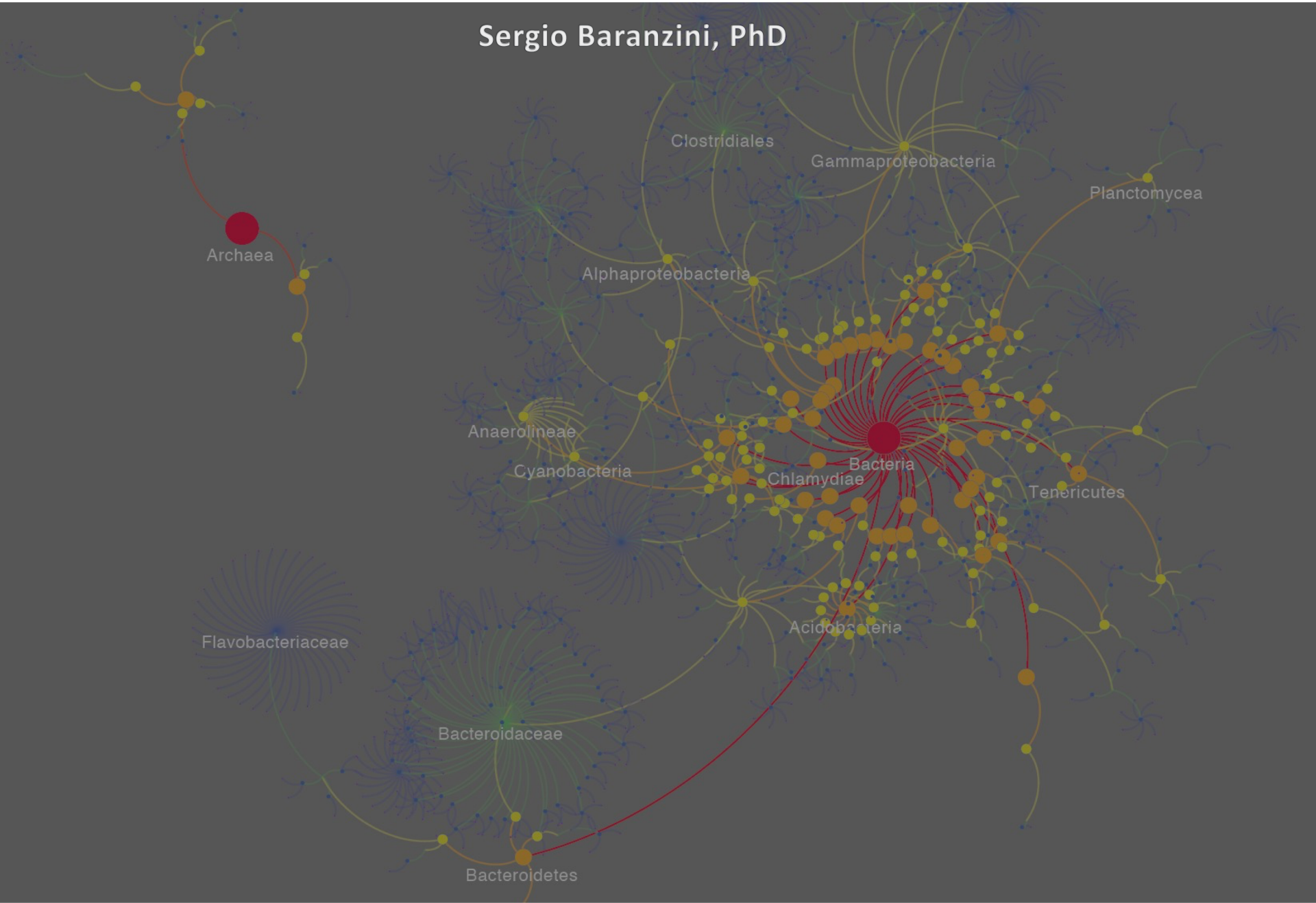
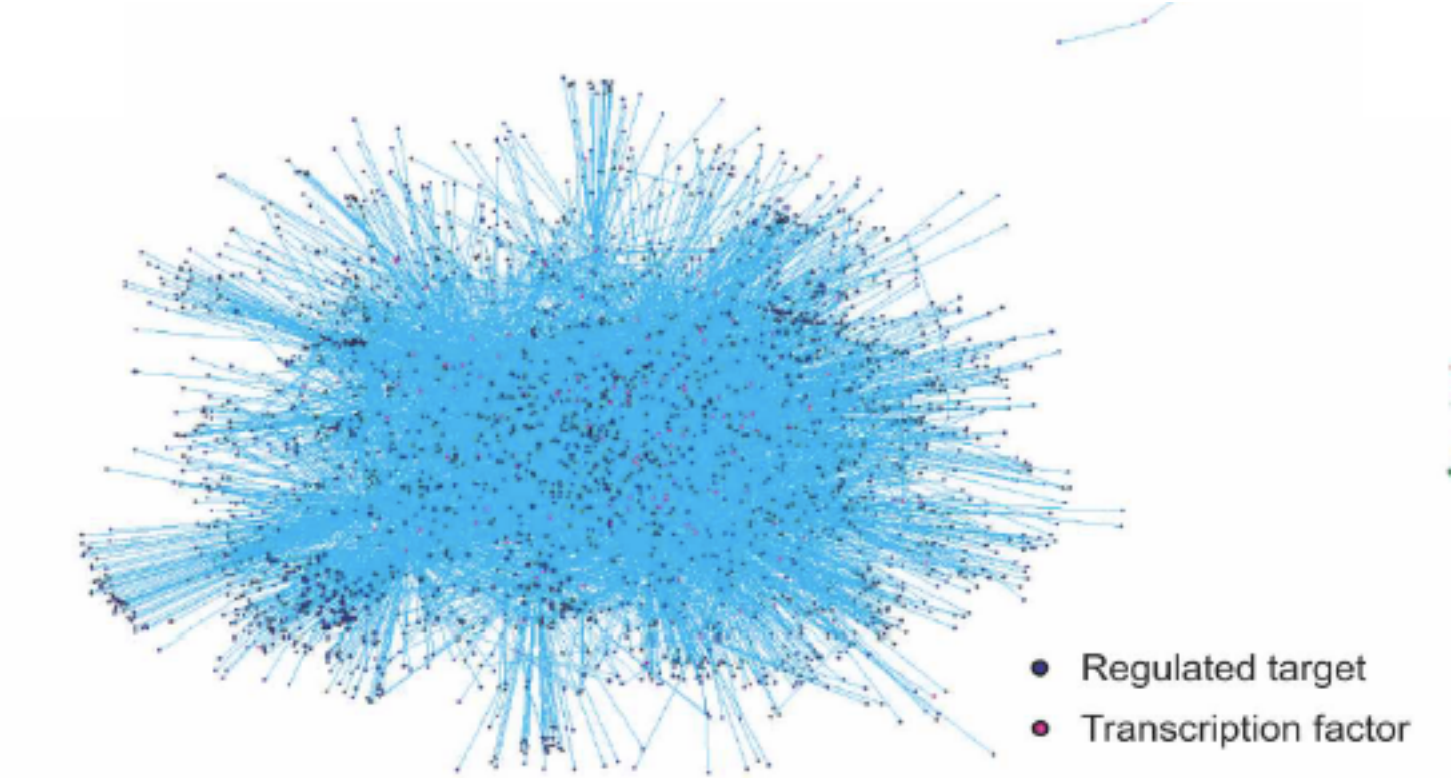
# Network biology



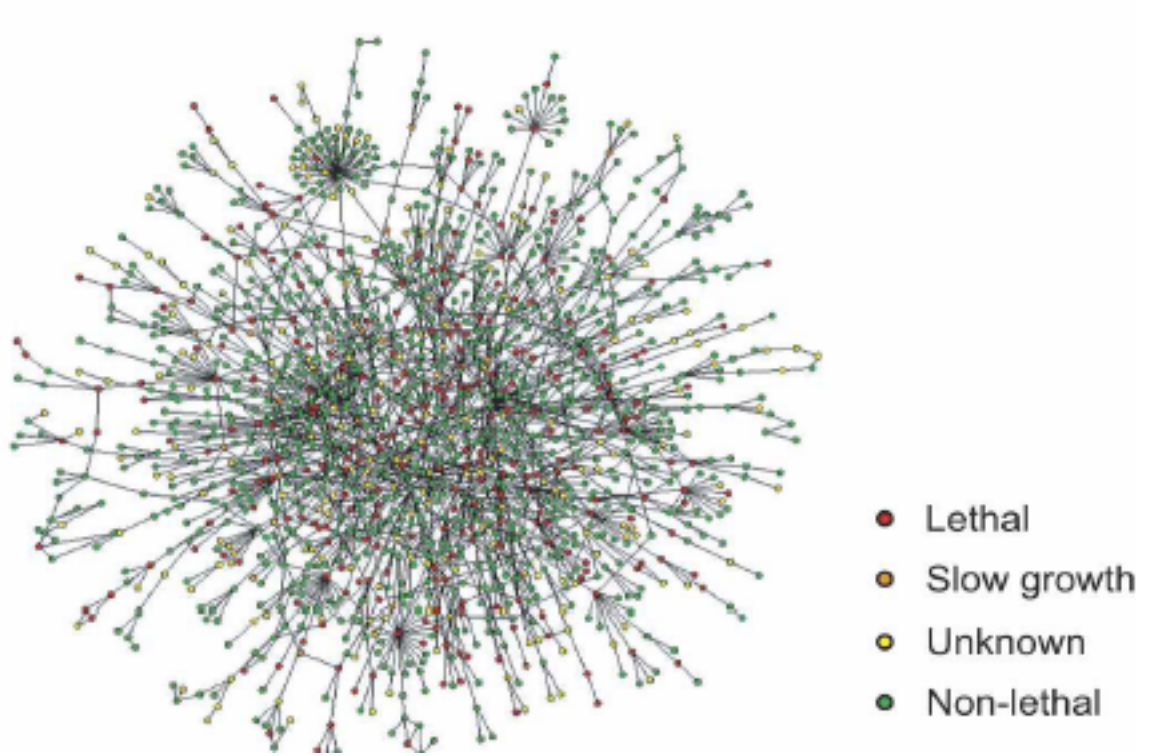
# Examples of network biology

- Genetic
- Co-expression
- Protein interaction
- Metabolic
- TF binding
- Phosphorylation

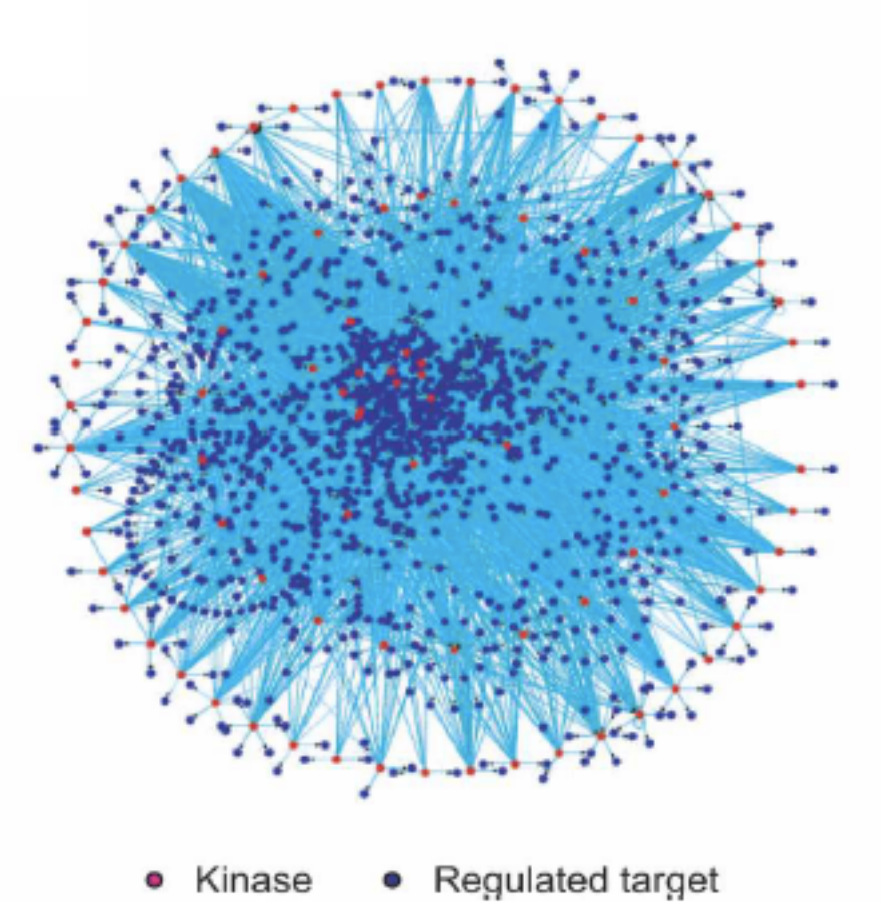
# TF binding net



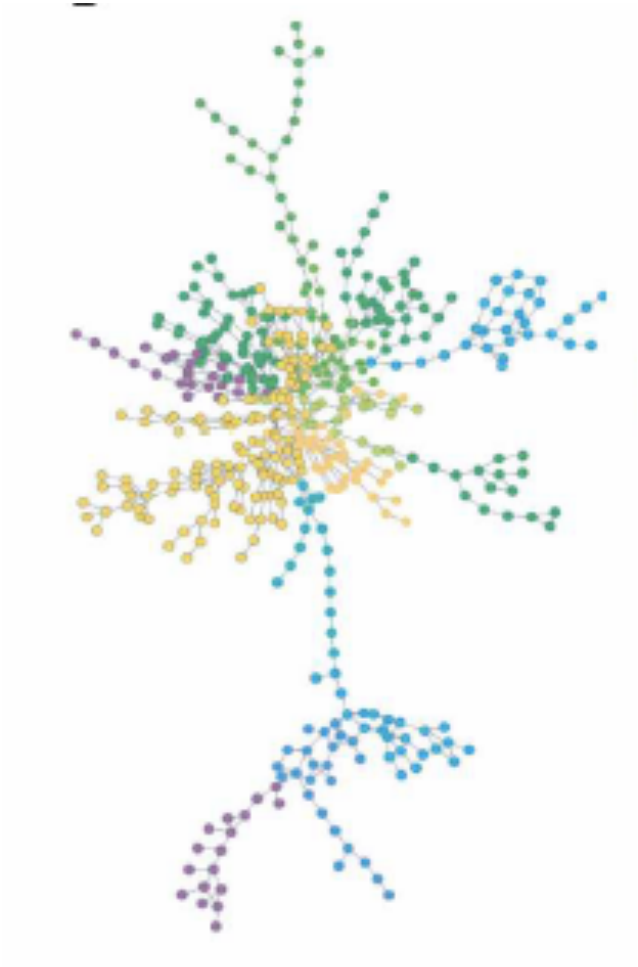
# Protein interaction net



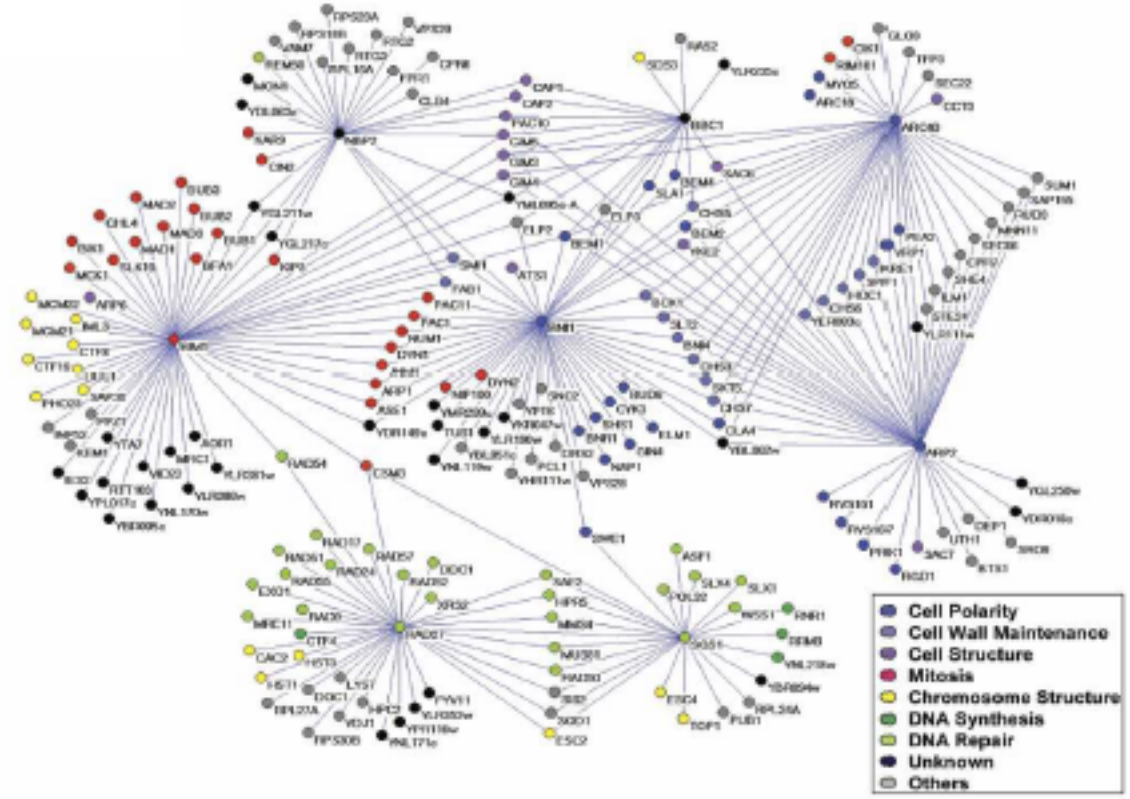
# Phosphorylation net



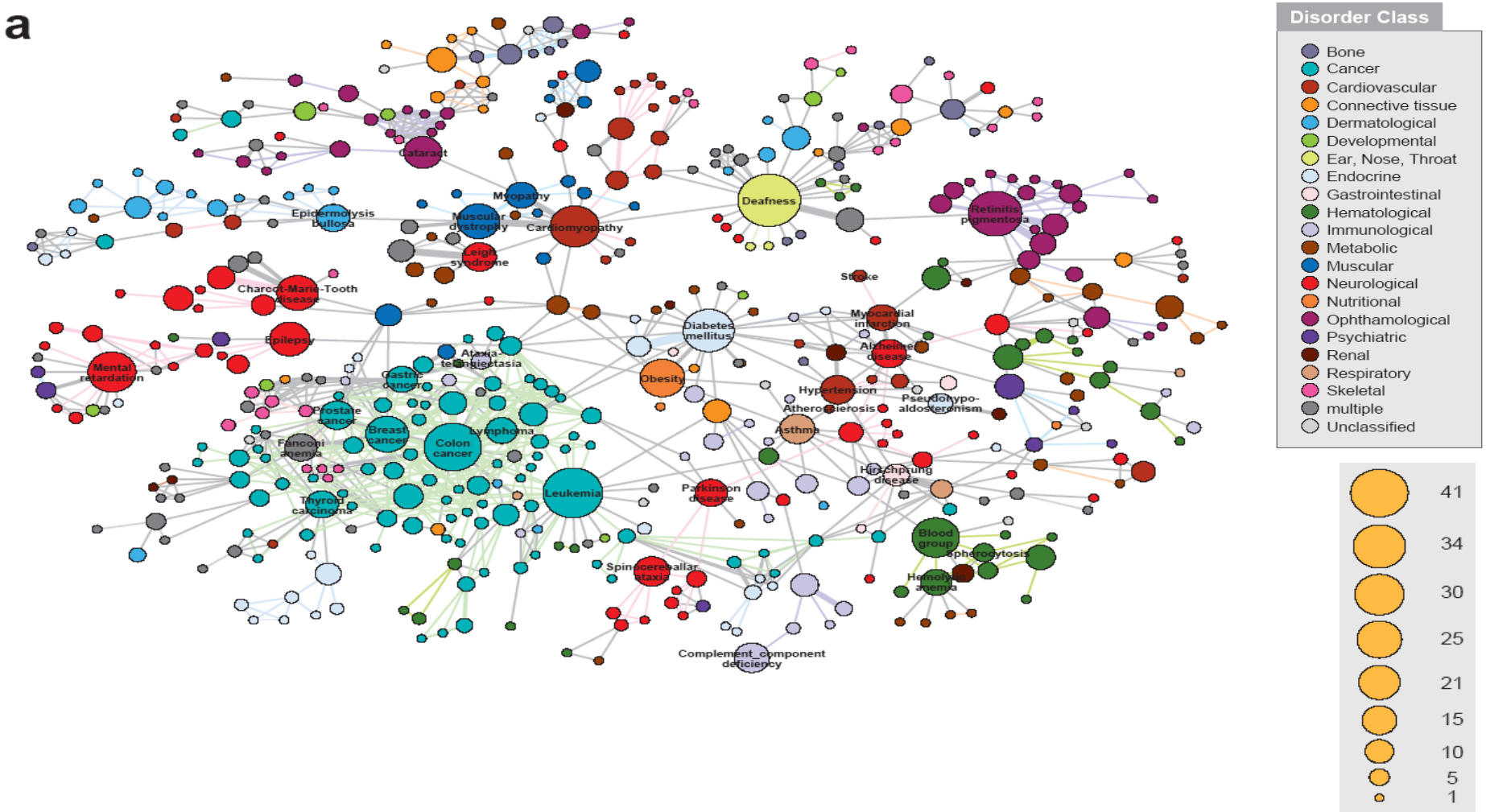
# Metabolic net



# Genetic (synthetic lethality)



# Bi-partite nets: The diseasome



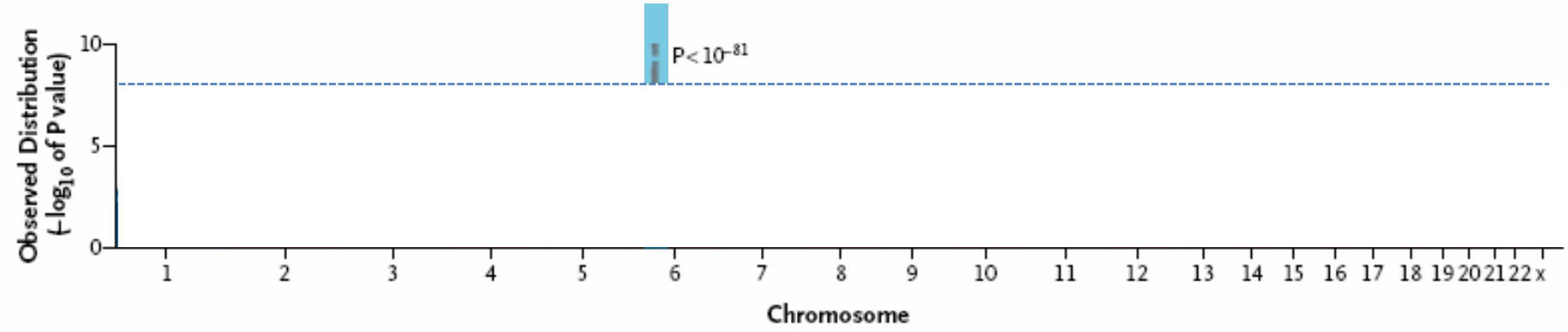

Proc Natl Acad Sci U S A. 2007 May 22;104(21):8685-90

# Other bi-partite nets

- Drug target net (Yildirim et al. Nat Biotechnol. 2007 Oct;25(10):1119-26.)
- Drug interactions (Campillos et al. Science. 2008 Jul 11;321(5886):263-6)
- Drug repositioning (Keiser et al. Nature. 2009 Nov 12; 462(7270):175-81.)
- Disease-symptoms (Zhou et al. Nat Commun. 2014 Jun 26;5:4212. )

# First GWAS in multiple sclerosis (MS)

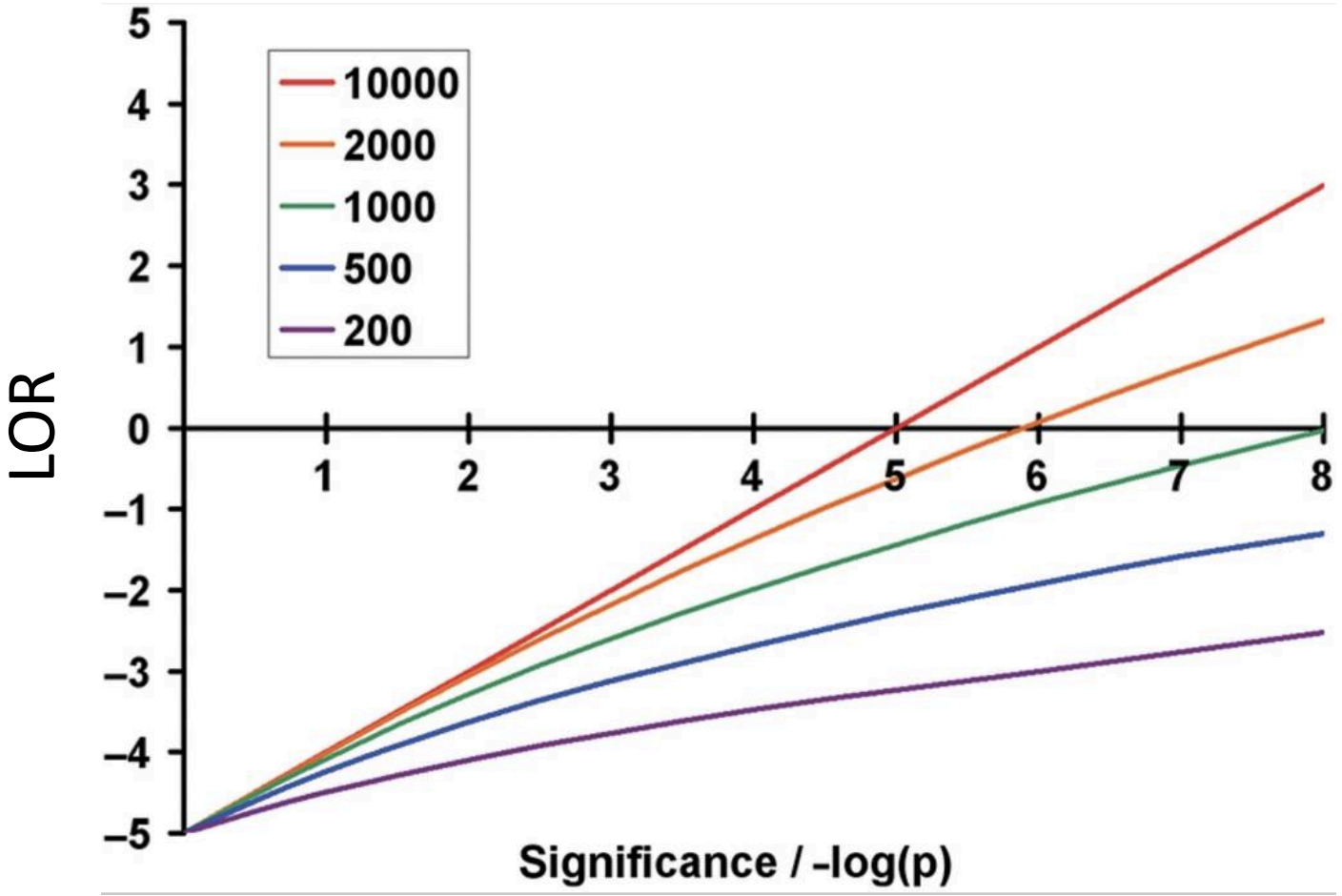
1000 cases



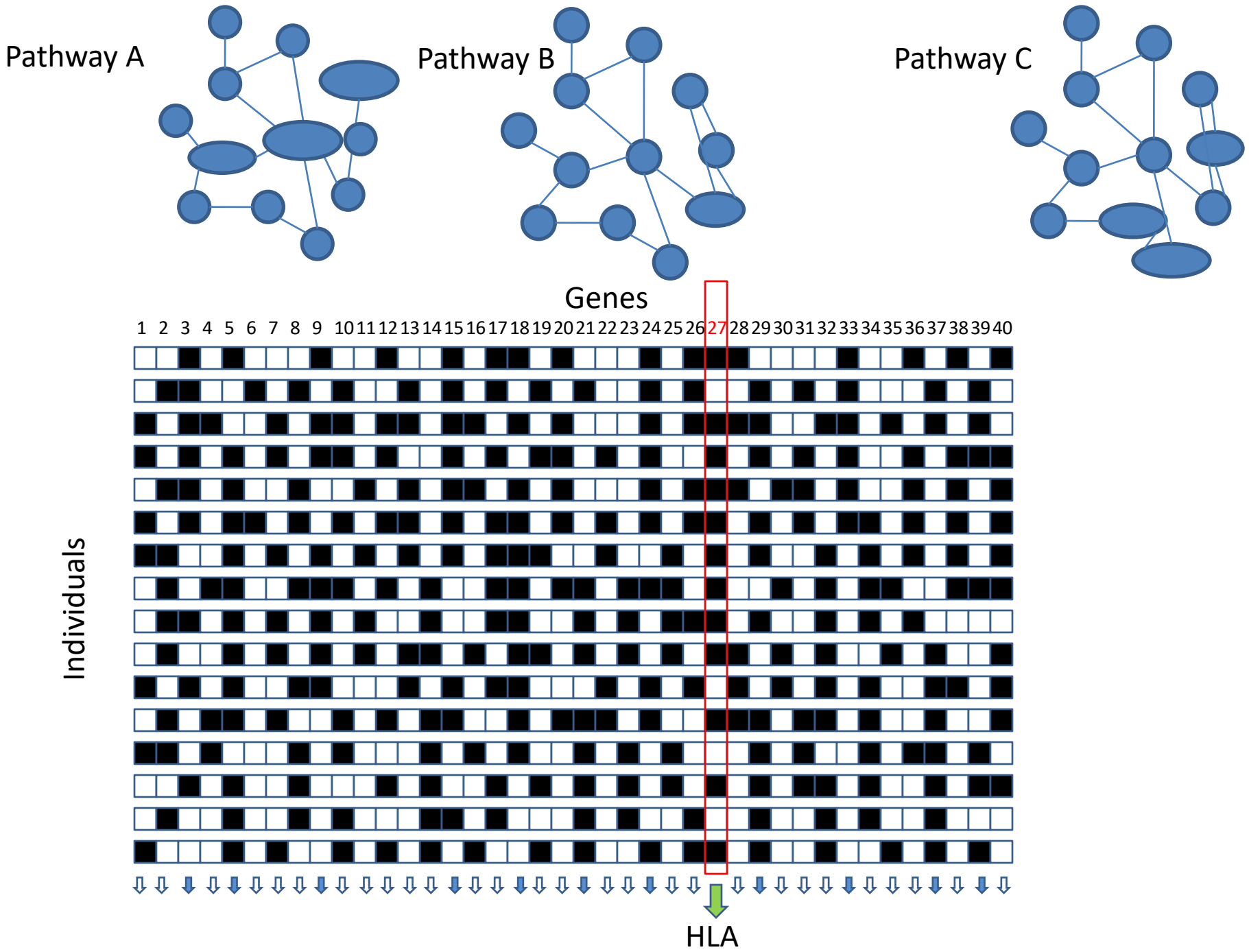
# The problem


Did you recognize this gene?
I'm not sure
GWAS
http://biocomicals.blogspot.com

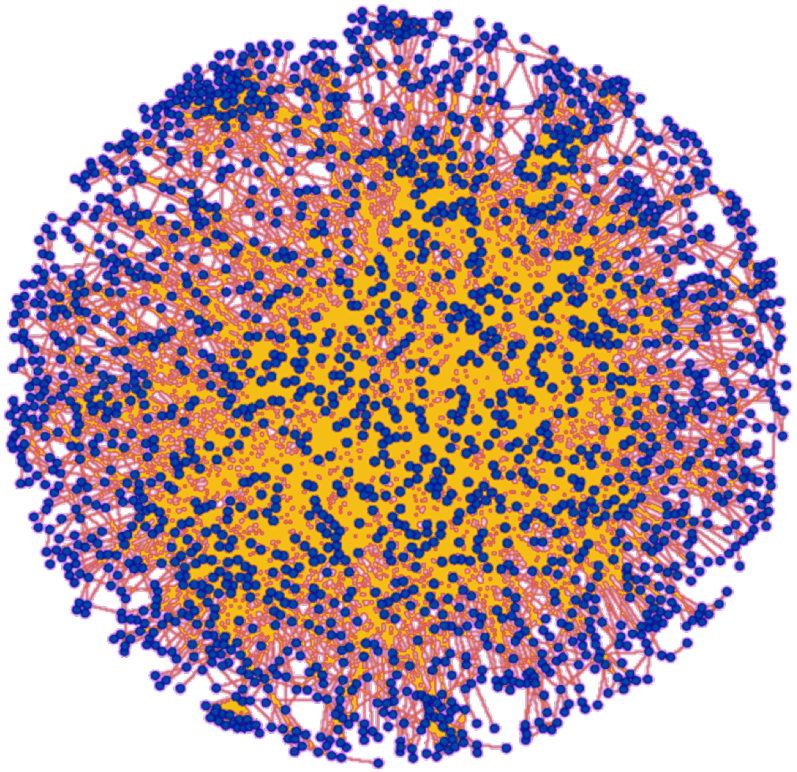
# Posterior odds of GWAS associations




Sawcer SJ, Brain. 2008


Pathway A
Pathway B
Pathway C
Genes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Individuals
HLA

Hypothesis: Overlapping statistical evidence of association with physical evidence of interaction should discover modestly associated genes



Human protein interaction network

ARTICLE

# Network-Based Multiple Sclerosis Pathway Analysis with GWAS Data from 15,000 Cases and 30,000 Controls

International Multiple Sclerosis Genetics Consortium[1,*]

Multiple sclerosis (MS) is an inflammatory CNS disease with a substantial genetic component, originally mapped to only the human leukocyte antigen (HLA) region. In the last 5 years, a total of seven genome-wide association studies and one meta-analysis successfully identified 57 non-HLA susceptibility loci. Here, we merged nominal statistical evidence of association and physical evidence of interaction to conduct a protein-interaction-network-based pathway analysis (PINBPA) on two large genetic MS studies comprising a total of 15,317 cases and 29,529 controls. The distribution of nominally significant loci at the gene level matched the patterns of extended linkage disequilibrium in regions of interest. We found that products of genome-wide significantly associated genes are more likely to interact physically and belong to the same or related pathways. We next searched for subnetworks (modules) of genes (and their encoded proteins) enriched with nominally associated loci within each study and identified those modules in common between the two studies. We demonstrate that these modules are more likely to contain genes with bona fide susceptibility variants and, in addition, identify several high-confidence candidates (including *BCL10*, *CD48*, *REL*, *TRAF3*, and *TEC*). PINBPA is a powerful approach to gaining further insights into the biology of associated genes and to prioritizing candidates for subsequent genetic studies of complex traits.

# Reading assignment

SCIENTIFIC REPORTS

OPEN

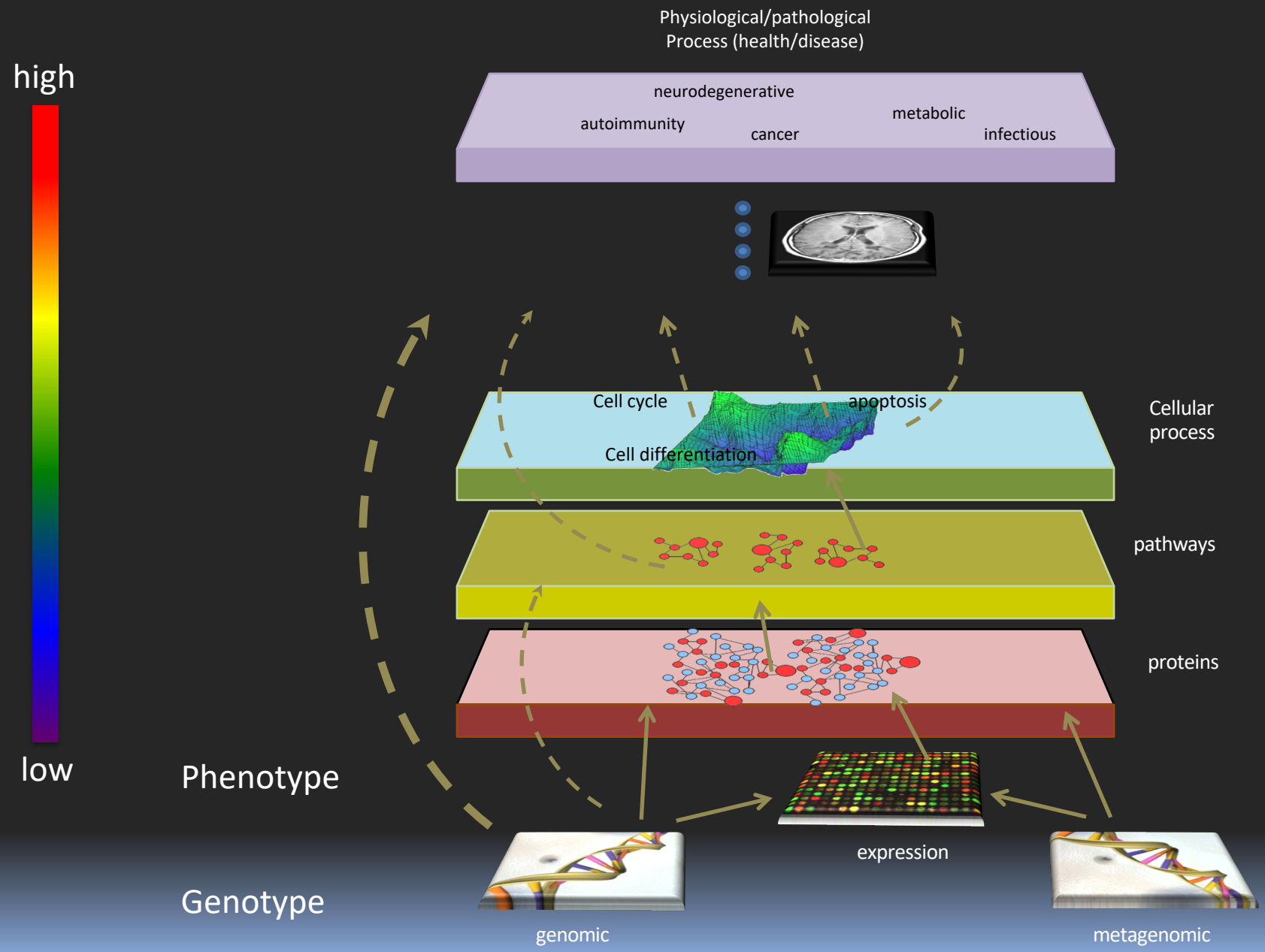
## Tissue Specificity of Human Disease Module

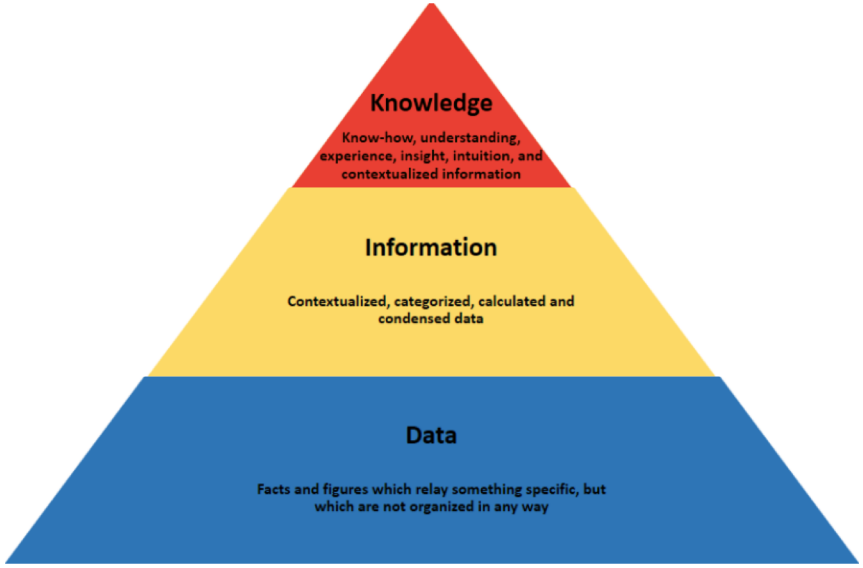
Maksim Kitsak[1,2], Amitabh Sharma[1,2,3], Jörg Menche[1,2,4,5], Emre Guney[1,2,6], Susan Dina Ghiassian[1,2,7], Joseph Loscalzo[8] & Albert-László Barabási[1,2,4,7,8,9]

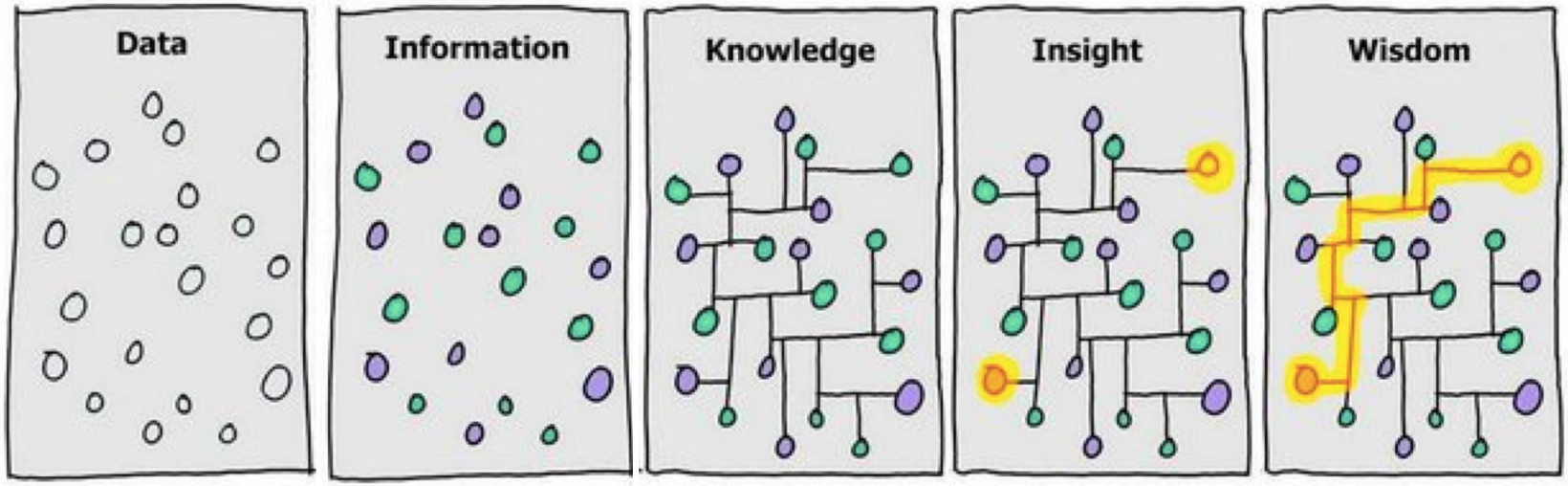
Received: 16 May 2016
Accepted: 20 September 2016
Published: 17 October 2016

Genes carrying mutations associated with genetic diseases are present in all human cells; yet, clinical manifestations of genetic diseases are usually highly tissue-specific. Although some disease genes are expressed only in selected tissues, the expression patterns of disease genes alone cannot explain the observed tissue specificity of human diseases. Here we hypothesize that for a disease to manifest itself in a particular tissue, a whole functional subnetwork of genes (disease module) needs to be expressed in that tissue. Driven by this hypothesis, we conducted a systematic study of the expression patterns of disease genes within the human interactome. We find that genes expressed in a specific tissue tend to be localized in the same neighborhood of the interactome. By contrast, genes expressed in different tissues are segregated in distinct network neighborhoods. Most important, we show that it is the integrity and the completeness of the expression of the disease module that determines disease manifestation in selected tissues. This approach allows us to construct a disease-tissue network that confirms known and predicts unexpected disease-tissue associations.

# Hierarchical organization of biological complexity

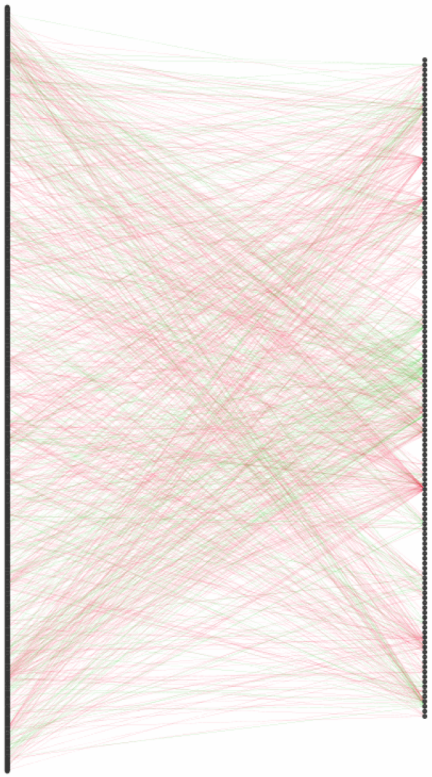



Knowledge
Know-how, understanding, experience, insight, intuition, and contextualized information
Information
Contextualized, categorized, calculated and condensed data
Data
Facts and figures which relay something specific, but which are not organized in any way

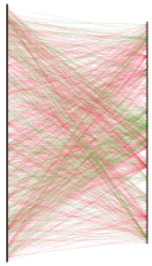

Data
Information
Knowledge
Insight
Wisdom

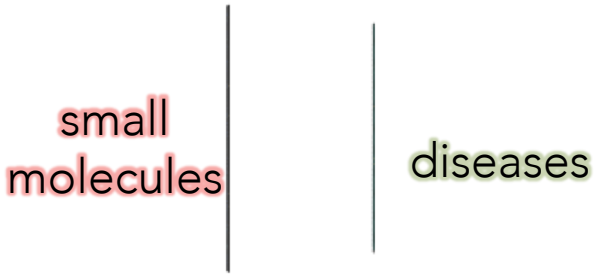
# Imagine...

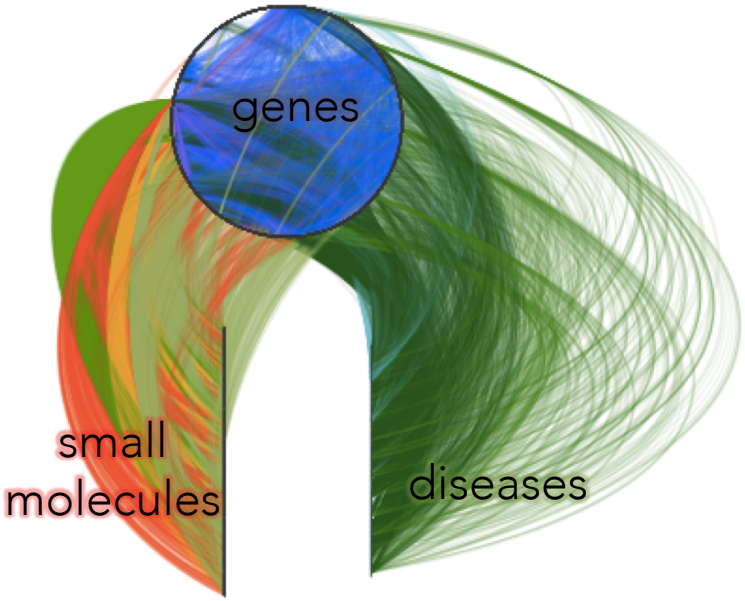
1,538 approved
small molecule compounds

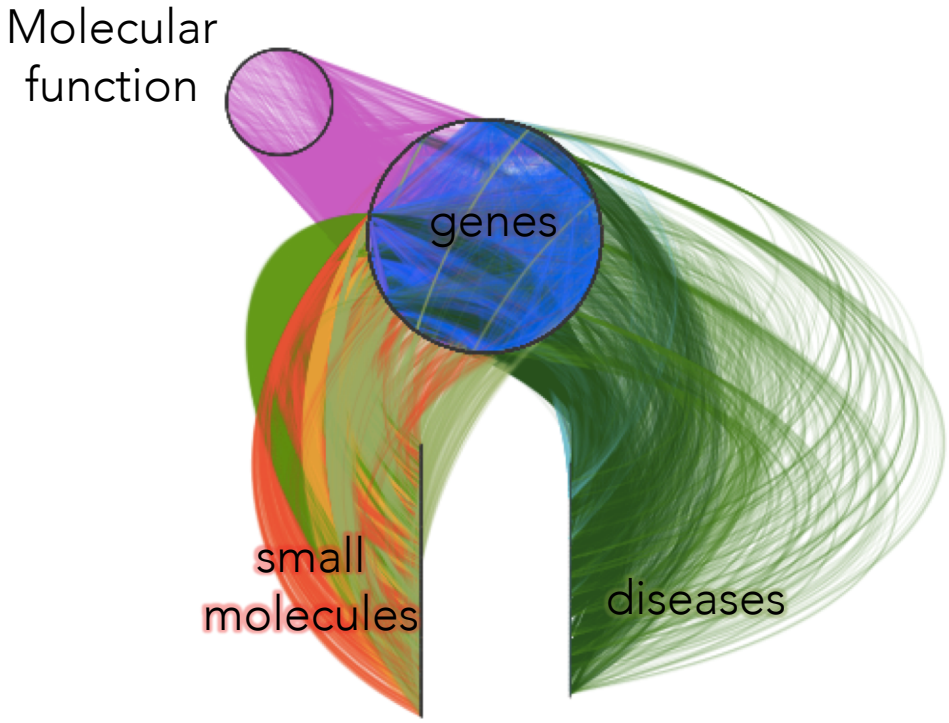


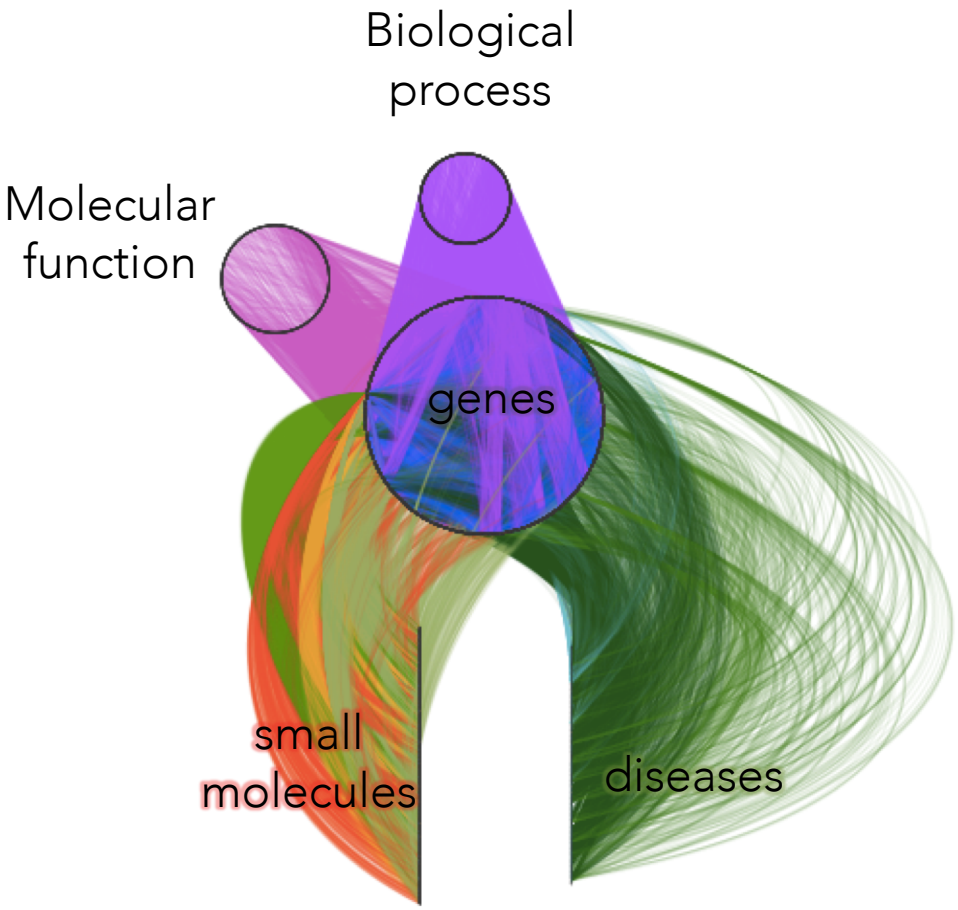
136 complex diseases

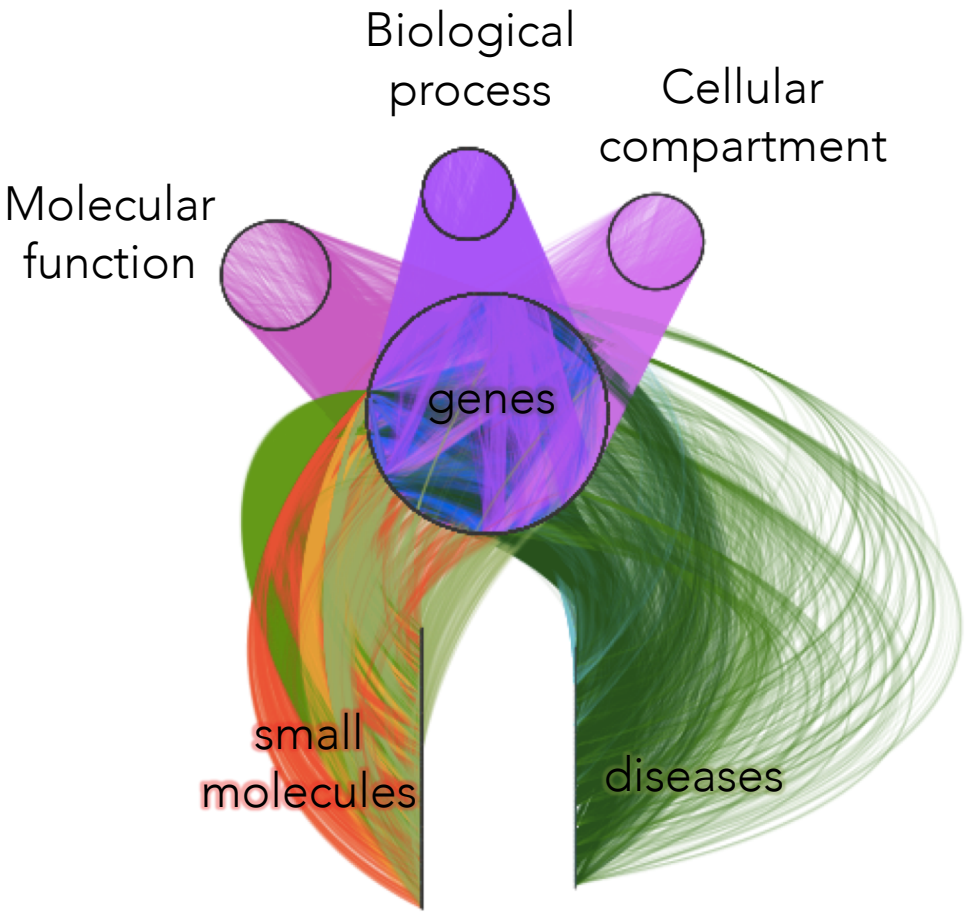


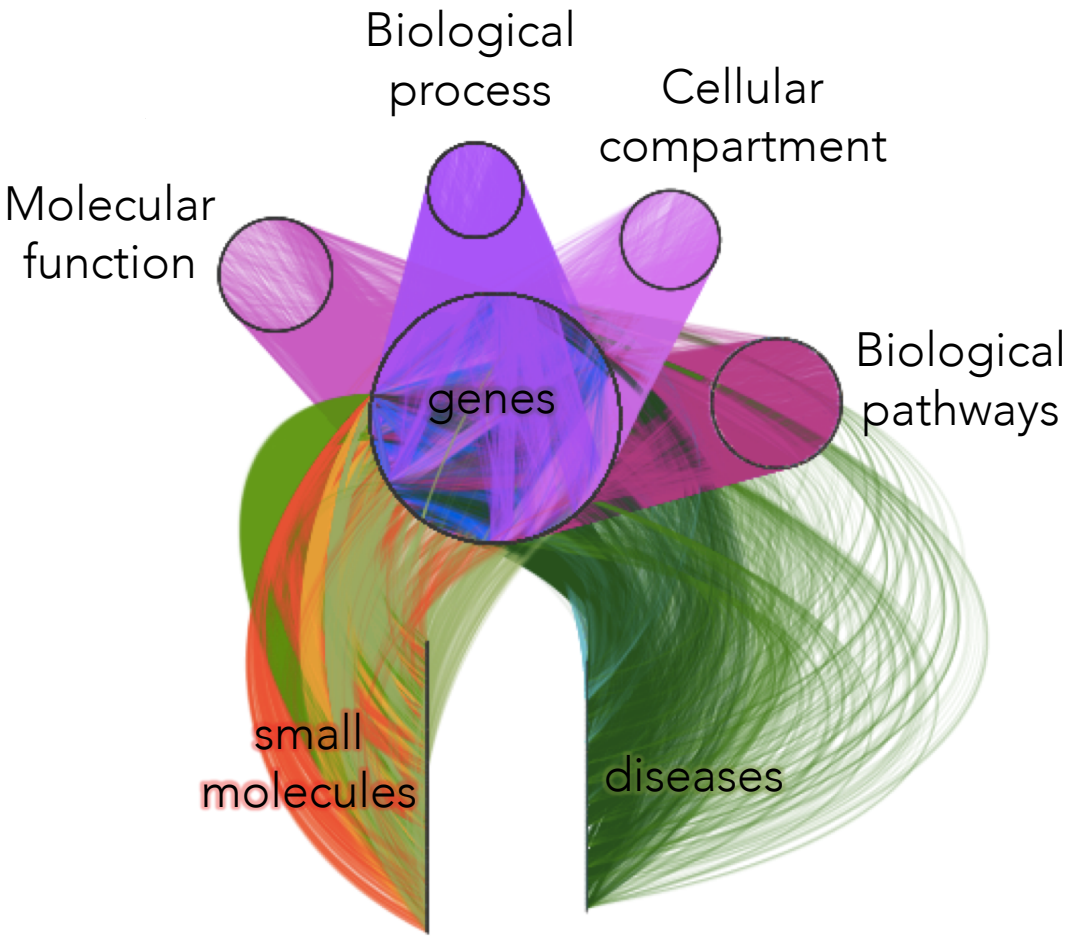

small
molecules
diseases

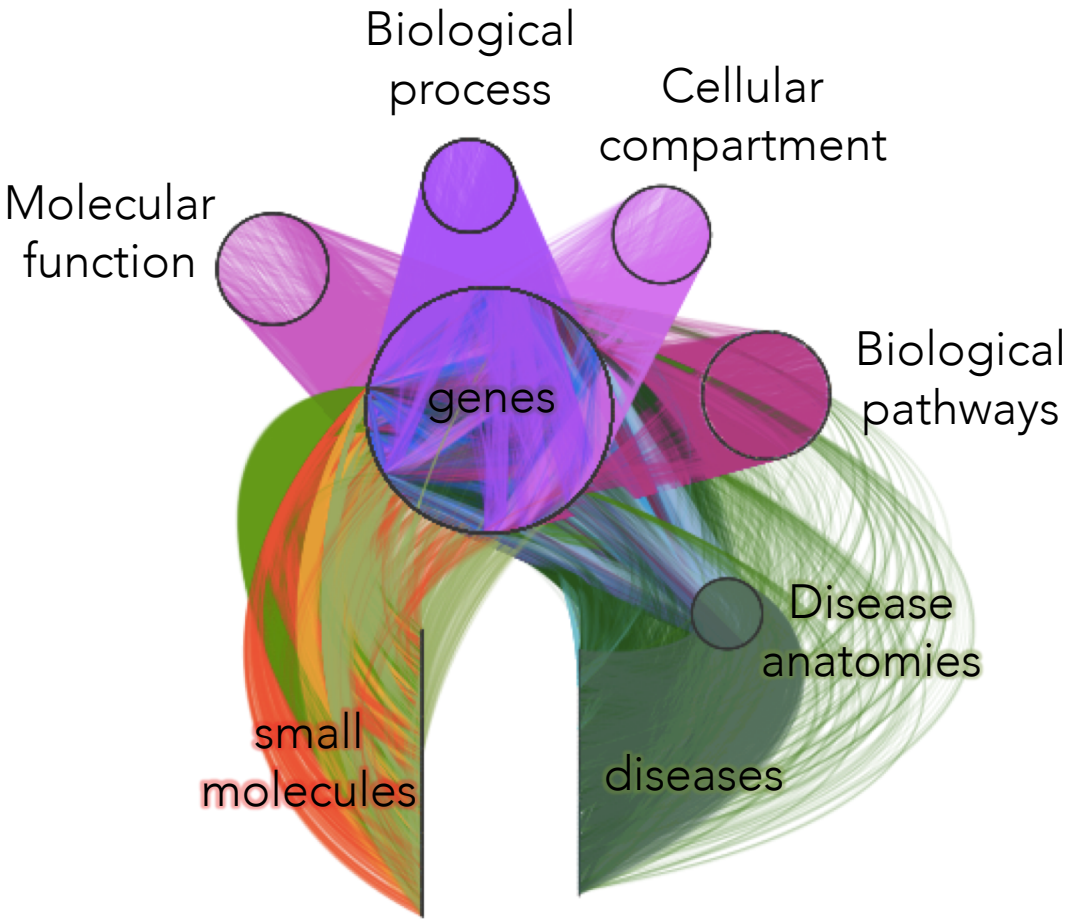

genes
small
molecules
diseases

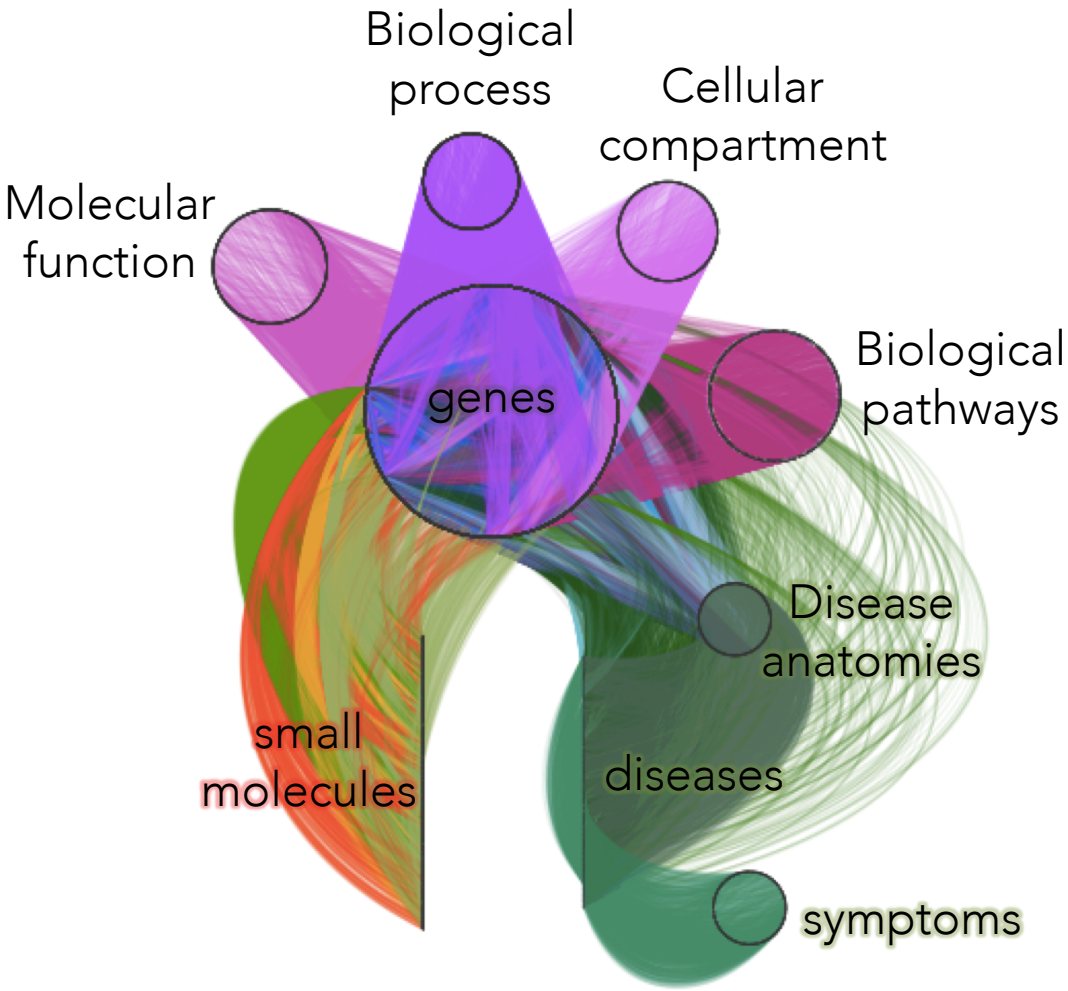

Molecular function
genes
small molecules
diseases

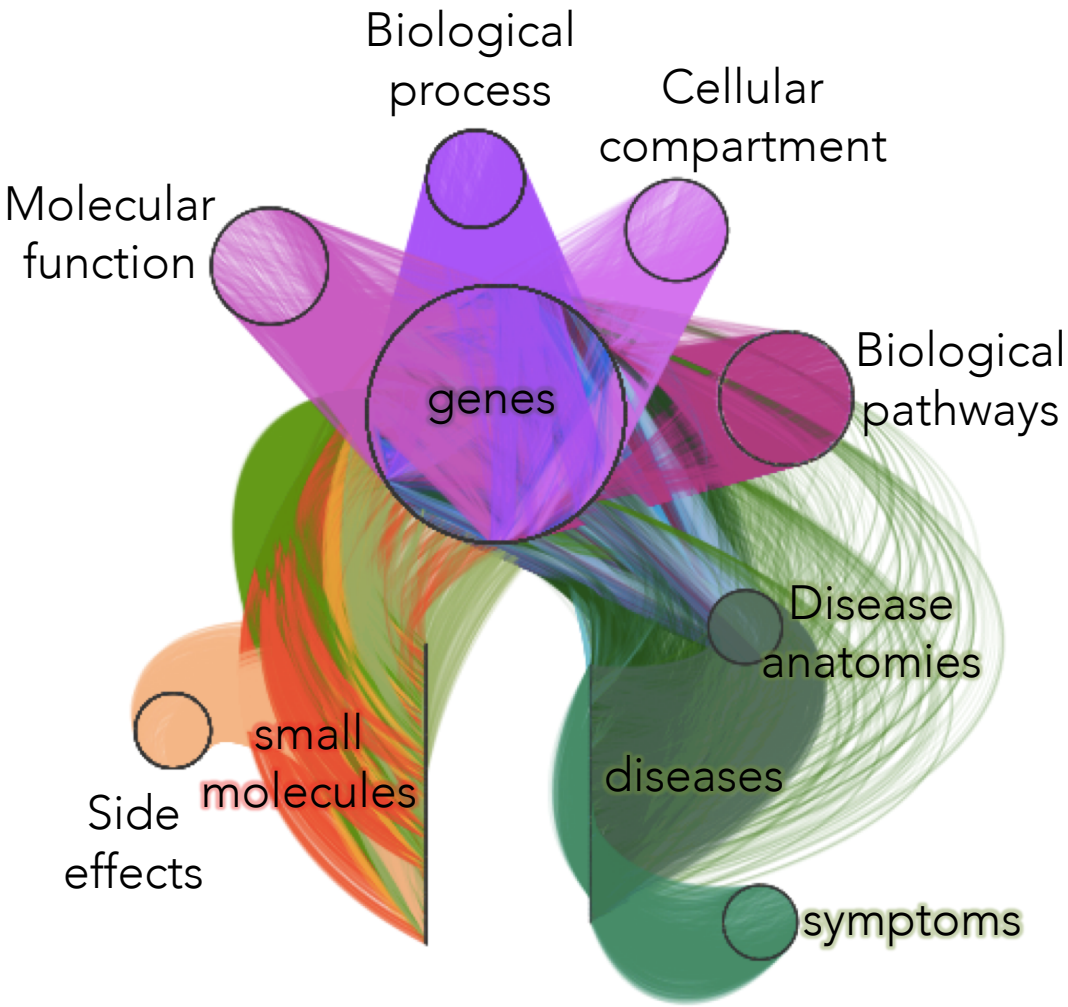

Biological process
Molecular function
genes
small molecules
diseases

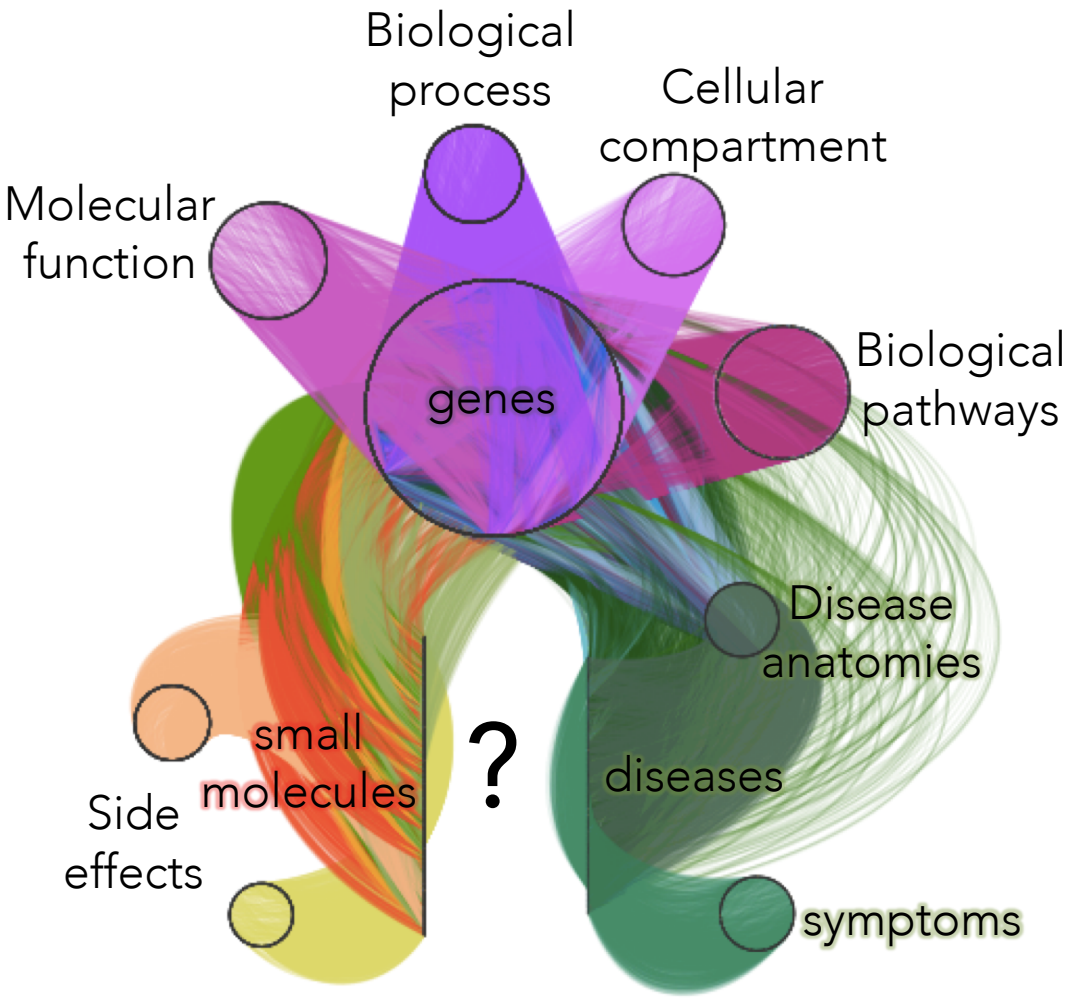

Biological process
Cellular compartment
Molecular function
genes
small molecules
diseases


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


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


Biological process
Cellular compartment
Molecular function
Biological pathways
genes
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)**

# Hetionet v1.1 characteristics

11 node types

| Metanode | Abbr | Nodes |
|---|---|---|
| Anatomy | A | 402 |
| Biological Process | BP | 11,381 |
| Cellular Component | CC | 1,391 |
| Compound | C | 1,552 |
| Disease | D | 137 |
| Gene | G | 20,945 |
| Molecular Function | MF | 2,884 |
| Pathway | PW | 1,822 |
| Pharmacologic Class | PC | 345 |
| Side Effect | SE | 5,734 |
| Symptom | S | 438 |

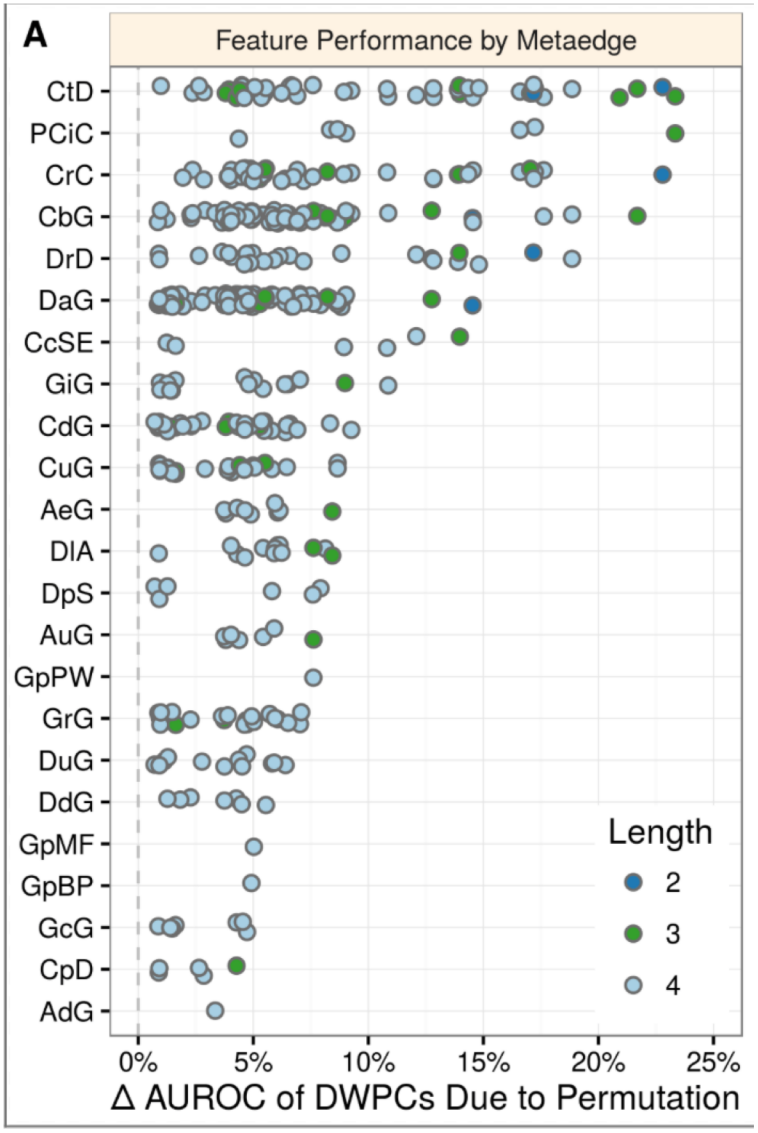
24 edge types

| metaedge | Abbr | Edges |
|---|---|---|
| Anatomy–downregulates–Gene | AdG | 102,240 |
| Anatomy–expresses–Gene | AeG | 526,407 |
| Anatomy–upregulates–Gene | AuG | 97,848 |
| Compound–binds–Gene | CbG | 11,571 |
| Compound–causes–Side Effect | CcSE | 138,944 |
| Compound–downregulates–Gene | CdG | 21,102 |
| Compound–palliates–Disease | CpD | 390 |
| Compound–resembles–Compound | CrC | 6,486 |
| Compound–treats–Disease | CtD | 755 |
| Compound–upregulates–Gene | CuG | 18,756 |
| Disease–associates–Gene | DaG | 12,623 |
| Disease–downregulates–Gene | DdG | 7,623 |
| Disease–localizes–Anatomy | DlA | 3,602 |
| Disease–presents–Symptom | DpS | 3,357 |
| Disease–resembles–Disease | DrD | 543 |
| Disease–upregulates–Gene | DuG | 7,731 |
| Gene–covaries–Gene | GcG | 61,690 |
| Gene–interacts–Gene | GiG | 147,164 |
| Gene–participates–Biological Process | GpBP | 559,504 |
| Gene–participates–Cellular Component | GpCC | 73,566 |
| Gene–participates–Molecular Function | GpMF | 97,222 |
| Gene–participates–Pathway | GpPW | 84,372 |
| Gene→regulates→Gene | Gr>G | 265,672 |
| Pharmacologic Class–includes–Compound | PCiC | 1,029 |

# 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

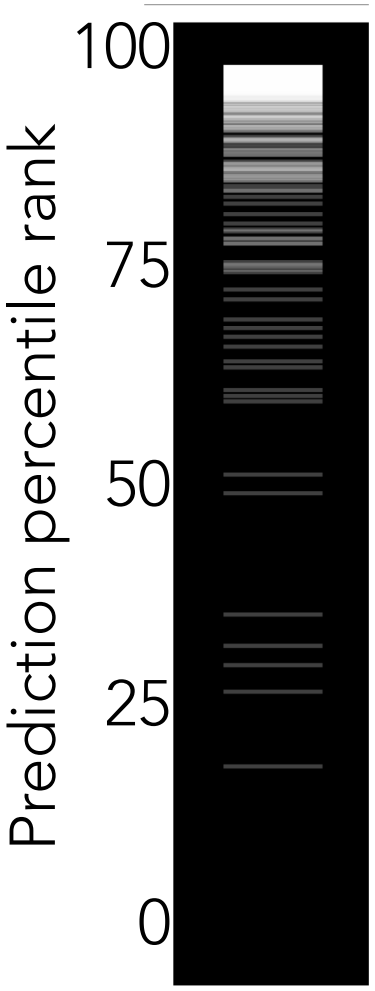


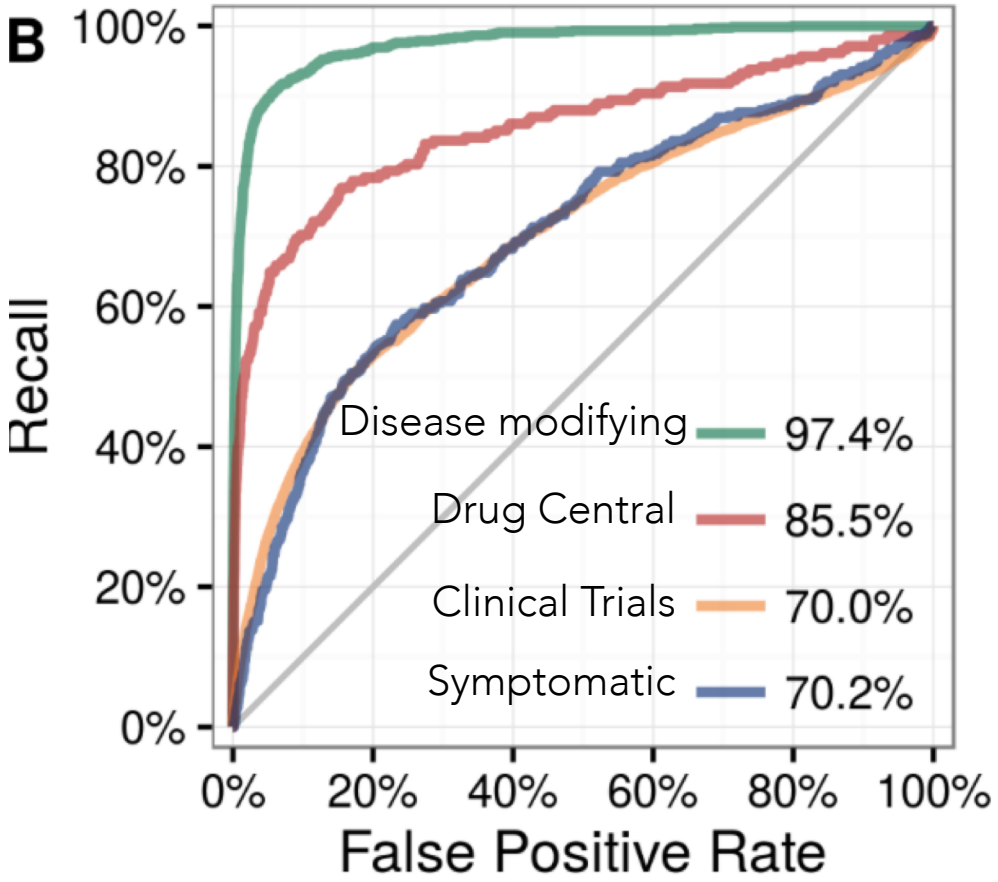
# Top predictions

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 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!

# Prediction Performance

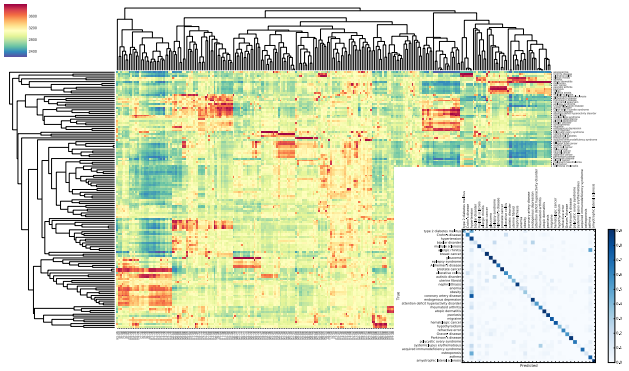


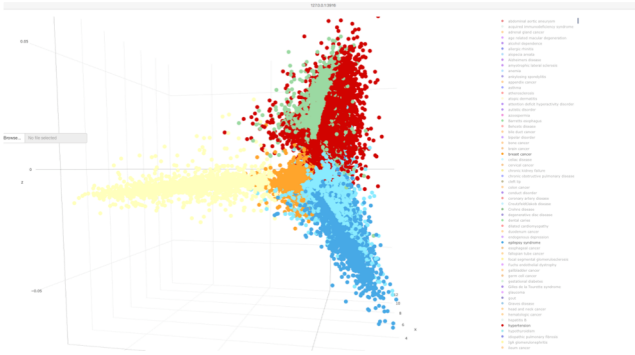
# SPOKE Applications Coming Soon

## Disease Patterns

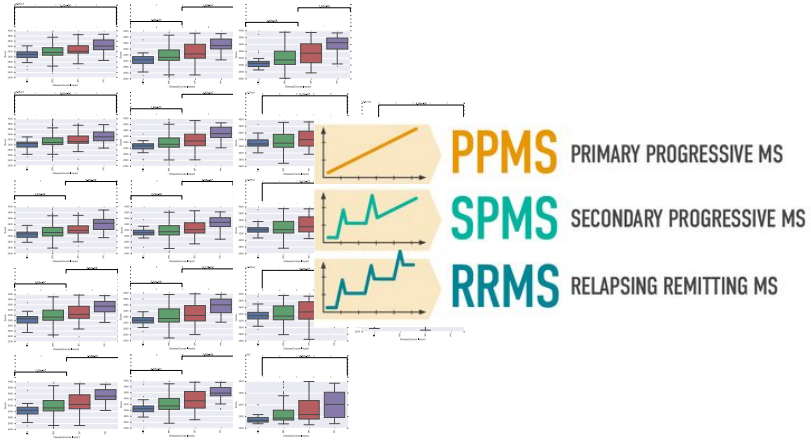
## Disease Prediction



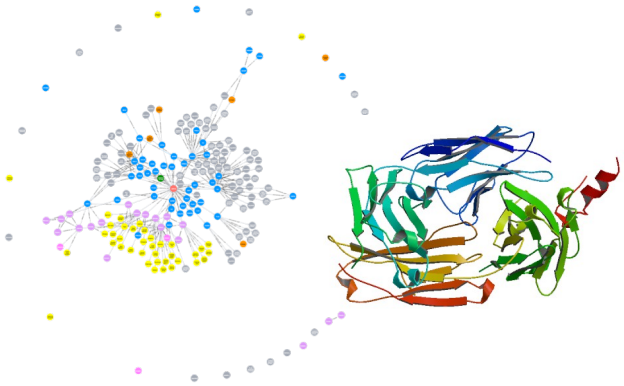
## Patient Stratification



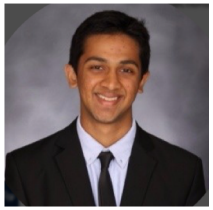
## Multiple Sclerosis Disease Course




## Drug Signatures

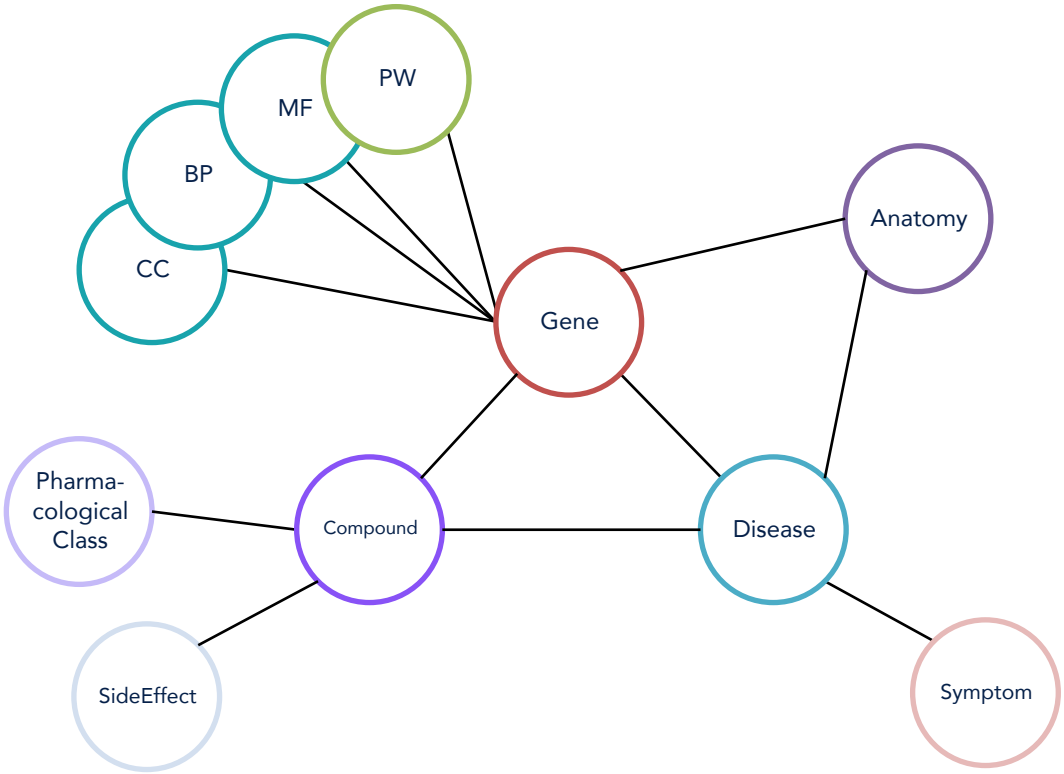


Charlotte Nelson

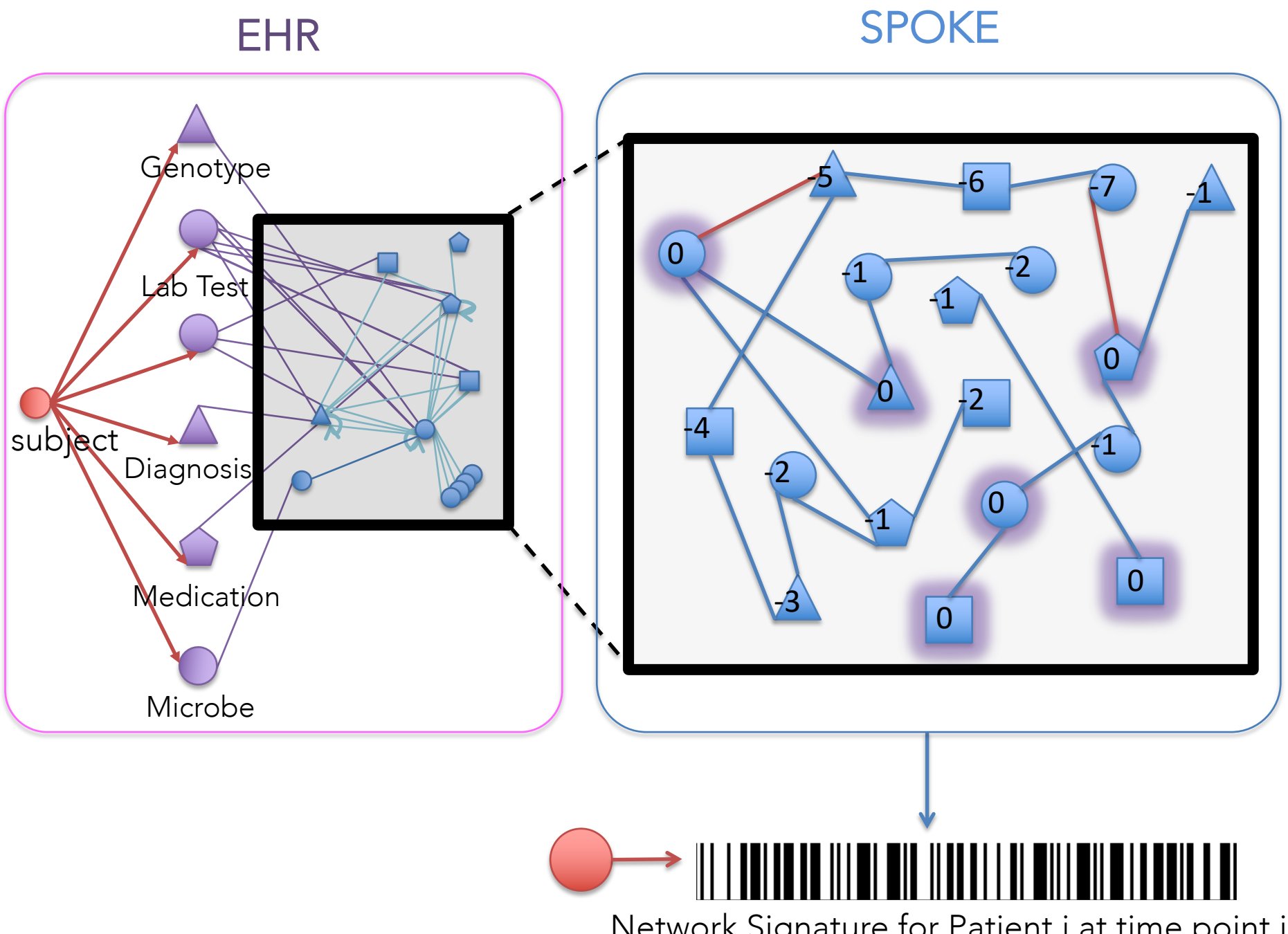


Krish Bharat

# Hetionet -> SPOKE - Scalable PrecisiOn Medicine Knowledge Engine

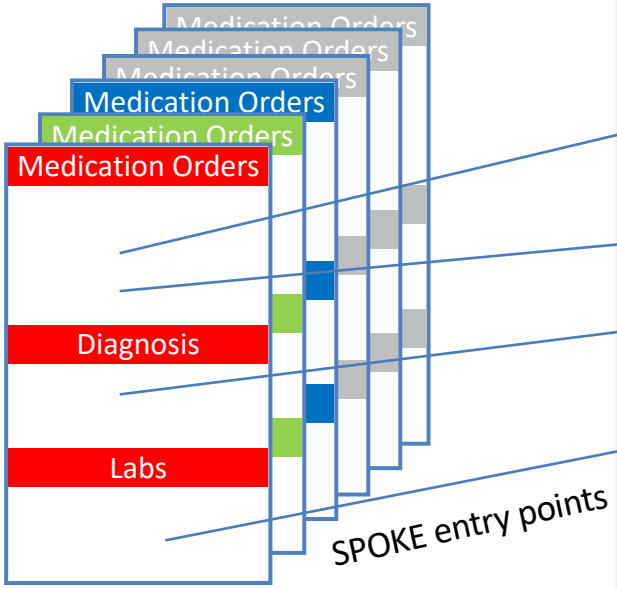


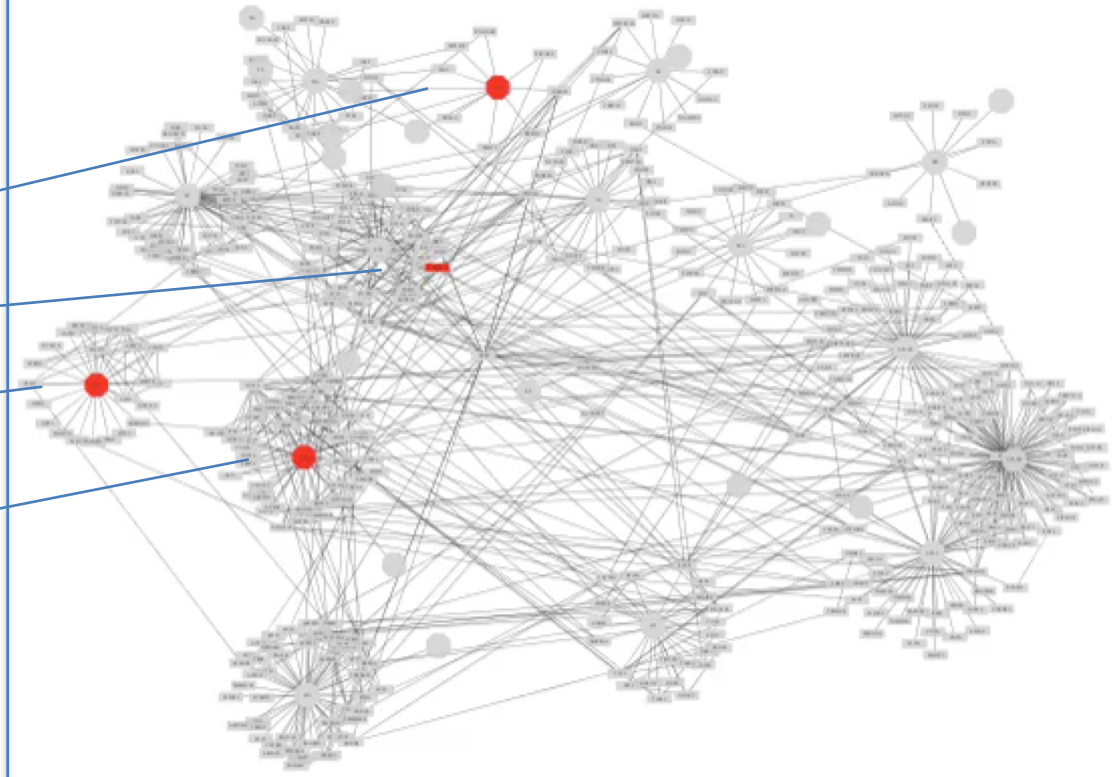

**1,941,858 nodes (12 types)**
**2,464,273 relationships (26 types)**


EHR
SPOKE
Genotype
Lab Test
subject
Diagnosis
Medication
Microbe
-5
-6
-7
-1
0
-1
-2
-1
0
0
-2
-4
-1
-2
0
-1
0
-3
0
Network Signature for Patient i at time point j

# UCSF EHR (800k)

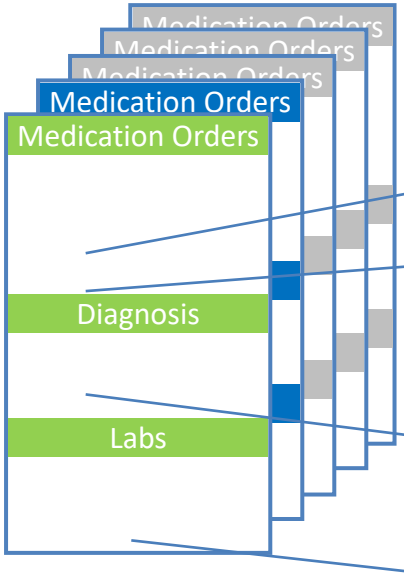
# SPOKE

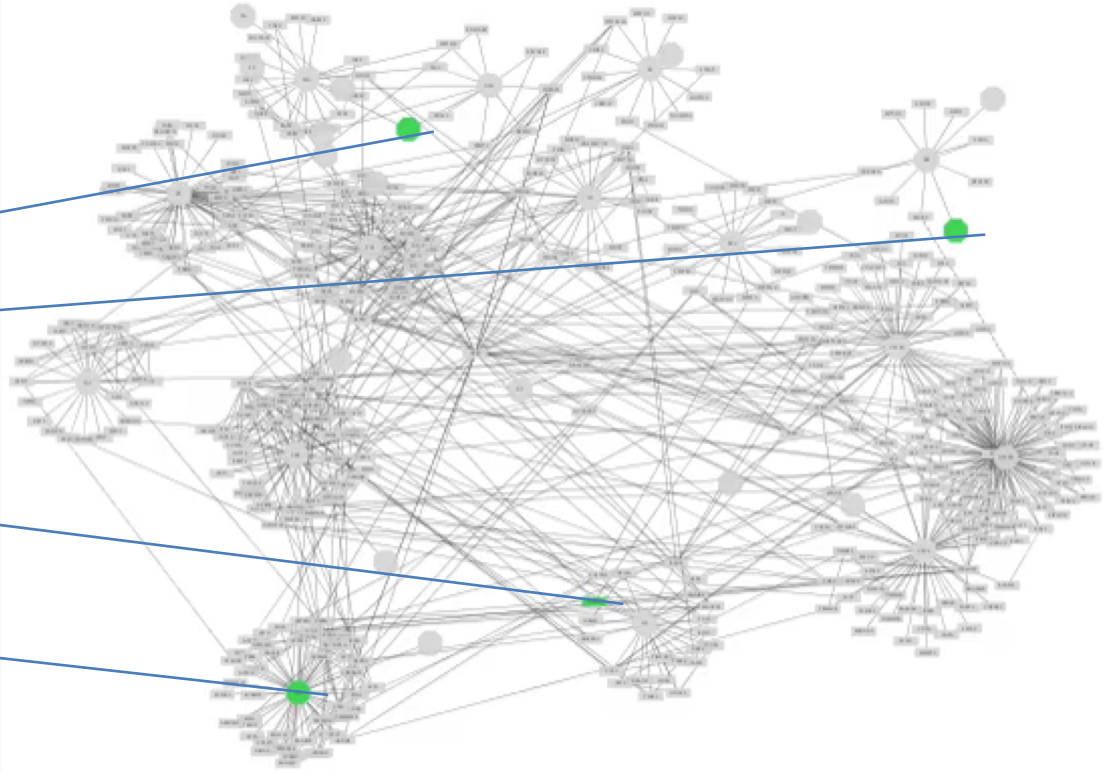


EHR

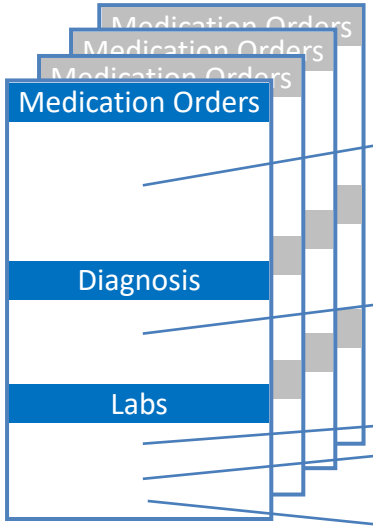
SPOKE

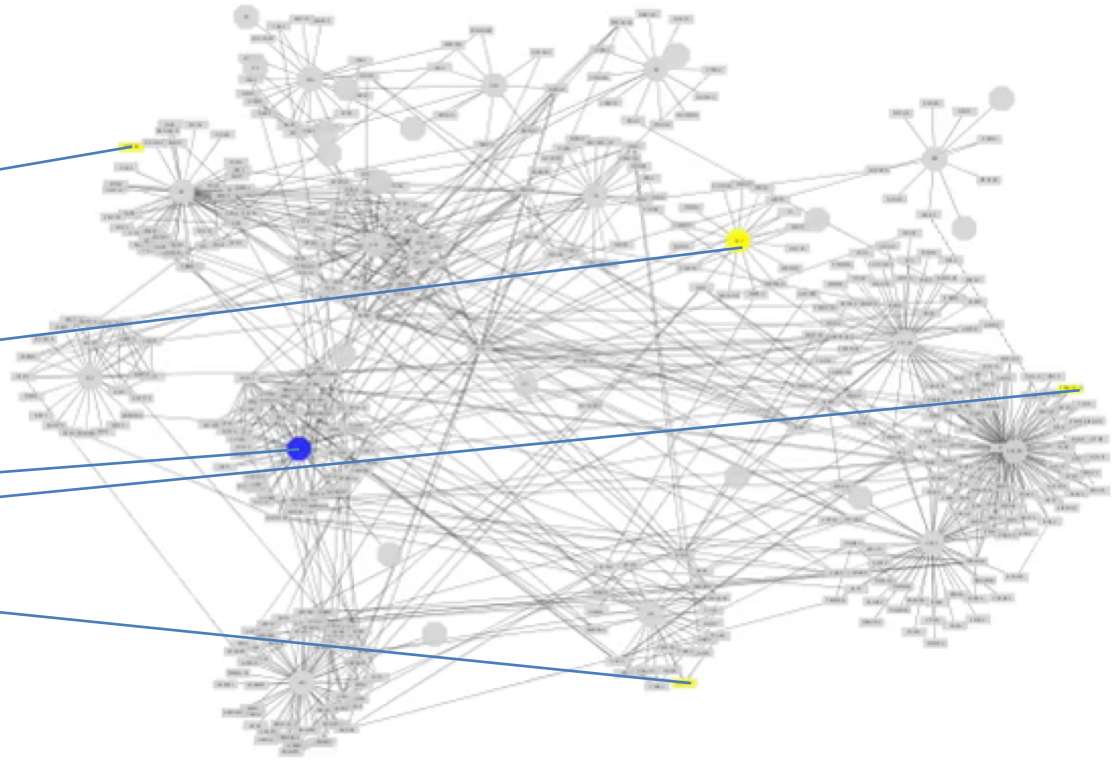

Medication Orders
Medication Orders
Medication Orders
Medication Orders
Diagnosis
Labs



EHR

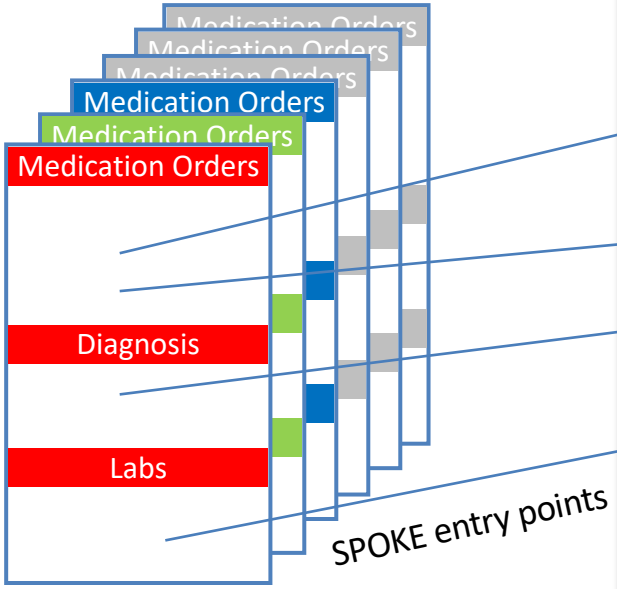
SPOKE

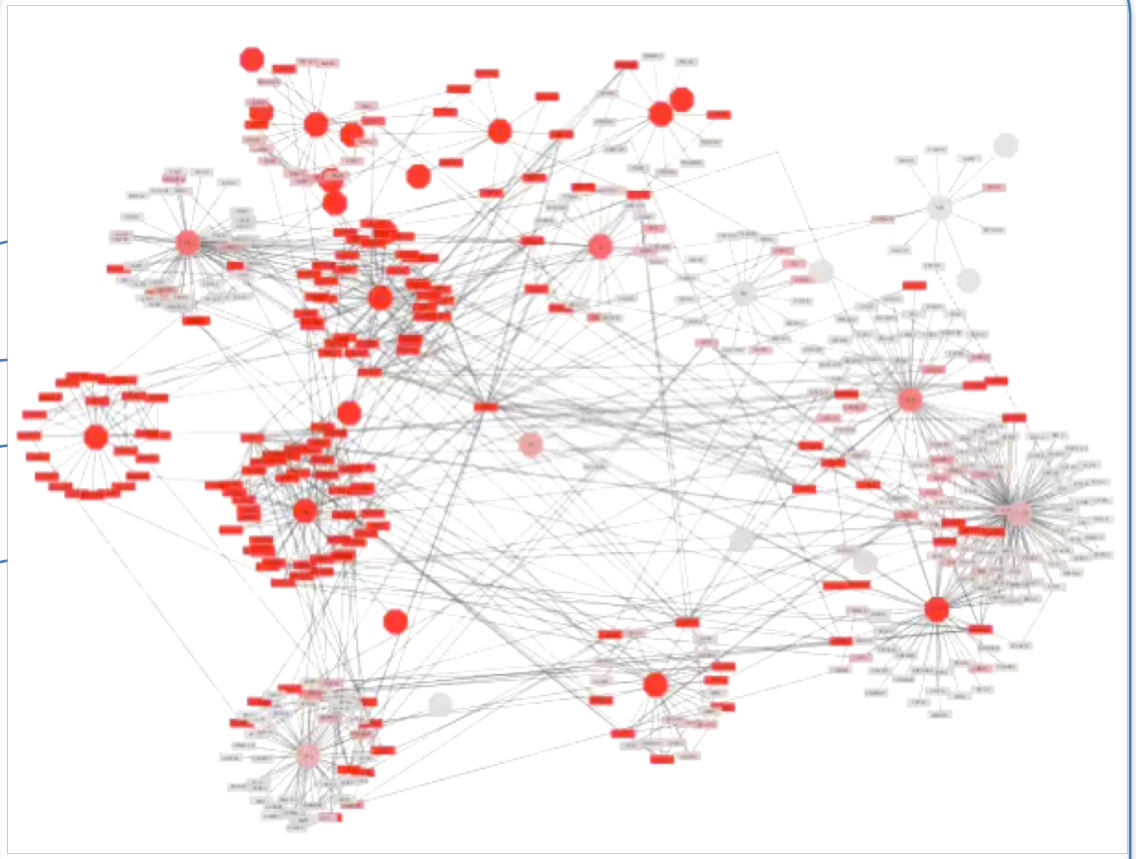


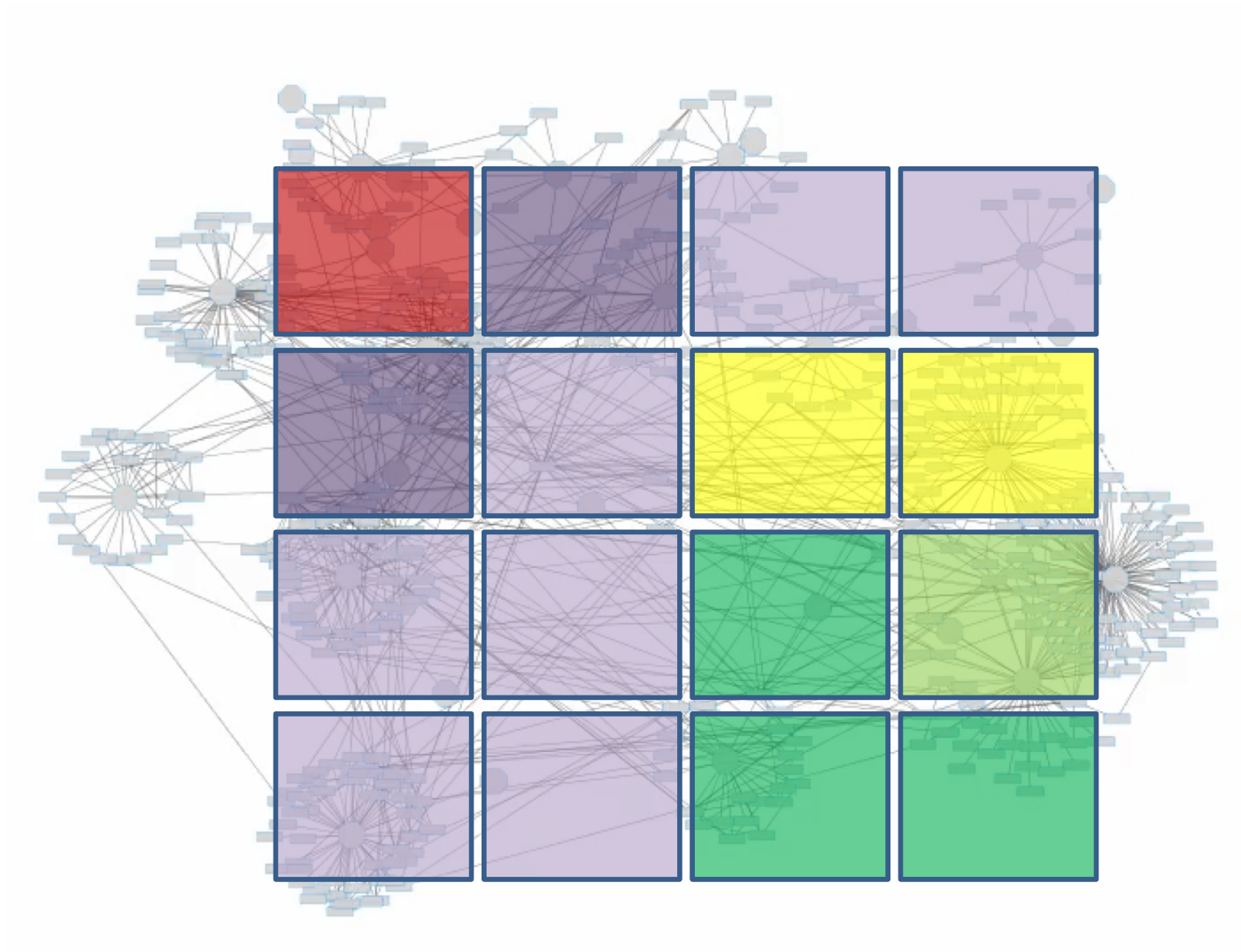



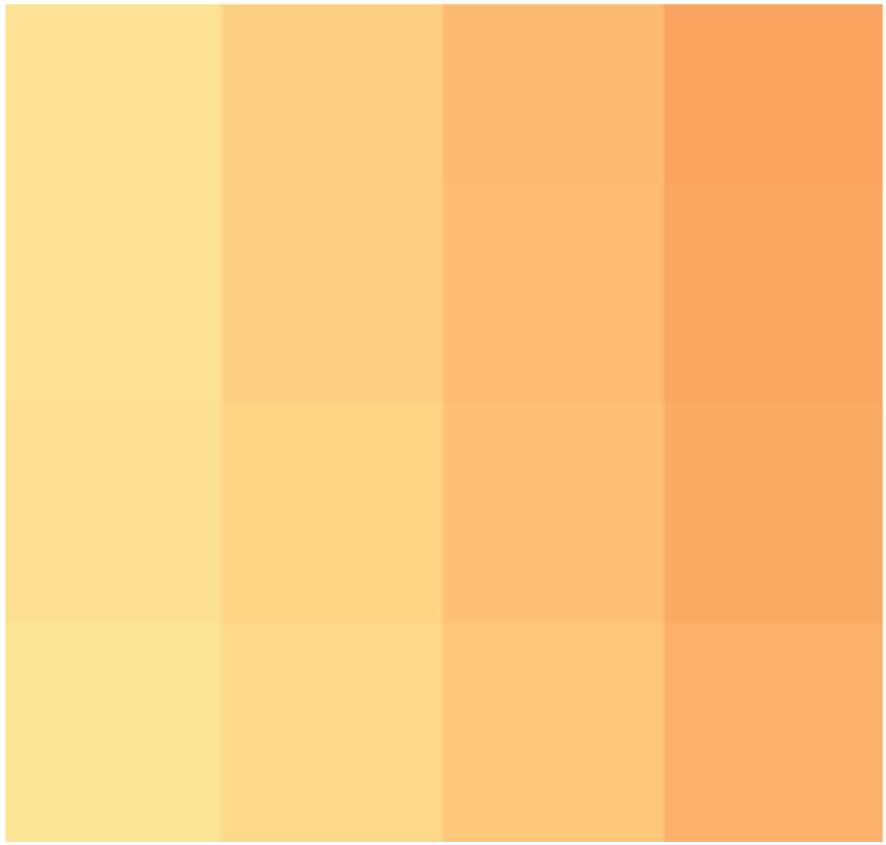
EHR

SPOKE


Medication Orders
Medication Orders
Medication Orders
Medication Orders
Medication Orders
Diagnosis
Labs
SPOKE entry points

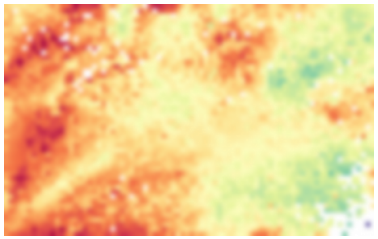




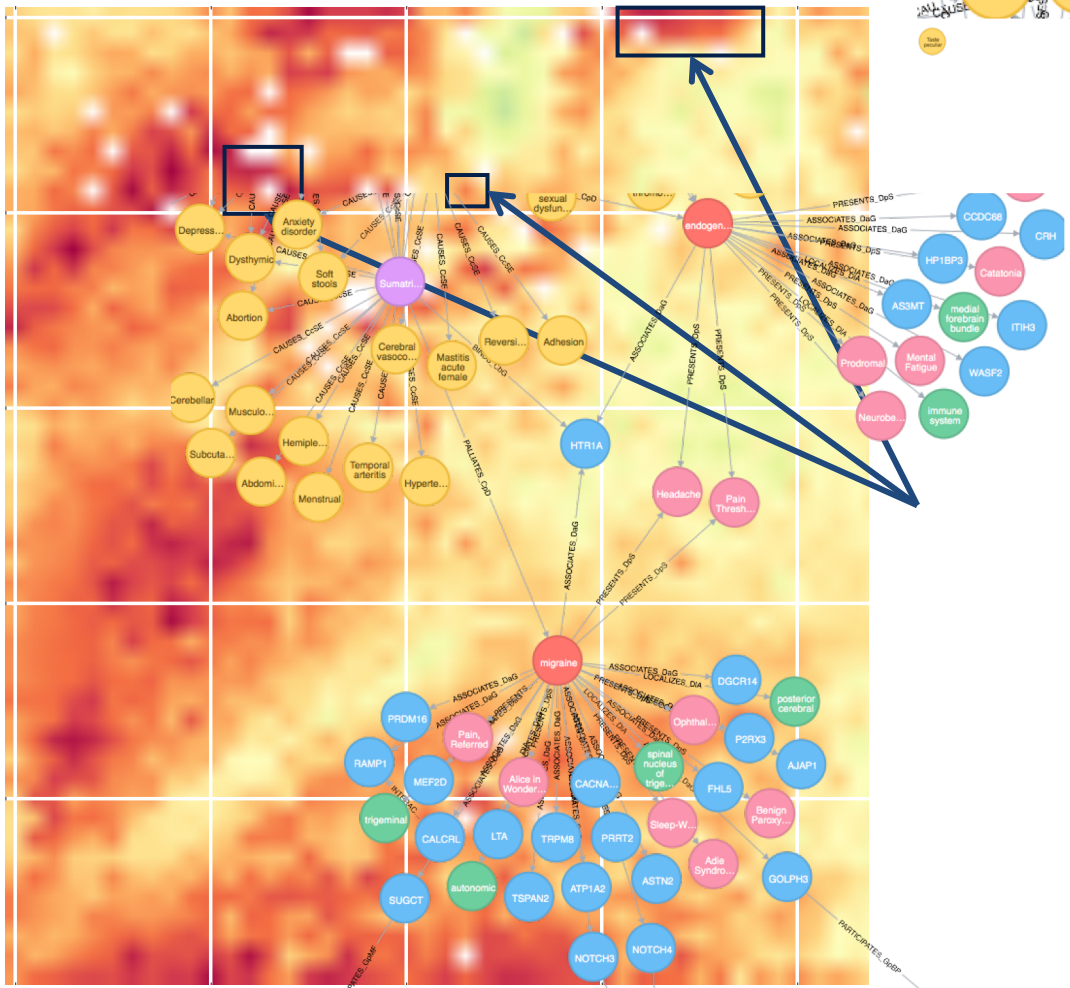


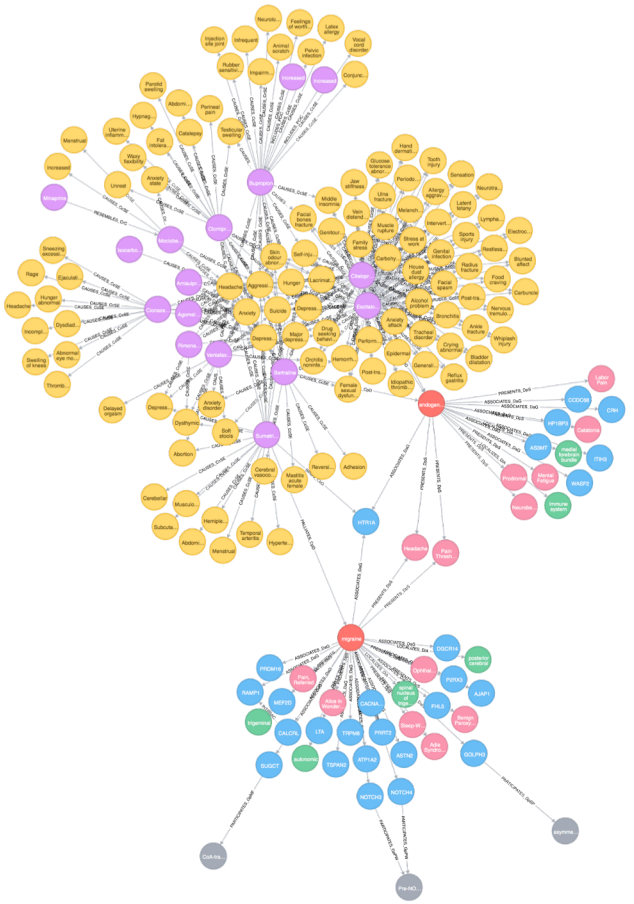
# Most Similar Diseases to the Average PTSD Patient at UCSF

PTSD (n~4000)

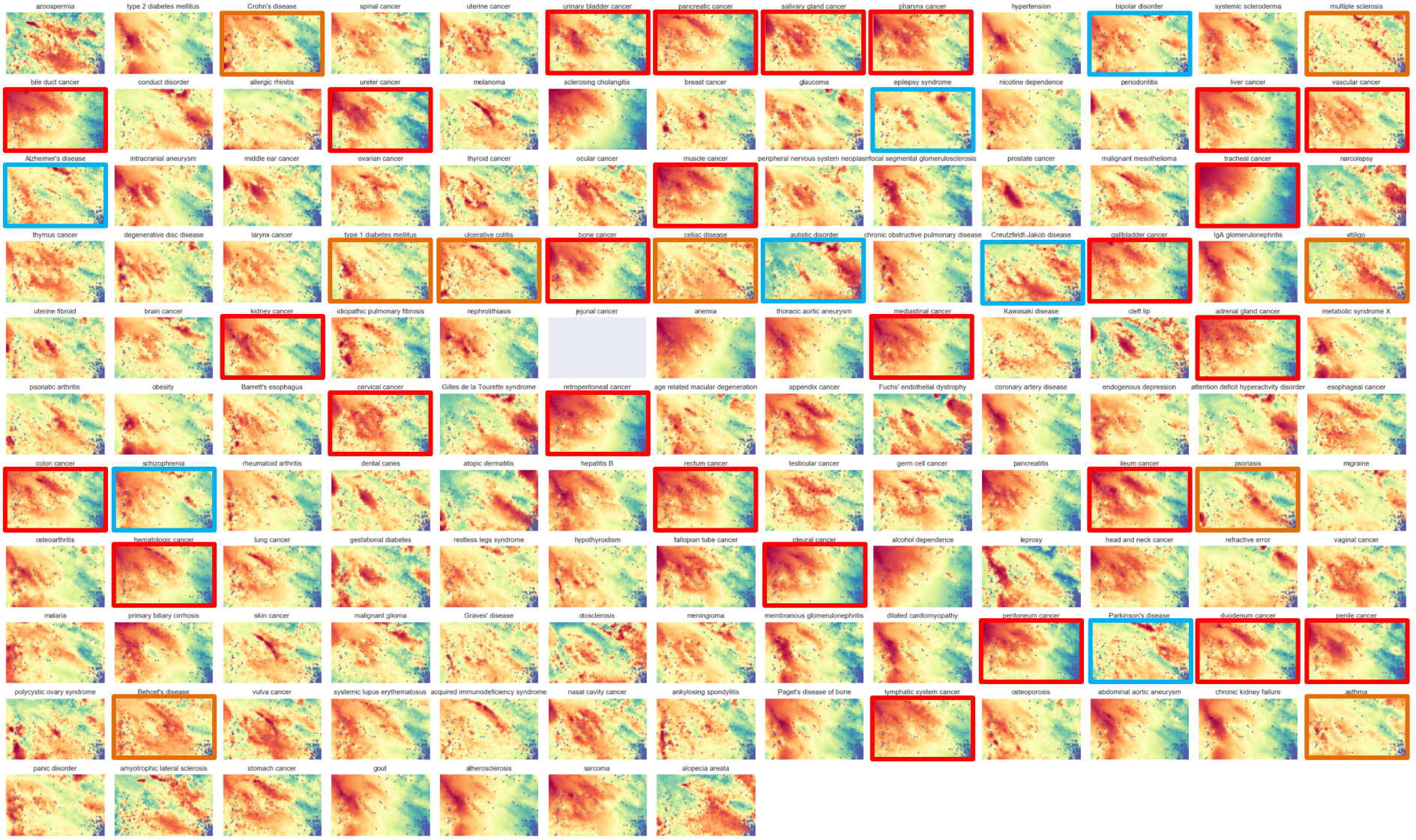


# Pixels That are Significantly [illegible] & Unique to PTSD

# 137 diseases

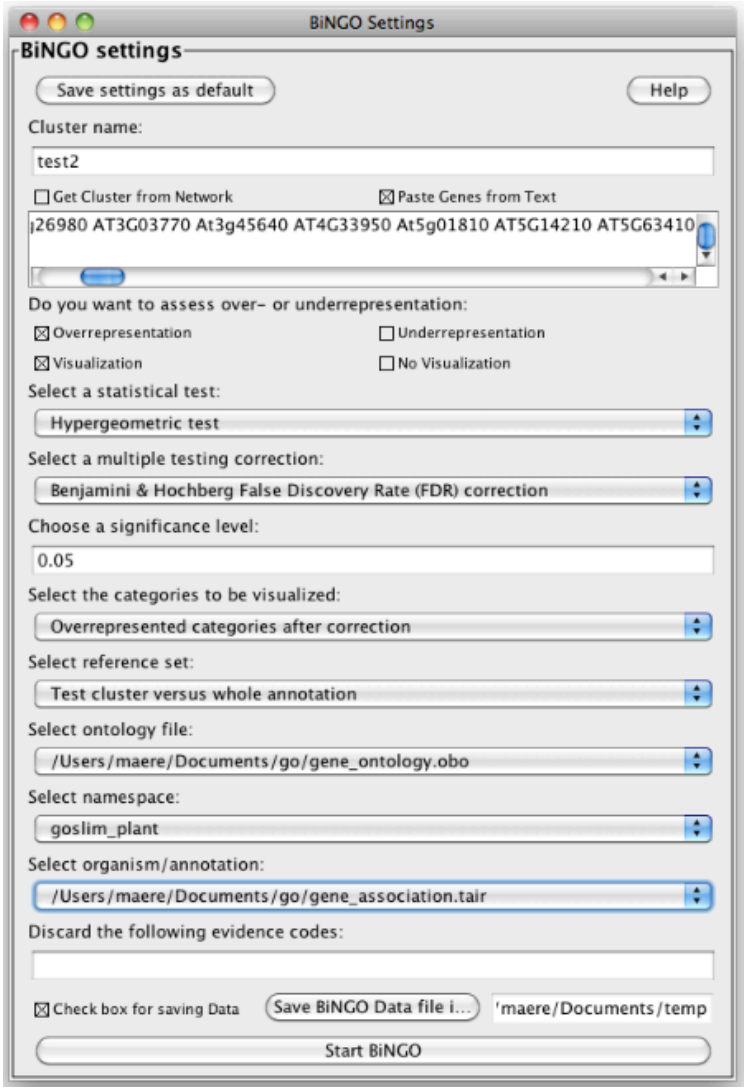



Cytoscape

# Cytoscape Apps

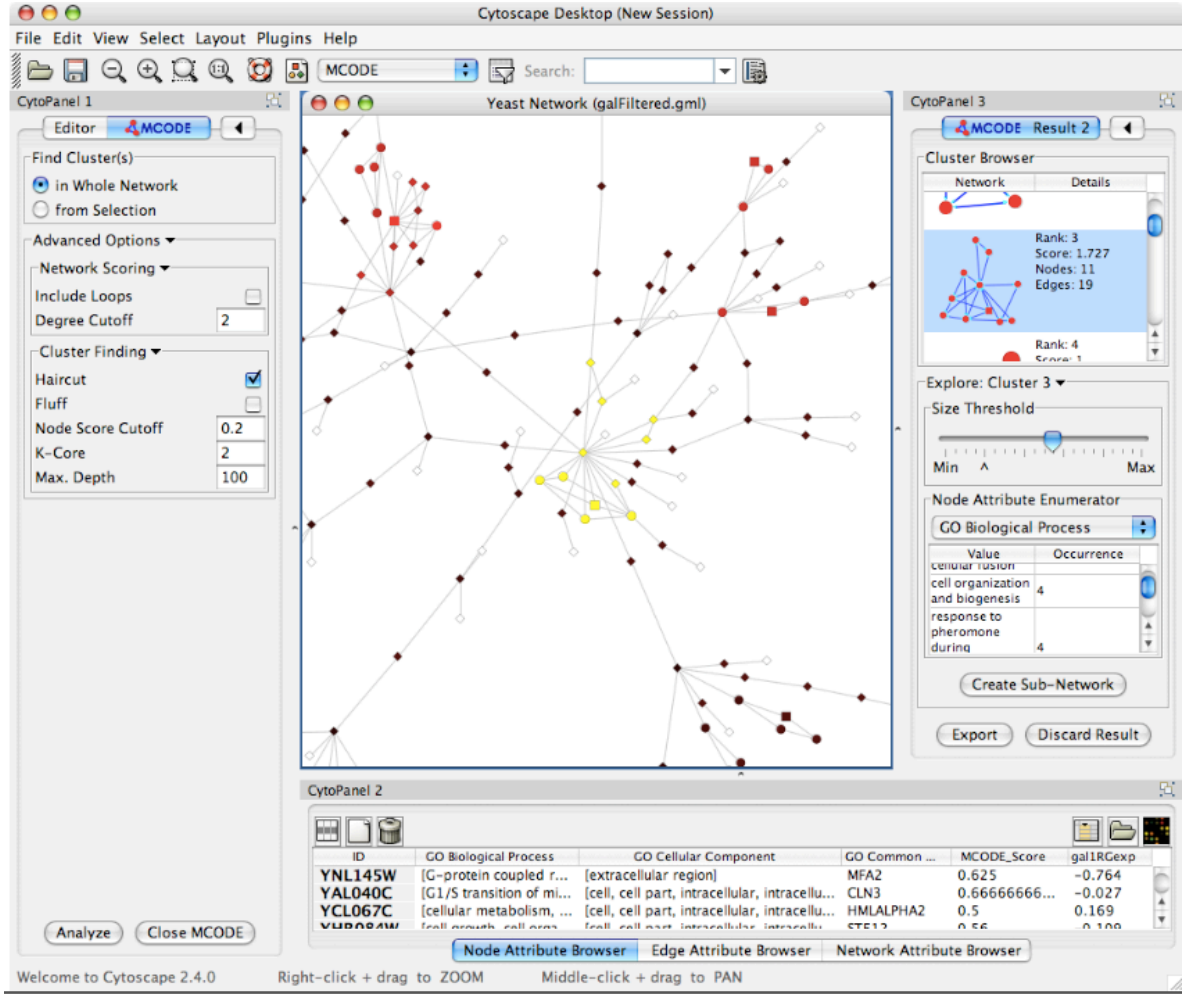
- jActive modules
- MCODE
- BINGO
- GeneMANIA
- PINBPA
- iCTNet

# BINGO: Gene ontology analysis



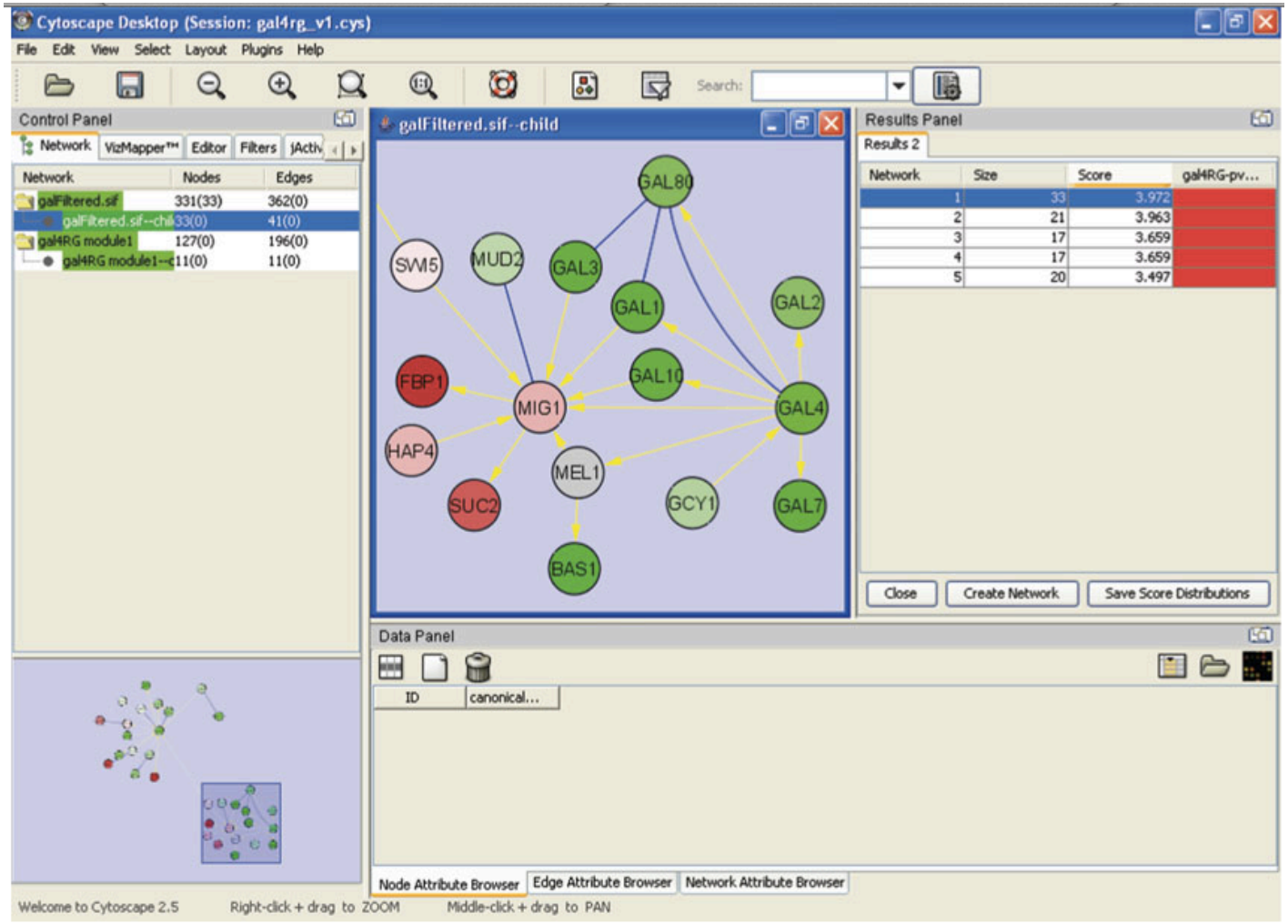
# MCODE

(Clusters a given network based on topology to find densely connected regions)



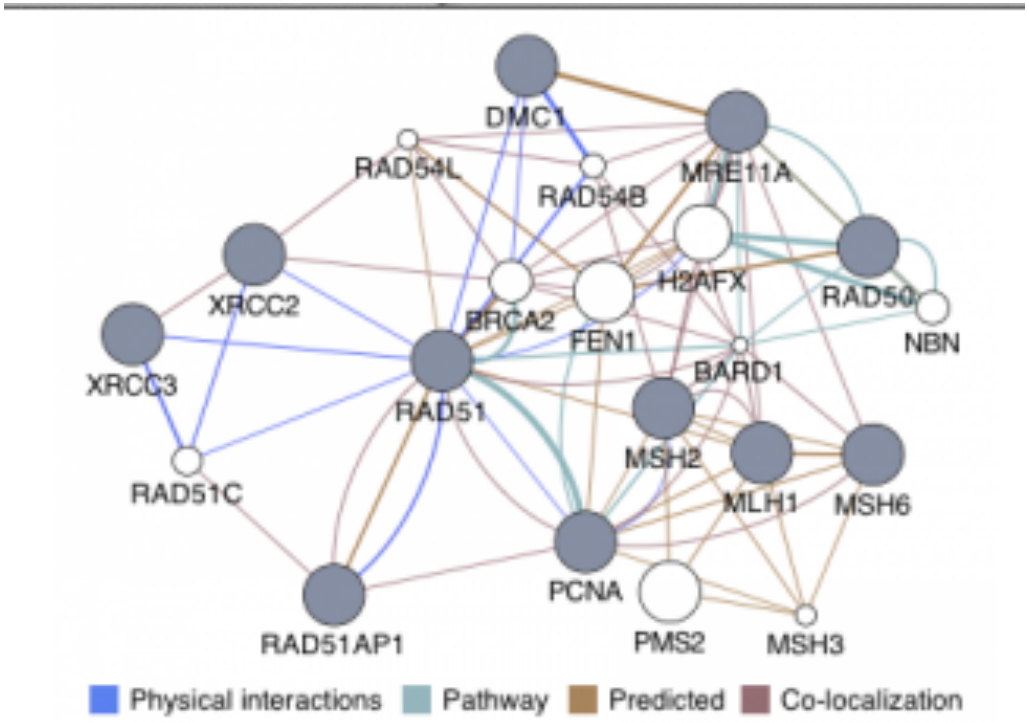
# jActive modules

Finds clusters where member nodes show significant changes in expression levels



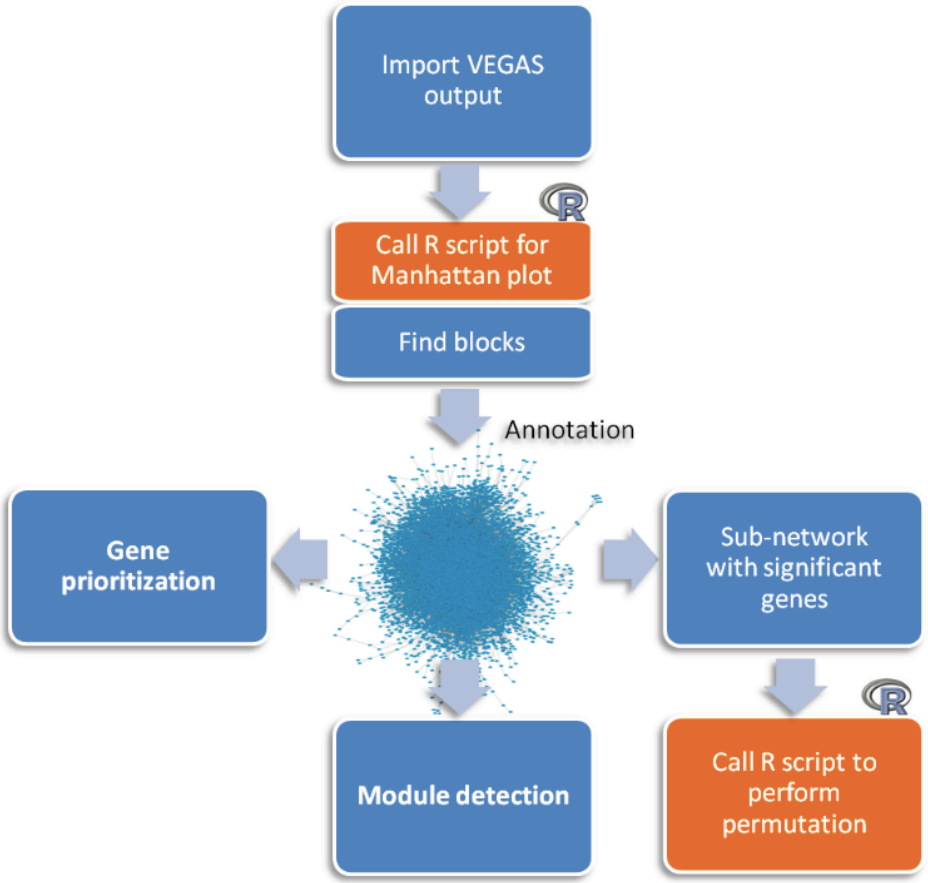
# GeneMANIA

Imports interaction networks from public databases from a list of genes with their annotations and putative functions.



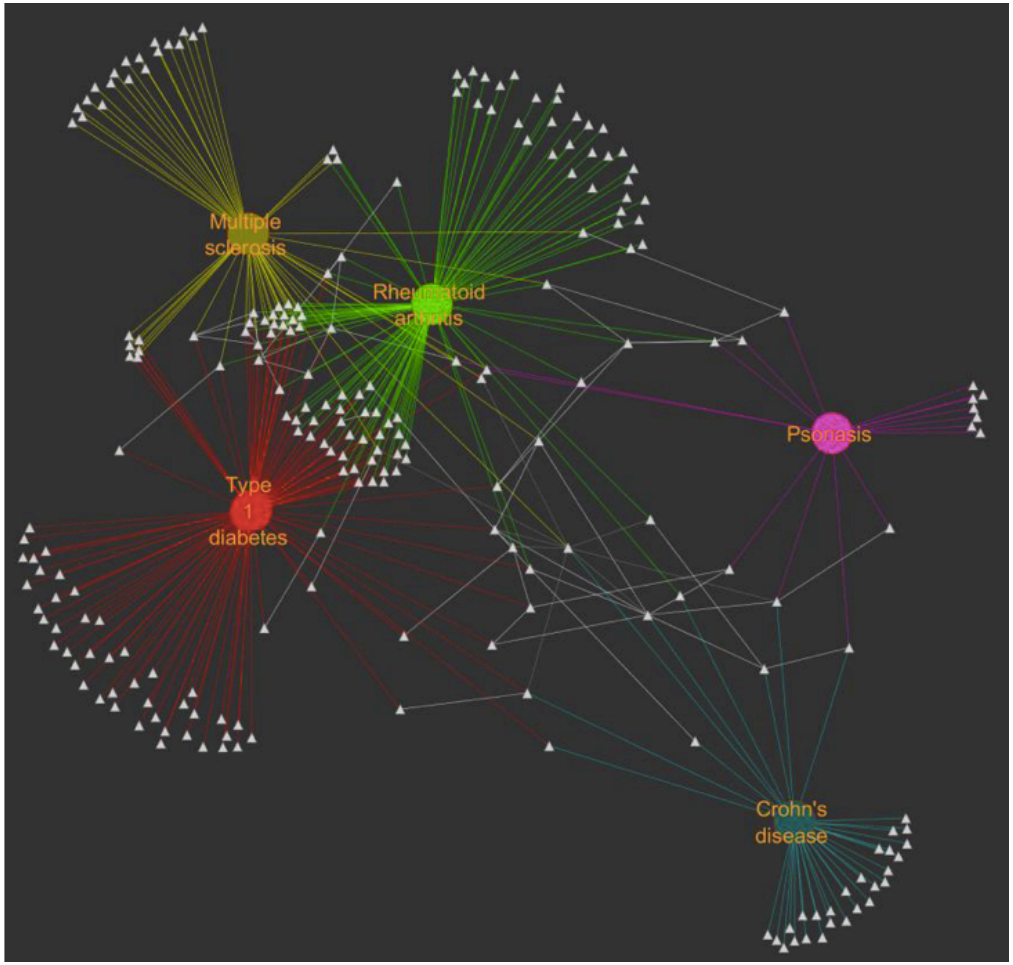
# PINBPA

## Protein-interaction-network-based Pathway Analysis of GWAS



# iCTNet:

assembles human disease, tissue, gene and drug-target interactions



# Reading assignment

SCIENTIFIC REPORTS

OPEN

## Tissue Specificity of Human Disease Module

Maksim Kitsak[1,2], Amitabh Sharma[1,2,3], Jörg Menche[1,2,4,5], Emre Guney[1,2,6], Susan Dina Ghiassian[1,2,7], Joseph Loscalzo[8] & Albert-László Barabási[1,2,4,7,8,9]

Received: 16 May 2016
Accepted: 20 September 2016
Published: 17 October 2016

Genes carrying mutations associated with genetic diseases are present in all human cells; yet, clinical manifestations of genetic diseases are usually highly tissue-specific. Although some disease genes are expressed only in selected tissues, the expression patterns of disease genes alone cannot explain the observed tissue specificity of human diseases. Here we hypothesize that for a disease to manifest itself in a particular tissue, a whole functional subnetwork of genes (disease module) needs to be expressed in that tissue. Driven by this hypothesis, we conducted a systematic study of the expression patterns of disease genes within the human interactome. We find that genes expressed in a specific tissue tend to be localized in the same neighborhood of the interactome. By contrast, genes expressed in different tissues are segregated in distinct network neighborhoods. Most important, we show that it is the integrity and the completeness of the expression of the disease module that determines disease manifestation in selected tissues. This approach allows us to construct a disease-tissue network that confirms known and predicts unexpected disease-tissue associations.

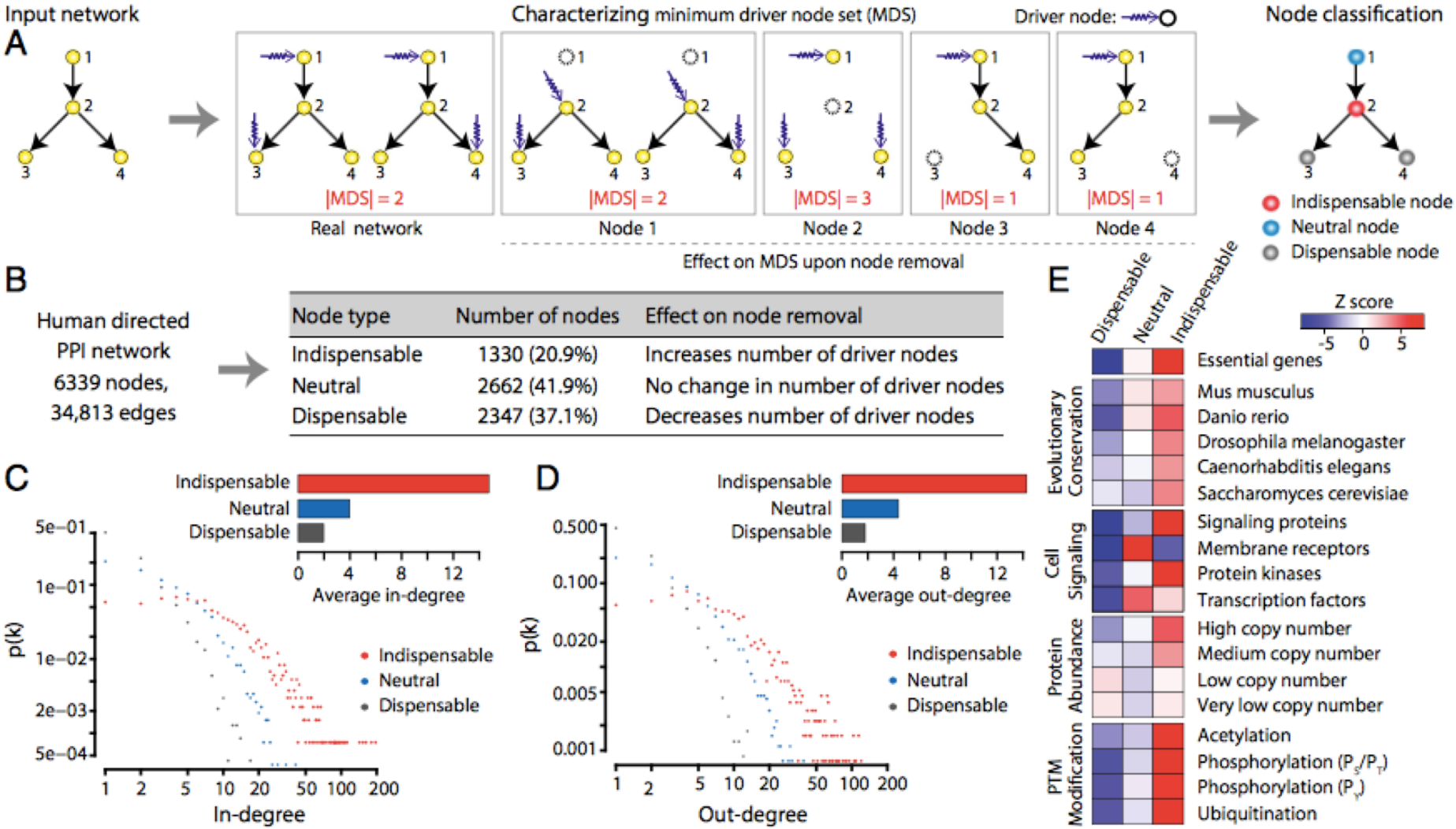
PNAS

# Controllability analysis of the directed human protein interaction network identifies disease genes and drug targets

**Arunachalam Vinayagam[a,1], Travis E. Gibson[b], Ho-Joon Lee[c,2], Bahar Yilmazel[d,e], Charles Roesel[d,e,3], Yanhui Hu[a,d], Young Kwon[a], Amitabh Sharma[b,f,g], Yang-Yu Liu[b,f,g,1], Norbert Perrimon[a,h,1], and Albert-László Barabási[f,g,1]**

[a]Department of Genetics, Harvard Medical School, Boston, MA 02115; [b]Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115; [c]Department of Systems Biology, Harvard Medical School, Boston, MA 02115; [d]Drosophila RNAi Screening Center, Department of Genetics, Harvard Medical School, Boston, MA 02115; [e]Bioinformatics Program, Northeastern University, Boston, MA 02115; [f]Center for Complex Network Research, Department of Physics, Northeastern University, Boston, MA 02115; [g]Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115; and [h]Howard Hughes Medical Institute, Harvard Medical School, MA 02115

Fig 1


Input network
A
Characterizing minimum driver node set (MDS)
Driver node:
Node classification
|MDS| = 2
|MDS| = 2
|MDS| = 3
|MDS| = 1
|MDS| = 1
Real network
Node 1
Node 2
Node 3
Node 4
Indispensable node
Neutral node
Dispensable node
Effect on MDS upon node removal
B
Human directed PPI network 6339 nodes, 34,813 edges
Node type | Number of nodes | Effect on node removal
Indispensable | 1330 (20.9%) | Increases number of driver nodes
Neutral | 2662 (41.9%) | No change in number of driver nodes
Dispensable | 2347 (37.1%) | Decreases number of driver nodes
C
Indispensable
Neutral
Dispensable
Average in-degree
p(k)
In-degree
D
Average out-degree
Out-degree
E
Dispensable
Neutral
Indispensable
Z score
Essential genes
Evolutionary Conservation
Mus musculus
Danio rerio
Drosophila melanogaster
Caenorhabditis elegans
Saccharomyces cerevisiae
Cell Signaling
Signaling proteins
Membrane receptors
Protein kinases
Transcription factors
Protein Abundance
High copy number
Medium copy number
Low copy number
Very low copy number
PTM Modification
Acetylation
Phosphorylation (PS/PT)
Phosphorylation (PY)
Ubiquitination

Kim *et al. BMC Bioinformatics* (2019) 20:328
https://doi.org/10.1186/s12859-019-2897-z

BMC Bioinformatics

RESEARCH ARTICLE Open Access

# Identification of critical connectors in the directed reaction-centric graphs of microbial metabolic networks

Eun-Youn Kim[1], Daniel Ashlock[2] and Sung Ho Yoon[3*]

# Questions

- 1- What is the difference between bridging centrality and betweenness centrality?
- 2- Can you identify whose position might represent a node of high betweenness centrality in a large, hierarchical organization such as UCSF?
- 3- Why nodes with high degree tend not to be important for information flow?

- **Challenge question 1:** What is the difference between the concept of "cascade number" and network controllability?

- **Challenge question 2:** Would the results of this analysis change significantly if authors used controllability instead of cascade number to identify essential reactions? Why?