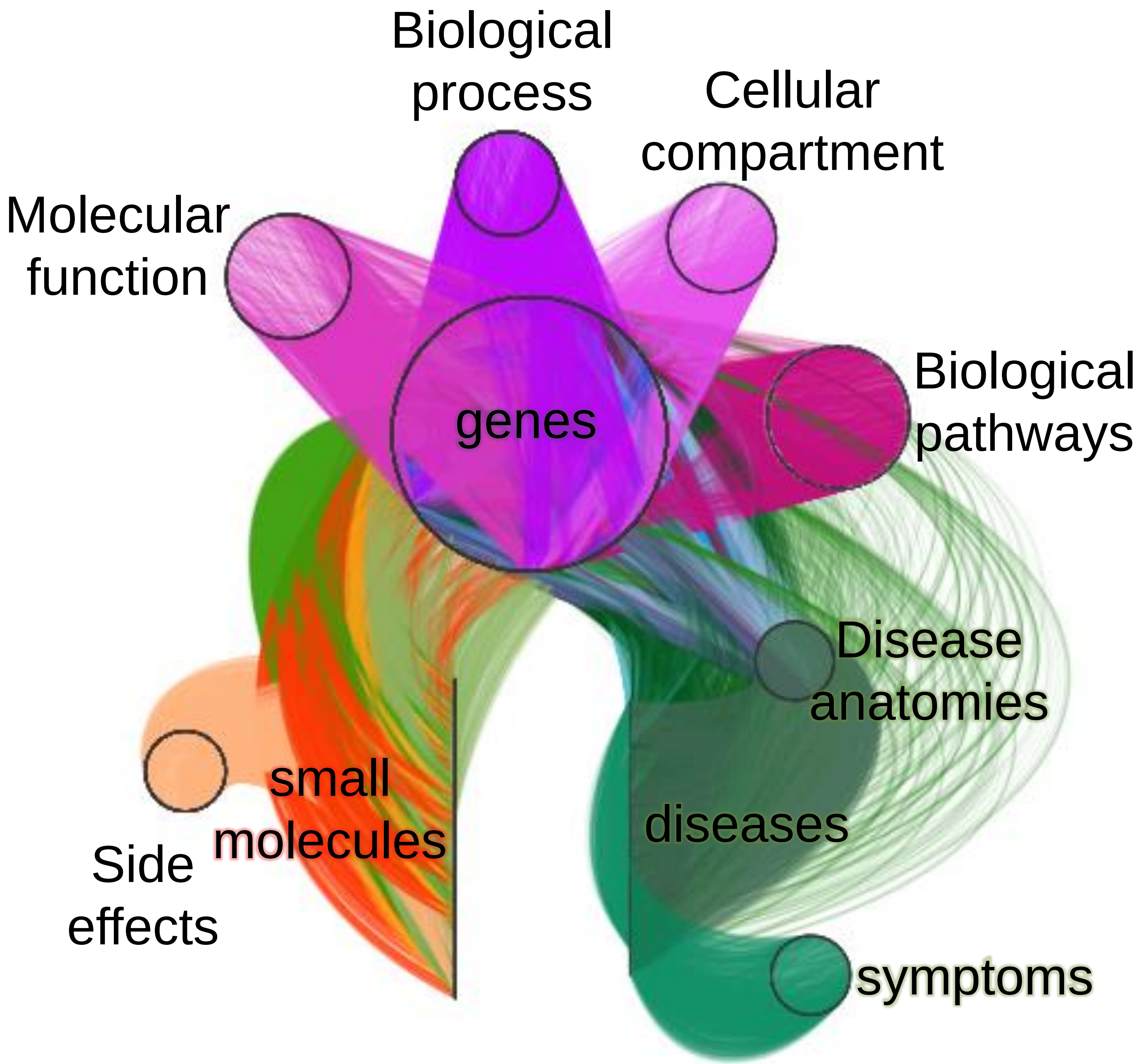


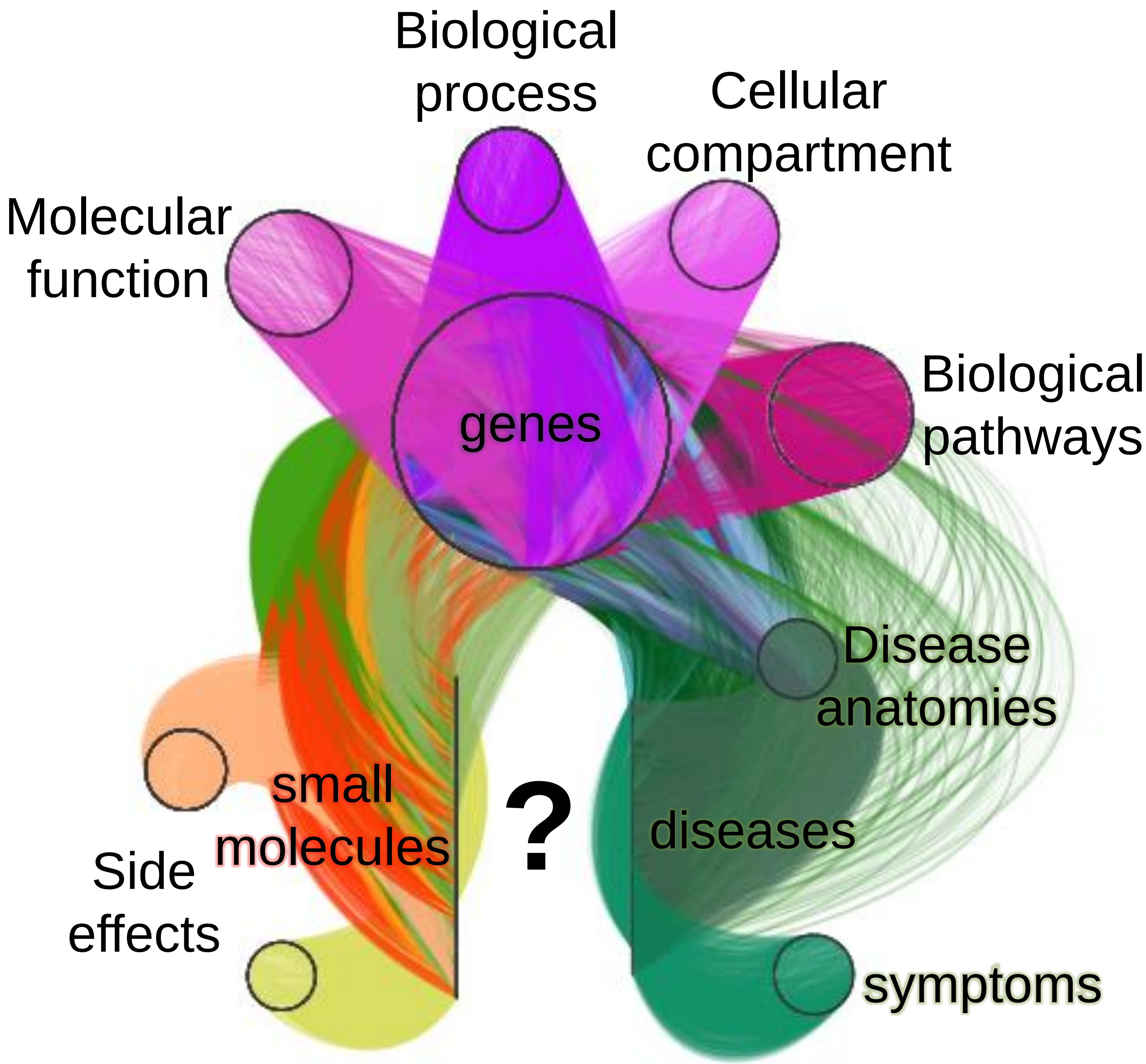
Biological process
Cellular compartment
Molecular function
genes
Biological pathways
small molecules
diseases


Biological process
Cellular compartment
Molecular function
genes
Biological pathways
Disease anatomies
small molecules
diseases



Biological process
Cellular compartment
Molecular function
genes
Biological pathways
Disease anatomies
small molecules
diseases
symptoms


Biological process
Cellular compartment
Molecular function
genes
Biological pathways
Disease anatomies
small molecules
Side effects
diseases
symptoms

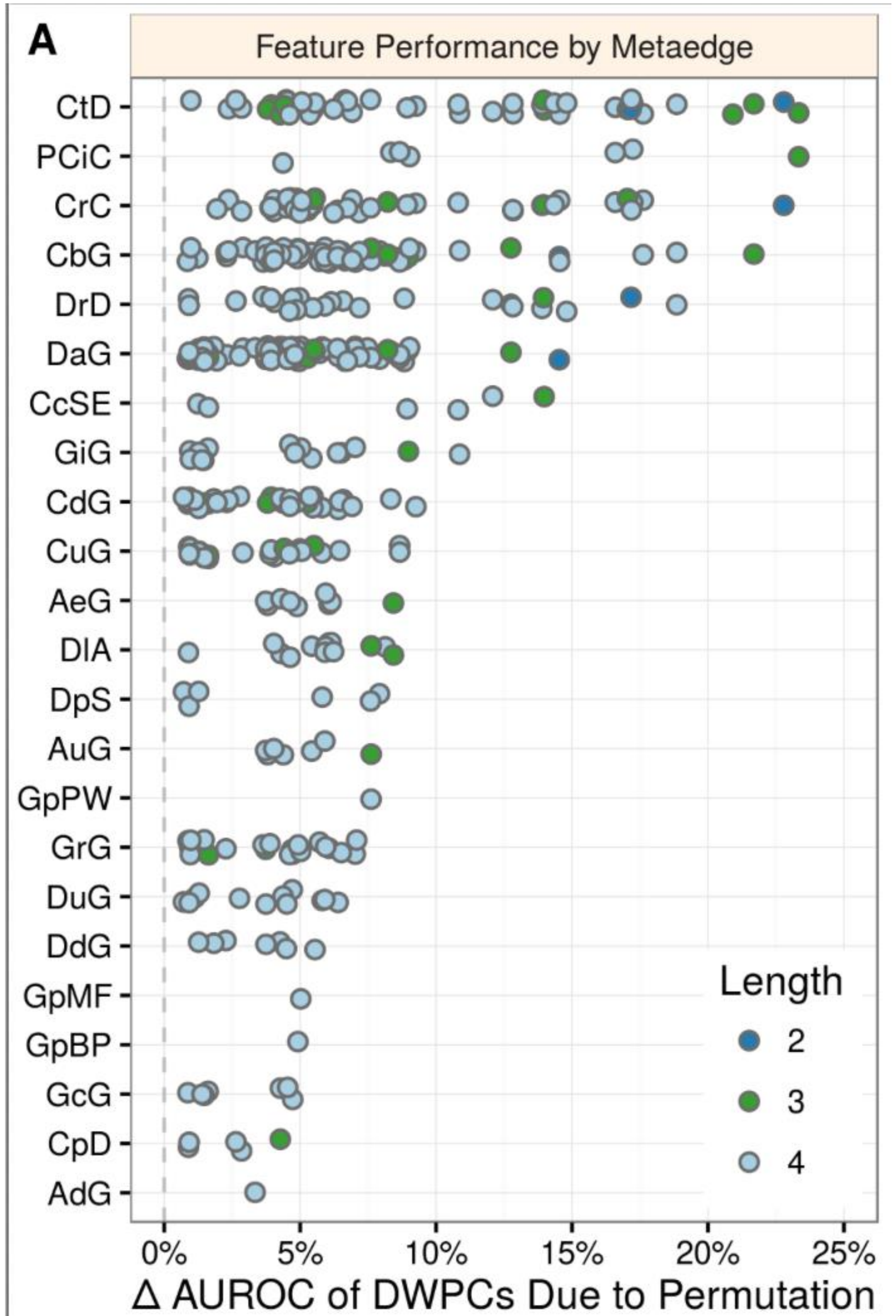


47,031 nodes (11 types)
2,250,197 relationships (24 types)

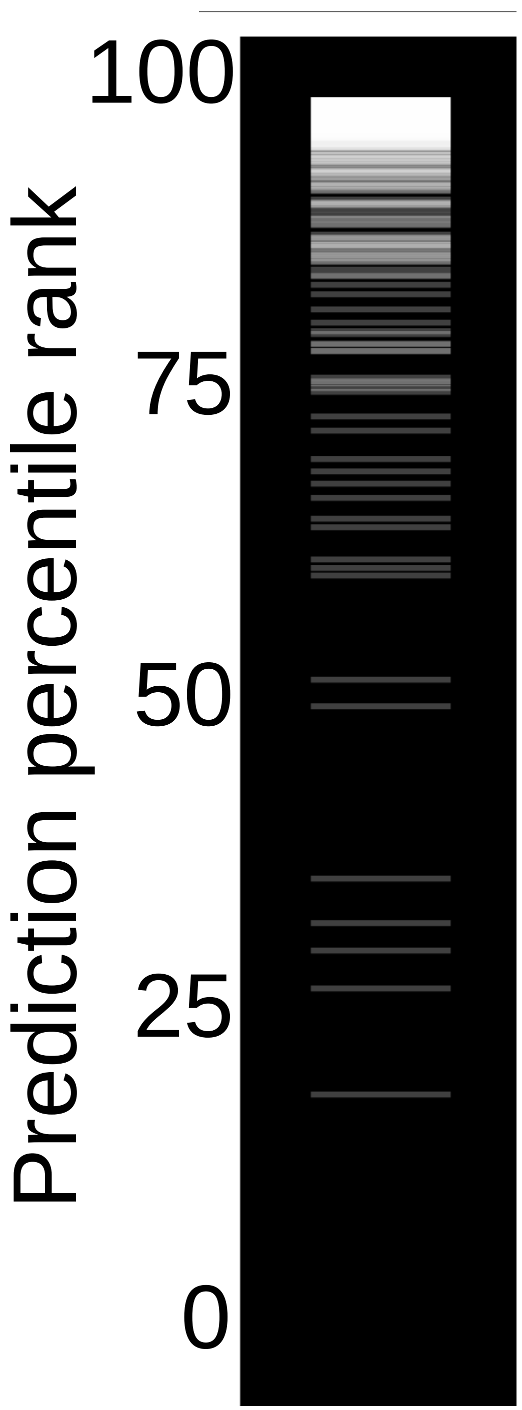
Pipeline summary

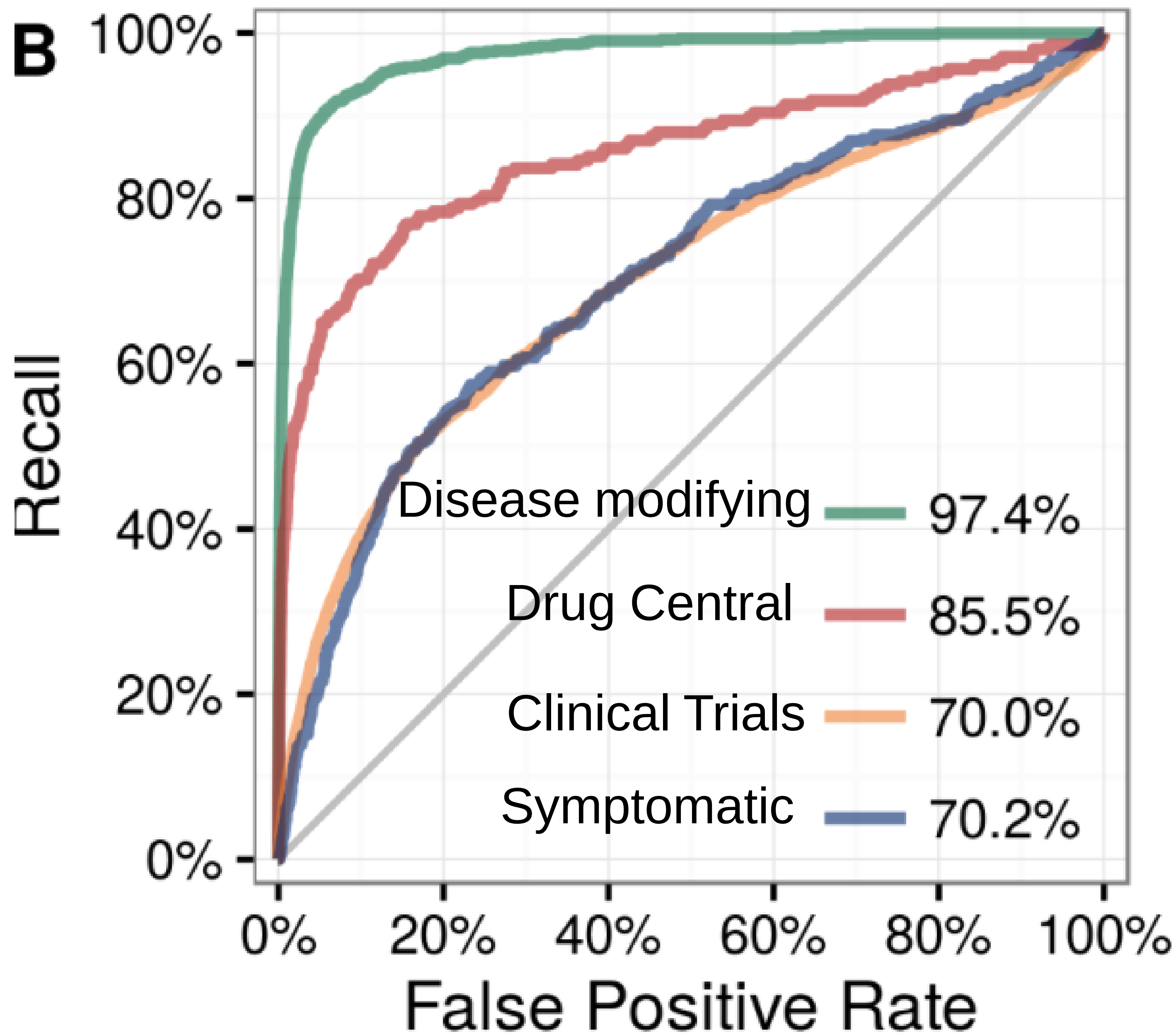
- Created Hetionet v1.0 — an integrative network with 2,250,197 relationships of 24 types.
- Extracted features from the network (to quantify the prevalence of specific path types between each compound and disease). 46.8M paths!
- Fitted regularized regression model (to translate from network-based features to a probability of treatment for a given compound–disease pair).
- Permuted the network (to reduce false positives)

Feature contribution



Prediction Performance



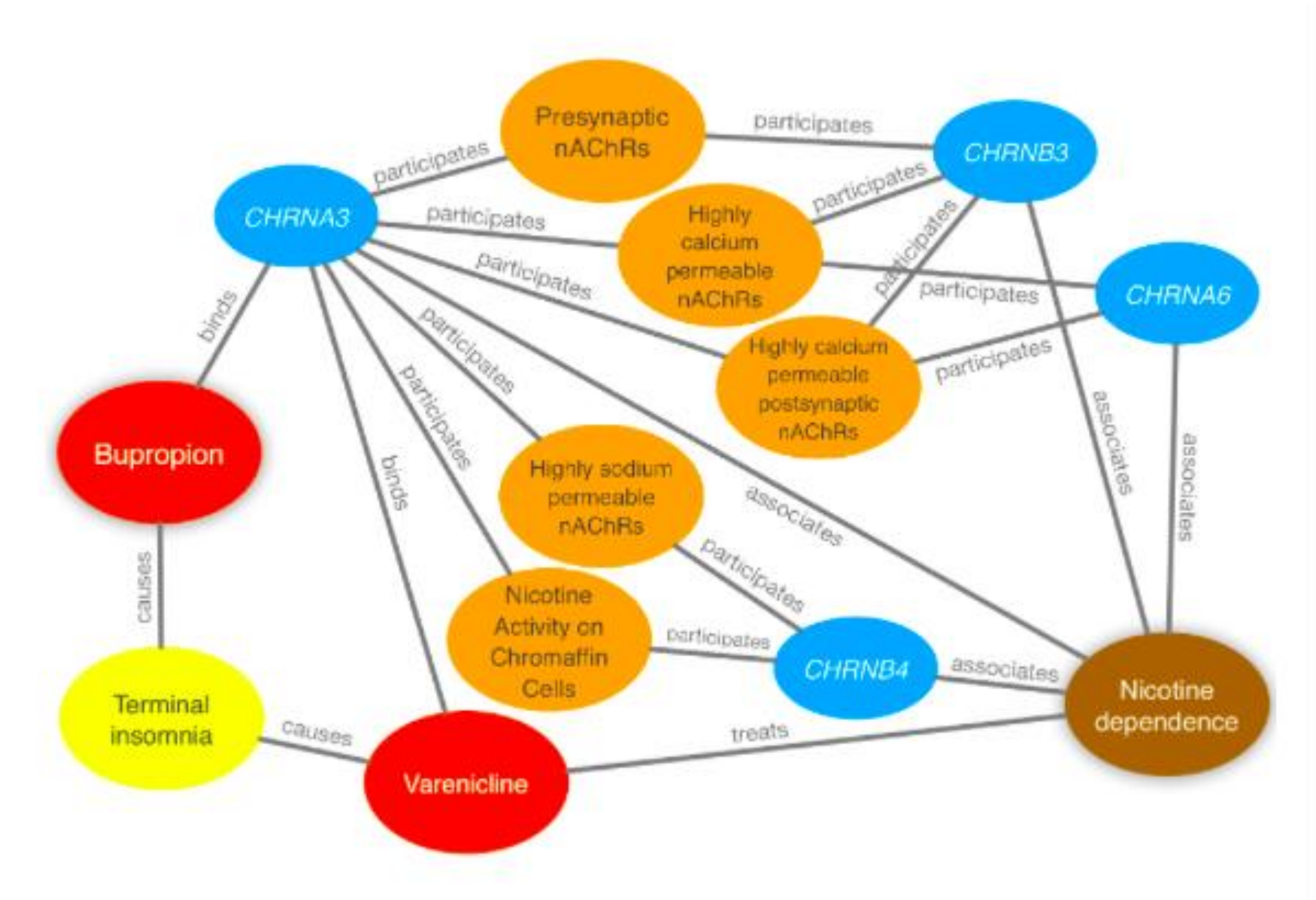


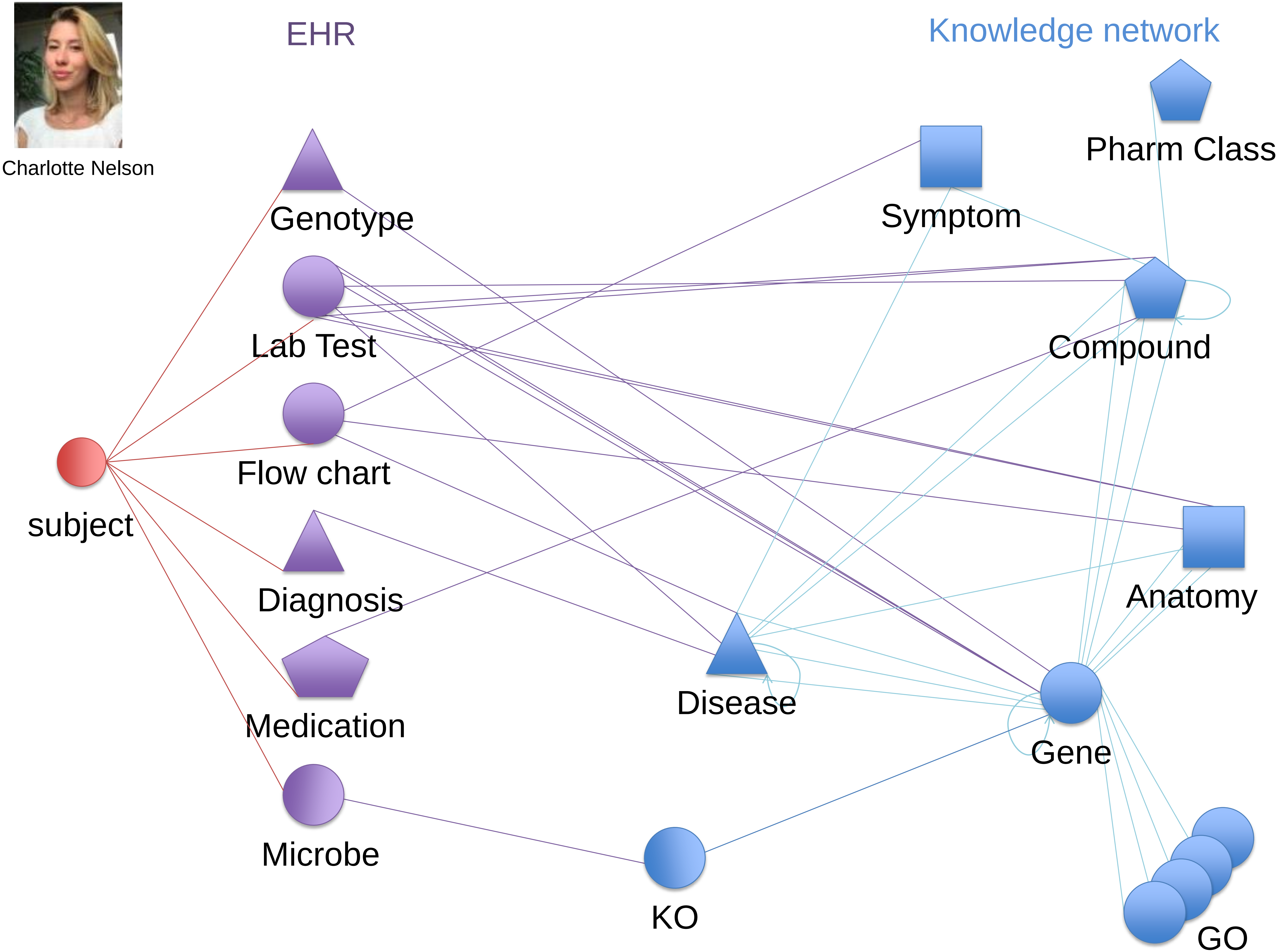
Top predictions

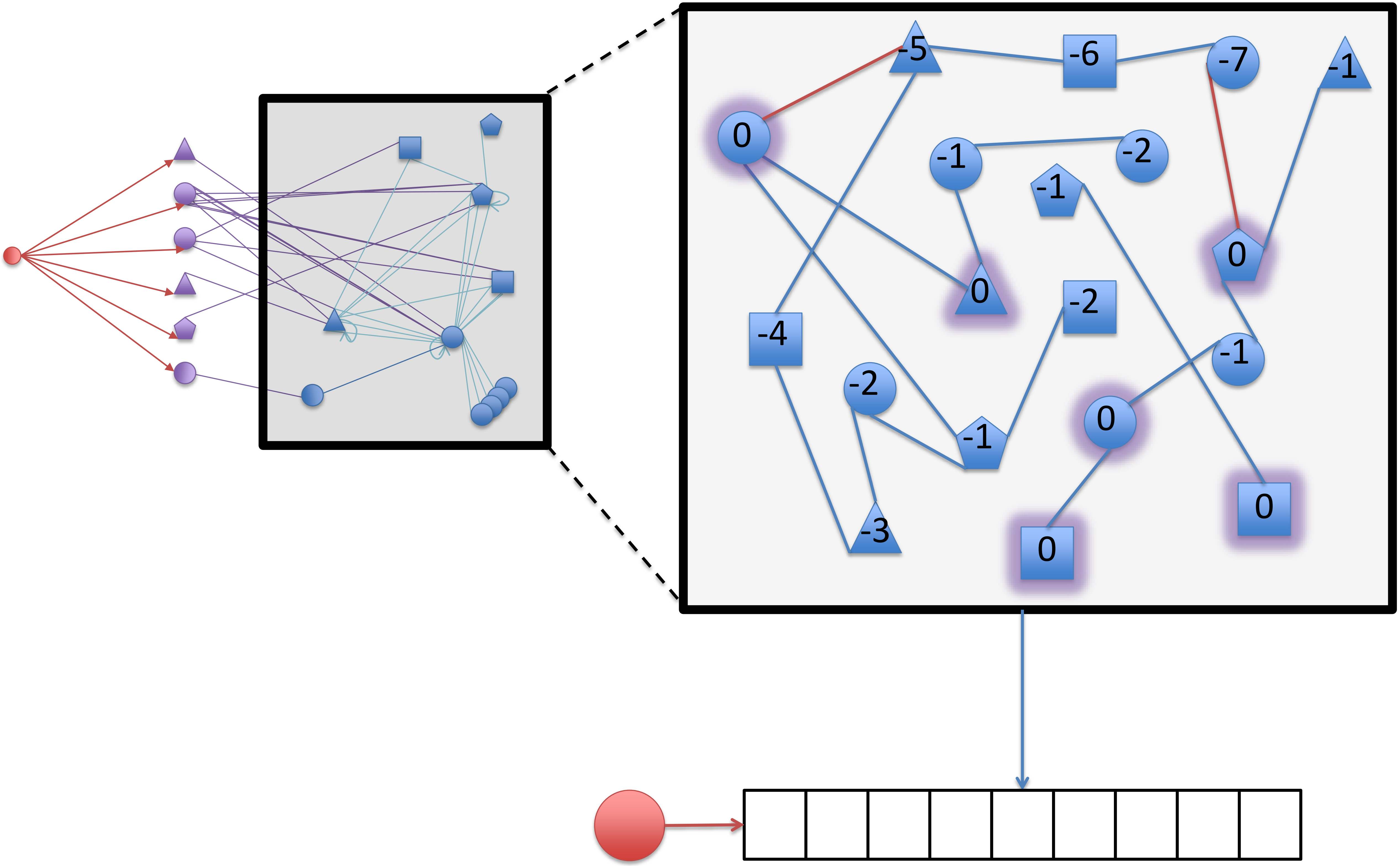
	A	B	C	D	E	F	G	H
1	compound_name	disease_name	category	prediction	compound_percentile	disease_percentile	prior_prob	n_trials
2	Pamidronate	osteoporosis	DM	88.69%	100.00%	100.00%	3.89%	0
3	Alendronate	osteoporosis	DM	88.50%	100.00%	99.93%	3.89%	68
4	Risedronate	osteoporosis	DM	88.11%	100.00%	99.87%	3.89%	0
5	Esomeprazole	Barrett's esophagus	DM	82.31%	100.00%	100.00%	0.23%	7
6	Ibandronate	Paget's disease of bone		80.26%	100.00%	100.00%	0.72%	0
7	Glyburide	type 2 diabetes mellitus	DM	79.96%	100.00%	100.00%	5.88%	26
8	Omeprazole	Barrett's esophagus	DM	78.68%	100.00%	99.93%	0.23%	11
9	Alendronate	Paget's disease of bone	DM	76.98%	99.26%	99.93%	1.49%	2
10	Etidronic acid	Paget's disease of bone	DM	74.88%	100.00%	99.87%	1.49%	2
11	Pamidronate	Paget's disease of bone	DM	72.96%	99.26%	99.80%	1.49%	0
12	Furosemide	hypertension	DM	71.71%	100.00%	100.00%	28.33%	4
13	Risedronate	Paget's disease of bone	DM	71.71%	99.26%	99.74%	1.49%	0
14	Ibandronate	osteoporosis	DM	71.47%	99.26%	99.80%	1.92%	39
15	Etidronic acid	osteoporosis	DM	68.64%	99.26%	99.74%	3.89%	15
16	Bumetanide	hypertension	DM	68.42%	100.00%	99.93%	11.06%	0
17	Olsalazine	Crohn's disease		66.53%	100.00%	100.00%	0.72%	0
18	Aminophylline	asthma	DM	64.97%	100.00%	100.00%	10.43%	3
19	Methotrexate	lung cancer	DM	61.48%	100.00%	100.00%	41.76%	0
20	Paricalcitol	osteoporosis		60.35%	100.00%	99.67%	0.00%	0
21	Topiramate	epilepsy syndrome	DM	60.27%	100.00%	100.00%	10.13%	35
22	Ethotoin	epilepsy syndrome		58.85%	100.00%	99.93%	0.00%	0
23	Losartan	hypertension	DM	57.33%	100.00%	99.87%	28.33%	79

Between 30-300 fold increase over null!

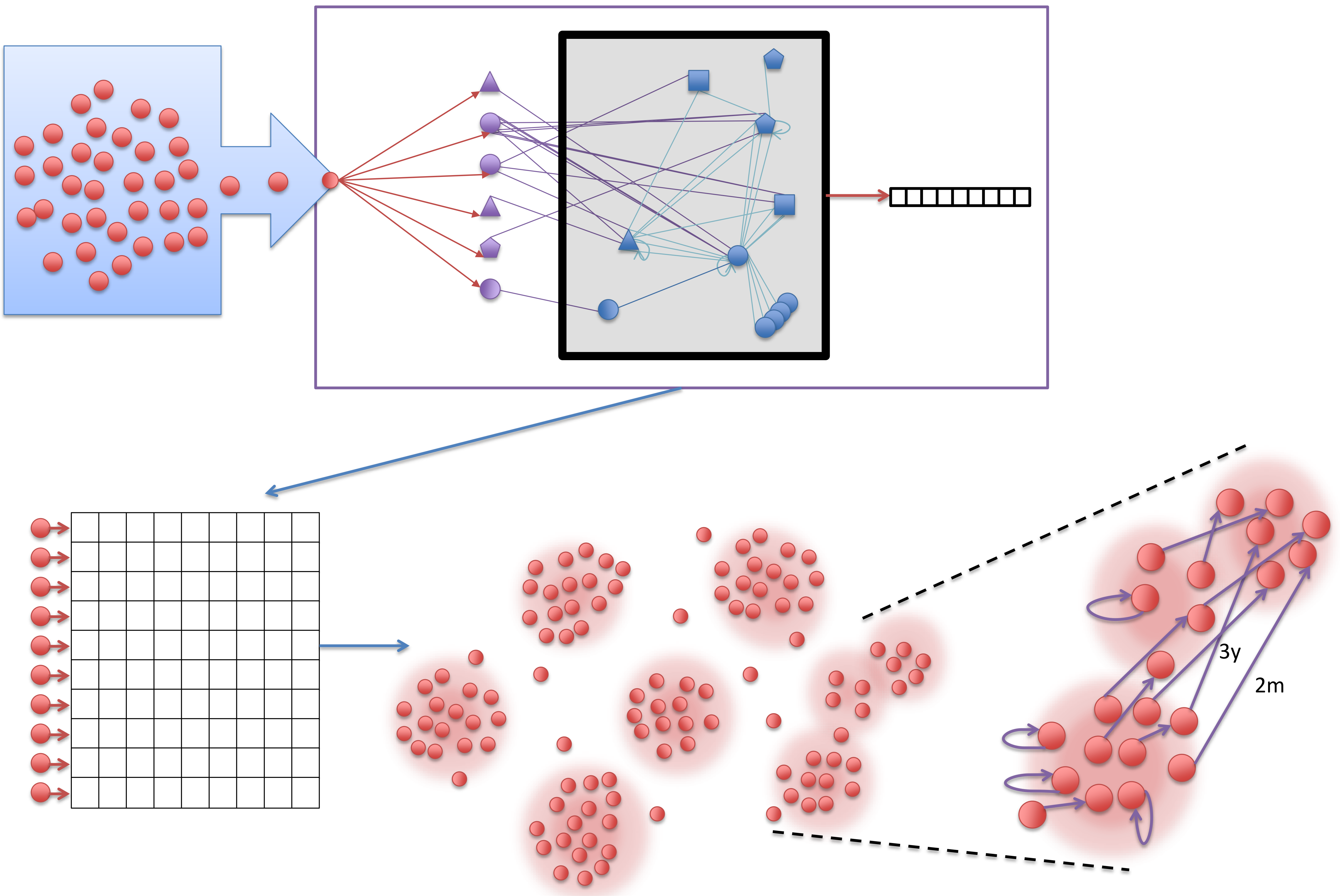
Bupropion may treat nicotine dependence




Charlotte Nelson
EHR
Knowledge network
Genotype
Lab Test
Flow chart
subject
Diagnosis
Medication
Microbe
KO
Disease
Gene
GO
Symptom
Pharm Class
Compound
Anatomy



Network Signature for Patient i at time point j



Find hubs by clustering network signatures

UCSF Knowledge Network

Home | Monthly Meetings | Documents

UCSF Knowledge Network*

Principal Investigators: Sourav Bandyopadhyay, Sergio Baranzini & Michael Keiser

*subject to renaming and new graphics

SPOKE
Scalable PrecisiOn medicine Knowledge Engine

UCSF University of California, San Francisco | About UCSF | Search UCSF | UCSF Medical Center

Precision Medicine at UCSF | About | Elements of Precision Medicine | Programs | News | Resources

Home > Content > Computational Health Sciences

Computational Health Sciences

Computational Health Sciences: Atul Butte, MD, PhD

Precision Medicine at UCSF | About | Elements of Precision Medicine

Home > UCSF Precision Medicine Platform Committee

UCSF Precision Medicine Platform Committee

Leadership

Chair: Keith Yamamoto, PhD

UCSF SPOKE Team

Sourav Bandyopadhyay

Michael Keiser

Sharat Israni

Charlotte Nelson

Krish Bharat

Scooter Morris

Bioscience breakout

- Brief background of OKN, meeting goals, and discussion topics
- What is the state-of-the-art for open knowledge networks in BIO
- What are the main gaps/challenges, why do they exist, and how do we address them
- What is different about BIO from other domains (geosciences, finance, etc)
- What does BIO share with other domains
- Possible next steps

Goals

- OKN
- To gather *experts'* opinions on KN

State of the art

- Watson (IBM)
- SNORKEL (Stanford)
- NCATS (NIH)
- SPOKE (UCSF)
- Others?

Challenges to BIO OKN

- Experts on multiple disciplines are needed

Why is BIO unique?

- Scale
  - Several orders of magnitude
- Complexity
  - unmatched