Network Medicine: Systems biology based approach to human disease

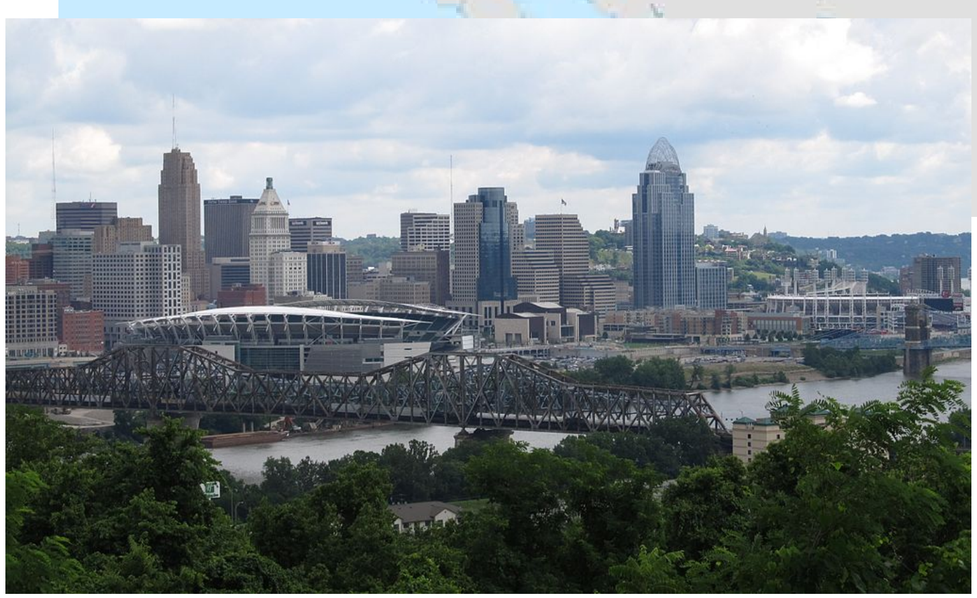
Anil Jegga
Division of Biomedical Informatics
Cincinnati Children's Hospital & Medical Center
Departments of Pediatrics & Computer Sciences
University of Cincinnati, USA
http://anil.cchmc.org

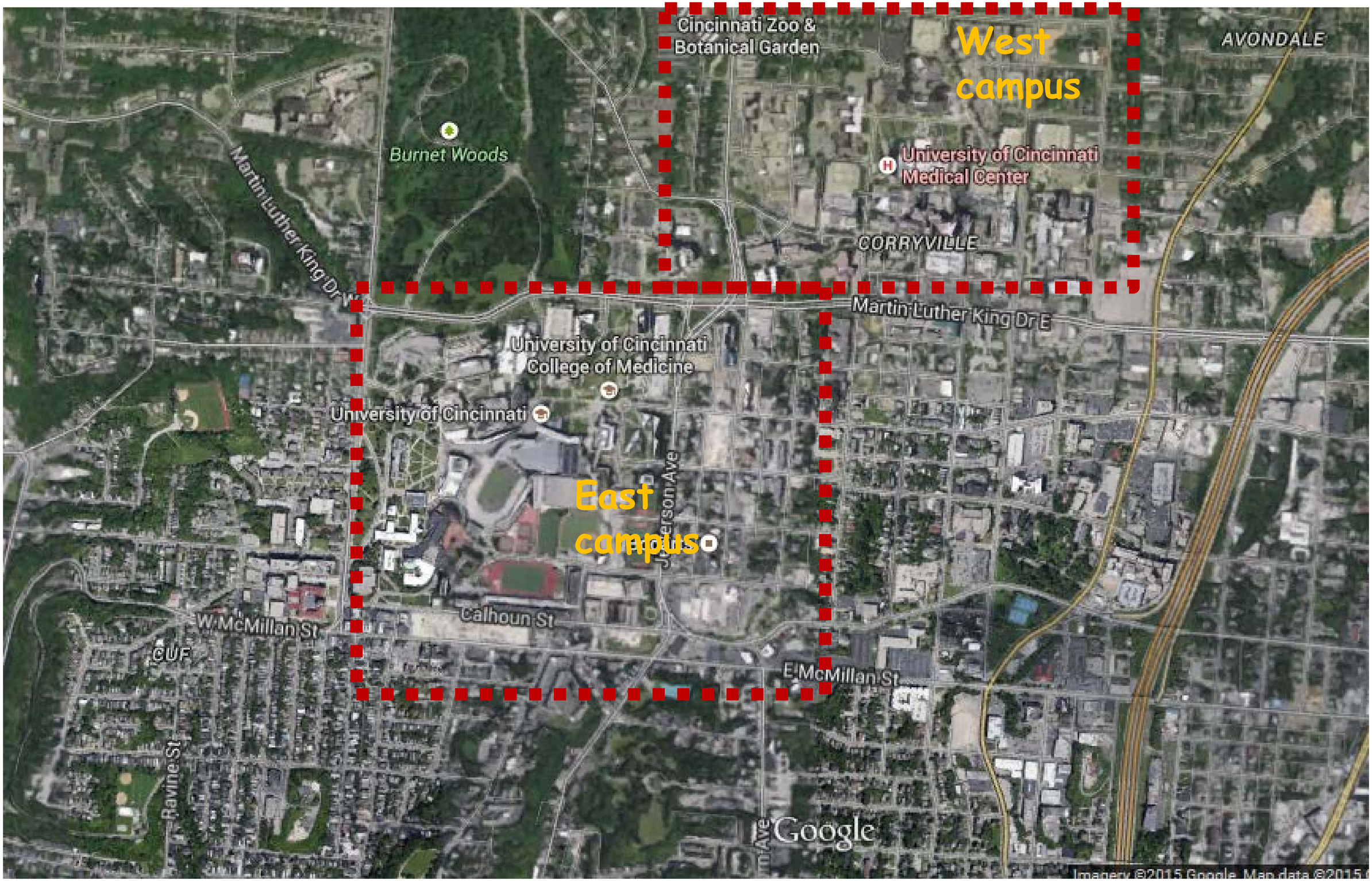
Acknowledgements

Cheng Zhu
Minlu Zhang
Chao Wu
Divya Sardana
Jagadeesh Patchala
Jing Chen
Eric Bardes
Bruce Aronow

Cincinnati Children's®

Cincinnati,
Ohio




Cincinnati Zoo &
Botanical Garden
West
campus
AVONDALE
Burnet Woods
Martin Luther King Dr W
University of Cincinnati
Medical Center
CORRYVILLE
Martin Luther King Dr E
University of Cincinnati
College of Medicine
University of Cincinnati
East
campus
Jefferson Ave
Calhoun St
W McMillan St
CUF
E McMillan St
Ravine St
Google
Imagery ©2015 Google, Map data ©2015

About Cincinnati Children's

- Cincinnati Children's, a nonprofit academic medical center established in **1883**, is one of the oldest and most distinguished pediatric hospitals in the United States.
- With nearly 600 registered beds, Cincinnati
- Children's had more than 1.1 million patient encounters and served patients from all 50 states in the USA and **53 countries** in fiscal 2013.

Clinical Facility

- CHRF – Children's Hospital Research Foundation
- Sabin oral **polio vaccine** to the development of surfactant to the latest in genomics and molecular medicine, Cincinnati Children's has a rich history of research achievement.
- **>million square feet** of lab space and core services such as transgenic mouse models, high-throughput DNA analysis, biomedical informatics, a pluripotent stem cell facility, viral vector development and much more.
- **Research funding** has increased more than **400 percent** from $49 million in 2000 to more than $200 million in FY 2014. - **Rank No. 2 in research grants** from the NIH to pediatric institutions.

Outline

- **Brief Overview**
- **Rare/Orphan Disease Networks**
  - Gene-based networks
  - Functional linkage networks
  - Literature-based networks
  - Characteristics of orphan disease causing mutant genes
- **Candidate gene discovery and prioritization**
  - Functional enrichment-based approaches
  - Model organism phenotype
- **Drug repositioning**
  - Gene- and pathway- based networks
  - Phenotype-based networks

Facts & Figures

1. **Orphan Diseases** (OD) or Rare diseases (RD) are life-threatening or chronically debilitating diseases with a **low prevalence** and a **high level of complexity**.
2. A disease is termed as orphan or rare if there is a prevalence of **<200,000** (or 1 in 1500) in the US each year (US Rare Disease Act of 2002).
3. **~7000** distinct, known RDs or ODs.
4. About **30 million** (~1 in 10; cumulative prevalence) people in the US and about **350 million** people globally suffer from a rare disease.

5. Most of the rare diseases are genetic diseases, the others being rare cancers, auto-immune diseases, congenital malformations, toxic and infectious diseases.
6. More than 50% of rare diseases affect children (NIH) and about 30% of rare disease patients die before the age of five (Orphanet)
7. A disease may be considered rare in one part of the world, or in a particular group of people, but still be common in another (e.g., tuberculosis, malaria).

Orphan Disease - Genes

1. Exact cause for many ODs remains unknown.
2. For a significant number of ODs, the problem can be traced to single gene mutations (genetic diseases). Many of these are inheritable.
3. Environmental factors also known to play a role.
4. A significant number are infectious or parasitic diseases – Some major progress in genomics recently. E.g., malaria, schistosomiasis, leishmaniasis, trypanosomiasis (also classified as neglected diseases).
5. About 2900 of the known ~7000 human ODs have at least one known gene (orphan disease causing mutant gene – **ODMG**) mutation associated with them.

Rare Diseases – India (2011)

Countries in South Asia	Rare Diseases and Disorders Population [1-8]
Afghanistan	1,530,006
Bangladesh	9,151,081
Bhutan	44,099
India	72,611,605
Maldives	19,037
Nepal	1,589,670
Pakistan	10,999,800
Sri Lanka	1,216,656

- Rare Disease: fewer than **100 patients per 100,000** population
- Ultra rare disease: fewer than **2 patients per 100,000**

http://www.rarediseasesindia.org

States (India)	Rare Diseases and Disorders Population[4]
Andhra Pradesh (AP & TS)	5,079,932
Arunachal Pradesh	82,957
Assam	1,870,156
Bihar	6,228,278
Chhattisgarh	1,532,412
Goa	87,463
Gujarat	3,623,018
Haryana	1,521,185
Himachal Pradesh	411,391
Jammu and Kashmir	752,936
Jharkhand	1,977,974
Karnataka	3,667,842
Kerala	2,003,261
Madhya Pradesh	4,355,854
Maharashtra	6,742,378
Manipur	163,305
Meghalaya	177,840
Mizoram	65,461
Nagaland	118,836
Orissa	2,516,841
Punjab	1,662,254
Rajasthan	4,117,261
Sikkim	36,461
Tamil Nadu	4,328,337
Tripura	220,262
Uttar Pradesh	11,974,891
Uttarakhand	607,005
West Bengal	5,480,864

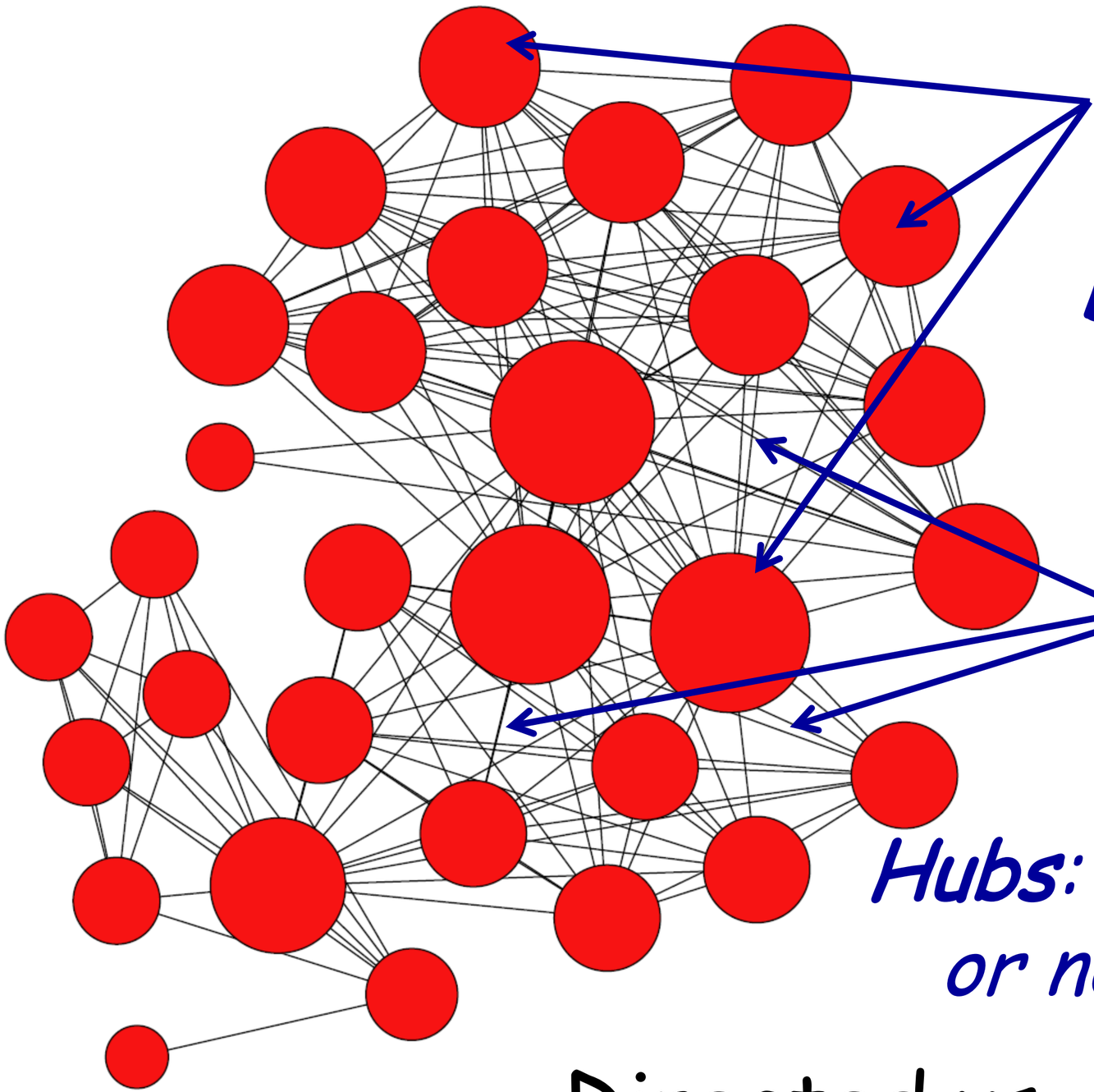
Orphan Disease Networks

Motivation

- Relatively few efforts have addressed scientific or technical questions across a spectrum of orphan diseases.
- Finding common genes, pathways, and targets is critical if we have to make progress in orphan disease research.
- Studies of biological networks can identify common pathways or processes for multiple orphan diseases that are related.
  - Understanding such molecular basis could provide opportunities for interventions that are beneficial for an array of related orphan diseases.
  - Drug repositioning or repurposing
    - Common drug for orphan disease
    - Orphan drug for another orphan disease
    - Orphan drug for common disease

How are ODs and ODMGs different from more common diseases and non-ODMGs (or those causing common diseases)?

Networks - Few Basics



Nodes/vertices

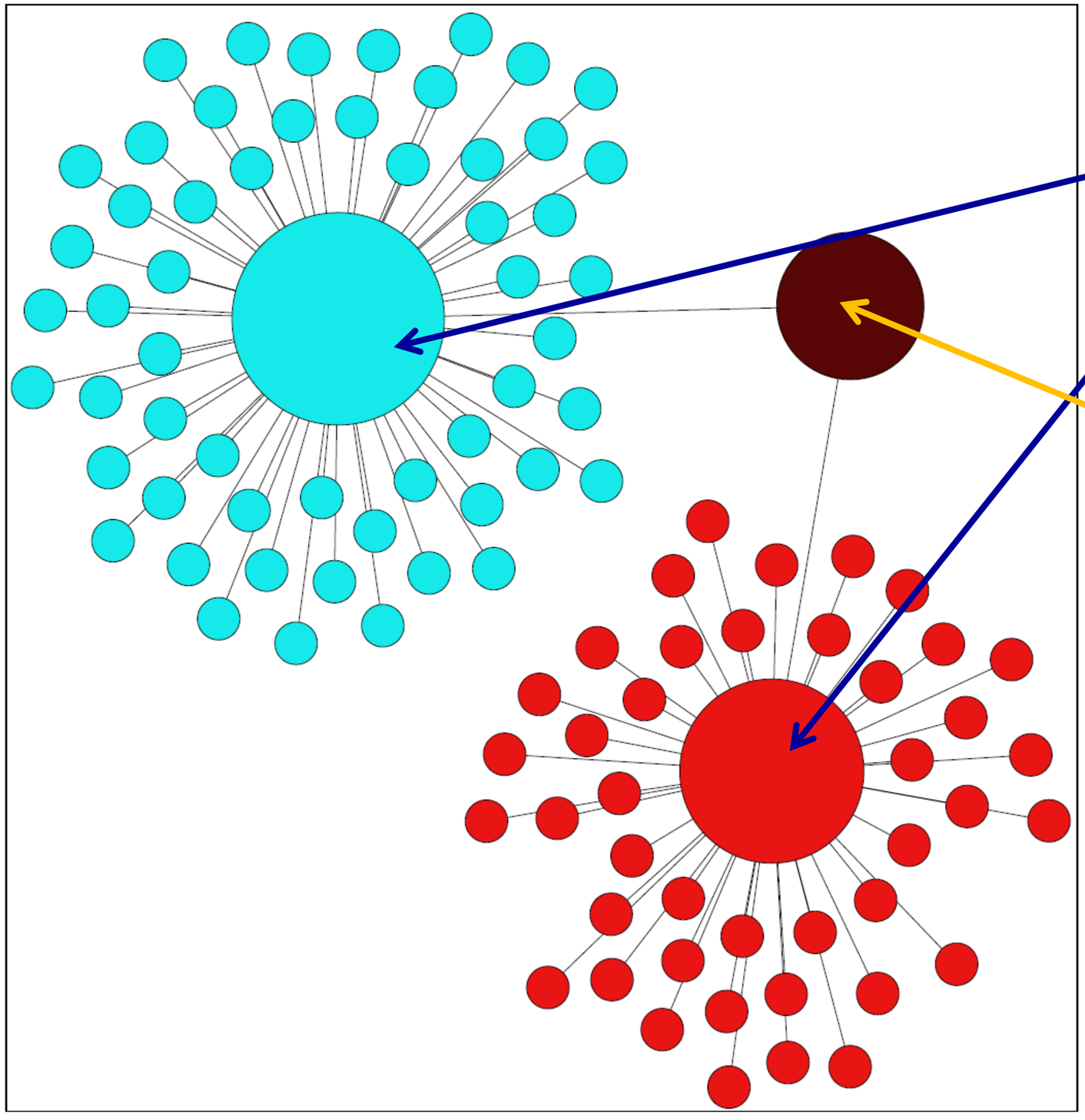
***Degree**: No. of edges or connections a node has*

Edges or connections
- Connect nodes

***Hubs**: Highly connected nodes or nodes with high degree*

Directed vs. undirected network:
Edges have directions or are undirected

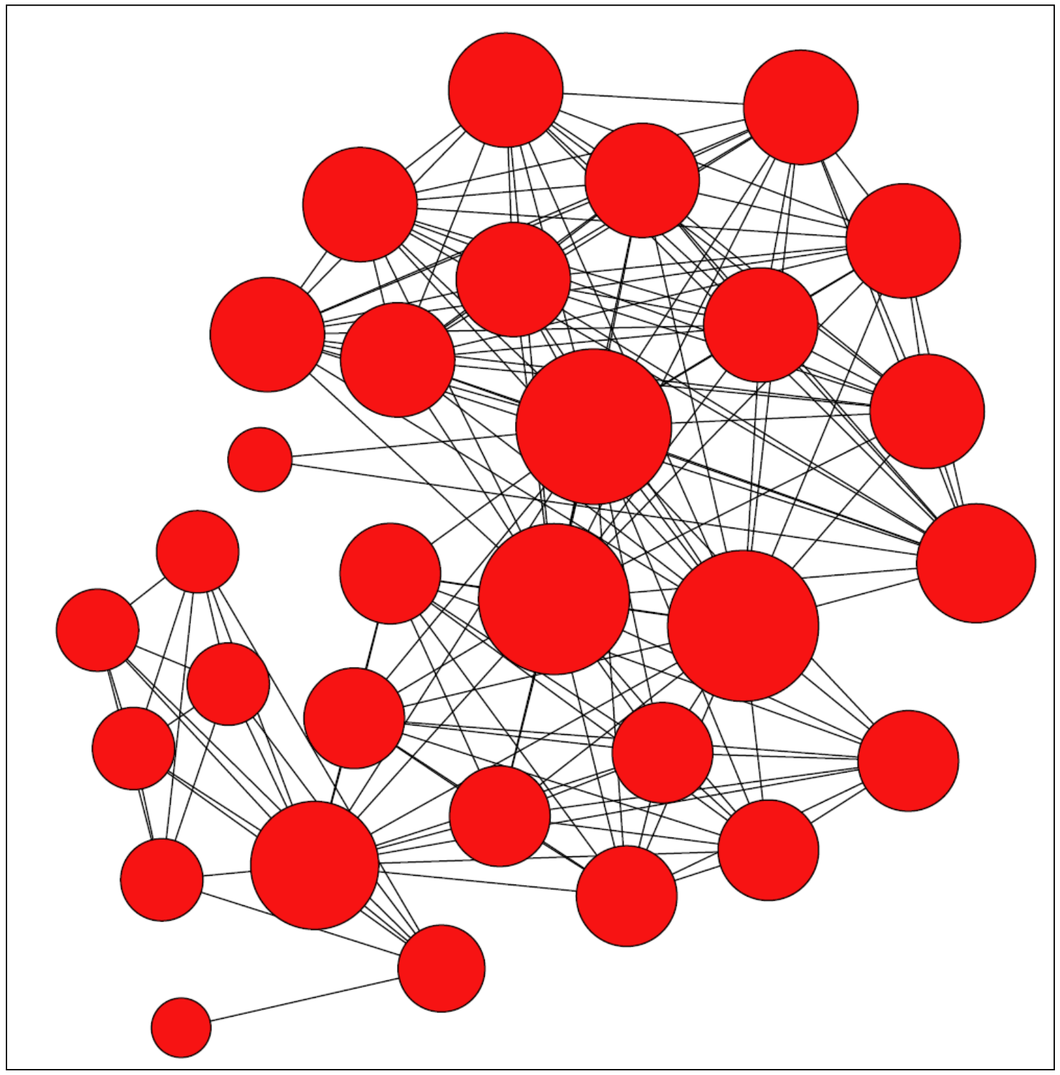
Networks - Few Basics



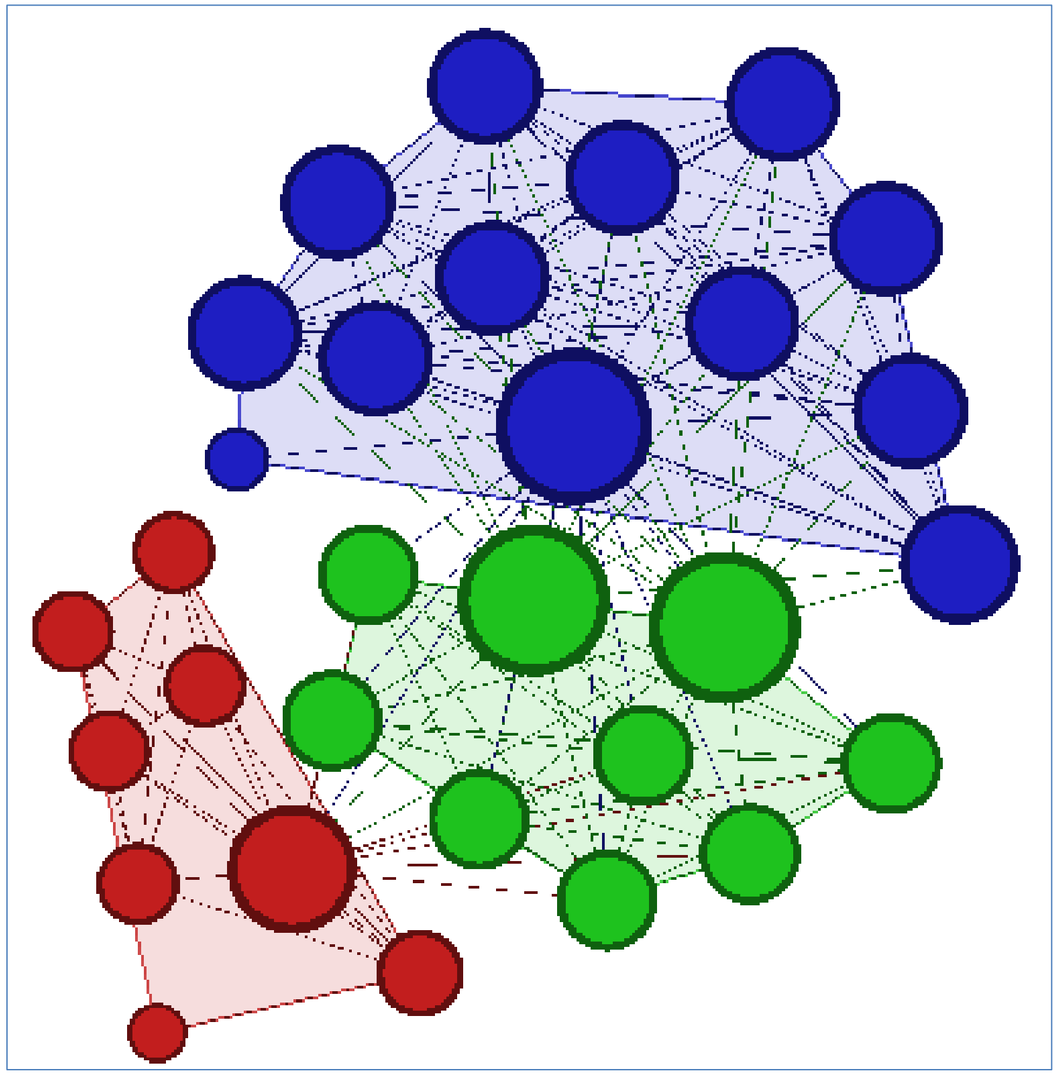
Hubs

Super-Hubs or bottlenecks

Networks - Few Basics



Sub-network or connected component or loosely connected network

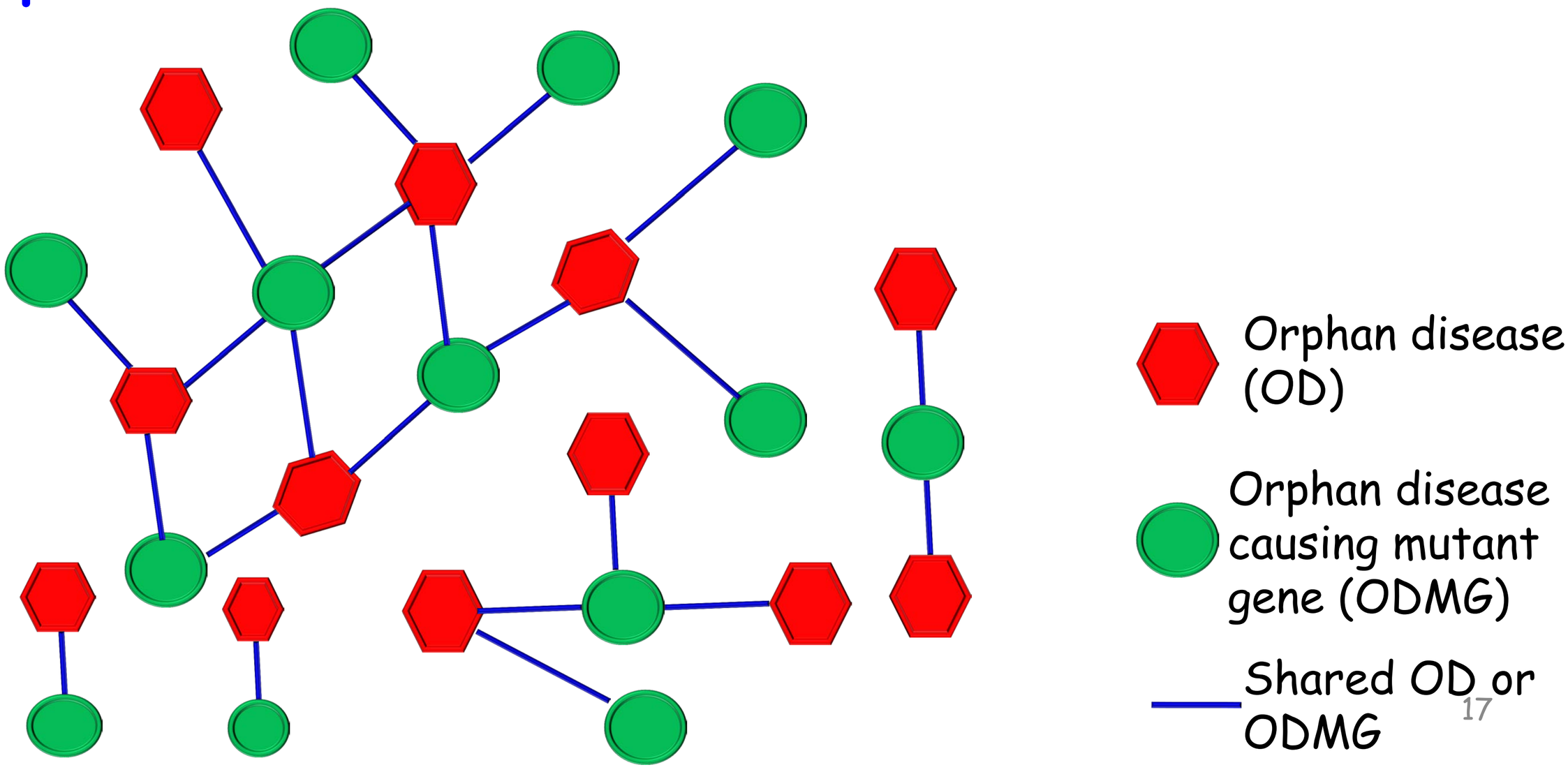


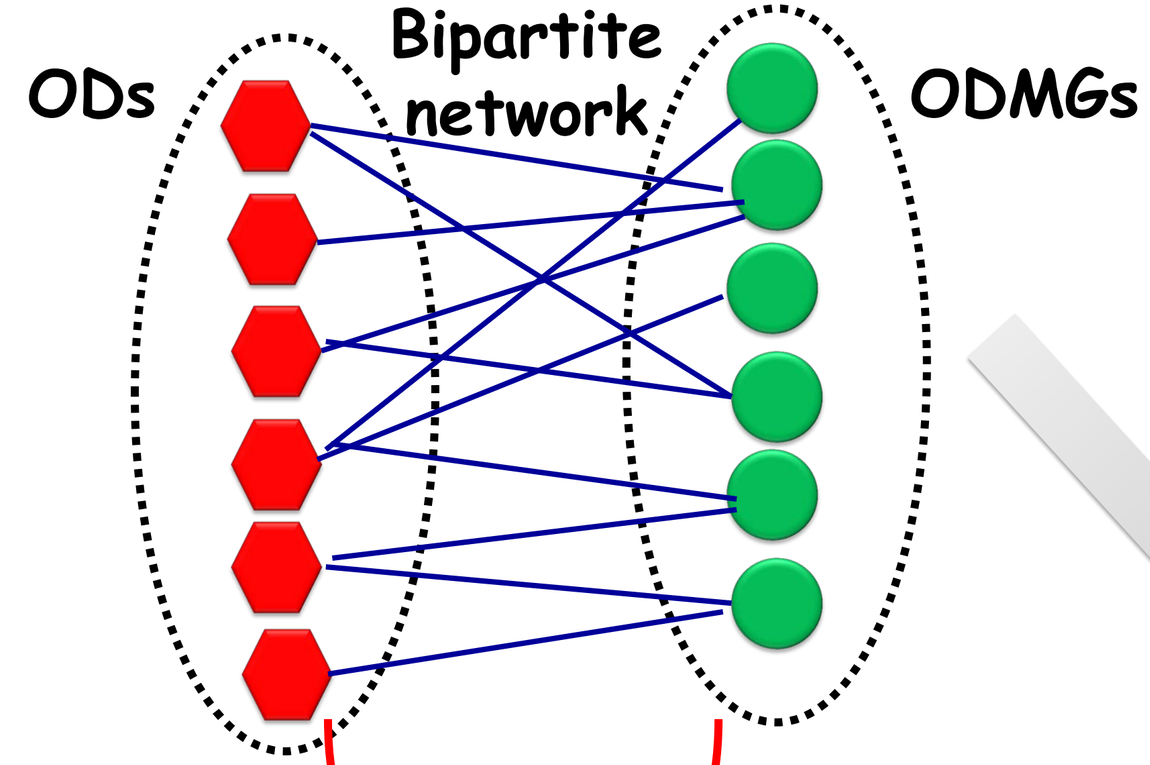
Three Communities or modules or highly interconnected networks

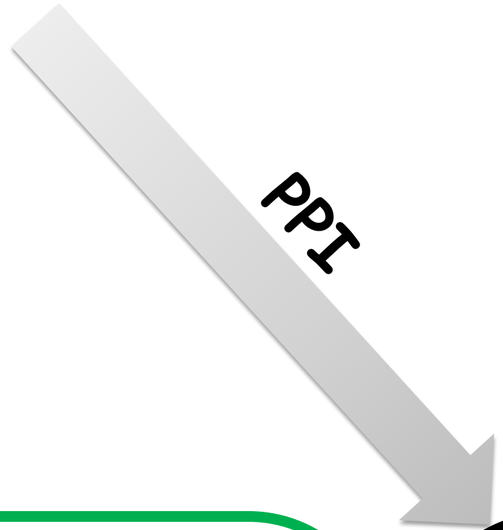
Constructing the networks

An orphan disease and an OD-causing mutant gene are connected by a link if mutations in that gene are implicated in that disorder.

The list of ODs, Orphan Disease-causing Mutant Genes (ODMG), and associations between them were obtained from the **Orphanet database** and the **OMIM**.





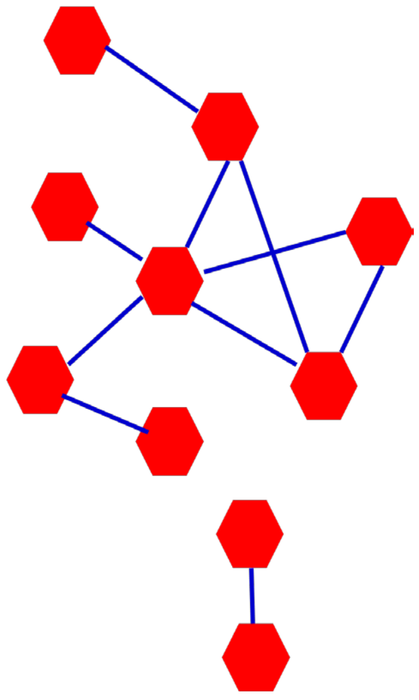


Orphan disease (OD)

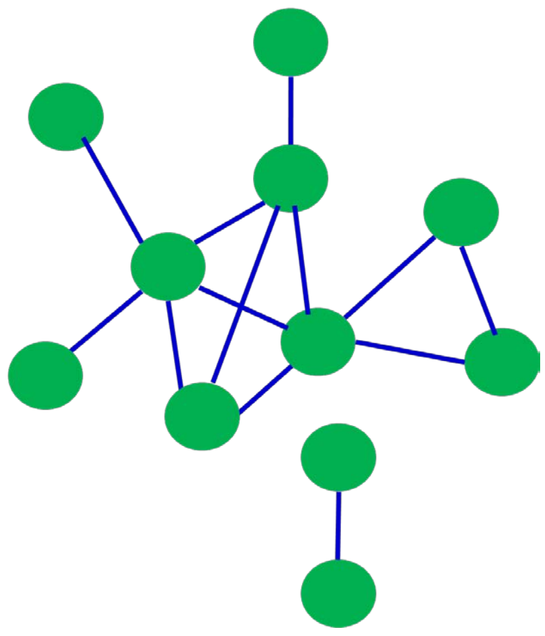
Orphan disease causing mutant gene (ODMG)

Shared OD or ODMG

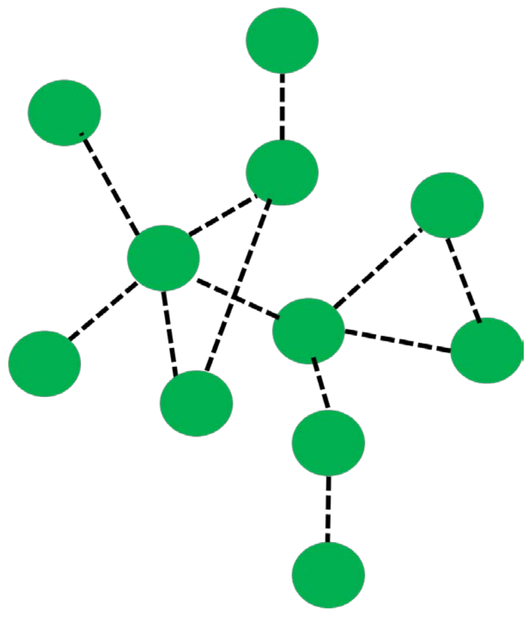
PPI



ODN – Orphan Disease Network
Edge = shared gene



ODMGN – Orphan Disease causing Mutant Gene Network
Edge = shared OD



ODMGI – ODMG Interactome
Edge = Protein interaction

Orphan Diseases

3 Types of Networks:

1. Orphan Disease Network (ODN):
   - ❖ Nodes = Orphan diseases
   - ❖ Edge = shared gene(s)

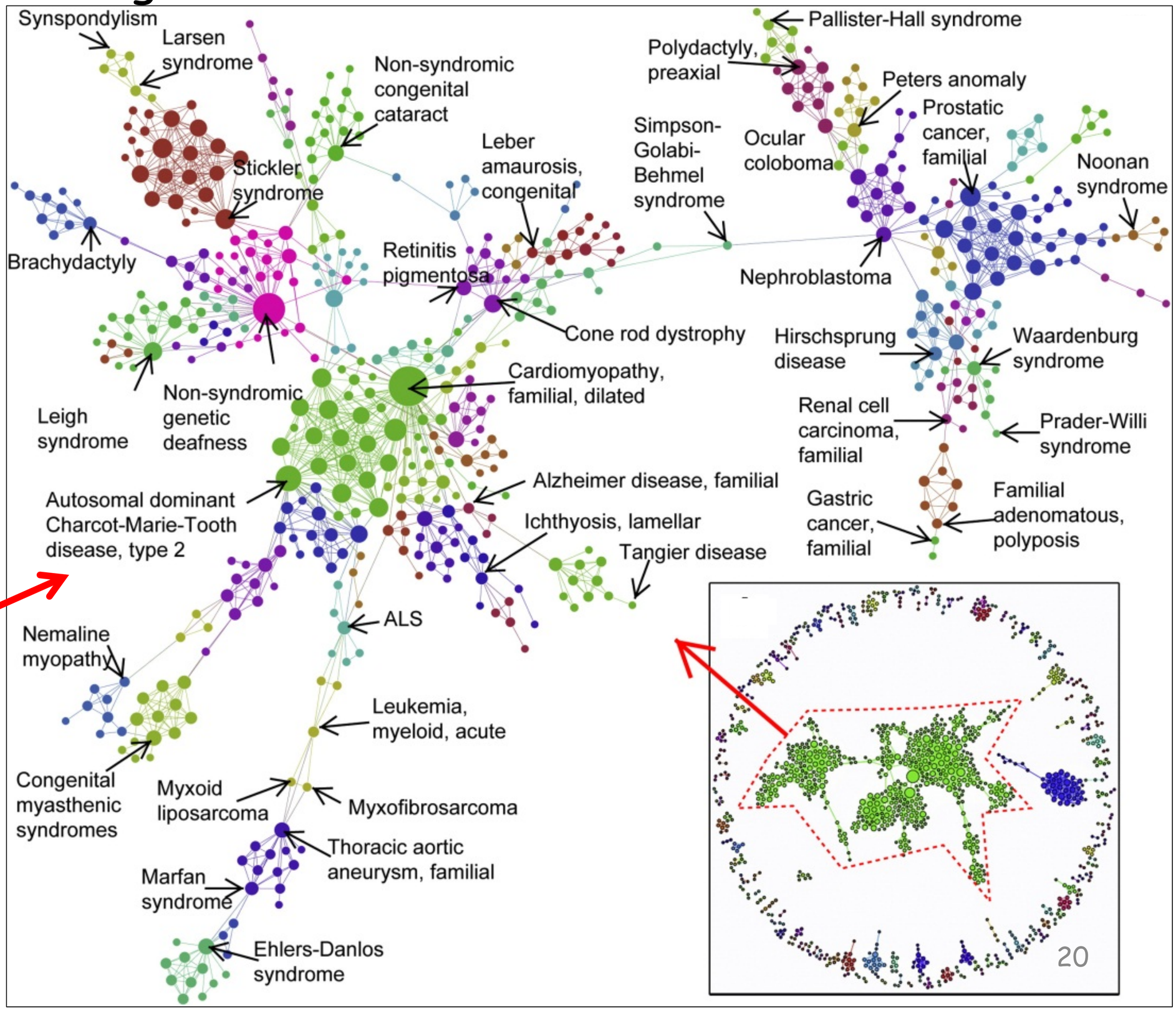
2. Orphan Disease Gene Network (ODMGN):
   - ❖ Nodes = Orphan disease-causing mutant genes
   - ❖ Edge = shared orphan disease(s)

3. Orphan Disease Gene Interactome (ODMGI)
   - ❖ Nodes = Orphan disease-causing mutant genes
   - ❖ Edge = Protein-protein interaction

Gene-based Orphan Disease Network (ODN)

1170/1772 (~66%) ODs are connected to at least another OD through at least one gene.

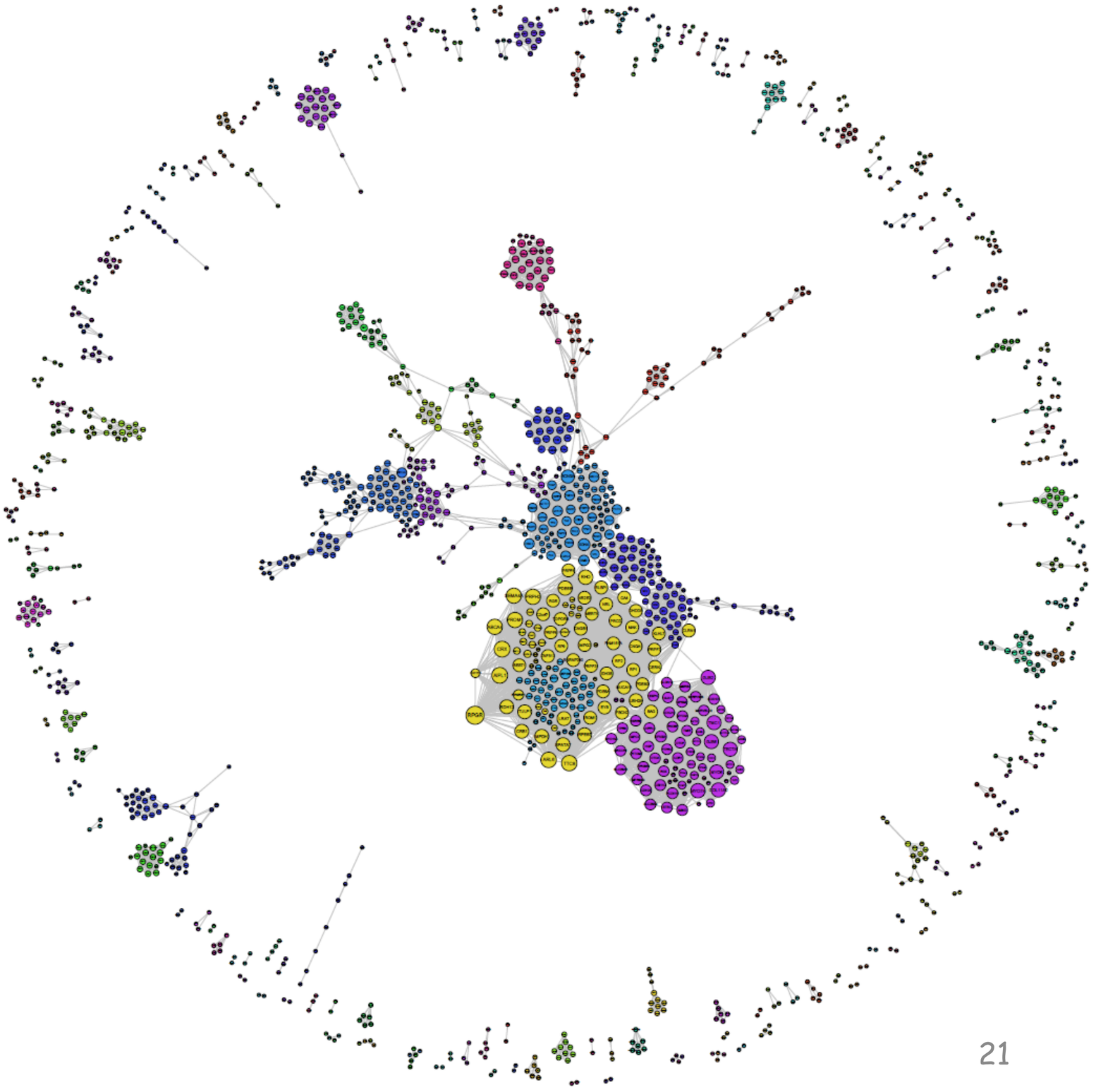
- **1170 nodes (ODs)**
- **2259 edges (shared genes)**
- **184 connected components**
- **Largest connected component has 530 ODs & 1396 edges**
- **274 closely connected modules or communities**



Constructing Orphan Disease causing Mutant Gene Network (ODMGN)

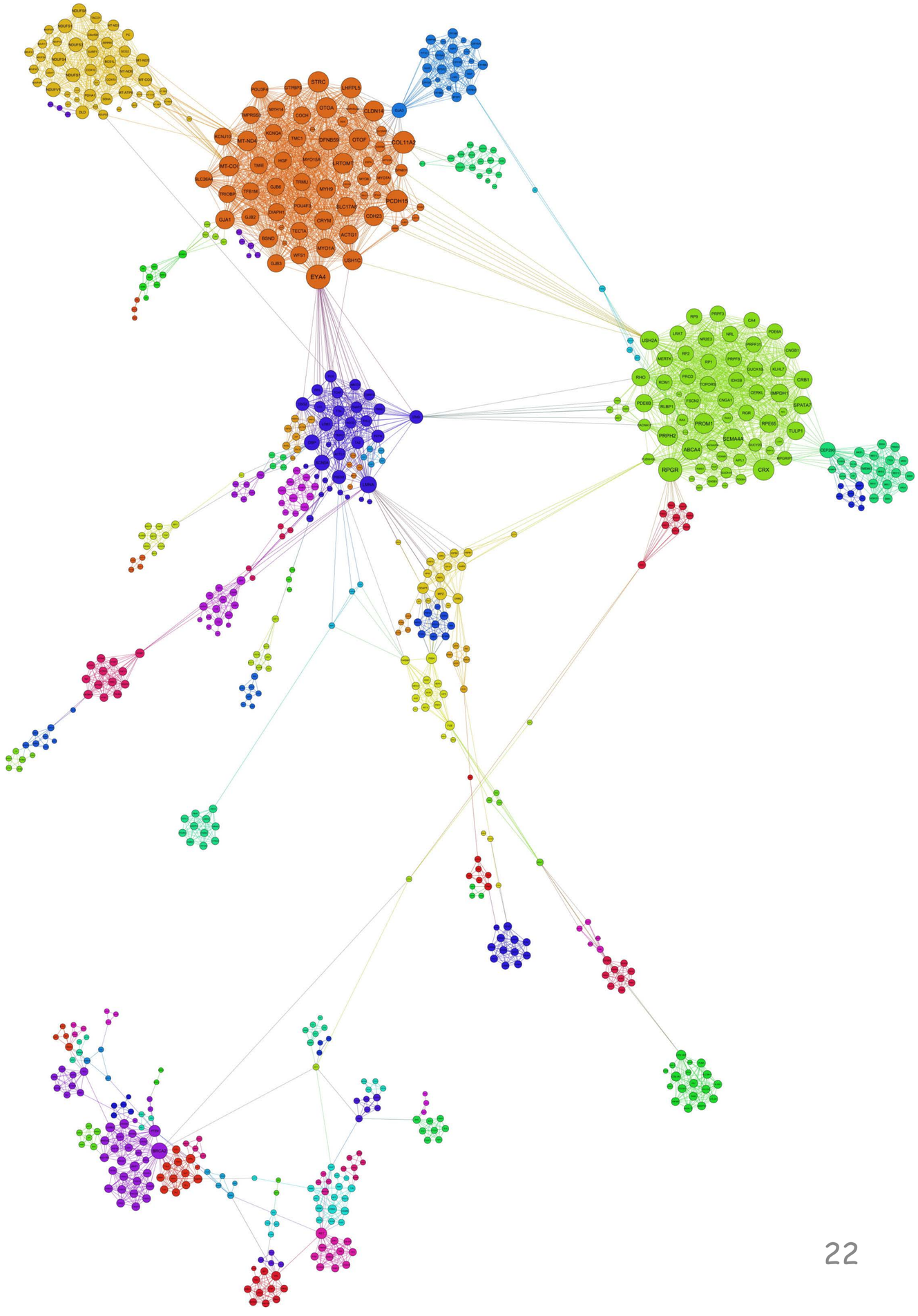
1521/2124 ODMGs are connected to at least another ODMG through at least one OD.

- **1521 nodes (ODMG)**
- **6855 edges (shared ODs)**
- **183 connected components**
- **Largest connected component has 734 nodes & 4817 edges**
- **277 communities**



Largest subnetwork or connected component of ODMGN based on shared OD

- High connectivity among different orphan diseases or OD-causing mutant genes - Infer the common mechanism and targeted pathways.
- Find candidates for **drug repositioning** or **drug repurposing** (i.e., to extrapolate or suggest novel applications for already approved drugs), especially when one or more than one orphan disease in the community has an approved drug.



Constructing ODMG Interactome (ODMGI)

Previous studies on disease gene networks

- Disease genes are nonessential
- Show no tendency to encode hubs

Are ODMGs also similar to common disease-causing mutant genes?

ODMG interactome (ODMGI)

Human protein interactome – Resources used

12,260 proteins
70,576 interactions

1. U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F.H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr and S. Koeppen, *et al.* A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, **122** (2005), pp. 957–968
2. J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze and N. Ayivi-Guedehoussou, *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437** (2005), pp. 1173–1178.
3. A.K. Ramani, R.C. Bunescu, R.J. Mooney and E.M. Marcotte, Consolidating the set of known human protein-protein interactions in preparation for large-scale mapping of the human interactome. *Genome Biol.*, **6** (2005), p. R40.
4. T.S. Prasad, K. Kandasamy and A. Pandey, Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol. Biol.*, **577** (2009), pp. 67–79
5. G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G.R. Gopinath, G.R. Wu and L. Matthews, *et al.* Reactome: A knowledgebase of biological pathways. *Nucleic Acids Res.*, **33** Database issue (2005), pp. D428–D432.
6. C. Alfarano, C.E. Andrade, K. Anthony, N. Bahroos, M. Bajec, K. Bantoft, D. Betel, B. Bobechko, K. Boutilier and E. Burgess, *et al.* The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.*, **33** Database issue (2005), pp. D418–D424.

ODMGs have high connectivity

- 507/1811 (28%) ODMGs are hubs in PPI network which is higher than 20% cutoff definition for all hubs
- Average degree of ODMGs in the PPIN (15.4) is significantly higher than that of other proteins in the PPIN (10.8).
- Previous studies, in contrast, reported a weak correlation between hubs and disease genes

Summary: ODMG vs other proteins (Interactome minus ODMG) in the PPI network

	Average degree	Average betweenness
ODMG in PPI (1811)	15.39	3.97E+04
Other proteins in PPI (10449)	10.84	12706.09
All proteins (12260)	11.51	1.67E+04

degree difference	Wilcoxon rank sum test
ODMG_in_PPI vs Others_in_PPI	2.20E-16

betweenness difference	Wilcoxon rank sum test
ODMG_in_PPI vs Others_in_PPI	2.20E-16

ODMGs encode proteins that tend to be essential

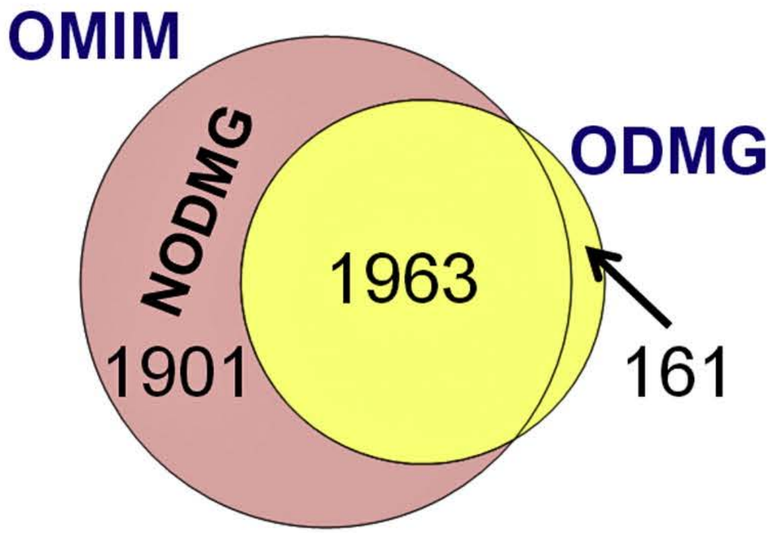
Direct comparison with essential genes to confirm that ODMGs tend to encode hub proteins and therefore could be essential.

- ~36% (765/2124) of the ODMGs are essential genes whose ortholog gene knockout in mice is lethal.
- This is much higher than the 22% (398/1777) of essential genes in the disease network reported by a previous study (Goh et al. 2007).
- ~18% (376/2124) of the ODMGs cause premature deaths in mouse ortholog gene knockout models.
- Together ~43% (907/2124) of the 2124 ODMGs result in either premature death and/or lethality in mouse gene knockout models.
- This is even more significant and specific to ODs because Goh et al.'s diseasome comprised several ODs, and the reported 22% is probably due to the presence of some of the ODs and related genes in their dataset.

ODMG Vs. Non-ODMG (NODMG)

- Separated all ODMGs from the entire set of OMIM disease genes (Morbid Map of the OMIM database), resulting in two classes of disease genes: **2124 ODMGs** and **1901 non-ODMGs (NODMG)** or common disease genes.
- NODMGs: Genes whose mutant forms are not associated with any orphan disease (based on current knowledge).
- Compared to NODMGs, ODMGs are significantly enriched for lethality and mitochondrion, as well as premature death ($p < 1.0 \times 10^{-5}$; Fisher's exact test).
- A total of 765 (~36% of 2124) of ODMGs are essential, whereas only 10% (192/1901) of NODMGs are essential.

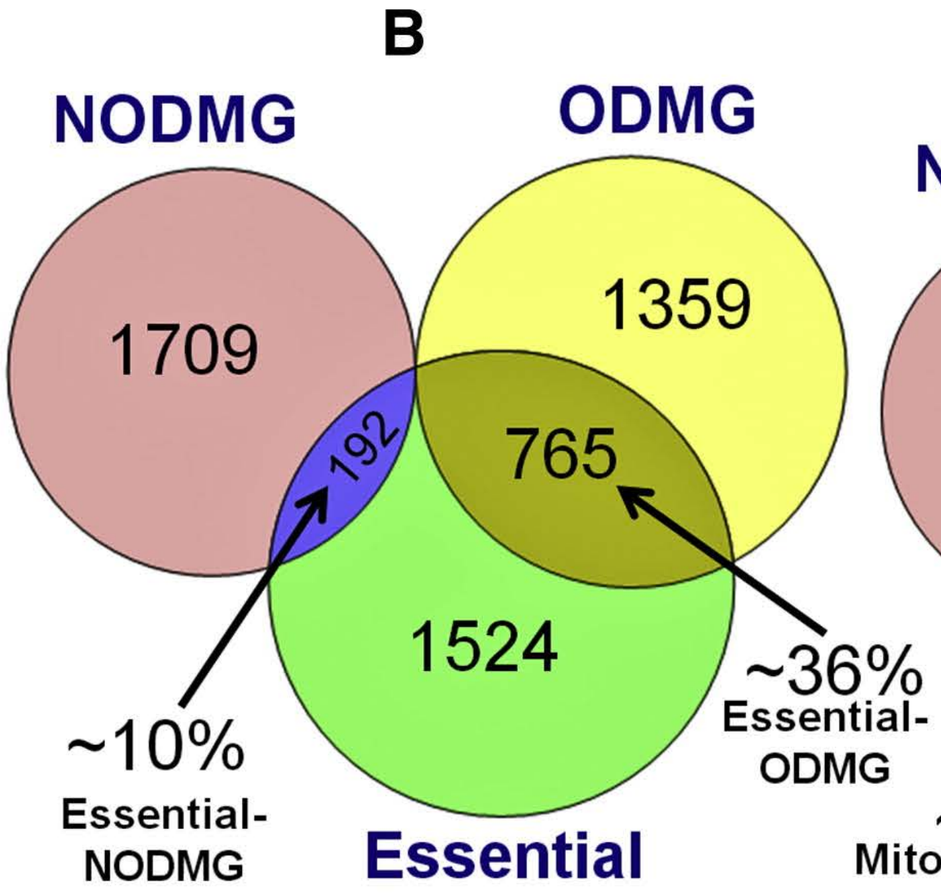
A



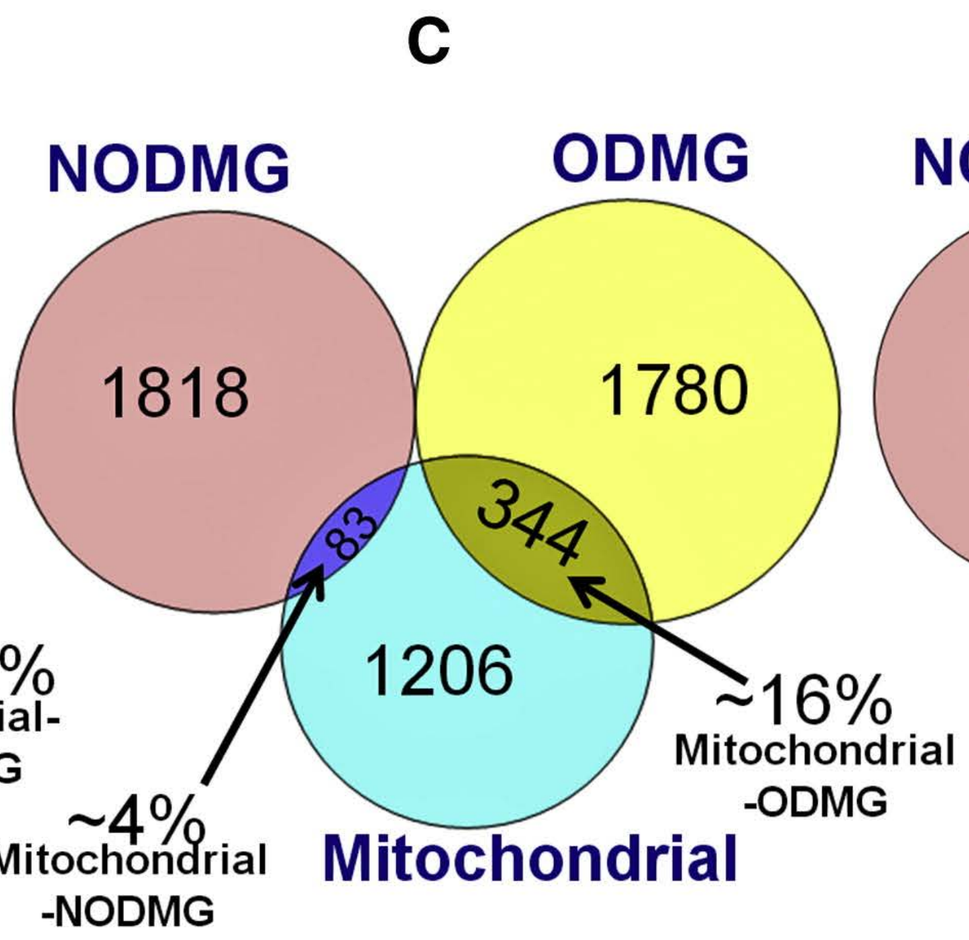
E

		ODMG	NODMG	Fisher's exact test
Essentiality	Yes	765	192	p < 1e-5
	No	1359	1709	
Mitochondrial	Yes	344	83	p < 1e-5
	No	1780	1818	
Premature Death	Yes	376	82	p < 1e-5
	No	1748	1819	

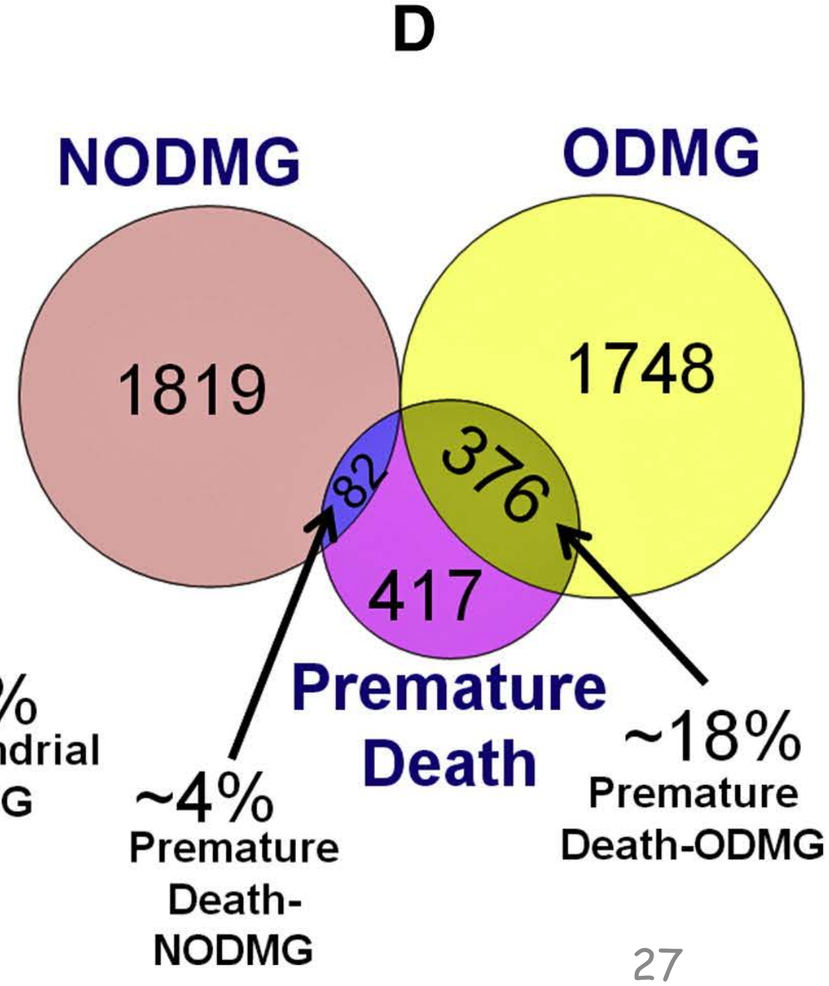
B



C



D



How specific is this finding?

- Overlap of essential genes with the **entire set of disease genes** from OMIM Morbid Map (as in Goh et al. 2007 but with updated disease and essential gene lists)
- 920 (**24%**) essential disease genes, which is similar to the original **22%** reported by Goh et al. 2007.

Confirms two things:

- Findings of Goh et al., whose study was based on **all disease genes**, still hold good despite the increase in the database sizes of human disease genes (from 1776 to 3864) and the essential genes (from 1267 to 2481).
- **It also strengthens our conclusion that the enrichment of essential genes is something specific to ODMGs because the percentage of essential ODMGs is higher when compared to either NODGMs or all disease genes from OMIM.**

Partitioning disease genes as ODMG and NODMG
Is it justifiable or just oversimplification?

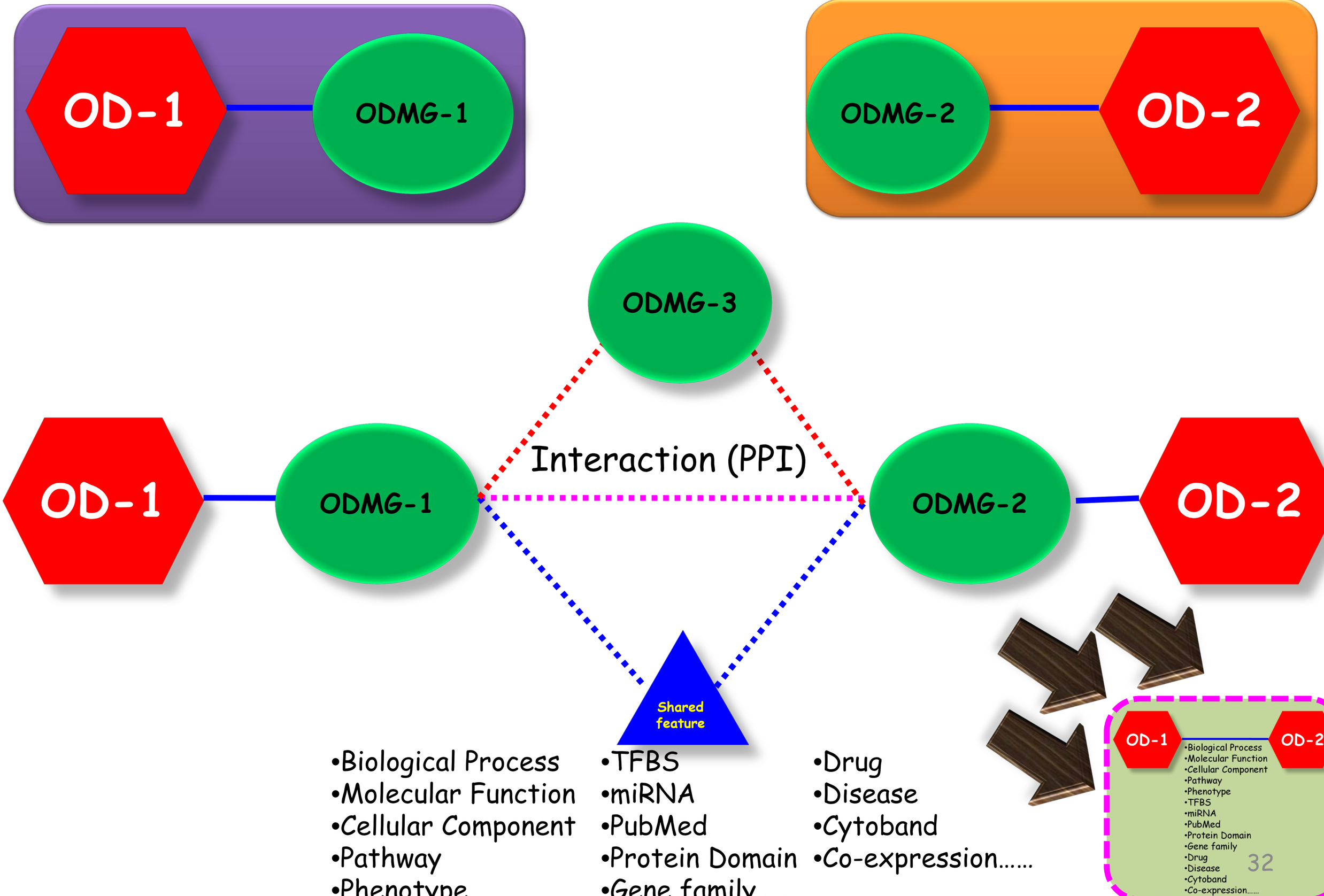
1. Helped in gaining insights into the relationship between the orphan disease characteristics (rare, lethal, and syndromic in nature) and the underlying causal mutant gene.
2. By an evolutionary argument, the partition could explain the *rarity of orphan diseases* in a population because mutations in hubs might not be compatible with survival and hence less likely to be maintained in a population.

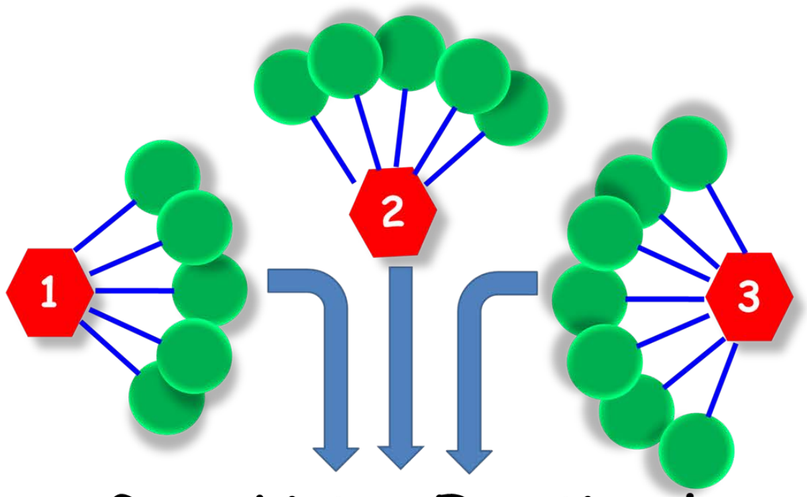
Partitioning disease genes as ODMG and NODMG
Is it justifiable or just simplification?

3. The partition could also explain the *severity and lethality* associated with most of the ODs because mutations in hubs could have wider repercussions and larger consequences on entire system than those in non-hubs. Additionally, functional enrichment analysis of ODMGs showed that a majority result in premature deaths or are lethal in the orthologous mouse gene knockout models.
4. Because hubs through their multiple interacting proteins connect heterogeneous cellular processes, the partition might explain the *complex phenotypic or syndromic nature of ODs* that have an impact on multiple physiological systems.

Orphan Diseases Functional Enrichment Analysis

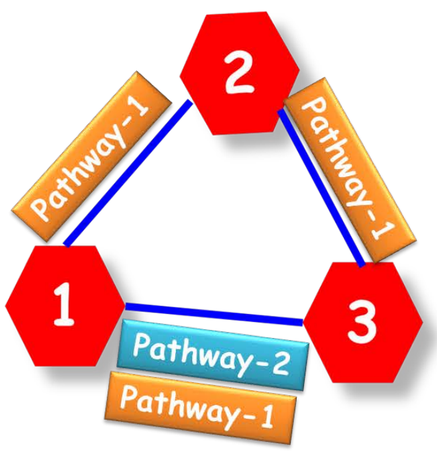
Two orphan diseases may not share genes but may share pathways, processes, phenotype, TFBS, miRNA, etc.





Gene List - Functional Enrichment Analysis

Enriched Pathways



3 ODs not sharing any ODMGs
i.e., 3 ODs are not *connected* in gene-based ODN

ToppFun (ToppGene Suite)
(functional enrichment analyses)
http://toppgene.cchmc.org

- Pathway 1 is enriched for all the 3 ODs
- Pathway 2 is shared between two of the 3 ODs

3 ODs which were not connected in gene-based network (ODN) actually share common pathways.

Function-based ODN Vs. Gene-based ODN

- The gene-based OD network (153 OD nodes and 191 edges; an edge indicates shared ODMG) is **largely different** from various function-based OD networks
  - BP-based OD network (176 OD nodes and 2244 edges; edges are shared BP or **B**iological **P**rocess terms)
  - CC-based OD network (153 OD nodes and 1135 edges; edges are shared CC or **C**ellular **C**omponent terms)
  - MP-based OD network (155 OD nodes and 745 edges; edges are shared MP or **M**ammalian **P**henotype terms)
  - pathway-based OD network (159 OD nodes and 511 edges; edges are shared pathways)
- The node agreement between the gene-based ODN and function-based ODNs was higher (Jaccard indices ranged from **0.647 to 0.732**)
- The edge agreement was much lower (Jaccard indices ranged from **0.0592 to 0.162**)

Literature-based ODN Vs. gene-based ODN

- Regenerated the ODN with the edge as a shared published article instead of a shared gene.
- To avoid potential false positives, we used the corresponding OMIM records of ODs, which summarize results from publications about gene-disease relationships, instead of mining literature.
- Specifically, we used the **cited literature** (the links to PubMed records for the references cited in an OMIM entry) in the OMIM records.
- For 1461 ODs there is a corresponding OMIM record (obtained from Orphanet). Of the 1475 mapped OMIM records, 1370 had at least one cited article (indicated by presence of at least one PubMed ID). We used this subset of 1370 ODs to compare the gene-based OD network with the literature-based OD network.

#612219

EWING SARCOMA; ES

Other entities represented in this entry:

NEUROEPITHELIOMA, PERIPHERAL, INCLUDED; PNE, INCLUDED

ASKIN TUMOR, INCLUDED

Phenotype Gene Relationships

Location	Phenotype	Phenotype MIM number	Gene/Locus	Gene/Locus MIM number
22q12.2	Neuroepithelioma	612219	EWSR1	133450
22q12.2	Ewing sarcoma	612219	EWSR1	133450

TEXT

A number sign (#) is used with this entry because the Ewing sarcoma family of tumors (ESFT) involve translocations of the EWS gene (133450) on chromosome 22q12 with various members of the ETS (see 164720) family of transcription factors.

Description

The Ewing sarcoma family of tumors (primitive neuroectodermal tumors; PNET) comprise morphologically heterogeneous tumors that are characterized by nonrandom chromosomal translocations involving the EWS gene on chromosome 22q12 and one of several members of the ETS family of transcription factors. The tumors include Ewing sarcoma, peripheral neuroepithelioma, and Askin tumor. In approximately 90% of cases of ESFT, the FLI1 gene (193067) on chromosome 11 is the fusion partner of EWS; in approximately 10%, the EWS fusion partner is the ERG gene (165080) on chromosome 22. Many other ETS family members have been identified as fusion partners of EWS, but these cases are rare (Khoury, 2005).

Clinical Features

Ewing Sarcoma

Ewing sarcoma is a highly malignant, metastatic, primitive small round cell tumor of bone and soft tissue that affects children and adolescents. It was first described by Ewing (1921) as a diffuse endothelioma of bone.

In a study of 5 Ewing sarcoma cell lines established from 4 patients, Turc-Carel et al. (1984) found a consistent reciprocal translocation t(11;22)(q24;q12). In 4 patients, Aurias et al. (1984) studied fresh tumor cells derived by biopsy of primary or metastatic tumors. Abnormal karyotypes with translocations involving 22q12 were found in all. In 2 cases, t(11;22)(q24;q12) was found. Histologic differentiation of ES from several other childhood tumors is often difficult; the marker chromosome may be very useful to precise diagnosis.

Among 13 cases of Ewing sarcoma, Douglass et al. (1986) found that 9 had t(11;22) and that 2 additional cases had only a deleted chromosome 22. Griffin et al. (1986) could distinguish the cytologically indistinguishable tumor-related t(11;22) by doing in situ hybridization with probes for the constant region of the lambda light chain located at 22q11 and the ETS1 oncogene (164720) located at 11q23.3-q24.

REFERENCES

1. Aurias, A., Rimbaut, C., Buffe, D., Zucker, J.-M., Mazabraud, A. **Translocation involving chromosome 22 in Ewing's sarcoma: a cytogenetic study of four fresh tumors.** Cancer Genet. Cytogenet. 12: 21-25, 1984. [PubMed: 6713357, related citations] [Full Text: Pubget]

2. Bielack, S. S., Paulussen, M., Kohler, G. **A patient with two Ewing's sarcomas with distinct EWS fusion transcripts. (Letter)** New Eng. J. Med. 350: 1364-1365, 2004. [PubMed: 15044653, related citations] [Full Text: Atypon, Pubget]

3. Budarf, M., Sellinger, B., Griffin, C., Emanuel, B. S. **Comparative mapping of the constitutional and tumor-associated 11;22 translocations.** Am. J. Hum. Genet. 45: 128-139, 1989. [PubMed: 2741943, related citations] [Full Text: Pubget]

4. Burchill, S. A., Wheeldon, J., Cullinane, C., Lewis, I. J. **EWS-FLI1 fusion transcripts identified in patients with typical neuroblastoma.** Europ. J. Cancer 33: 239-243, 1997. [PubMed: 9135495, related citations] [Full Text: Elsevier Science, Pubget]

5. Douglass, E. C., Valentine, M., Green, A. A., Hayes, F. A., Thompson, E. I. **t(11;22) and other chromosomal rearrangements in Ewing's sarcoma.** J. Nat. Cancer Inst. 77: 1211-1213, 1986. [PubMed: 3467113, related citations] [Full Text: Pubget]

6. Ewing, J. **Diffuse endothelioma of bone.** Proc. N. Y. Path. Soc. 21: 17-24, 1921.

Literature-based ODN Vs. gene-based ODN

- Although a large number of common nodes exist between the gene- and literature-based ODNs, common edges are fewer.
- literature-based ODN identified additional relationships for those diseases sharing no known disease genes but having potential functional links between their corresponding disease gene sets.
- A large number (672 edges; ~72%) share no known disease genes, and their relationships are identified solely on the basis of literature-connectivity.
  - Tay-Sachs disease (mutant HEXA and GM2A) and Sandhoff syndrome (mutant HEXB) do not share any disease genes and hence are not connected in shared-gene-based studies. However, Tay-Sachs disease and Sandhoff disease are connected in the literature-based OD network, which is not surprising because these two disorders arise because of the failure of the same metabolic pathway.
  - Rubinstein-Taybi syndrome (CREBP and EP300 mutants) and ICF syndrome (mutant DNMT3B), which are both syndromes of chromatin modeling
  - ornithine transcarbamylase deficiency, arginosuccinic aciduria, and citrullinemia, which are all urea cycle disorders
  - Prader-Willi syndrome and Angelman syndrome, which are both genomic-imprinting disorders (paternal and maternal)
  - Lathosterolosis, Smith-Lemli-Opitz syndrome, and Greenberg dysplasia, which are all inborn errors of cholesterol synthesis.

Summary

1. A large number of orphan disease-causing mutant genes are **essential**. In confirmation of this finding, we also found that OD-causing mutant genes tend to be **topologically important** in the protein interactome.
2. Functional enrichment analysis of those genes in which mutations cause ODs showed that a majority result in **premature death** or are **lethal** in the orthologous mouse gene knockout models.
3. Analyzing these functionally-linked OD networks, we identified several additional OD-OD relations that are both phenotypically similar and phenotypically diverse.
4. Surprisingly, we also observed that the wiring of the gene-based and other feature-based OD networks are largely different; this suggests that the **relationship between ODs cannot be fully captured by the gene-based networks alone**.

AJHG

Volume 88, Issue 6, 10 June 2011, Pages 755–766

Article

The Orphan Disease Networks

Minlu Zhang[1, 3, 5], Cheng Zhu[1, 5], Alexis Jacomy[4], Long J. Lu[1, 2, 3], Anil G. Jegga[1, 2, 3], .

[1] Department of Computer Science, University of Cincinnati, Cincinnati, OH 45229, USA

[2] Department of Pediatrics, University of Cincinnati, Cincinnati, OH 45229, USA

[3] Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

[4] Gephi, WebAtlas, Paris 75011, France

Orphan Diseasome

http://research.cchmc.org/od

HOME PAGE | PROJECT DESCRIPTION | NETWORKS/MAPS | CONTACT | RESOURCES | PUBLICATION

Home page

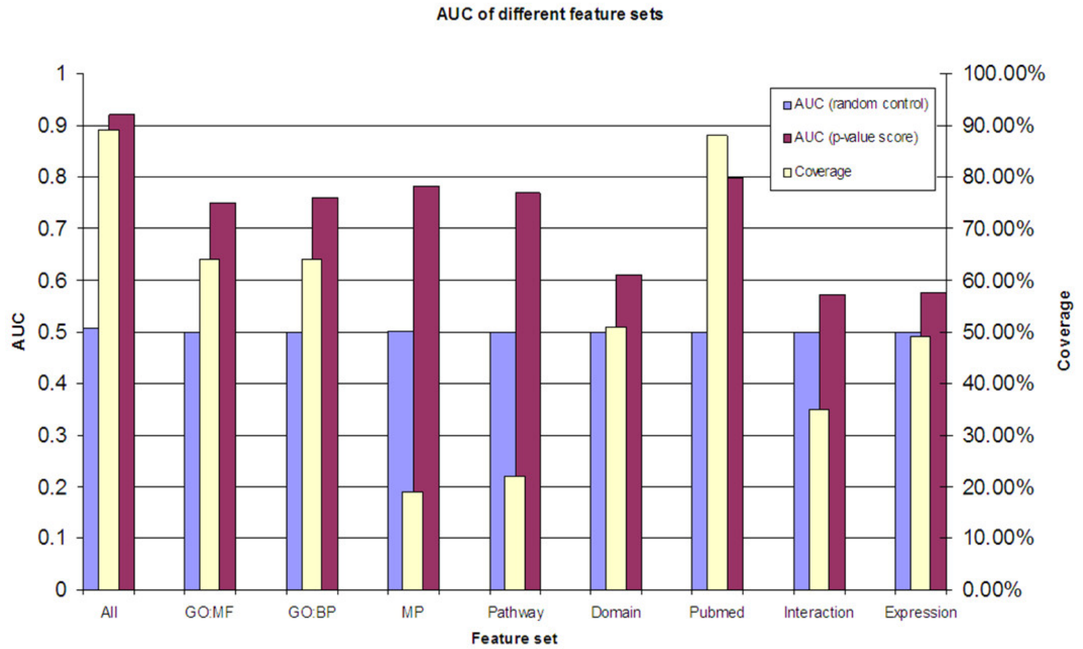
The Orphan Diseasome web site allows investigators to explore the orphan disease (OD) or rare disease relationships based on shared genes and shared enriched features (e.g., Gene Ontology Biological Process, Cellular Component, Pathways, Mammalian Phenotype). Additionally, users can also explore the networks of orphan disease causal genes where the nodes are orphan disease genes (ODG) while the edge represents shared OD or a protein-protein interaction.

To start exploring the Orphan Diseasome, click on the "Networks/Maps" (top navigation bar).

Orphan Disease Genes
Computational Discovery & Prioritization

Orphan Disease Gene Discovery and Ranking Using Model organism data

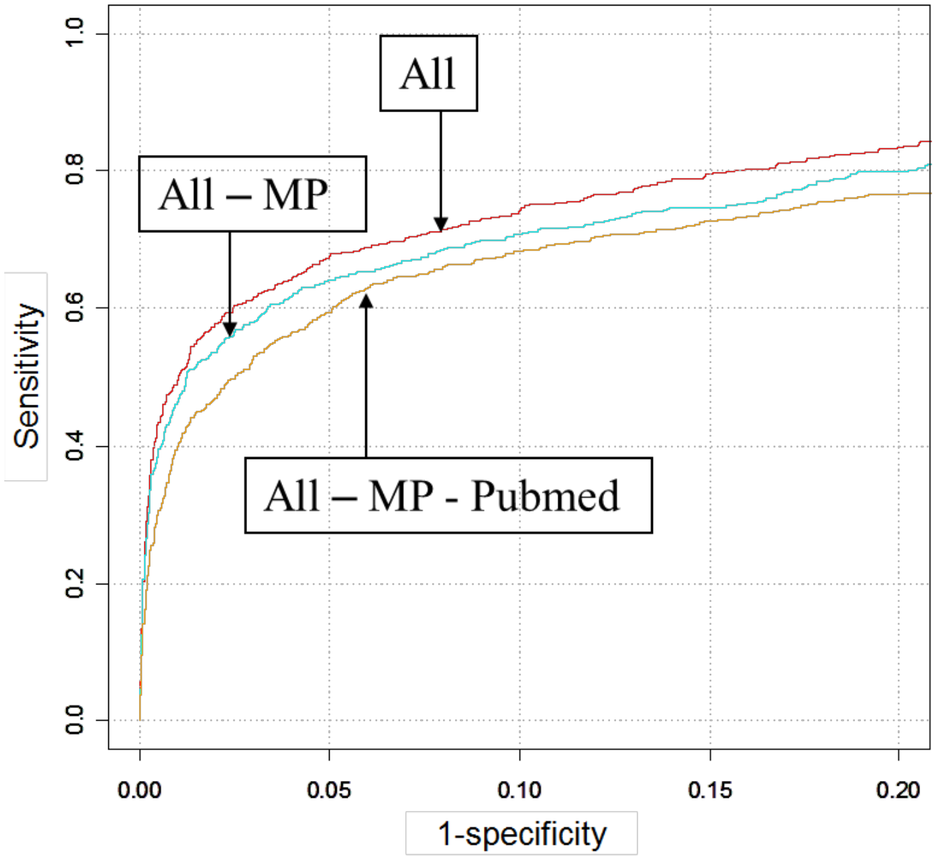
- The incorporation of phenotype information for mouse orthologs of human genes greatly improves the human disease candidate gene analysis and prioritization (Chen et al., 2007)



AUC of different feature sets.

- Red bars - AUC scores based on each feature set
- blue bars - corresponding random controls
- Yellow bars - Coverage of each feature set in the whole genome.

For example, mouse phenotype (MP) has AUC score 0.78 and covers 19% of genes in the whole genome.

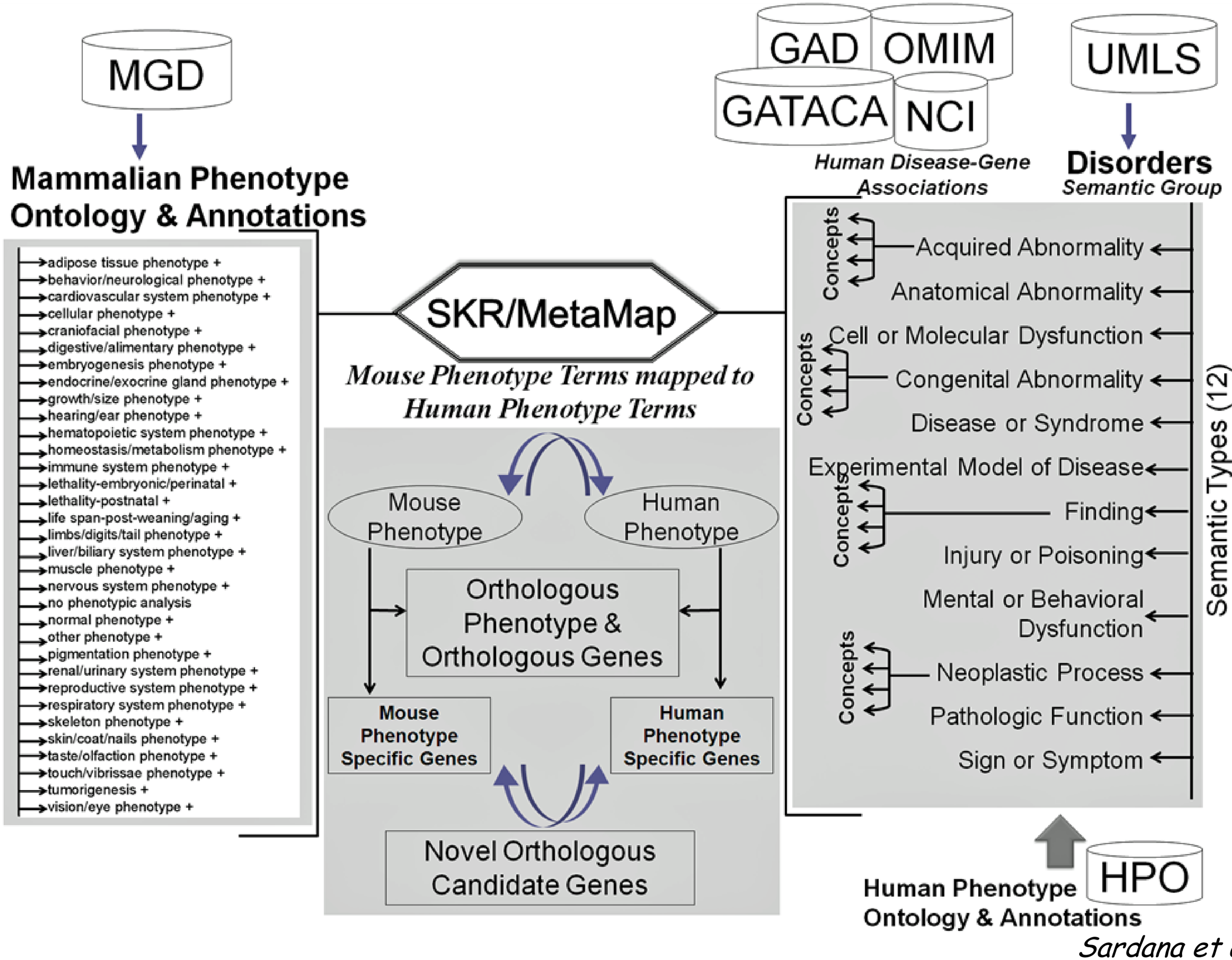


Random gene cross-validation: Leave-one-feature-out - Overall performance

- All features: 0.913
- All - MP: 0.893
- All - MP - PubMed: 0.888

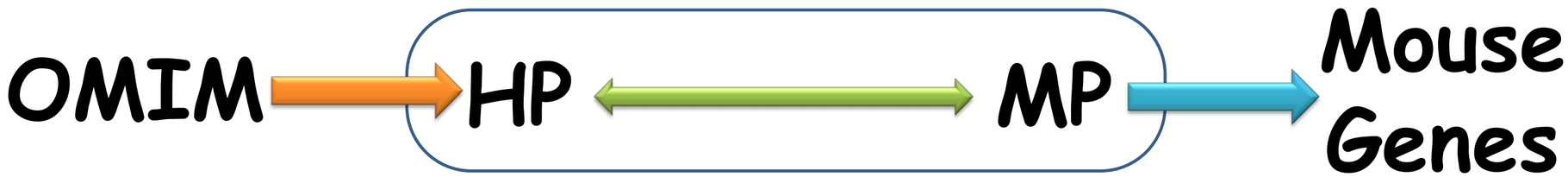
Chen *et al.* 2007

PhenoHM: http://phenome.cchmc.org



Sardana et al., 2010

Orphan Disease Gene Discovery and Ranking: Using Model organism data



#190900
TRITANOPIA: BLUE COLORBLINDNESS
caused by heterozygous mutation in the **OPN1SW** gene

- Are the human orthologs of these mouse genes novel candidates for tritanopia?
- Can these phenotype-matched mouse genes be used as a training set to rank candidate genes for tritanopia?

Smedley et al., 2013

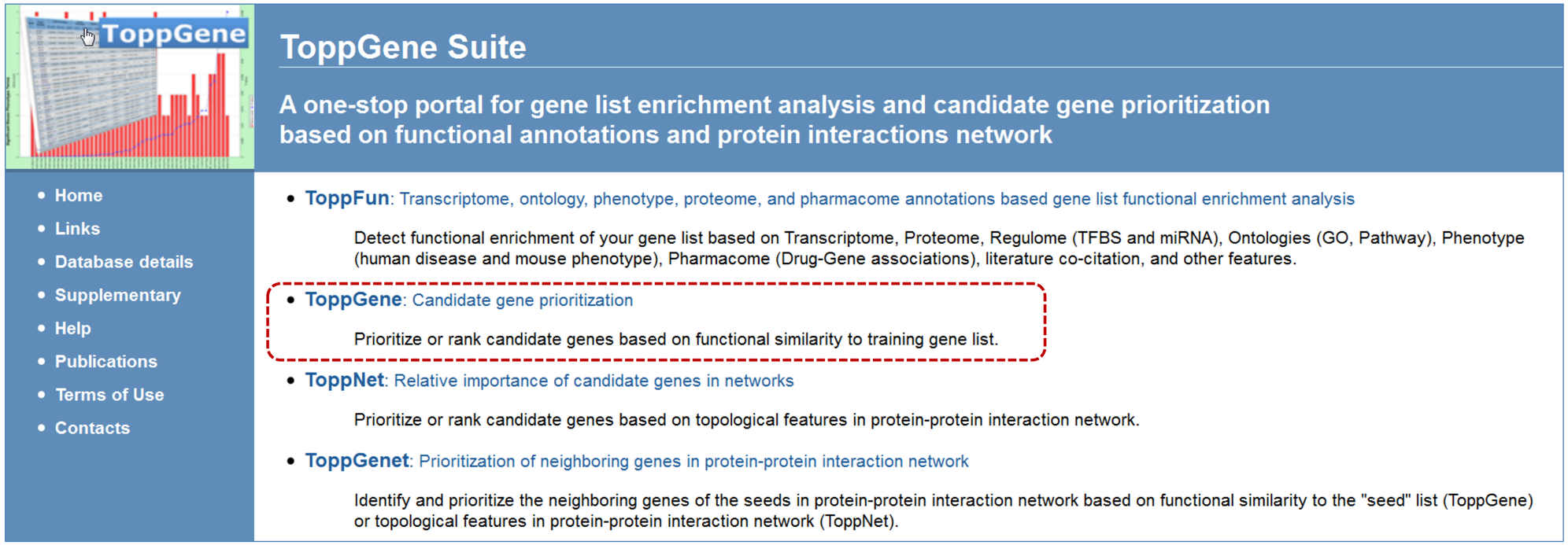
http://www.sanger.ac.uk/resources/databases/phenodigm

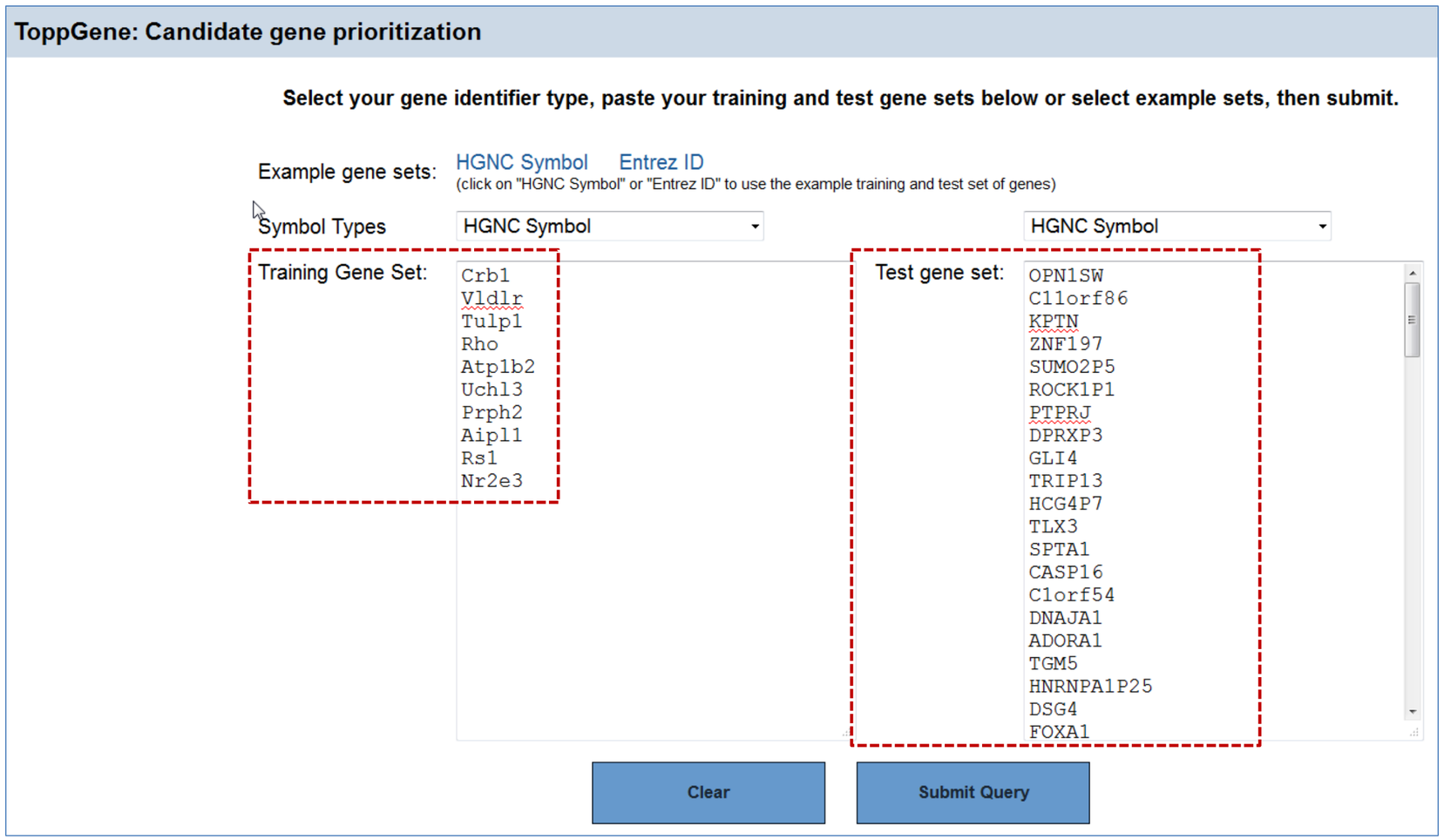
Chao et al., under preparation

ToppGene Suite: http://toppgene.cchmc.org

Training set:
Crb1
Vldlr
Tulp1
Rho
Atp1b2
Uchl3
Prph2
Aipl1
Rs1
Nr2e3

Test set:
OPN1SW + 99 random genes





Training set (10 / 10)

Entered	Human Symbol	Gene ID
Crb1	CRB1	23418
Vldlr	VLDLR	7436
Tulp1	TULP1	7287
Rho	RHO	6010
Atp1b2	ATP1B2	482
Uchl3	UCHL3	7347
Prph2	PRPH2	5961
Aipl1	AIPL1	23746
Rs1	RS1	6247
Nr2e3	NR2E3	10002

Test set (100 / 100)

Entered	Human Symbol	Gene ID
OPN1SW	OPN1SW	611
C11orf86	C11orf86	254439
KPTN	KPTN	11133
ZNF197	ZNF197	10168
SUMO2P5	SUMO2P5	100526738
ROCK1P1	ROCK1P1	727758
PTPRJ	PTPRJ	5795
DPRXP3	DPRXP3	503644
GLI4	GLI4	2738
TRIP13	TRIP13	9319
HCG4P7	HCG4P7	353004
TLX3	TLX3	30012
SPTA1	SPTA1	6708
CASP16	CASP16	197350
C1orf54	C1orf54	79630
DNAJA1	DNAJA1	3301
ADORA1	ADORA1	134
TGM5	TG	
HNRNPA1P25	HN	
DSG4	DS	
FOXA1	FO	

Training parameters

Feature	Correction	p-Value cutoff	Gene Limits
All	Bonferroni	0.05	1 ≤ *n* ≤ 1500
GO: Molecular Function	Bonferroni	0.05	1 ≤ *n* ≤ 1500
GO: Biological Process	Bonferroni	0.05	1 ≤ *n* ≤ 1500
GO: Cellular Component	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Human Phenotype	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Mouse Phenotype	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Domain	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Pathway	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Pubmed	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Interaction	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Cytoband	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Transcription Factor Binding Site	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Gene Family	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Coexpression	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Coexpression Atlas	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Computational	Bonferroni	0.05	1 ≤ *n* ≤ 1500
MicroRNA	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Drug	Bonferroni	0.05	1 ≤ *n* ≤ 1500
Disease	Bonferroni	0.05	1 ≤ *n* ≤ 1500

Test parameter

Random sampling size: 1500 (6% of genome)

Min. feature count: 2

Home | Modify Query | Start prioritization

Input Parameters [Hide Detail]

Number of genes in training set:	10
Number of genes in test set:	100

Correction and Cutoff:

category	Correction	Cutoff	Min	Max
GO: Molecular Function	Bonferroni	0.05	1	1500
GO: Biological Process	Bonferroni	0.05	1	1500
GO: Cellular Component	Bonferroni	0.05	1	1500
Human Phenotype	Bonferroni	0.05	1	1500
Mouse Phenotype	Bonferroni	0.05	1	1500
Domain	Bonferroni	0.05	1	1500
Pathway	Bonferroni	0.05	1	1500
Pubmed	Bonferroni	0.05	1	1500
Interaction	Bonferroni	0.05	1	1500
Cytoband	Bonferroni	0.05	1	1500
Transcription Factor Binding Site	Bonferroni	0.05	1	1500
Gene Family	Bonferroni	0.05	1	1500
Coexpression	Bonferroni	0.05	1	1500
Coexpression Atlas	Bonferroni	0.05	1	1500
Computational	Bonferroni	0.05	1	1500
MicroRNA	Bonferroni	0.05	1	1500
Drug	Bonferroni	0.05	1	1500
Disease	Bonferroni	0.05	1	1500

Random sampling size in analysis:	1500
Minimun feature count in test set:	2
Analysis took:	1 seconds
Analysis finished at:	Tue Nov 05 15:58:54 EST 2013

Training Results [Expand All] [Download All] [Sparse Matrix] Display pValues and Scores as Scientific (4 sign

1: GO: Molecular Function [Display Chart]

	ID	Name	Source	P-value	Term in Query	Term in Genome
1	GO:0008020	G-protein coupled photoreceptor activity		4.417E-4	2	7
2	GO:0009881	photoreceptor activity		3.209E-3	2	18
3	GO:0005546	phosphatidylinositol-4,5-bisphosphate binding		1.627E-2	2	40
4	GO:0001918	farnesylated protein binding		4.843E-2	1	1

2: GO: Biological Process [Display Chart]

	ID	Name	Source	P-value	Term in Query	Term in Genome
1	GO:0007602	phototransduction		4.502E-8	5	66
2	GO:0060041	retina development in camera-type eye		1.291E-7	6	205
3	GO:0009583	detection of light stimulus		1.746E-7	5	86
4	GO:0007601	visual perception		2.941E-7	6	235
5	GO:0050953	sensory perception of light stimulus		3.094E-7	6	237

Show 17 more annotations

3: GO: Cellular Component [Display Chart]

	ID	Name	Source	P-value	Term in Query	Term in Genome
1	GO:0001917	photoreceptor inner segment		1.856E-5	3	31
2	GO:0001750	photoreceptor outer segment		1.141E-2	2	49
3	GO:0060342	photoreceptor inner segment membrane		2.064E-2	1	1
4	GO:0045177	apical part of cell		3.858E-2	3	397

4: Human Phenotype [Display Chart]

	ID	Name	Source	P-value	Term in Query	Term in Genome
1	HP:0000512	Abnormal electroretinogram		1.167E-6	6	98
2	HP:0000510	Retinitis pigmentosa		2.782E-6	6	113
3	HP:0007703	Abnormal retinal pigmentation		4.791E-5	6	181
4	HP:0000662	Night blindness		9.856E-5	5	99
5	HP:0008051	Abnormality of the retinal pigment epithelium		1.040E-4	6	206

Show 14 more annotations

5: Mouse Phenotype [Display Chart]

	ID	Name	Source	P-value	Term in Query	Term in Genome
1	MP:0003731	abnormal retinal outer nuclear layer morphology		1.296E-16	10	105
2	MP:0001004	abnormal retinal photoreceptor morphology		1.769E-15	10	135
3	MP:0003728	abnormal retinal photoreceptor layer morphology		2.773E-15	10	141
4	MP:0003730	abnormal photoreceptor inner segment morphology		7.486E-15	8	40
5	MP:0003729	abnormal photoreceptor outer segment morphology		1.351E-14	9	86

Show 51 more annotations

Enriched features of training set

Test Results [Hide Detail] [Download] [Show Network]

Rank (net)	Gene Symbol	Gene ID	GO: Molecular Function		GO: Biological Process		GO: Cellular Component		Average score	Overall P-value
			Score	pValue	Score	pValue	Score	pValue		
1	OPN1SW	611	7.887E-1	1.000E-6	9.982E-1	1.308E-3	7.468E-1	1.000E-6	4.182E-1	4.187E-12
2	CLCN2	1181	0.000E0	5.677E-1	8.989E-1	1.046E-2	3.593E-1	6.540E-3	1.801E-1	1.221E-2
3	ADORA1	134	5.282E-1	1.308E-3	9.727E-1	4.578E-3	3.593E-1	6.540E-3	1.420E-1	1.534E-2
4	KPTN	11133	7.303E-2	2.943E-2	3.050E-1	1.236E-1	3.029E-1	1.700E-2	5.675E-2	2.869E-2
5	EIF5A	1984	7.303E-2	2.943E-2	4.418E-1	8.633E-2	3.593E-1	6.540E-3	6.243E-2	2.925E-2
6	L1CAM	3897	7.303E-2	2.943E-2	7.331E-1	3.401E-2	3.593E-1	6.540E-3	1.254E-1	3.221E-2
7	CDK14	5218	7.303E-2	2.943E-2	3.570E-1	1.020E-1	9.989E-2	6.802E-2	4.416E-2	3.381E-2
8	CIB2	10518	0.000E0	5.677E-1	8.971E-1	1.046E-2	0.000E0	6.645E-1	1.284E-1	3.748E-2
9	RPS6KA4	8986	7.303E-2	2.943E-2	7.331E-1	3.401E-2	2.075E-2	1.622E-1	5.907E-2	5.464E-2
10	SH3BGRL2	83699	7.303E-2	2.943E-2	0.000E0	6.037E-1	2.075E-2	1.622E-1	8.526E-3	5.807E-2
									7.823E-2	6.239E-2

Ranked list of test set genes

Supporting Details

Selected genes shown in **bold**

Feature	ID	Name	Genes
GO: Molecular Function	GO:0009881	photoreceptor activity	**OPN1SW** RHO TULP1
GO: Biological Process	GO:0007602	phototransduction	AIPL1 NR2E3 **OPN1SW** RHO RS1 TULP1
GO: Biological Process	GO:0009583	detection of light stimulus	AIPL1 NR2E3 **OPN1SW** RHO RS1 TULP1
GO: Biological Process	GO:0009582	detection of abiotic stimulus	AIPL1 NR2E3 **OPN1SW** RHO RS1 TULP1
GO: Biological Process	GO:0009581	detection of external stimulus	AIPL1 NR2E3 **OPN1SW** RHO RS1 TULP1
GO: Biological Process	GO:0007601	visual perception	AIPL1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
GO: Biological Process	GO:0050953	sensory perception of light stimulus	AIPL1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
GO: Biological Process	GO:0051606	detection of stimulus	AIPL1 NR2E3 **OPN1SW** RHO RS1 TULP1
GO: Biological Process	GO:0009416	response to light stimulus	AIPL1 NR2E3 **OPN1SW** RHO RS1 TULP1
GO: Biological Process	GO:0009314	response to radiation	AIPL1 NR2E3 **OPN1SW** RHO RS1 TULP1
GO: Biological Process	GO:0007600	sensory perception	AIPL1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
GO: Cellular Component	GO:0001750	photoreceptor outer segment	**OPN1SW** RHO TULP1
Human Phenotype	HP:0000512	Abnormal electroretinogram	AIPL1 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1
Human Phenotype	HP:0000479	Abnormality of the retina	AIPL1 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
Human Phenotype	HP:0000504	Abnormality of vision	AIPL1 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
Human Phenotype	HP:0001098	Abnormality of the fundus	AIPL1 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
Human Phenotype	HP:0004329	Abnormality of the posterior segment of the eye	AIPL1 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
Mouse Phenotype	MP:0008585	absent photoreceptor outer segment	AIPL1 ATP1B2 CRB1 **OPN1SW** PRPH2 RHO RS1 TULP1
Mouse Phenotype	MP:0004022	abnormal cone electrophysiology	NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
Mouse Phenotype	MP:0003729	abnormal photoreceptor outer segment morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3
Mouse Phenotype	MP:0001004	abnormal retinal photoreceptor morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0003728	abnormal retinal photoreceptor layer morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0005551	abnormal eye electrophysiology	AIPL1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1
Mouse Phenotype	MP:0006069	abnormal retinal neuronal layer morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0003727	abnormal retinal layer morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0005253	abnormal eye physiology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3
Mouse Phenotype	MP:0001325	abnormal retina morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0002864	abnormal ocular fundus morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0000965	abnormal sensory neuron morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0005195	abnormal posterior eye segment morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0000959	abnormal somatic sensory system morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0002752	abnormal somatic nervous system morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0002092	abnormal eye morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0005391	vision/eye phenotype	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR
Mouse Phenotype	MP:0002882	abnormal neuron morphology	AIPL1 ATP1B2 CRB1 NR2E3 **OPN1SW** PRPH2 RHO RS1 TULP1 UCHL3 VLDLR

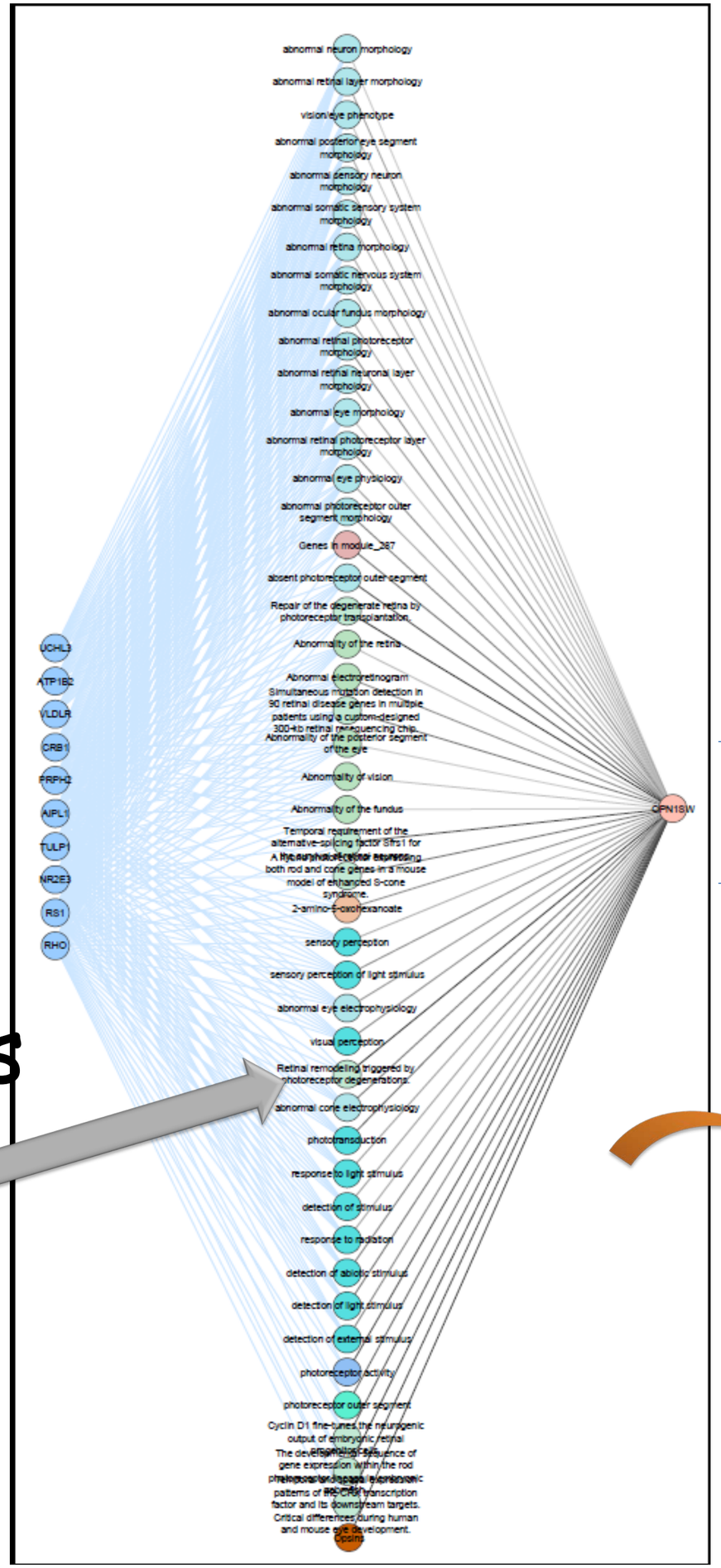
Shared features between training set and ranked test set gene

Training set genes

Ranked gene from Test set

Training set genes

Shared features between training set and ranked test set gene



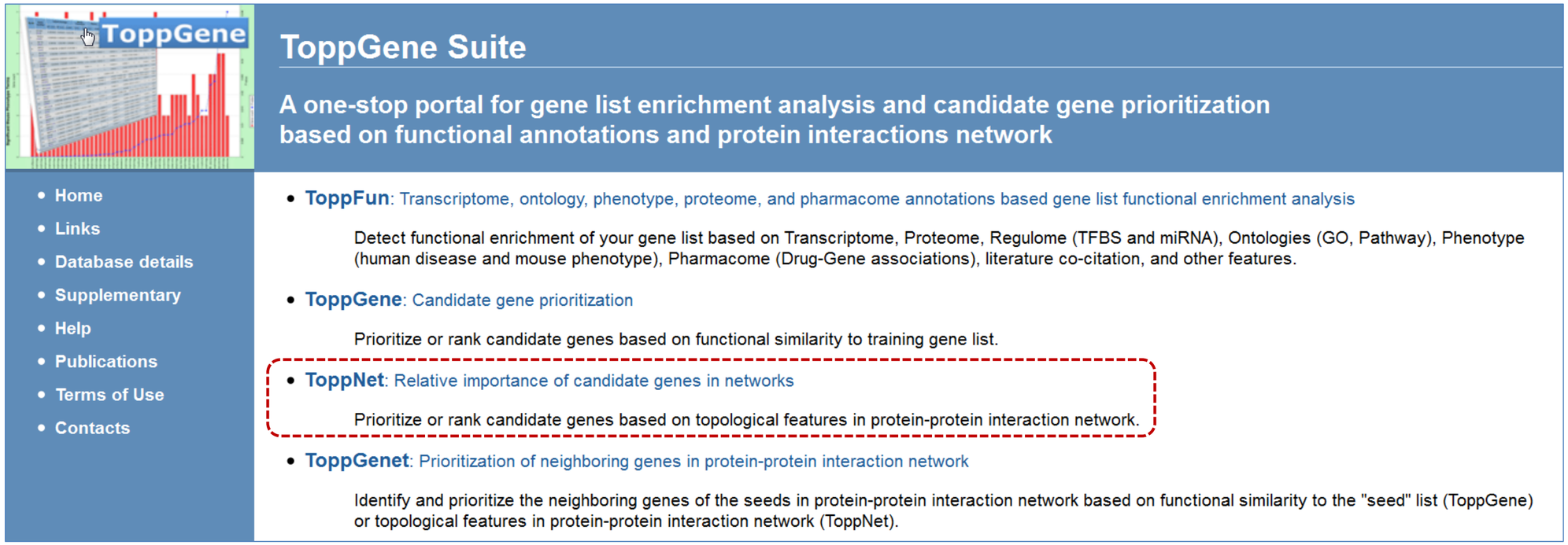
Ranked gene from Test set

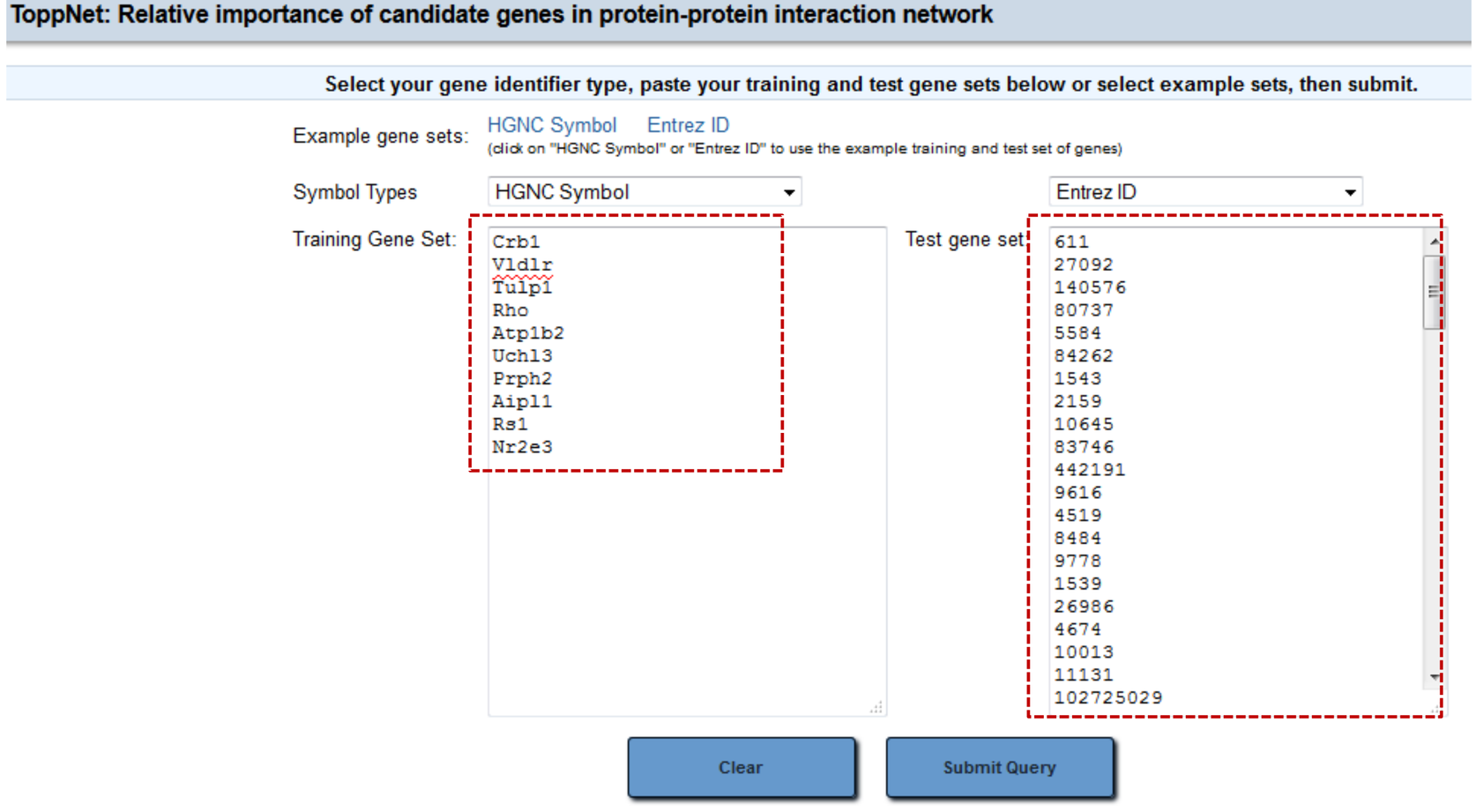
Cytoscape- or Gephi-compatible

ToppNet: http://toppgene.cchmc.org

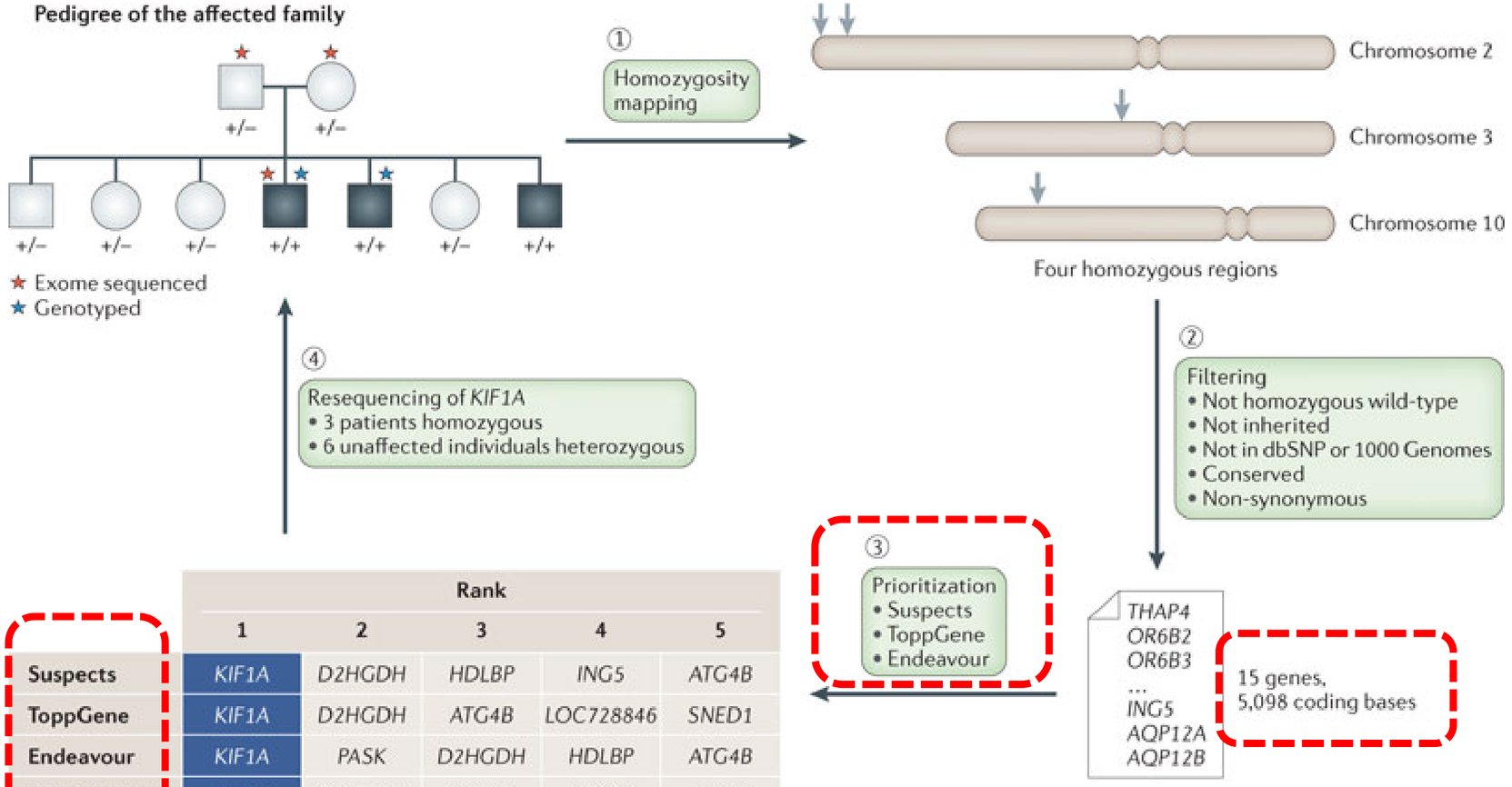
Training set:
Crb1
Vldlr
Tulp1
Rho
Atp1b2
Uchl3
Prph2
Aipl1
Rs1
Nr2e3

Test set:
OPN1SW + 99 random genes





- Familial case of **<u>hereditary spastic paraparesis</u>** (HSP) - Whole-exome sequencing
- four largest homozygous regions between two of the three affected brothers were considered to be potential disease loci, containing a total of 44 genes.
- After filtering step, 15 candidate genes remained.
- The list was then prioritized using three computational methods (namely, Suspects, ToppGene and Endeavour)
- The prioritization criteria were a list of 11 seed genes that were obtained through a review of the literature and are known to be associated with forms of HSP in which mutations lead to the core HSP phenotypic traits (that is, progressive lower-extremity spastic weakness, hypertonic urinary bladder disturbance and mild diminution of lower-extremity vibration sensation) but not to unrelated traits.
- The top-ranking gene from the prioritization was kinesin family member 1A (KIF1A).
- Sanger sequencing confirmed that KIF1A is the causative variant –(Ala255Val variant)



	Rank				
	1	2	3	4	5
Suspects	*KIF1A*	*D2HGDH*	*HDLBP*	*ING5*	*ATG4B*
ToppGene	*KIF1A*	*D2HGDH*	*ATG4B*	*LOC728846*	*SNED1*
Endeavour	*KIF1A*	*PASK*	*D2HGDH*	*HDLBP*	*ATG4B*
Combination	*KIF1A*	*D2HGDH*	*ATG4B*	*HDLBP*	*PASK*

Erlich et al., 2011
Moreau & Tranchevent, 2012

Nature Reviews | Genetics

GO Biological Process

Annotations:	10,887
Genes:	18,482
	Updated Nov 10, 2014

GO Cellular Component

Annotations:	1,529
Genes:	18,865
	Updated Nov 10, 2014

GO Molecular Function

Annotations:	3,381
Genes:	18,326
	Updated Nov 10, 2014

Human Phenotype

Annotations:	8,755
Genes:	4,145
	Updated Oct 30, 2014

Mouse Phenotype

Annotations:	8,331
Genes:	5,593
	Updated Nov 11, 2014

377k million records (gene-2-MP)

Pathways

Annotations:		3,633
	BioSystems: BIOCYC	319
	BioSystems: KEGG	362
	BioSystems: Pathway Interaction Database	186
	BioSystems: REACTOME	1,400
	BioSystems: WikiPathways	229
	GenMAPP	67
	MSigDB C2: BioCarta	217
	MSigDB C2: SigmaAldrich	10
	MSigDB C2: Signaling Gateway	8
	MSigDB C2: Signaling Transduction KE	28
	MSigDB C2: SuperArray	1
Jan 22, 2014	PantherDB	152
Aug 29, 2014	Pathway Ontology	321
	SMPDB	333
Genes:		10,567

Domains

Annotations:		15,281
	Gene3D	610
	InterPro	8,171
	PROSITE	1,566
	Pfam	4,115
	ProDom	138
	SMART	681
Genes:		16,924

Pubmed

Annotations:		618,793
	GeneRIF	391,304
	Pubmed	227,489
Genes:		33,563
		Updated Nov 10, 2014

1.6 million records (gene-2-PMID)

Interactions

Annotations:	15,886
Genes:	15,887
	Updated Oct 29, 2014

Cytoband

Annotations:	2,354
Genes:	34,661

TFBS

Annotations:	615
Genes:	9,770

miRNA

Annotations:		4,085
	MSigDB	313
	MicroRNA.org	2,200
	PITA	677
	PicTar	178
	TargetScan	249
	miRTarbase	324
	miRecords_TarBase	144
Genes:		19,844

Gene Families

Annotations:	151
Genes:	6,751

Coexpression

Annotations:		9,016
	GeneSigDB	3,515
	MSigDB C2: Aristoteles University of Thessaloniki	5
	MSigDB C2: Broad Institute	2,962
	MSigDB C2: Columbia University	1
	MSigDB C2: Dana-Farber Cancer Institute	16
	MSigDB C2: Giannina Gaslini Institute	2
	MSigDB C2: Johns Hopkins University School of Medicine	5
	MSigDB C2: Laboratoire CarMeN	2
	MSigDB C2: Michigan State University	6
	MSigDB C2: Stanford University	2
	MSigDB C2: Steinbeis Transfer Center for Proteome Analysis	1
	MSigDB C2: Telethon Institute for Child Health Research	4
	MSigDB C2: University Pierre and Marie Curie	2
	MSigDB C2: University of Liverpool	2
	MSigDB C2: University of Washington	392

Drugs

Annotations:		79,494
	Broad Institute CMAP	12,200
Oct 9, 2014	CTD	10,866
	CTD Marker	1,729
	CTD Therapeutic	1,986
Apr 18, 2013	Drug Bank	3,803
Feb 20, 2013	Stitch	48,910
Genes:		20,154

Disease

Annotations:		11,945
Aug 12, 2011	CTD	4,489
	Clinical Variations	4,406
	GWAS	291
	OMIM	2,759
Genes:		8,217

ToppGene Knowledgebase - Database Snapshot - ~12 million records

ToppGene - Google analytics:
Jan 15, 2014 – Jan 14, 2015
Cited ~600 times

ToppGene Suite for gene list enrichment analysis and candidate gene prioritization J Chen, EE Bardes, BJ Aronow, AG Jegga Nucleic acids research 37 (suppl 2), W305-W311	302	2009
Disease candidate gene identification and prioritization using protein interaction networks J Chen, B Aronow, A Jegga BMC bioinformatics 10 (1), 73	189	2009
Improved human disease candidate gene prioritization using mouse phenotype J Chen, H Xu, B Aronow, A Jegga BMC bioinformatics 8 (1), 392	121	2007

advanced

Sessions	Pages / Session	Avg. Session Duration	% New Sessions	Bounce Rate
27,929 % of Total: 100.00% (27,929)	6.55 Avg for View: 6.55 (0.00%)	00:08:13 Avg for View: 00:08:13 (0.00%)	29.10% Avg for View: 29.02% (0.30%)	21.28% Avg for View: 21.28% (0.00%)

Country	Sessions	Pages / Session	Avg. Session Duration
	27,929 % of Total: 100.00% (27,929)	6.55 Avg for View: 6.55 (0.00%)	00:08:13 Avg for View: 00:08:13 (0.00%)
1. United States	13,748 (49.22%)	6.88	00:08:06
2. Italy	2,088 (7.48%)	5.60	00:07:05
3. China	1,901 (6.81%)	7.04	00:09:37
4. United Kingdom	1,757 (6.29%)	5.97	00:09:10
5. India	1,390 (4.98%)	6.32	00:09:06
6. Canada	820 (2.94%)	7.81	00:09:04
7. Germany	804 (2.88%)	5.13	00:06:28
8. Spain	579 (2.07%)	6.82	00:10:54
9. Belgium	563 (2.02%)	8.32	00:11:46
10. France	457 (1.64%)	6.49	00:10:01

Teaching new tricks to old dogs

Drug Repositioning/ Repurposing

New Drug Development - Problems

- **Expensive** - 1 new drug - > ~$1 Billion
- **Time consuming** - >10 y - 10-15 years on average for an experimental drug to travel from the lab to patients.
- **Post-marketing drug failure**: Additional surfacing of drug-related adverse effects - withdrawal.
- **Decreased return** on investment by 50% in 10 y – patent expiry (generics, etc.)
- **High attrition** - Only five in 5,000 compounds that enter preclinical testing make it to human testing. Only 1 of these 5 tested in humans is approved

Drug Repositioning or Drug repurposing

- Also referred to as drug reprofiling or drug retasking
- **Reinvestigation of drug candidates that have not succeeded in previous advanced clinical trials, for reasons other than safety,**

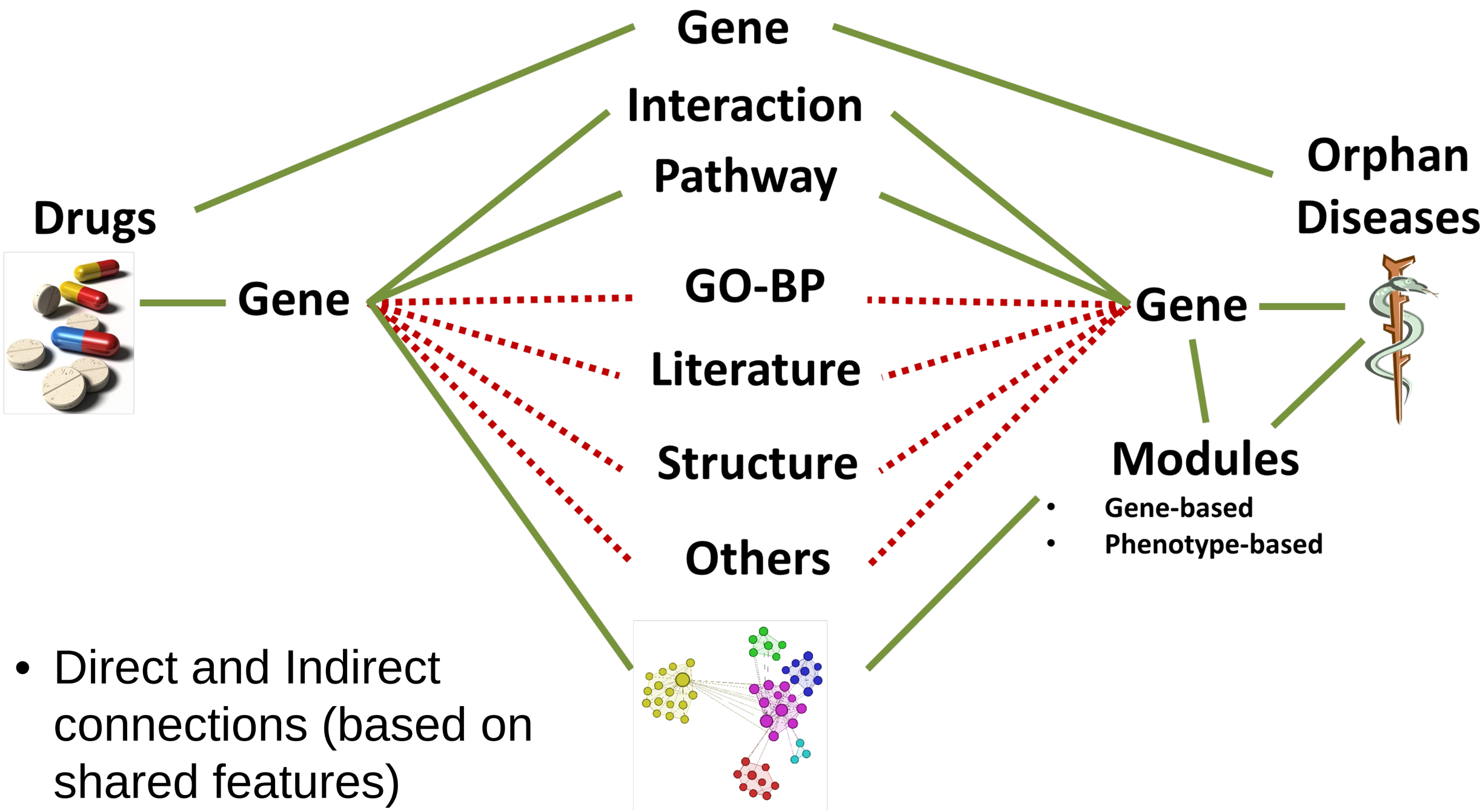
- **Search for potential new therapeutic applications of existing compounds**

Drug repositioning – Benefits & Examples

- Reduction of time and costs - Since the drug is already approved – initial timeline can be bypassed (1.5 to 2 years of preclinical and Phase I development time)
- Better/smart resource utilization
- De-Risking - lower development risk for investor
- Lower patient risk – known drug-related adverse effects

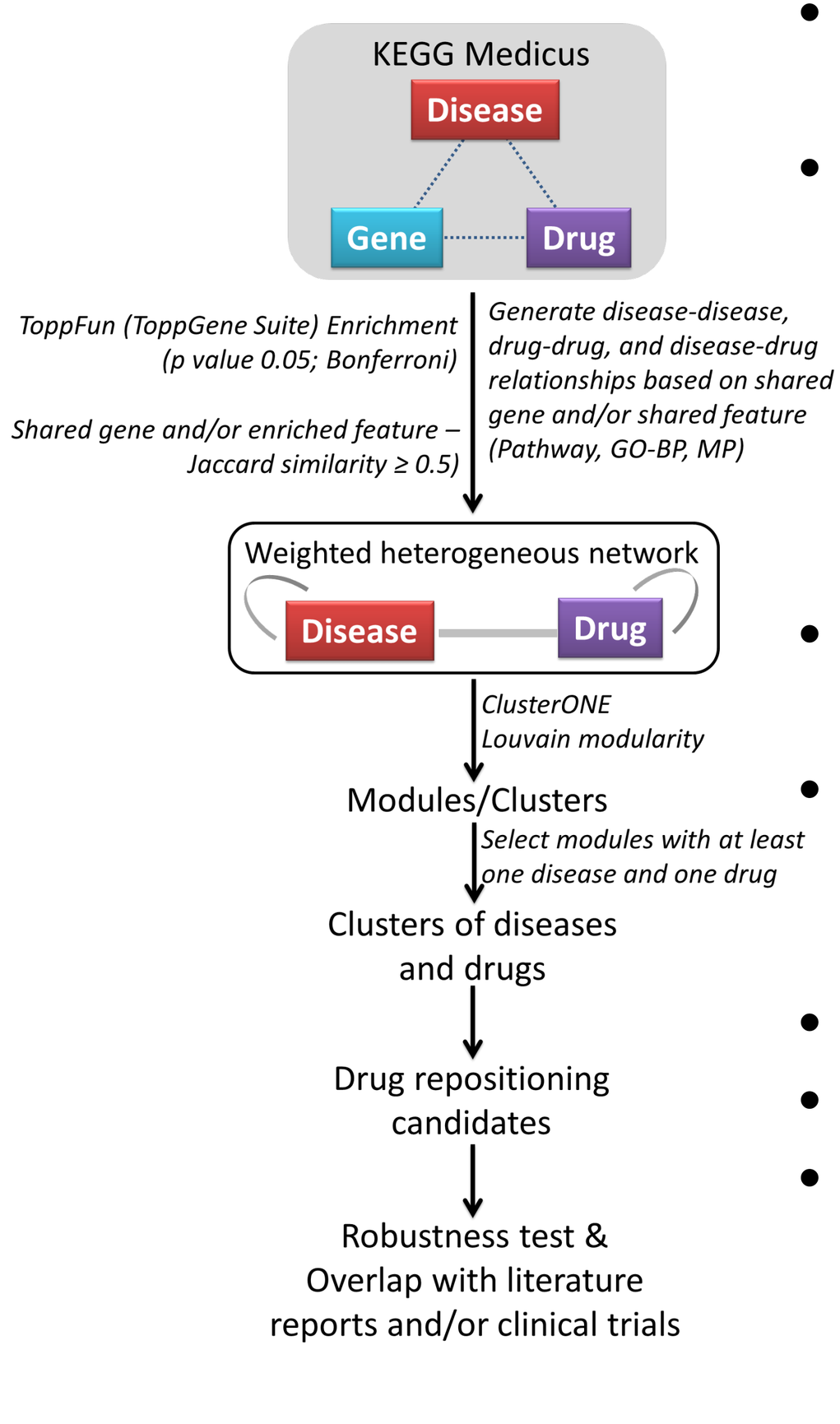
Sildenafil (Viagra): From failed antihypertensive to erectile dysfunction and to orphan disease
Thalidomide: From a dangerous drug to a promising start
Azidothymidine: Anti-cancer to AIDs
Ropinirole and **Pramipexole**: Parkinson's Drugs for Restless Legs Syndrome
Clioquinol: Antiprotozoal as a lead compound for neuroprotection
Finasteride: Prostate cancer to baldness/hair loss

Connecting approved drugs to orphan diseases



- Direct and Indirect connections (based on shared features)
- Module comparison
  - Shared genes
  - Shared enriched pathways

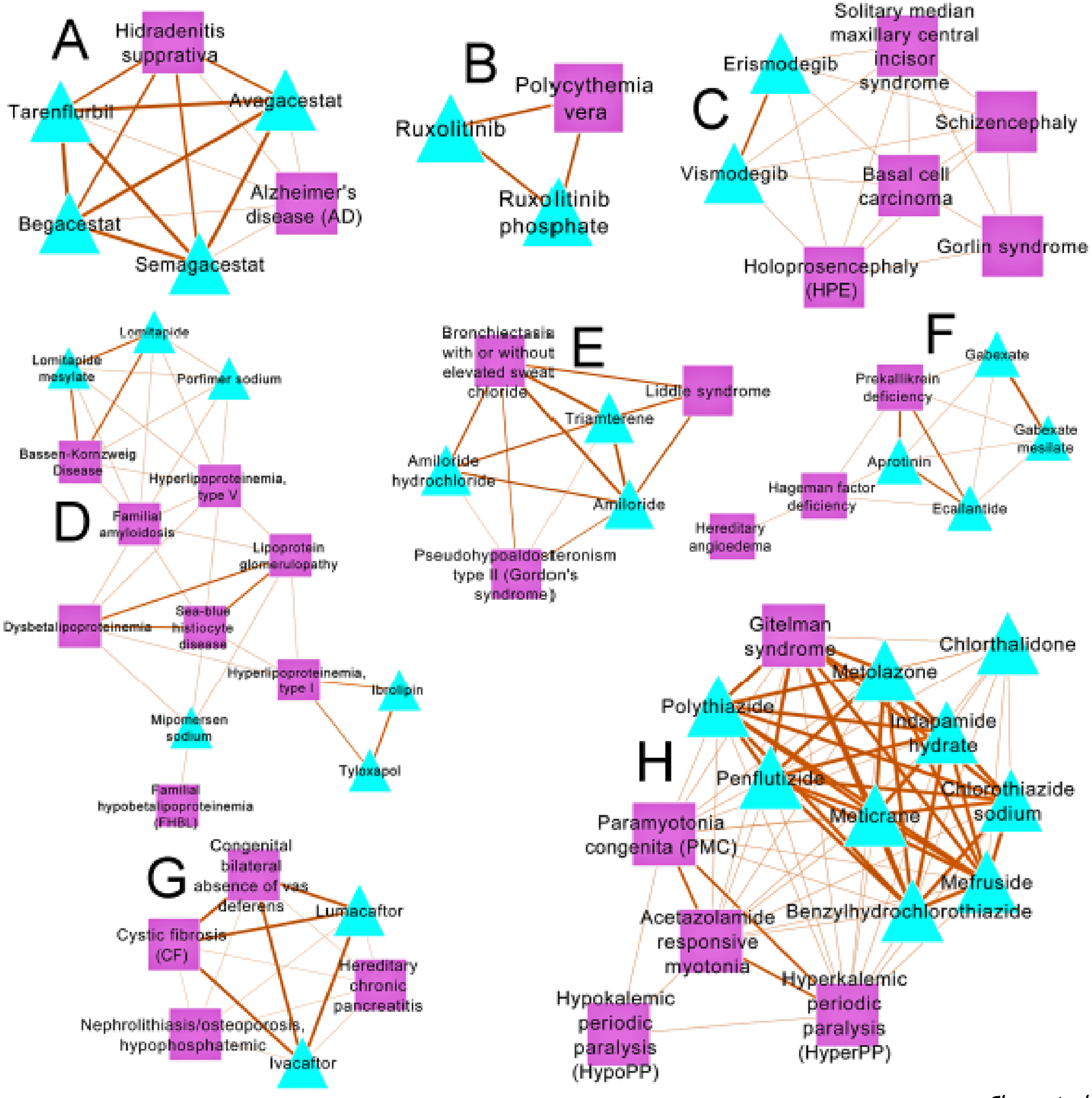
Drug repositioning - heterogeneous network clustering



- Known disease-gene and drug-target relationships from the KEGG database
- Weighted disease and drug heterogeneous network.
  - Nodes = drugs or diseases
  - Edges = shared gene, biological process, pathway, phenotype or a combination of these features
- Graph clustering of the weighted network to identify modules
- Assemble all possible drug-disease pairs (putative drug repositioning candidates) from these modules
- Validation:
- Test for robustness
- Overlap with drug indications that were either reported in published literature or investigated in clinical trials..

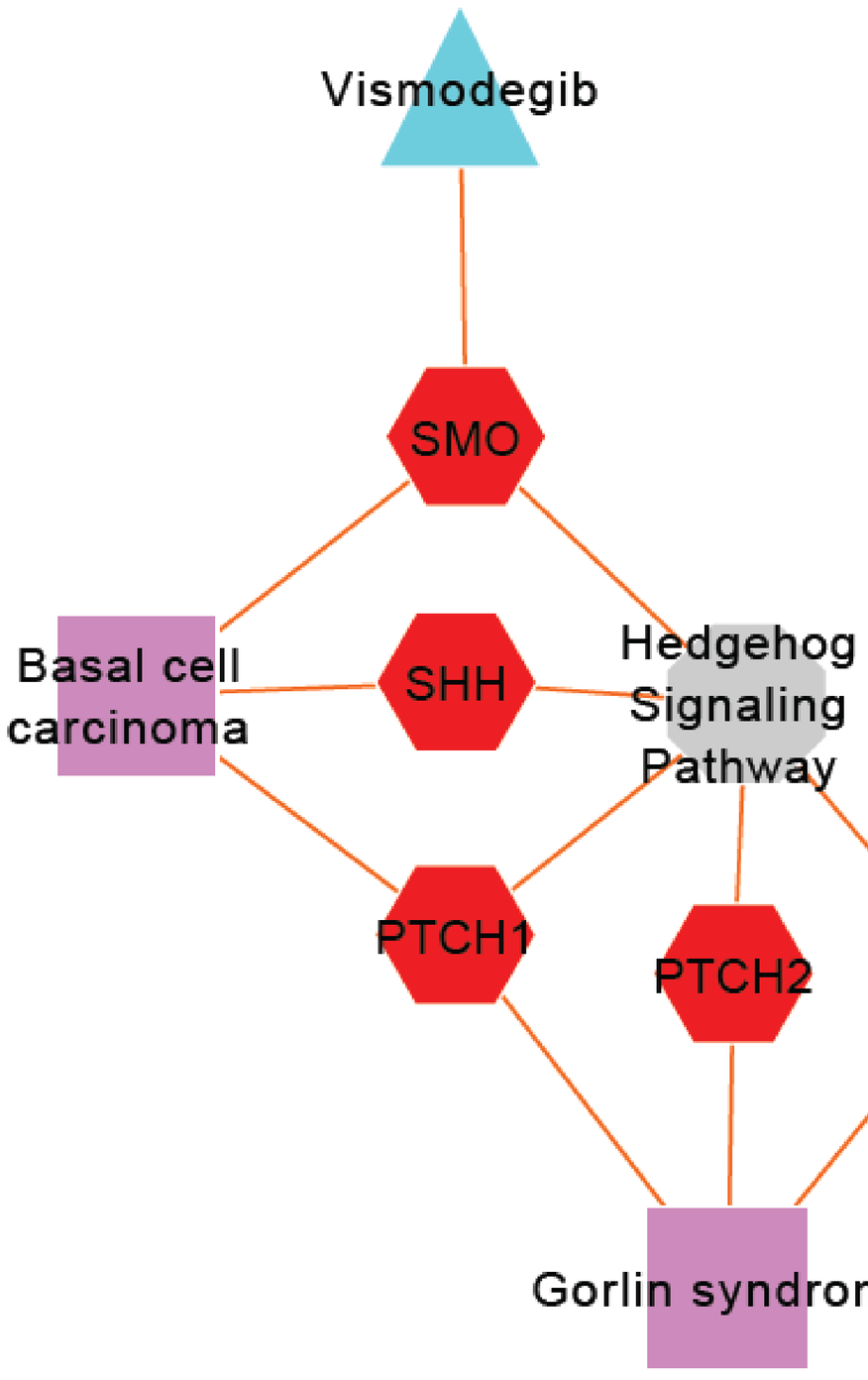
Chao et al., 2014 BMC Systems Biology

Network of clusters harboring some of the drug repositioning candidates



Chao et al., 2014 BMC Systems Biology

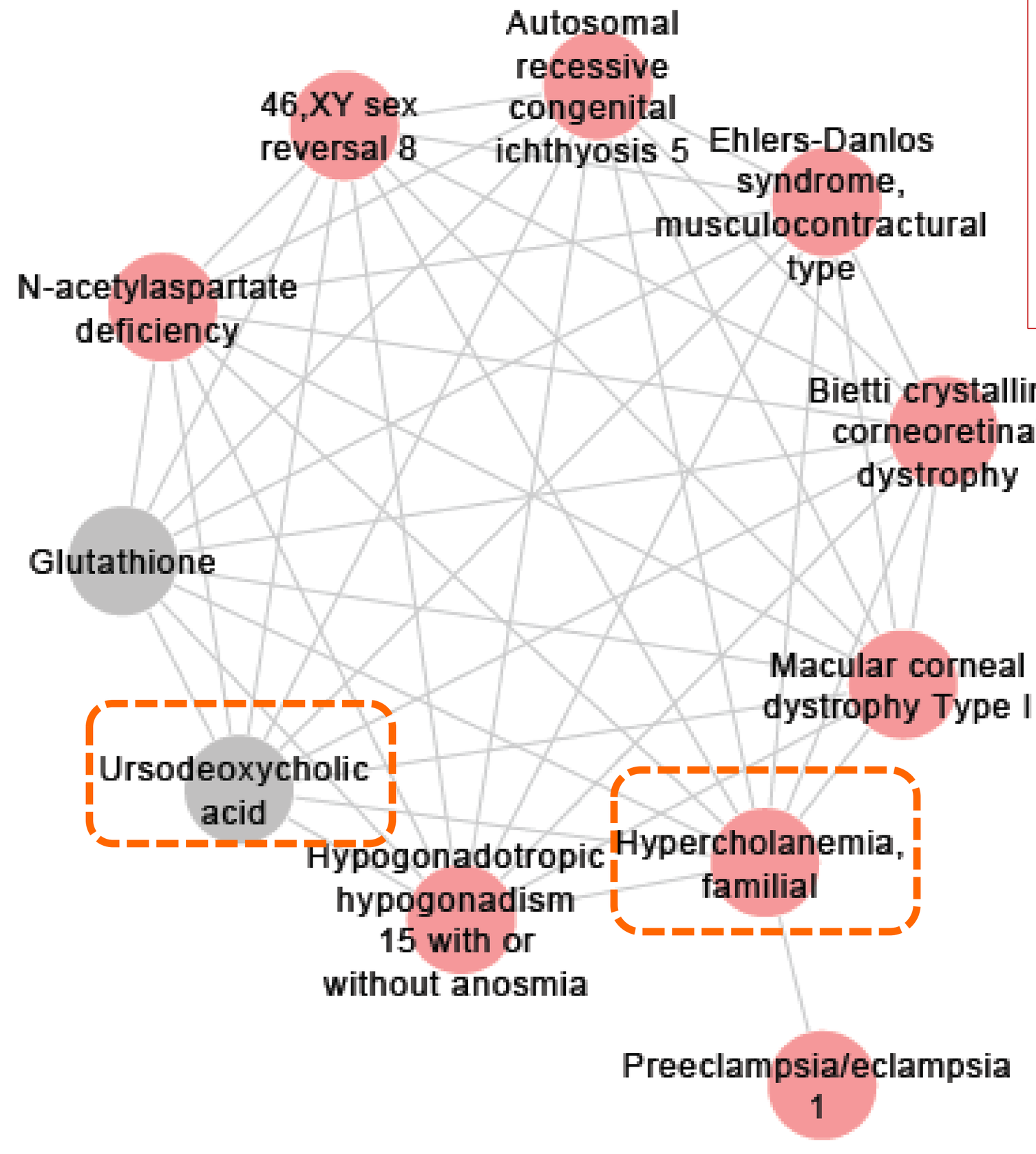
Vismodegib and Gorlin syndrome



- Cluster with drugs vismodegib and erismodegib and diseases basal cell carcinoma (BCC) and Gorlin syndrome.
- Vismodegib - oral inhibitor of the hedgehog pathway; first drug approved by the US FDA for the treatment of locally advanced and metastatic BCC
- Reported efficacy of vismodegib in patients with Gorlin syndrome (basal cell nevus syndrome), a rare autosomal dominant disorder in which those with the disease are prone to developing multiple BCCs at an early age (clinical trial NCT00957229).
- Although vismodegib and Gorlin syndrome do not share a common gene, they are still clustered together in our analyses because of the pathway-based connectivity (hedgehog signaling pathway)

Chao et al., 2014 BMC Systems Biology

All Diseases – Enrichment Networks & Modules



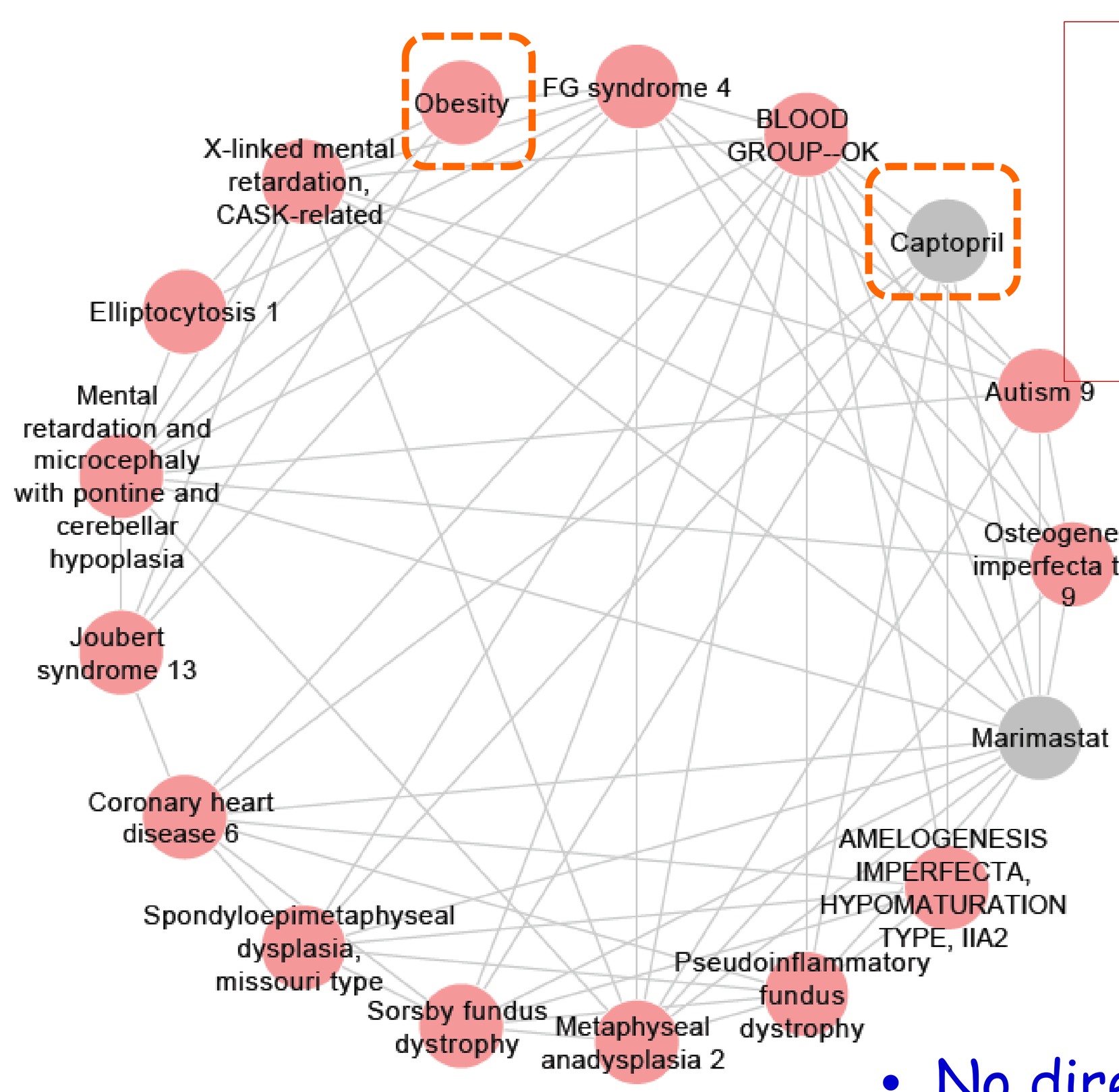
NCBI MedGen Concepts

⬇

ToppGene – Pathway enrichment

Familial hypercholanemia (FHCA) is a very rare genetic disorder characterized clinically by elevated serum bile acid concentrations, itching, and fat malabsorption reported in patients of Old Order Amish descent.

UDC (gall stones and PBC) – symptomatic treatment for FHCA



NCBI MedGen Concepts

↓

ToppGene – Pathway enrichment

Mice with diet induced obesity showed reduced food intake and body weight and improved insulin sensitivity following captopril (ACE inhibitor) treatment (Premaratna et al., 2011)

- No direct connectivity between obesity & captopril
- Part of same module

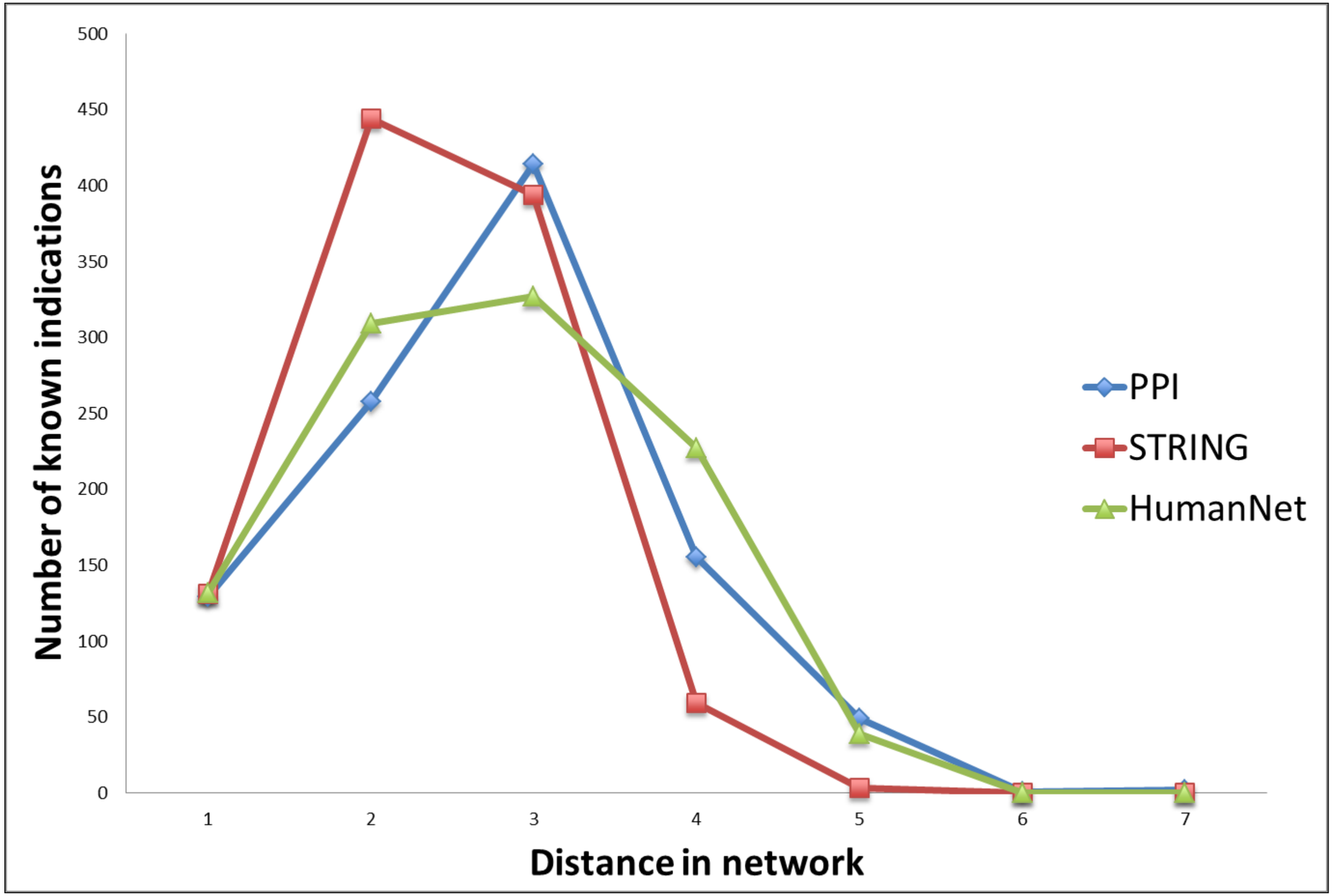
Drug repositioning candidate discovery
Random walk model

- Hypothesis: Drug targets tend to be located in proximity to the disease-associated genes in protein-protein interaction and association networks.
- Method: Random walk model
- Validation: Using known indications as a gold standard
- Results:
  - ❖ Overall area under the ROC curve of 0.95.
  - ❖ Of the 1041 known indications analyzed, about 92% (957 indications) were ranked among the top 20% suggesting that novel indications can be effectively identified by our approach.
- Robustness test
- Drug repositioning for ODs: 172 rare disorders to identify potential drug repositioning candidates.

Chao et al., (under preparation)

- 1976 known indications (disease-drug pairs) from Kegg Medicus
- Filter out diseases and drugs that do not have a known gene association in the Kegg database of disease genes and drug targets.
- 1041 known indications representing 203 diseases and 588 drugs
- Of the 1041 known indications (disease-drug pairs) only **132** pairs share at least one common gene (i.e., a disease-associated gene is also a drug target). Computed a distance measure between each of the known indication pairs in the human protein interactome.
- Calculated the shortest path for all known indications (i.e., shortest path between a known disease and drug pair) in the protein interactions network . Of the 1041 known indications, we were able to compute the shortest paths for **1008** disease-drug pairs. For the remaining pairs, we were unable to compute the shortest paths because their encoded proteins were either absent in the interactome or were not reachable (e.g., a disease protein and drug target present in two different connected components of the protein interactome).

Chao et al., (under preparation)

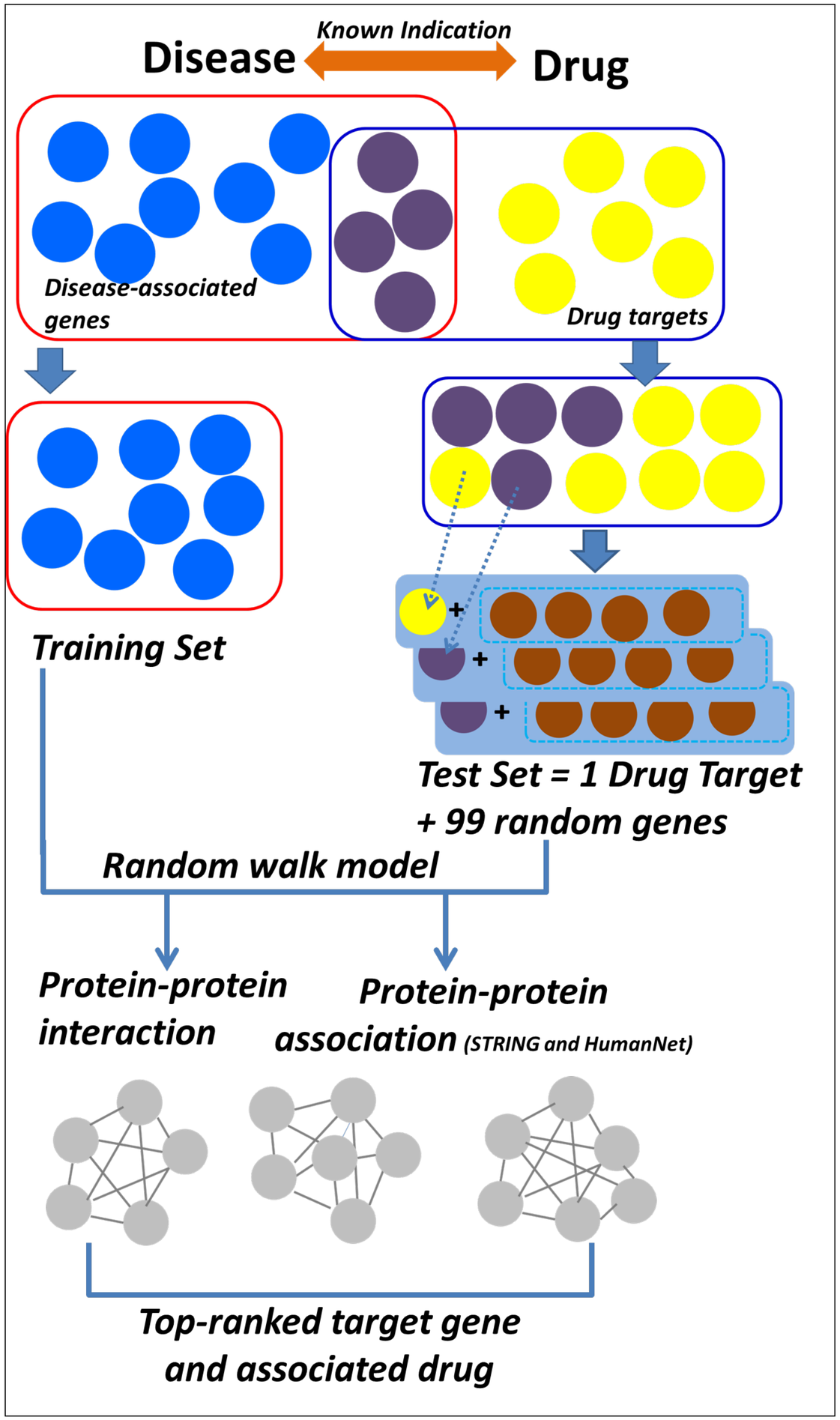


Average distance between a disease-drug of known indications in PPI is 3.75

Protein association networks:

- STRING: 3.38
- HumanNet: 3.74

Chao et al., (under preparation)

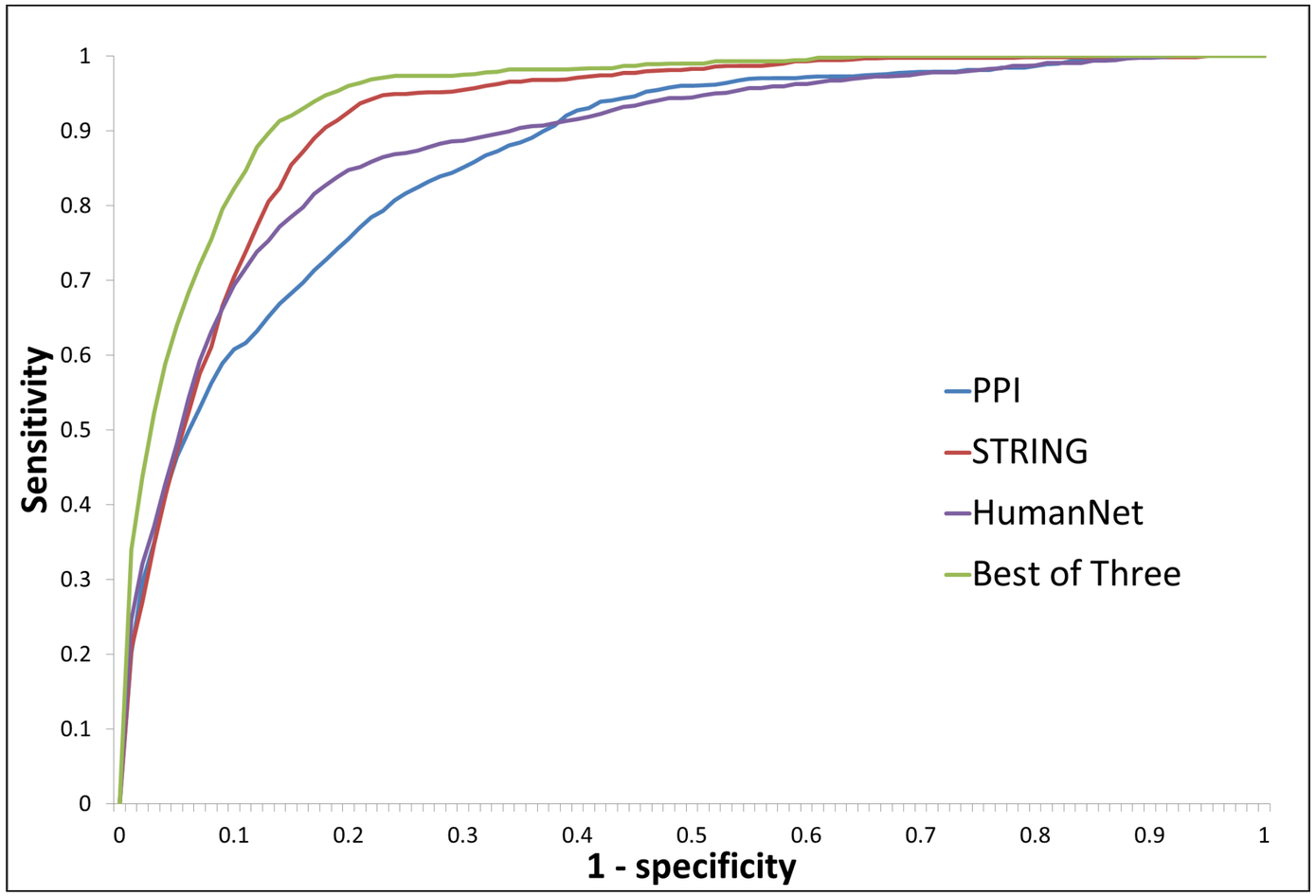


- Goal: Check whether using disease-associated genes as training set, the indicated drug can be identified by ranking the drug targets using random walk model.
- Known Indications from KEGG
- Training Set: Genes associated with a disease
- Test Set: Corresponding drug target gene plus 99 random genes
- Prioritization using Random Walk - Record the ranking of the drug target gene.

ODs:

- Training set: OD-associated genes
- Test set: Entire druggable gene set of KEGG and DrugBank
- Perform prioritization
- Retrieve the top ranked five genes and the drugs targeting these genes.

Chao et al., (under preparation)



1041 known indications

- ❑ 928 pairs in protein interactome
- ❑ 938 pairs in STRING
- ❑ 984 pairs in HumanNet

AUC scores:

- Protein interactome: 0.87
- HumanNet: 0.89
- STRING: 0.92 in STRING

Ranked in top 10%:

- Protein interactome: 564 pairs
- STRING: 654 pairs
- HumanNet: 683

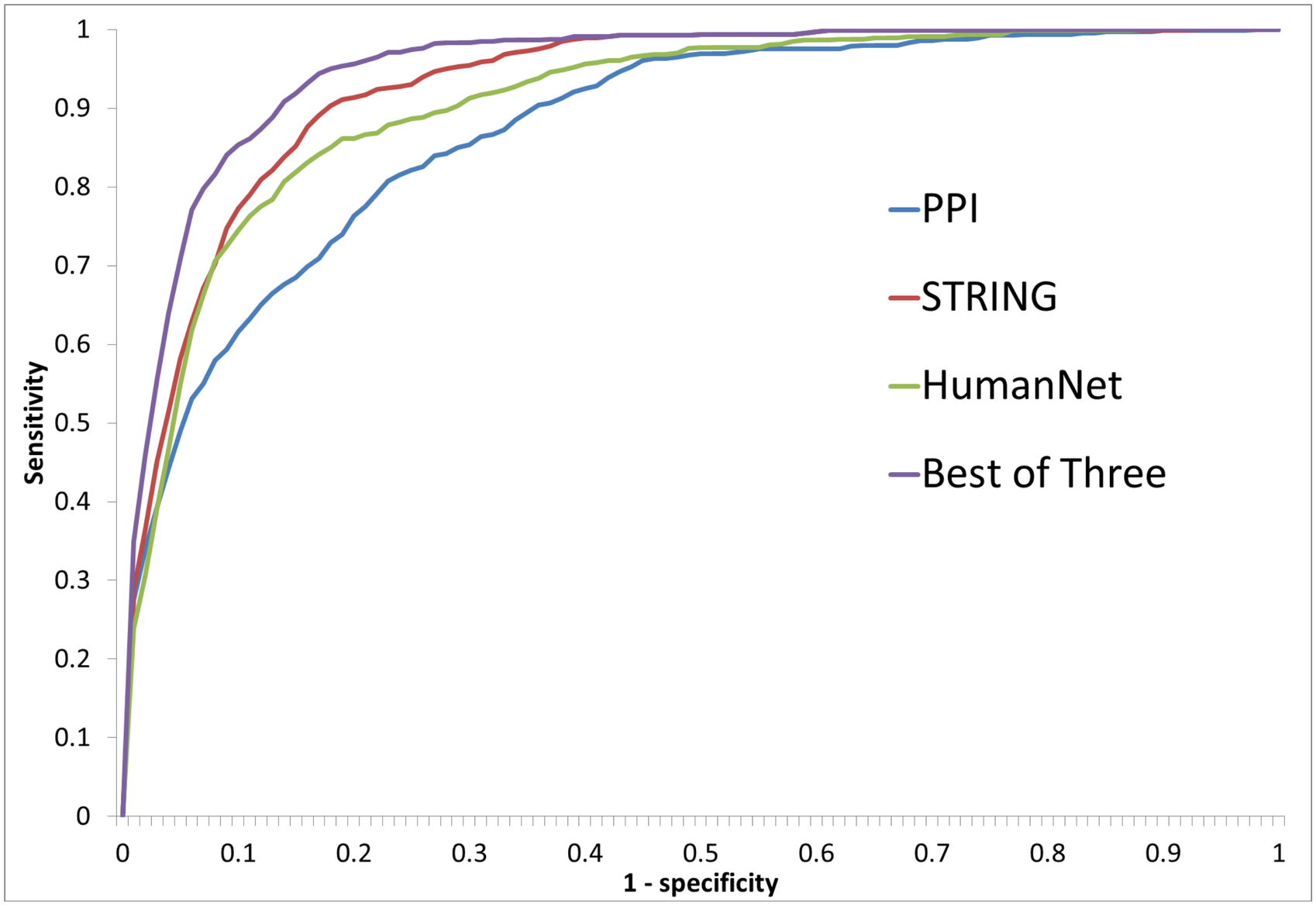
Combined (PPI+STRING+HumanNet)

AUC score: 0.95

Over 82% of the pairs had at least one drug-target gene ranked within top 10%

Chao et al., (under preparation)

- No "good" gold standard available for drug repositioning
- Clinical trials: generated a list of indications where a drug from our known indications is being investigated for a different disease
- Compiled 668 disease-drug pairs from clinical trials where an approved drug with a known indication was investigated as a therapeutic intervention for a different disease(s).



Prioritization analysis:

- PPI: 0.88
- STRING: 0.93
- HumanNet: 0.91

Chao et al., (under preparation)

OD networks based on similar phenotype (symptoms)

1. More than 4000 ODs <u>without</u> causal gene information.
2. Build a network of <u>**ALL**</u> diseases (OD and non-OD) based on shared phenotype.
3. Overlay this network with gene based disease (OD and non-OD) network.
4. Identify clusters of networks where:
   a. Edge represents *shared gene* only
   b. Edge represents *shared symptoms/phenotype* only
   c. Edge represents both *shared gene and phenotype*
5. Overlay the combined network with known drugs (drug – target database)

ODN

- Shared Genes
- Shared Phenotypes - Shared HP - Resnik Similarity score

ODN

- Node: ODs
- Edge:
  - Shared Genes
  - Shared Phenotypes
  - Both

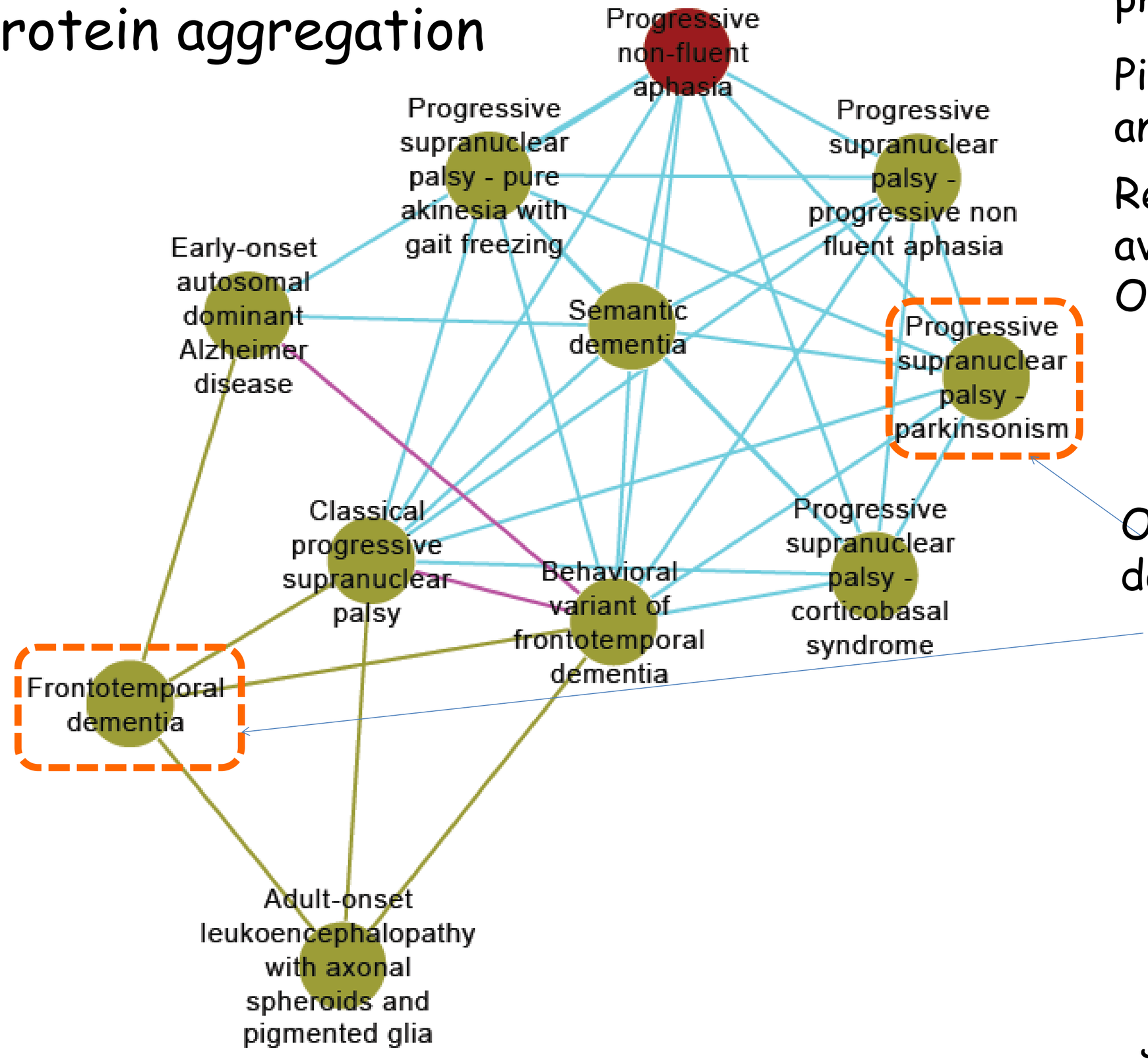
3209 nodes
13934 edges

Modules

Jegga et al., (under preparation)

Methylthioninium

Inhibitor of Tau protein aggregation



Blue edge: Shared gene

Green edge: Shared phenotype

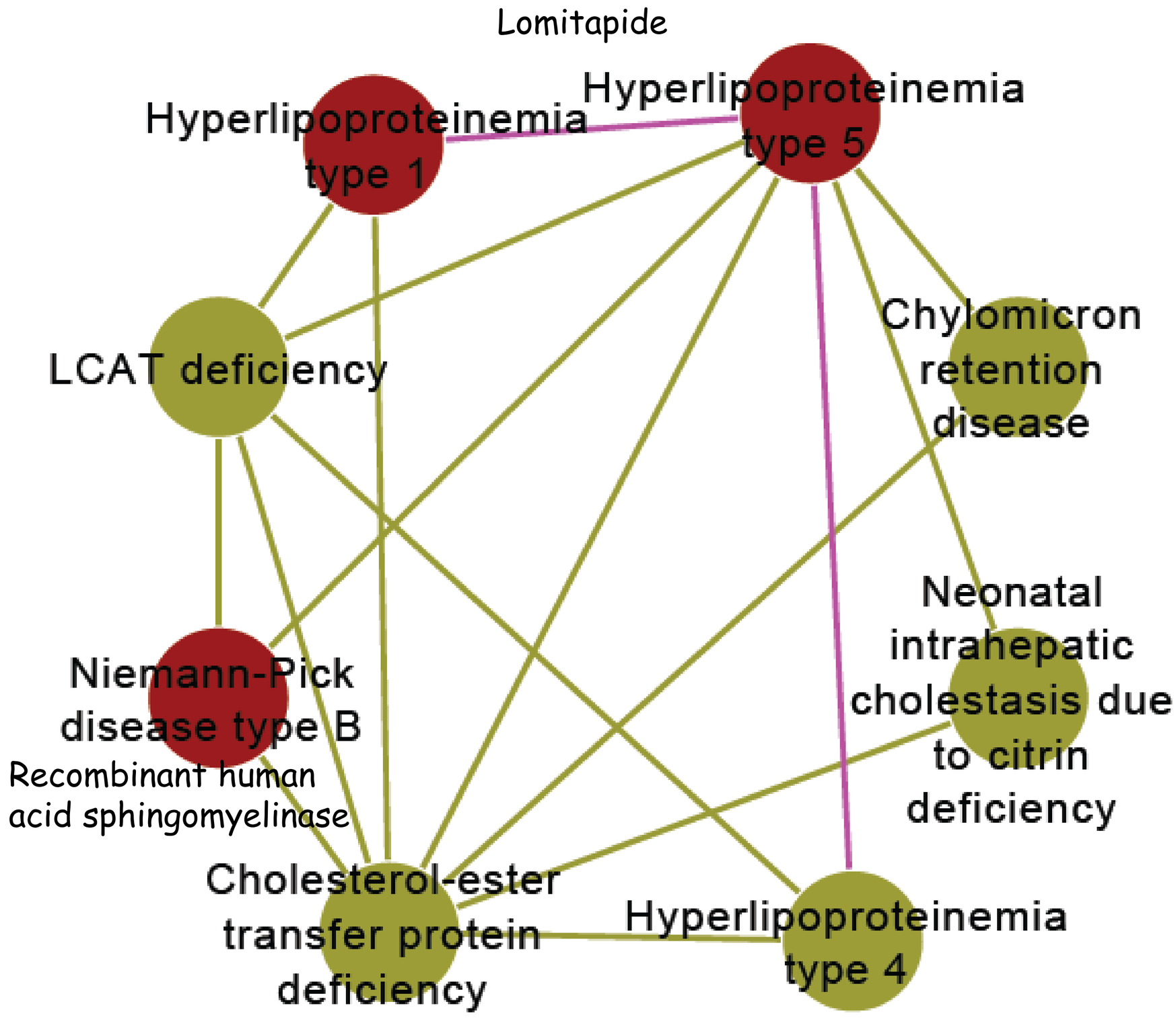
Pink edge: Shared gene and phenotype

Red nodes: Orphan drug available (as per Orphanet)

Orphan drug designation - Europe

Jegga et al., (under preparation)

Lomitapide



Blue edge: Shared gene

Green edge: Shared phenotype

Pink edge: Shared gene and phenotype

Red nodes: Orphan drug available (as per Orphanet)

Jegga et al., (under preparation)

ODN + non-ODN (OMIM X OMIM)

- Shared Genes
- Shared Phenotypes - Shared HP - Resnik Similarity score

Disease Network

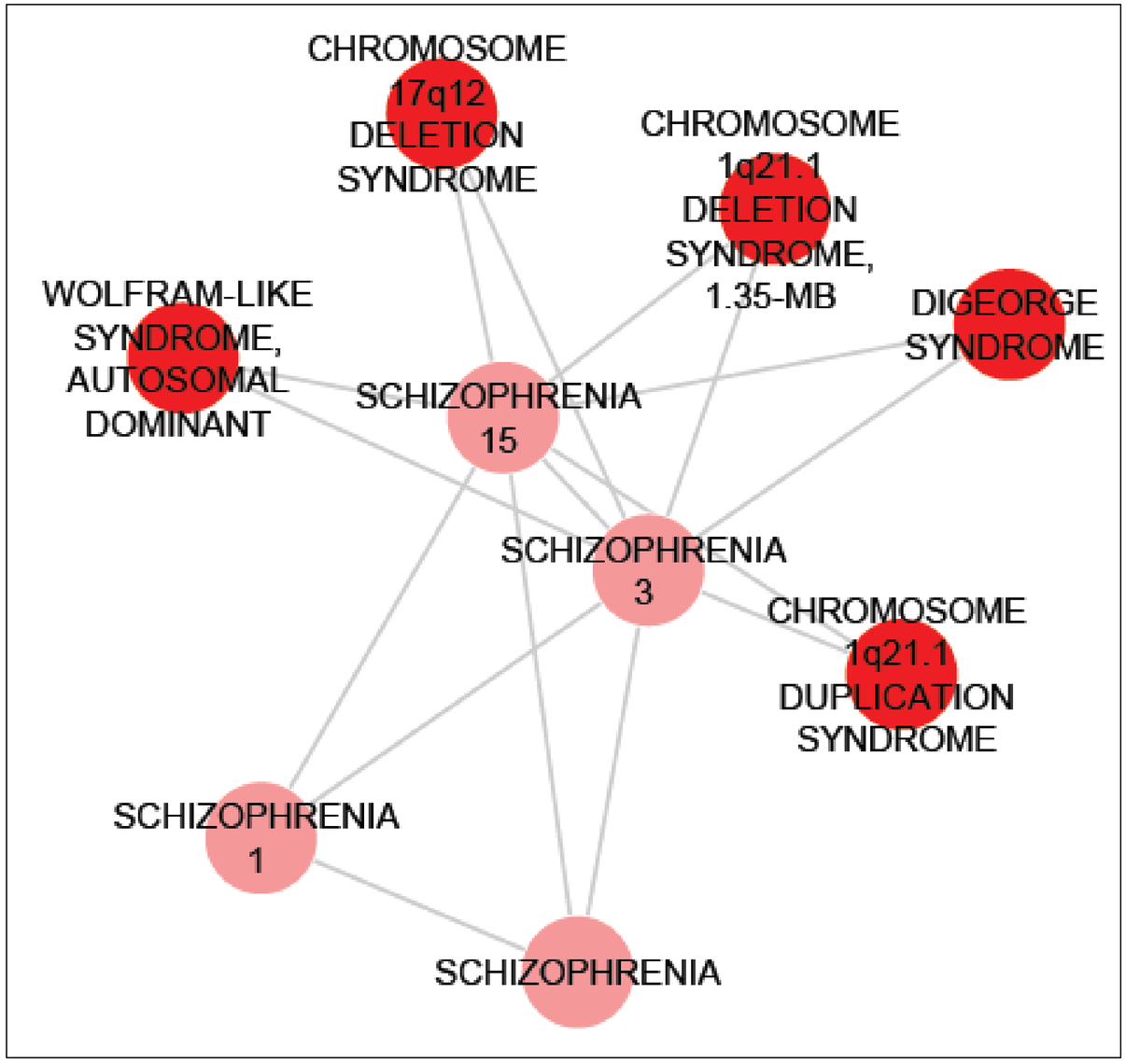
- Node: All Diseases
- Edge:
  - Shared Genes
  - Shared Phenotypes
  - Both

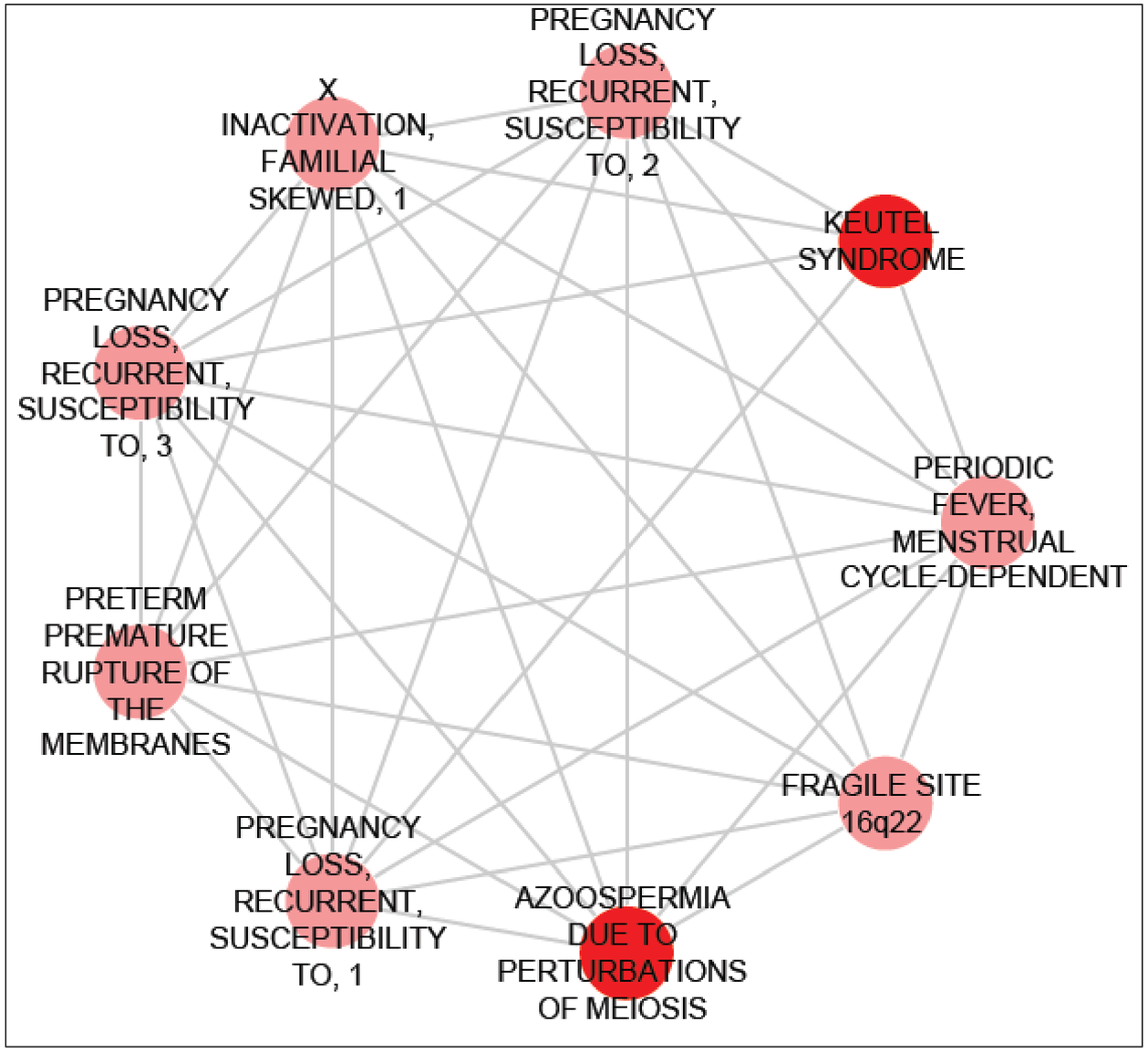
4735 Nodes
(3596 ODs)
32312 edges

→ Modules

Jegga et al., (under preparation)

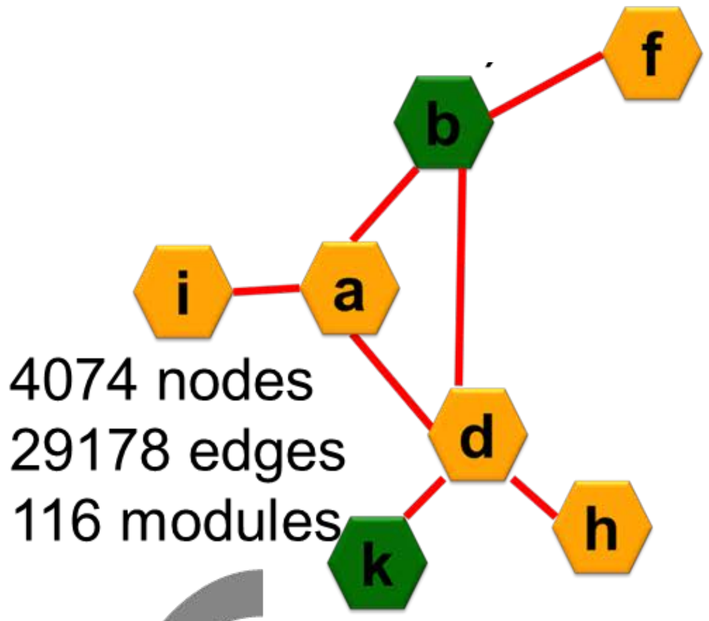
Co-clustering of OD and non-ODs





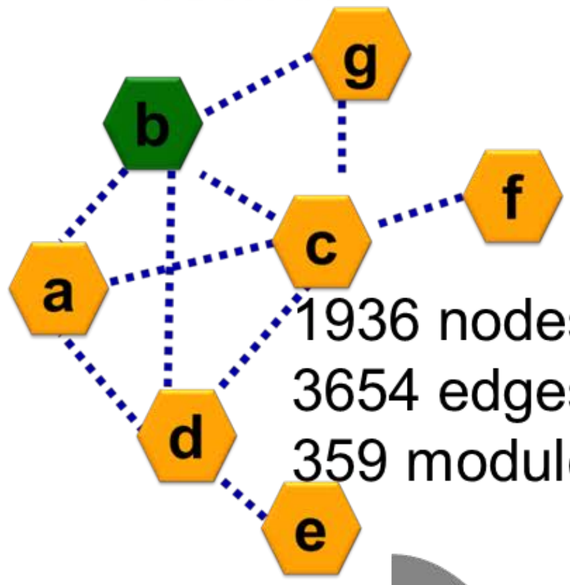
Red nodes are ODs
Edge: Shared phenotype

Phenotype-based disease network

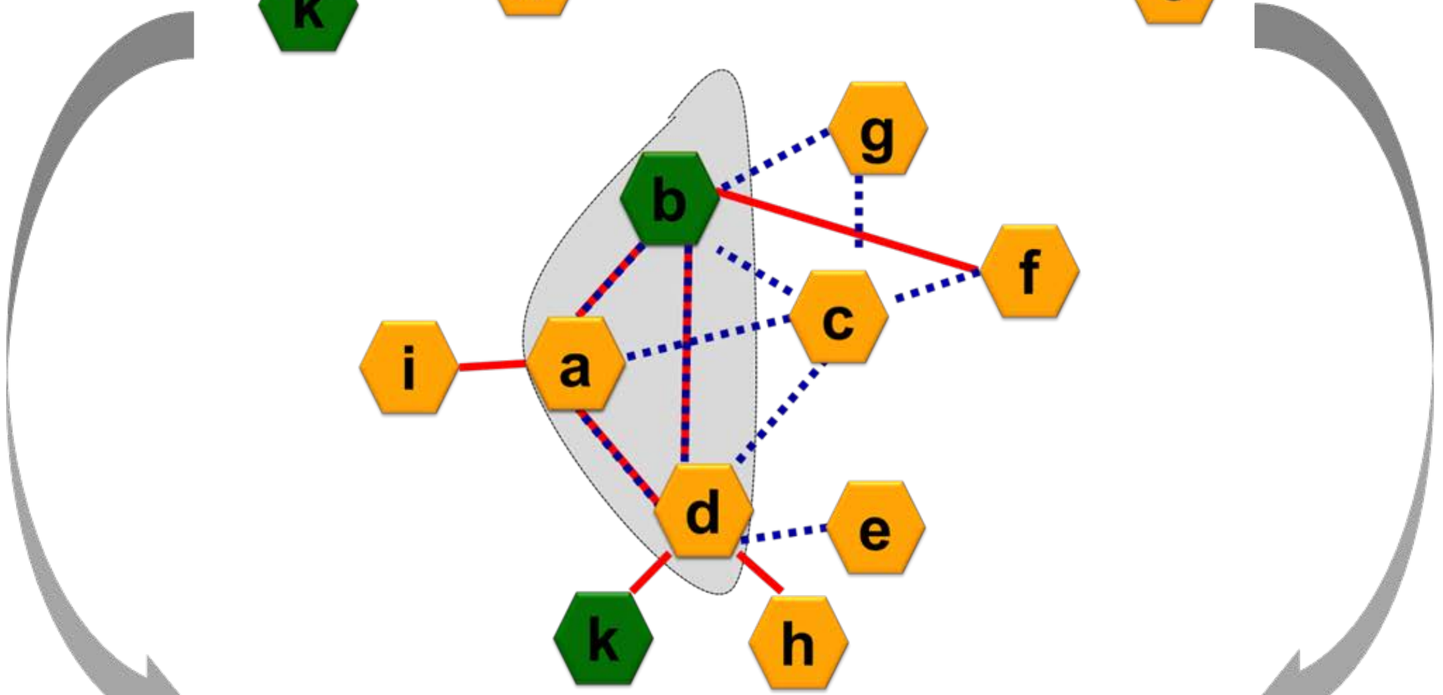


4074 nodes
29178 edges
116 modules

Gene-based disease network



1936 nodes
3654 edges
359 modules



Module comparison for gene overlap (Jaccard score) and identify if their encoded proteins are known drug targets.

Jegga et al., (under preparation)

Limitations & Future Work

1. Lack of a good gold standard for repositioned drugs. Validation (computational) is a challenge.
2. Nomenclature issues (OMIM-orphanet-HP-MP, etc. mappings)
3. Other metrics for computing module similarity - graph alignment-based approaches.
4. Incorporating additional features like gene expression data (both disease based and drug based; e.g. Connectivity Map), protein structural data (from PDB) and also additional functional linkage networks (weighted; shared GO terms, etc.).
5. Curation of repositioning candidates (literature search, etc.) – ranking/scoring.
6. EHR, where possible (OD and common disease networks).


www.CartoonStock.com
Zahn
"I was wrong...you can teach an old dog new tricks."

Thank You