Interactome networks for the system biology of complex disease

Tae Hyun Hwang, PhD
Biostatistics and Bioinformatics, Masonic cancer Center
University of Minnesota Twin Cities

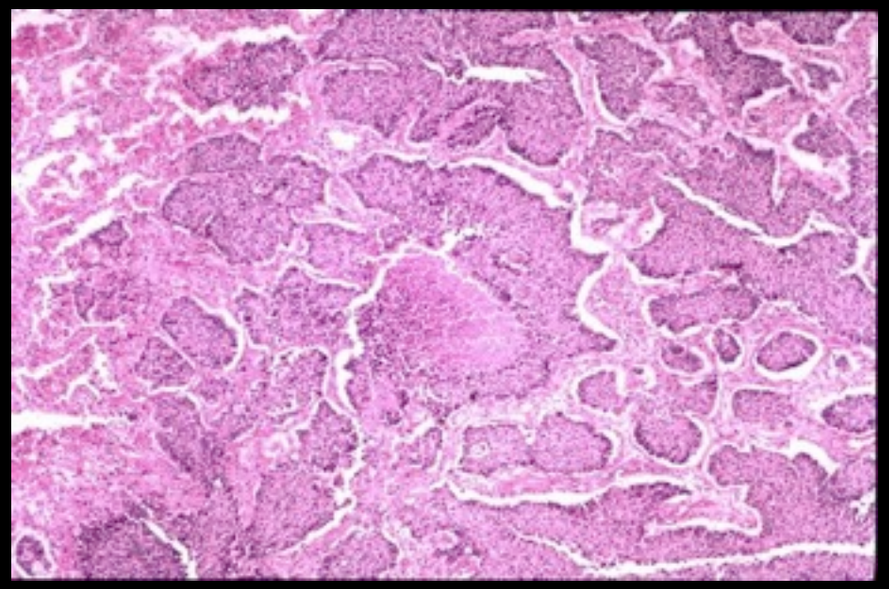
UNIVERSITY OF MINNESOTA
Driven to Discover℠

통계유전학 워크샵

NCI CCC
A Comprehensive Cancer Center Designated by the National Cancer Institute

Background

- **Phenotype**
  - The set of observable characteristics of an individual resulting from the interaction of its genotype with environment
  - Phenotypes could be either disease phenotypes or any observable characteristics



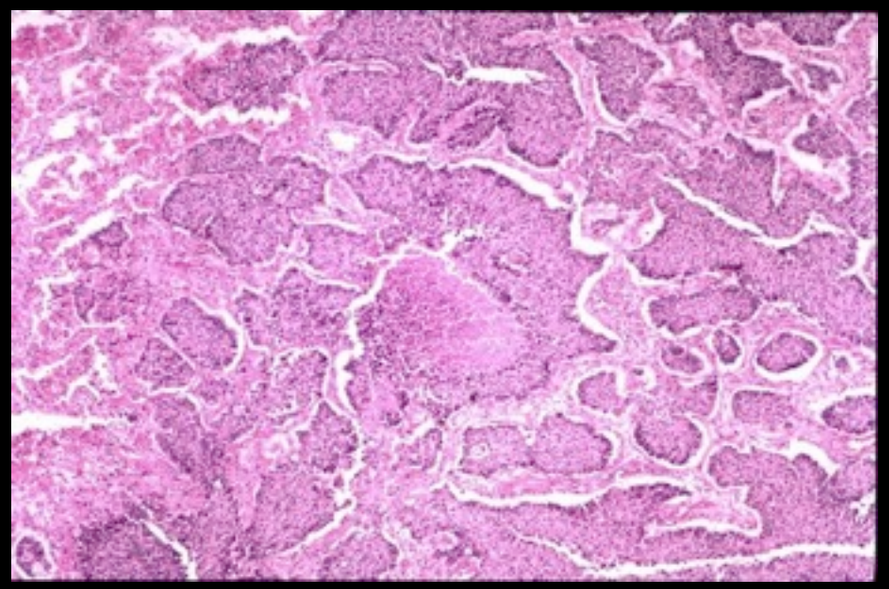
Cancer

Blond hair

Eye color

Background

- **Phenotype-genotype association**
  - Identify genetic variations affecting the phenotypic changes on a genome-scale
    - What/How can genetic variation affect to develop phenotypes



Cancer

Blond hair

Eye color

Applications and Significance

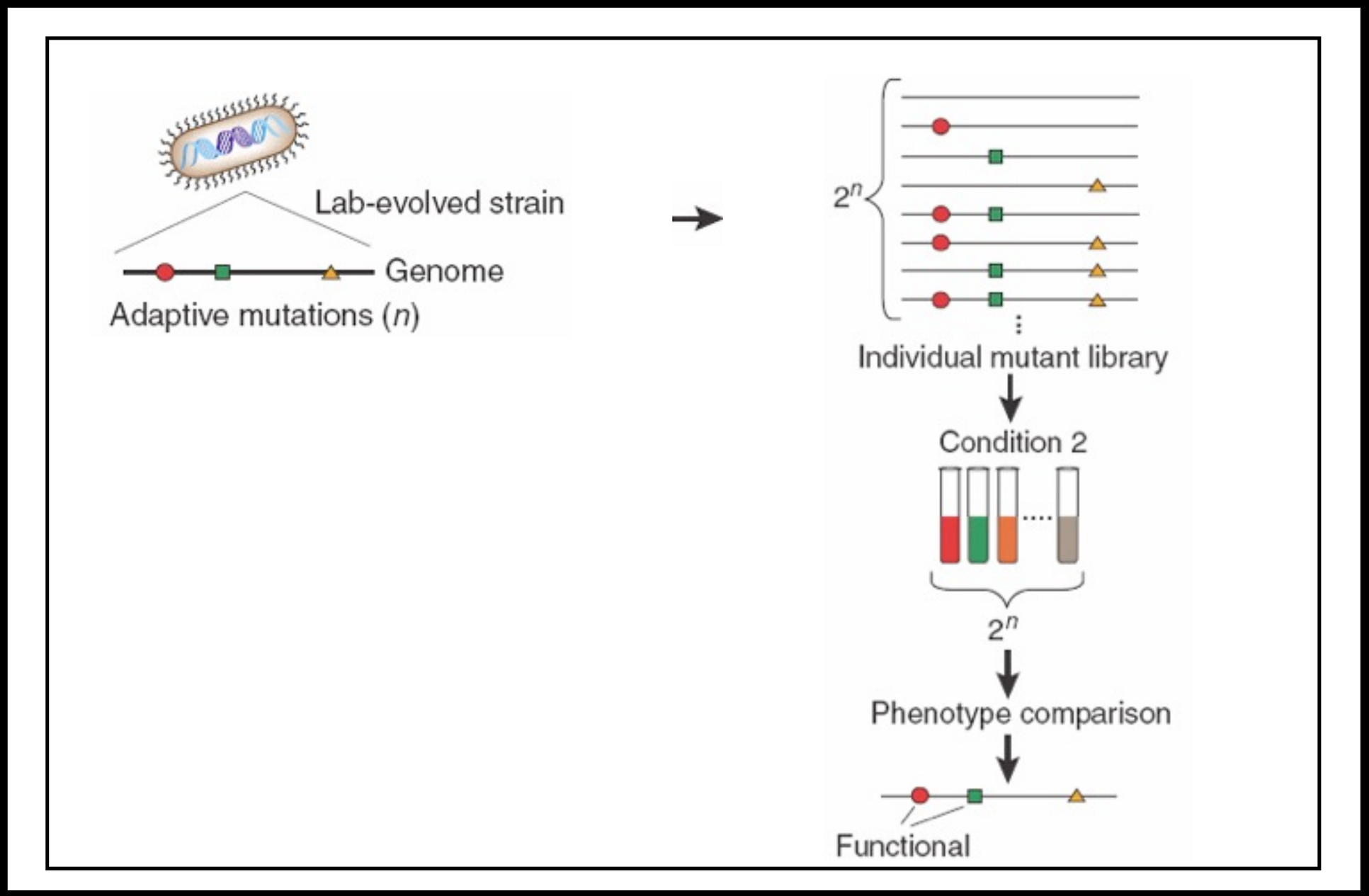
- **Understanding of how genome determines important phenotypes could lead to ...**
  - Find genes to develop new drug targets and treatments
  - Genetic engineering of yeast ethanol
  - Improve food production
  - and more ...

Time and Cost of Drug Discovery

- **From Genentech**
  - Time: ~ 10 years
  - Cost: ~ 2 billions (US dollors)
  - Human resource: ~ 200 PhDs
  - and more ...
    - But, no guarantee that candidate drug will be approved by FDA

Traditional approach

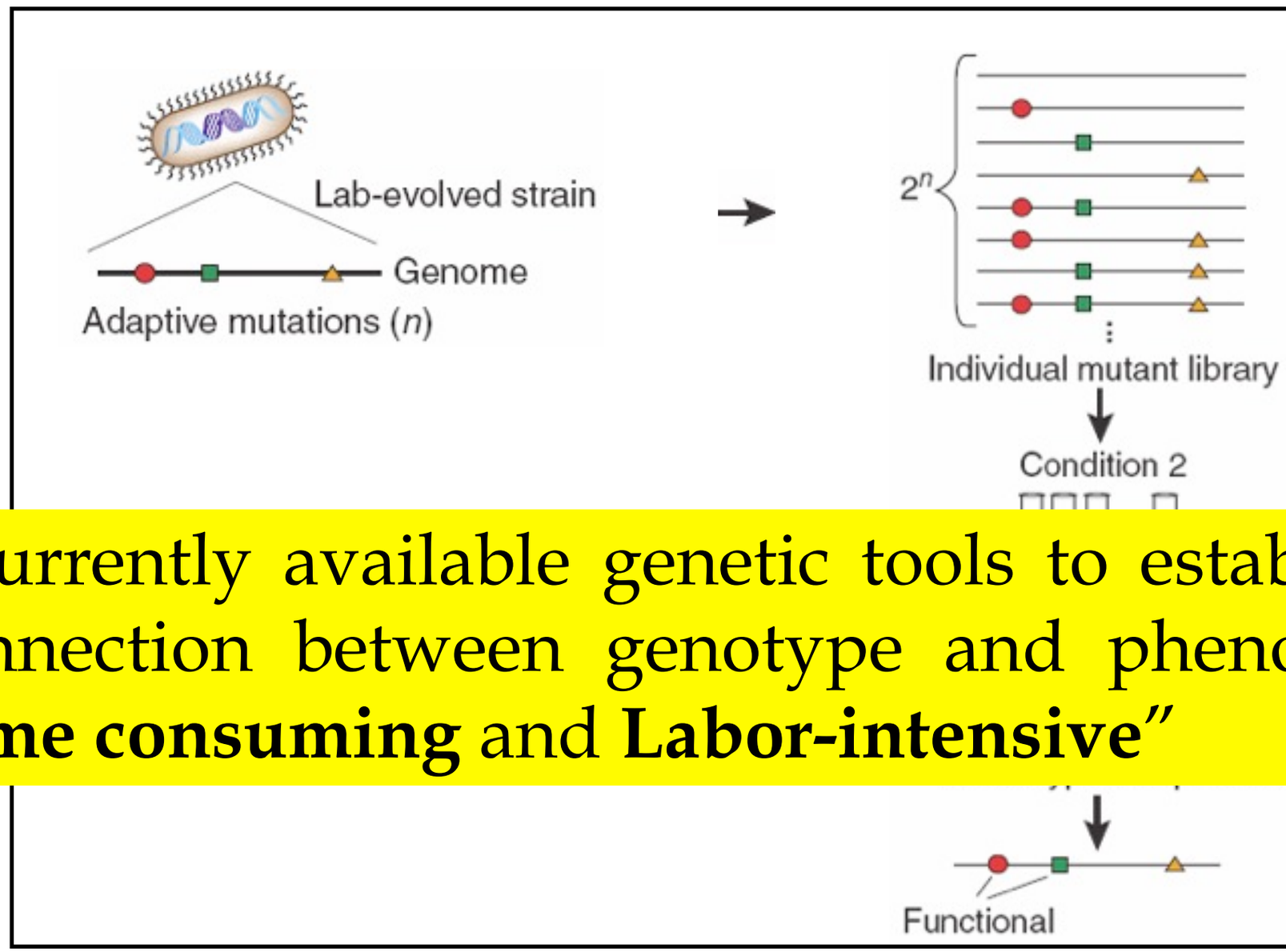
- **Phenotype-genotype association study**
  - Introduce genetic variations into model, and validate it *in vivo* and *in vitro*



[B Cho & B Palsson, “Probing the basis for genotype-phenotype relationships”, Nature methods 2009]

Traditional approach

- **Phenotype-genotype association study**
  - Introduce genetic variations into model, and validate it *in vivo* and *in vitro*

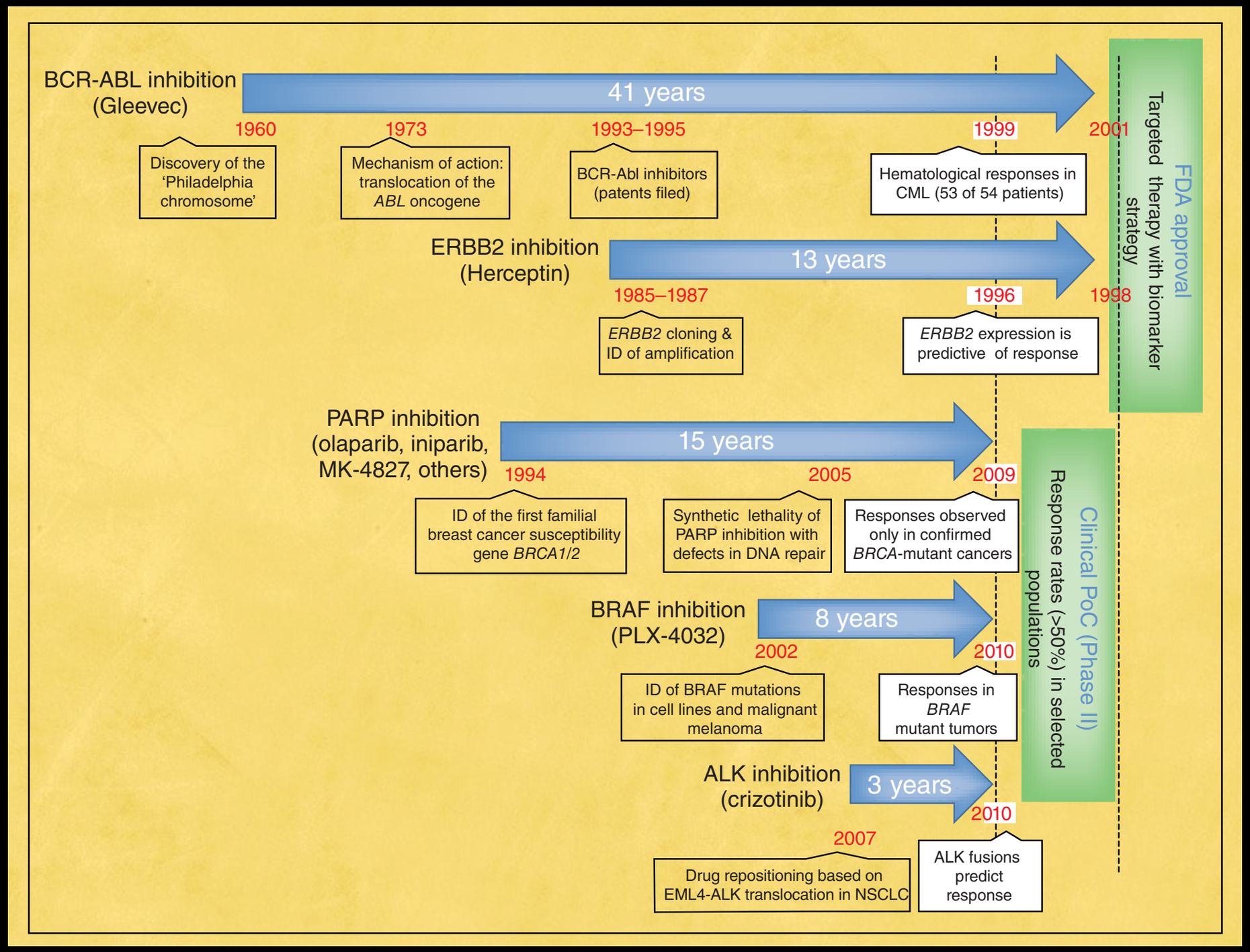


"Currently available genetic tools to establishing a connection between genotype and phenotype are **Time consuming** and **Labor-intensive**"

[B Cho & B Palsson, "Probing the basis for genotype-phenotype relationships", Nature methods 2009]

High-throughput approach

- **High-throughput technologies can improve the process of new drug development?**



Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. Nat Med. 2011; 17(3): 297-303.

Network-based approach

- **Network-based approach can help to boost disease gene discovery**
  - Disease gene and pathway discovery Cell

Mapping the NPHP-JBTS-MKS Protein Network Reveals Ciliopathy Disease Genes and Pathways Cell *145*, 513–528, May 13, 2011 ©2011 Elsevier Inc

  - Next-generation sequencing data analysis

Exome sequencing and disease-network analysis of a single family implicate a mutation in *KIF1A* in hereditary spastic paraparesis

Yaniv Erlich, Simon Edvardson, Emily Hodges, et al.

Genome Res. 2011 21: 658-664 originally published online April 12, 2011
Access the most recent version at doi:10.1101/gr.117143.110

  - Genomic data integration

Theory

Cell

An Integrated Approach to Uncover Drivers of Cancer Cell *143*, 1005–1017, December 10, 2010 ©2010 Elsevier Inc.

Network-based approach

- **Network-based approach can help to boost disease gene discovery**
  - Disease gene prioritization

Computational tools for prioritizing candidate genes: boosting disease gene discovery

NATURE REVIEWS | GENETICS

Nature Reviews Genetics | AOP, published online 3 July 2012; doi:10.1038/nrg3253

  - Functional enrichment analysis

BIOINFORMATICS **ORIGINAL PAPER**

Vol. 27 no. 19 2011, pages 2692–2699
doi:10.1093/bioinformatics/btr463

Systems biology

Advance Access publication August 8, 2011

Inferring disease and gene set associations with rank coherence in networks

  - and many (gene function prediction, drug target prediction, the pathological analysis of human disease)

Network medicine: a network-based approach to human disease

56 | JANUARY 2011 | VOLUME 12

Interactome Networks and Human Disease

Cell

Leading Edge

Review

Today's topic

- **Disease phenotype-gene association study**
  - Identify genetic variations affecting the phenotypic changes on a genome-scale
- **Applications**

1. Disease gene prediction
   - Predict candidate disease genes associated with a query disease phenotype
2. Predicting phenotypic/functional impact of candidate disease genes
   - Give a gene (or a set of genes), predict its target disease phenotypes/functions

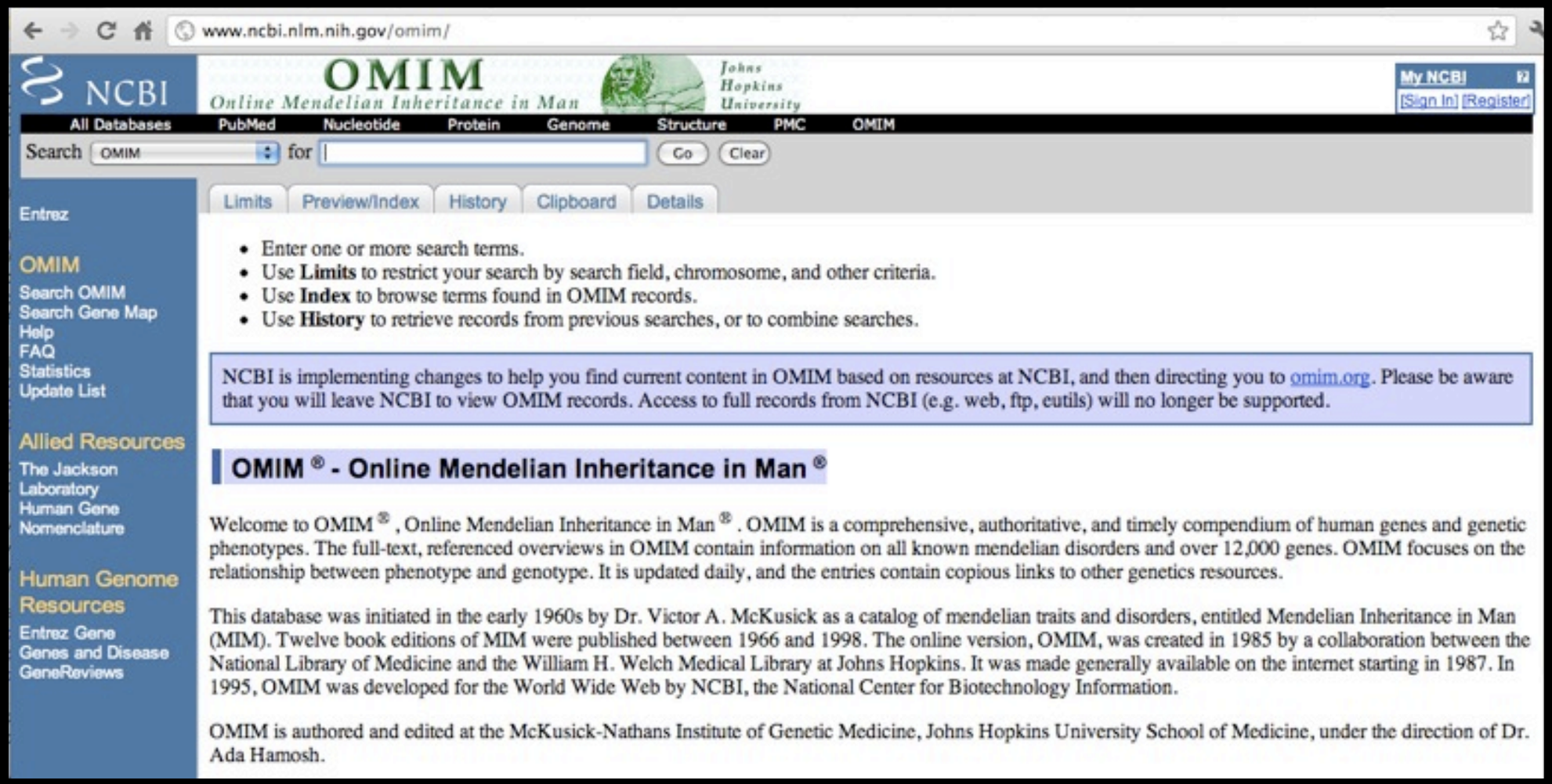
Today's topic

- **Disease phenotype-gene association study**
  - Identify genetic variations affecting the phenotypic changes on a genome-scale
- **Applications**

1. Disease gene prediction
   - Predict candidate disease genes associated with a query disease phenotype
2. Predicting phenotypic/functional impact of candidate disease genes
   - Give a gene (or a set of genes), predict its target disease phenotypes/functions

Disease gene prediction

- **Online Medelian Inheritance in Man Statistics (April 29, 2010)**
  - **3775 out of 6543 disease phenotypes** are still **not known** for their causative disease genes, and underlying genetic basis



http://www.ncbi.nlm.nih.gov/omim/

When we need computational approaches ?

- **Not enough experimental data**
  - Small sample size
- **Too many candidate biomarker**
  - > 1000 mutations from next-generation sequencing data
- **For the use of prior knowledge**
  - I know some important genes for phenotype X
- **and many ...**

Disease gene discovery methods

- **Data driven method**
  - Integrate experimental, sequence, and other biological data
    - Ex) Endeavour
- **Network-based method**
  - Use molecular interaction networks (e.g., protein-protein interaction networks)
- **Integrated network-based method**
  - Integrate multiple interactome networks data (e.g., disease network, disease-gene association network and PPI networks)

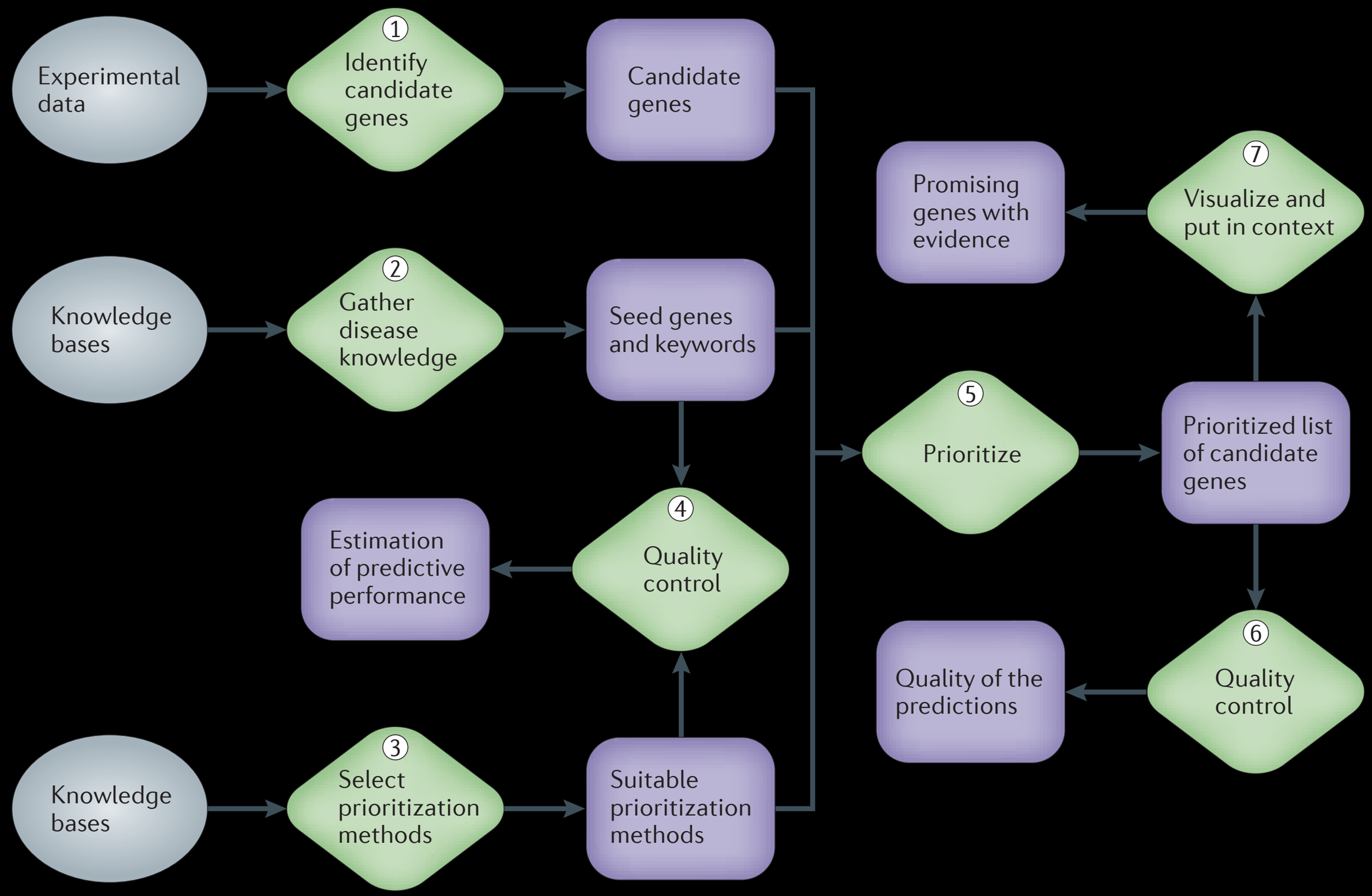
Disease gene discovery methods

- **Data driven method**
  - Integrate experimental, sequence, and other biological data
    - Ex) Endeavour
- **Network-based method**
  - Use molecular interaction networks (e.g., protein-protein interaction networks)
- **Integrated network-based method**
  - Integrate multiple interactome networks data (e.g., disease network, disease-gene association network and PPI networks)

Data driven method

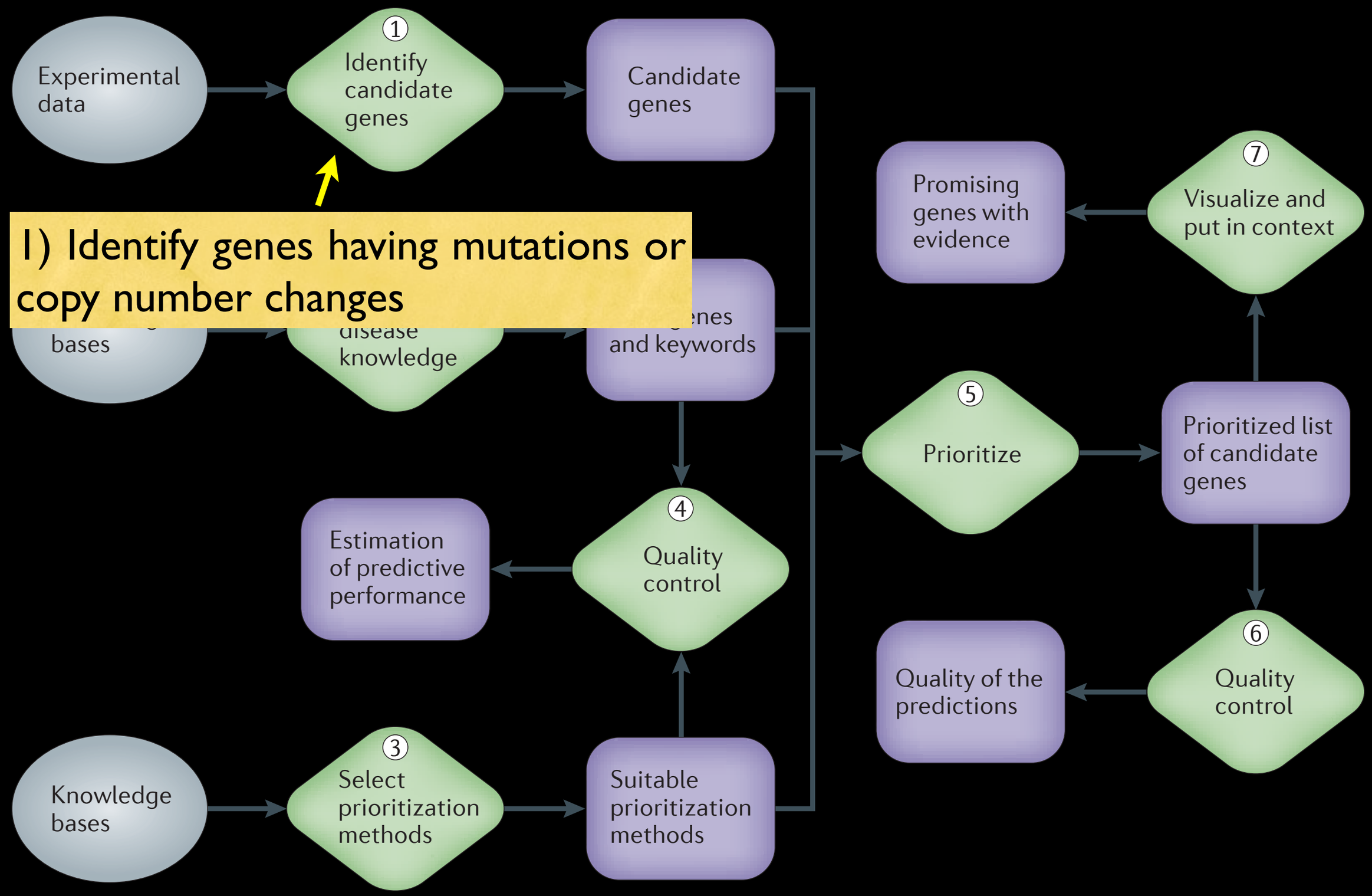
- **Background**
  - Various types of 'omics result in more information than we can easily handle
    - ✓ Gene ontology
    - ✓ Protein domain database
    - ✓ Literature
    - ✓ Sequence database
    - ✓ and etc
  - Leverage knowledge of each genomic resource to improve the ability for disease gene discovery

Workflow



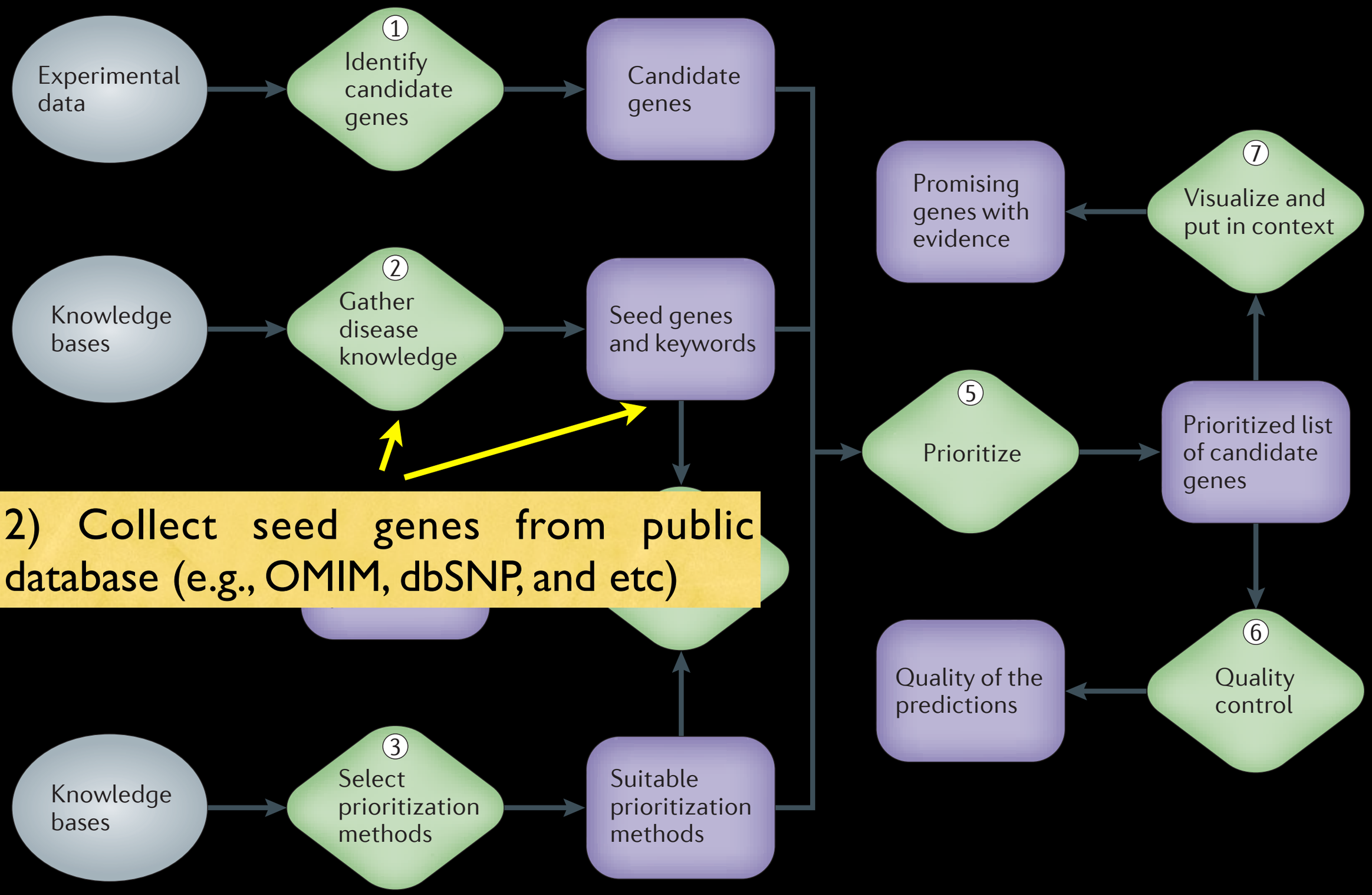
Y. Moreau and L. Tranchevent, Computational tools for prioritizing candidate genes: boosting disease gene discovery , Nat Rev Genet, 2012, dx.doi.org/10.1038/nrg3253

Workflow



Y. Moreau and L. Tranchevent, Computational tools for prioritizing candidate genes: boosting disease gene discovery , Nat Rev Genet, 2012, dx.doi.org/10.1038/nrg3253

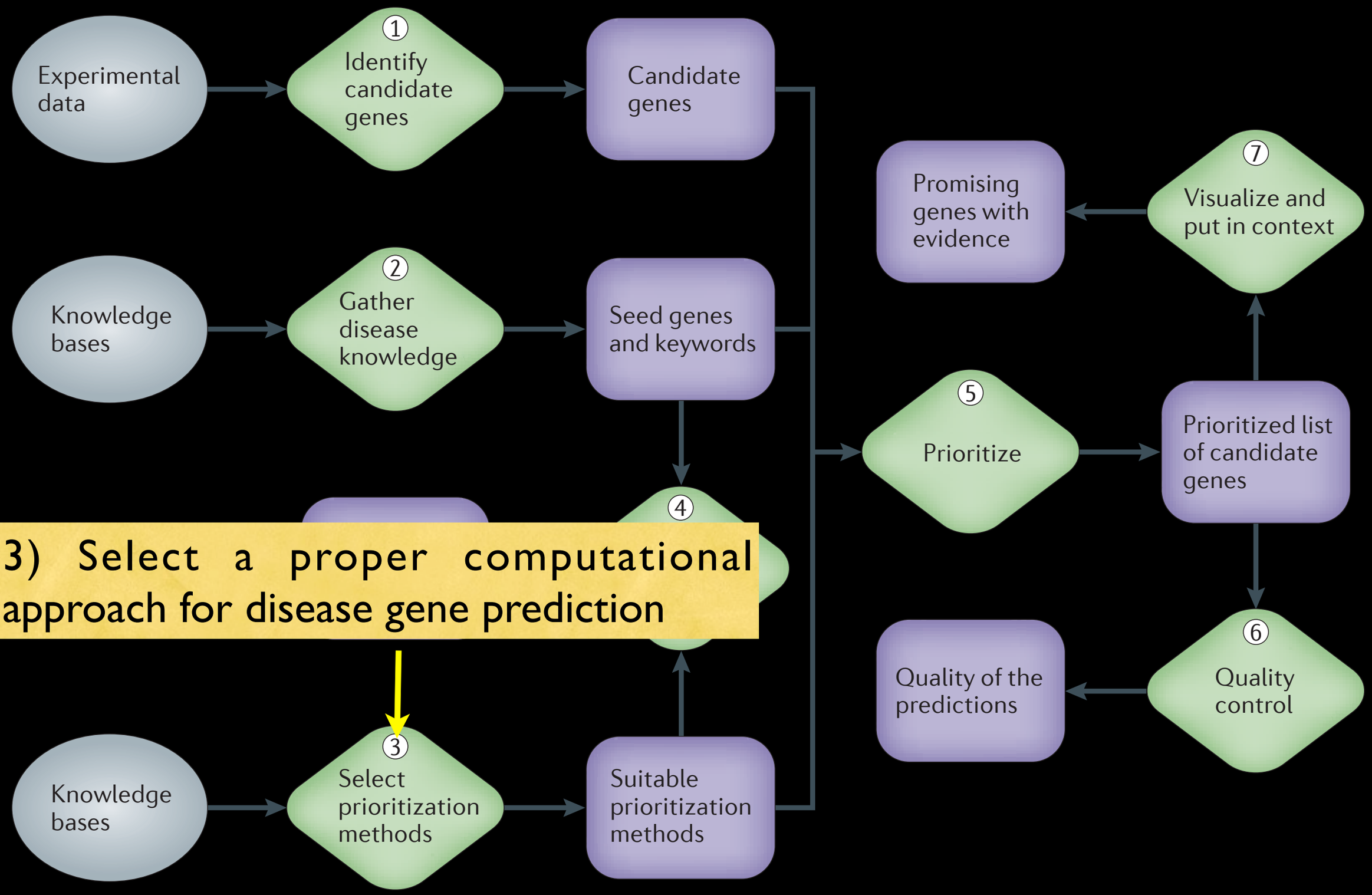
Workflow



2) Collect seed genes from public database (e.g., OMIM, dbSNP, and etc)

Y. Moreau and L. Tranchevent, Computational tools for prioritizing candidate genes: boosting disease gene discovery , Nat Rev Genet, 2012, dx.doi.org/10.1038/nrg3253

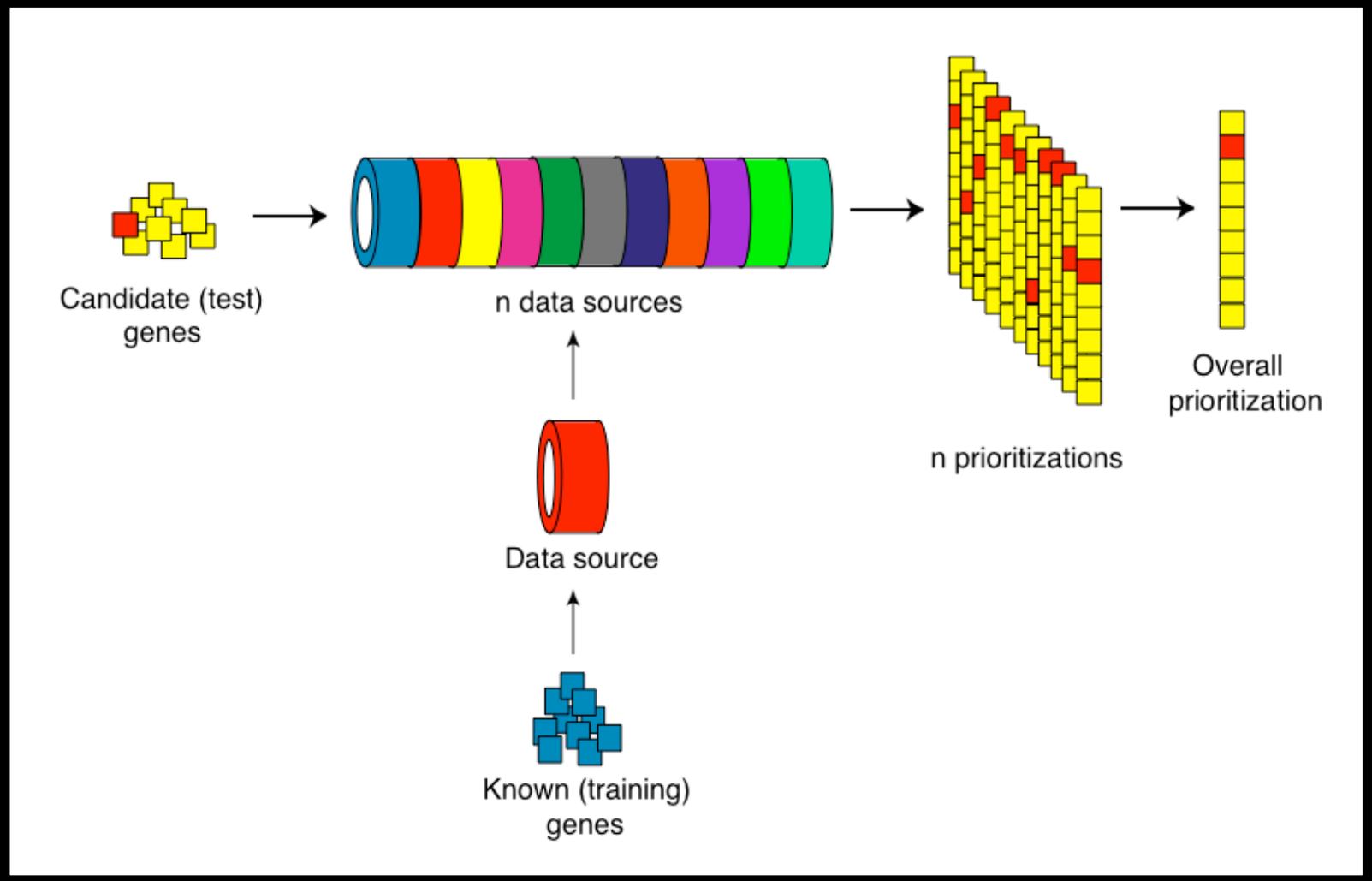
Workflow



Y. Moreau and L. Tranchevent, Computational tools for prioritizing candidate genes: boosting disease gene discovery , Nat Rev Genet, 2012, dx.doi.org/10.1038/nrg3253

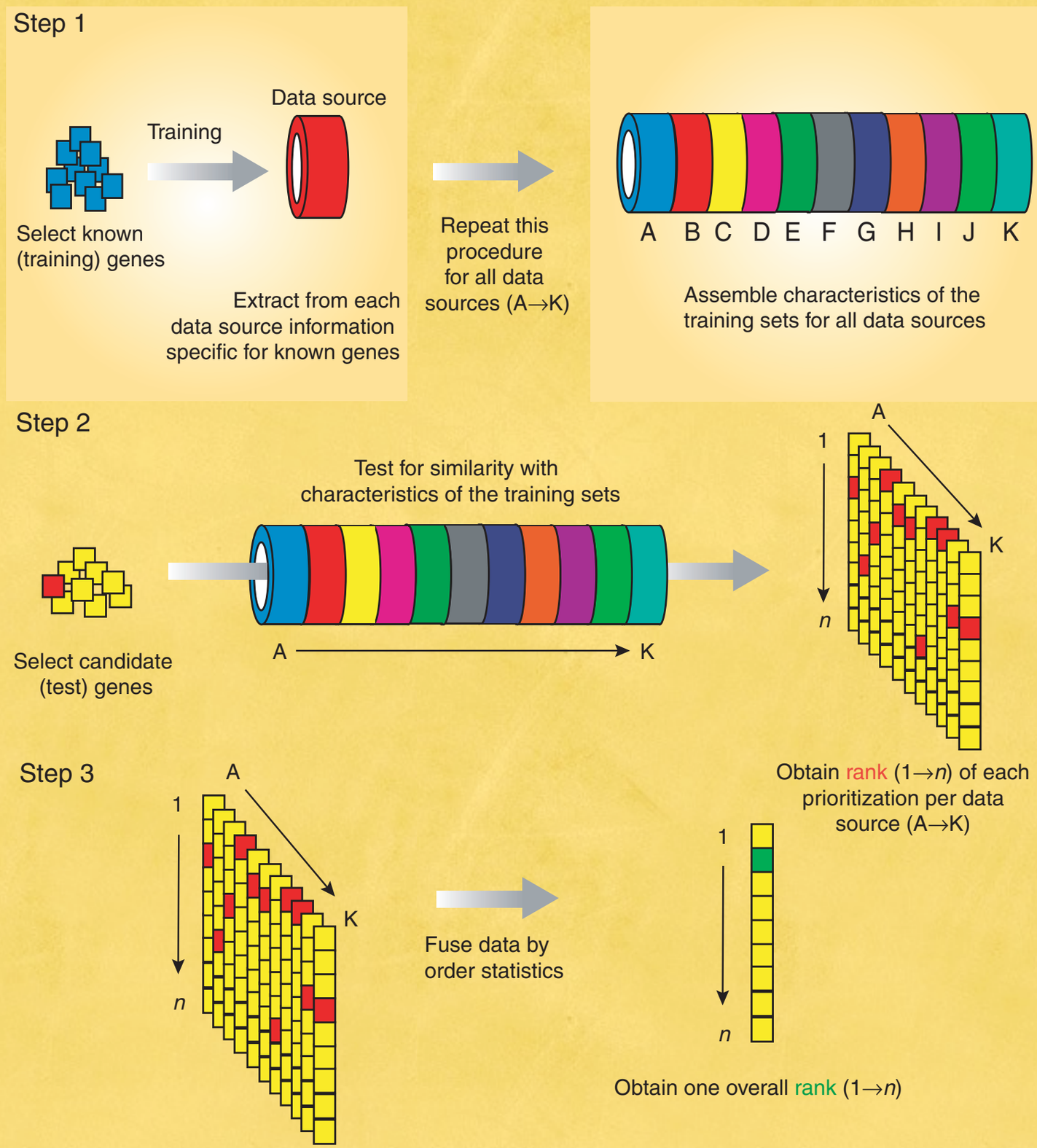
Data driven method

- **Endeavour**
  - Input: known genes (training), a set of candidate disease genes
  - Output: a list of ranked candidate genes



S. Aerts et al., Gene prioritization through genomic data fusion , Nat biotechnology 2006

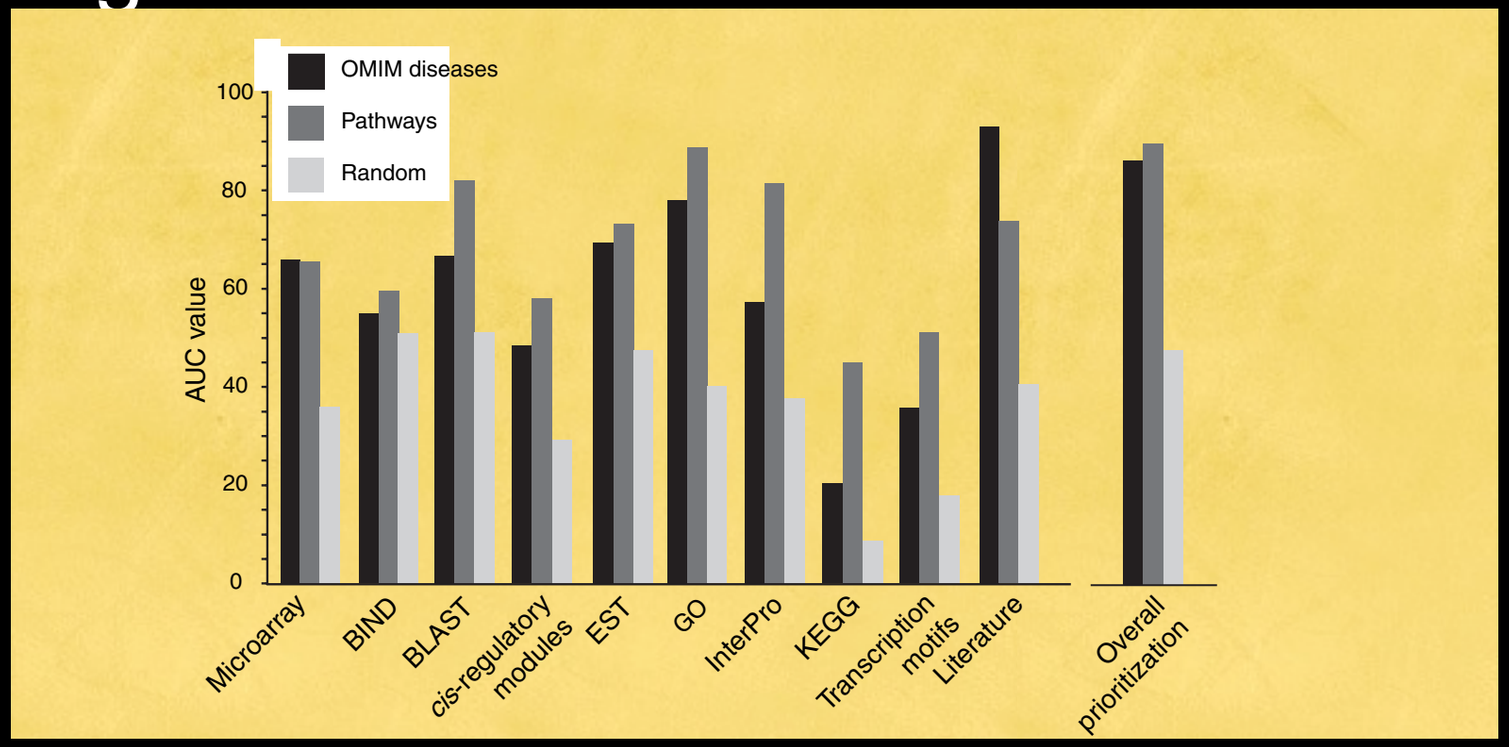
Endeavour workflow



S. Aerts et al., Gene prioritization through genomic data fusion , Nat biotechnology 2006

Data driven method

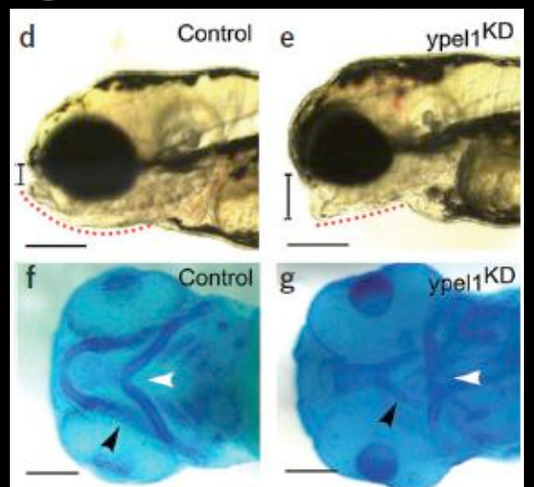
- **Validation**
  - Leave-one out cross validation
    - One gene is deleted from training genes (or known disease genes) as a test gene, and added to random test genes
    - Compare rankings of the test genes and random test genes



S. Aerts et al., Gene prioritization through genomic data fusion , Nat biotechnology 2006

Data driven method

- **Validation**
  - Leave-one out cross validation
    - One gene is deleted from training genes (or known disease genes) as a test gene, and added to random test genes
    - Compare rankings of the test genes and random test genes
  - Experimental validation
    - A knockdown of YPEL1 results in changes in the pharyngeal arches

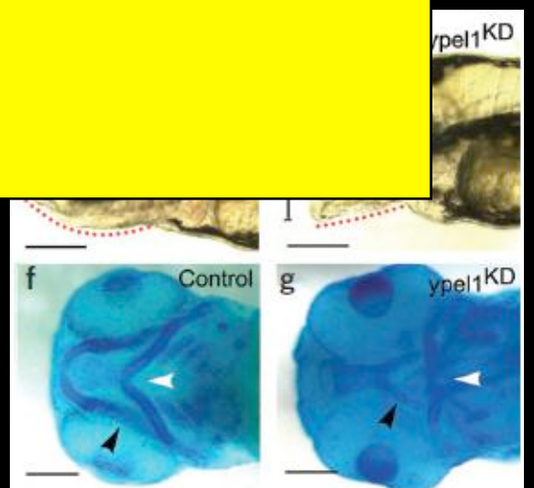


S. Aerts et al., Gene prioritization through genomic data fusion , Nat biotechnology 2006

Data driven method

- **Validation**
  - Leave-one out cross validation
    - One gene is deleted from training genes (or known disease genes) as a test gene, and added to random test genes
    - Compare rankings of the test genes and random test genes

✓Advantage: easy to perform analysis
✓Disadvantage:
1) not easy to find good training genes,
2) hard to interpret findings biologically



S. Aerts et al., Gene prioritization through genomic data fusion , Nat biotechnology 2006

Endeavour website

Endeavour Home Page

homes.esat.kuleuven.be/~bioiuser/endeavour/index.php

BIOInformatics

- Home
- Academic use
- Commercial use
- Help
- Publications
- Contact
- Links

Perform gene prioritization using our Web server

Perform gene prioritization using our Java server

Presentation

The identification of key genes involved in health and disease remains a formidable challenge. We develop novel bioinformatics to prioritize candidate genes underlying biological processes or diseases. Currently, our prioritization strategies are based on how similar a candidate gene is to a profile derived from genes already known to be involved in the processes. Data from multiple heterogeneous sources (coding sequence, gene expression, annotation, literature, regulatory information, etc.) are integrated, or fused, into a global ranking of the candidates. Ongoing research tackles the extension of these strategies to data from multiple organisms (cross-species data fusion) using more elegant machine learning strategies (kernel methods) and how to perform prioritization in the absence of a training set.

Endeavour is a software application for the computational prioritization of candidates genes, based on a set of training genes. It is made up of three stages: training, scoring and fusion. In the first stage, information about the training genes (genes already known to play a role in the process under study) are retrieved from numerous data sources in order to build models. It includes functionnal annotations, protein-protein interactions, regulatory information, expression data, sequence based data and literature mining data. In the second stage, the models are then used to score the candidate genes and to rank them according to their scores. Lastly, the rankings per data source are fused into a global ranking using order statistics. Endeavour is available for human, mouse, rat, fruit fly and worm.

Starting from a locus reported to be associated with DiGeorge syndrome and using Endeavour, we were able to propose YPEL1 as an interesting candidate. We further showed that YPEL1 knock-out zebrafish embryos exhibit features that are compatible with the human DiGeorge phenotypes. More recently, we have used Endeavour to optimize a genetic screen in *Drosophila melanogaster* in which we aimed at discovering novel *in vivo* interactions with the developmental gene *atonal*. Starting from 180 deficiency lines, we identified 12 positives loci harboring more than 1100 genes in total. These loci were prioritized using Endeavour and only the genes in the top 30% were assayed resulting in the identification of 12 positive genes. Researchers have also used Endeavour to look for genes involved in cleft lip / cleft palate from aCGH data, and to analyze the proteome of adipocytes. Please browse our reference section to find a list of Endeavour related publications.

Data

Data from multiple heterogeneous sources are collected and integrated in our databases in order to perform gene prioritization. This includes sequence data (genomic sequences of the genes and protein sequences of their products), expression data (usually EST data or large data sets covering the expression of thousands of genes over a wide range of different tissues/samples), functional annotations (usually from ontologies designed to describe the function of the gene products, their cellular localization, and the biomolecular pathways they are involved in), protein-protein interaction networks (describing which products interact with which other products either physically or mation contained in the scientific literature, it

Softwares

We have implemented the basic algorithm into an application termed Endeavour. It is a Java based client that can be started via Java Web Start. More recently, we have developed a web version that is more user friendly. However, it does not include all the options available in the Java client. Both tools are using the same core and thus give exactly the same results when running the same prioritization. The development of the kernel based application (with an improved performance) is on its way and should be made available during fall this year.

Research

Waiting for homes.esat.kuleuven.be...

Open Endeavour website

Endeavour Home Page

homes.esat.kuleuven.be/~bioiuser/endeavour/index.php

BIOInformatics

- Home
- Academic use
- Commercial use
- Help
- Publications
- Contact
- Links

Perform gene prioritization using our Web server

Perform gene prioritization using our Java server

click

Presentation

The identification of key genes involved in health and disease remains a formidable challenge. We develop novel bioinformatics to prioritize candidate genes underlying biological processes or diseases. Currently, our prioritization strategies are based on how similar a candidate gene is to a profile derived from genes already known to be involved in the processes. Data from multiple heterogeneous sources (coding sequence, gene expression, annotation, literature, regulatory information, etc.) are integrated, or fused, into a global ranking of the candidates. Ongoing research tackles the extension of these strategies to data from multiple organisms (cross-species data fusion) using more elegant machine learning strategies (kernel methods) and how to perform prioritization in the absence of a training set.

Endeavour is a software application for the computational prioritization of candidates genes, based on a set of training genes. It is made up of three stages: training, scoring and fusion. In the first stage, information about the training genes (genes already known to play a role in the process under study) are retrieved from numerous data sources in order to build models. It includes functionnal annotations, protein-protein interactions, regulatory information, expression data, sequence based data and literature mining data. In the second stage, the models are then used to score the candidate genes and to rank them according to their scores. Lastly, the rankings per data source are fused into a global ranking using order statistics. Endeavour is available for human, mouse, rat, fruit fly and worm.

Starting from a locus reported to be associated with DiGeorge syndrome and using Endeavour, we were able to propose YPEL1 as an interesting candidate. We further showed that YPEL1 knock-out zebrafish embryos exhibit features that are compatible with the human DiGeorge phenotypes. More recently, we have used Endeavour to optimize a genetic screen in *Drosophila melanogaster* in which we aimed at discovering novel *in vivo* interactions with the developmental gene *atonal*. Starting from 180 deficiency lines, we identified 12 positives loci harboring more than 1100 genes in total. These loci were prioritized using Endeavour and only the genes in the top 30% were assayed resulting in the identification of 12 positive genes. Researchers have also used Endeavour to look for genes involved in cleft lip / cleft palate from aCGH data, and to analyze the proteome of adipocytes. Please browse our reference section to find a list of Endeavour related publications.

Data

Data from multiple heterogeneous sources are collected and integrated in our databases in order to perform gene prioritization. This includes sequence data (genomic sequences of the genes and protein sequences of their products), expression data (usually EST data or large data sets covering the expression of thousands of genes over a wide range of different tissues/samples), functional annotations (usually from ontologies designed to describe the function of the gene products, their cellular localization, and the biomolecular pathways they are involved in), protein-protein interaction networks (describing which products interact with which other products either physically or

mation contained in the scientific literature, it

Softwares

We have implemented the basic algorithm into an application termed Endeavour. It is a Java based client that can be started via Java Web Start. More recently, we have developed a web version that is more user friendly. However, it does not include all the options available in the Java client. Both tools are using the same core and thus give exactly the same results when running the same prioritization. The development of the kernel based application (with an improved performance) is on its way and should be made available during fall this year.

Research

Waiting for homes.esat.kuleuven.be...

Open Endeavour website

Endeavour

Candidate genes prioritization through genomic data fusion

You are here on the web client version of Endeavour. For an introduction, latest news, academic or commercial use, contact info, mailing list, please refer to the Endeavour project main page.

How to cite

- **ENDEAVOUR update: a web resource for gene prioritization in multiple species.** Tranchevent L., Barriot R., Yu S., Van Vooren S., Van Loo P., Coessens B., Aerts S., De Moor B., Moreau Y. *Nucleic Acids Research*, Web Server issue, vol. 36, no. 1, Jun. 2008, pp. 377-384. *Abstract*
- **Gene prioritization through genomic data fusion.** Aerts S., Lambrechts D., Maity S., Van Loo P., Coessens B., De Smet F., Tranchevent L-C., De Moor B., Marynen P., Hassan B., Carmeliet P. & Moreau Y. *Nature Biotechnology*. [2006 May;24(5):537-544. PMID: 16680138] *Abstract*

Prioritize candidate genes

A **manual** is available here. If this is your first visit, you may try out some examples:

- YPEL1 taken from our Nature biotech. paper on DiGeorge syndrome candidate genes.
- KCNJ5 taken from the Elbers et al review on obesity and diabetes.
- DFNB31 based on the Ebermann et al paper describing the discovery of a novel Usher gene.

Clicking on the gene name (YPEL1, KCNJ5 or DFNB31) will cause the training and the candidate genes to be loaded in the following wizard; this will also select the most appropriate data sources. Then you should go over the different steps ('Next' button, enabled once the training and candidate genes are loaded) to revise the settings and launch the prioritization.

Prioritize your candidates in 4 steps with the following wizard

1. Species | 2. Training genes | 3. Data sources used to build models | 4. Candidates

Candidate genes to prioritize (3 genes)

Gene (Reference ID)	Alias	Description
ENSG00000149311	ATM	Serine-protein kinase ATM (EC 2.7.11.1) (Ataxia telangiectasia mutated) (A-T, mutated). [Source:Uniprot/SWISSPROT;Acc:Q13315]
ENSG00000136997	MYC	Myc proto-oncogene protein (c-Myc) (Transcription factor p64). [Source:Uniprot/SWISSPROT;Acc:P01106]
ENSG00000141736	ERBB2	Receptor tyrosine-protein kinase erbB-2 precursor (EC 2.7.10.1) (p185erbB2) (C-erbB-2) (NEU proto-oncogene) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19) (CD340 antigen). [Source:Uniprot/SWISSPROT;Acc:P04626]

Add following candidates

ATM
MYC
ERBB2

Add

Full genome

« Previous | Launch prioritization

Ready

Useful resource

http://homes.esat.kuleuven.be/~bioiuser/gpp/index.php

Gene Prioritization Portal | www.cs.tau.ac.il/~roded/pri | How to Take a Screenshot in

homes.esat.kuleuven.be/~bioiuser/gpp/index.php

Gene Prioritization Portal

Home | Tools | Stats | Search | Contact

Gene prioritization

One of the major challenge in human genetics is to find the genetic cause underlying disorders in order to unravel the molecular basis of these diseases and eventually elaborate medical treatments. High-throughput technologies such as linkage analysis and association studies are usually able to associate a chromosomal region with a genetic condition. One can also use expression arrays and obtain a list of transcripts differentially expressed in a disease sample with respect to a reference sample. A common characteristic of these methods is usually the large number of genes returned: for instance hundreds of differentially expressed genes are often reported. The working hypothesis is often that only one or a few candidates are really of primary interest. Identifying the most promising candidates among such large lists of genes is a challenging and time consuming task.The bioinformatics community has therefore introduced the concept of gene prioritization that uses computational methods and genomic data to identify the most promising candidates.

The gene prioritization portal

In the last decade, many different strategies have been developed, some of which have been implemented into web applications and eventually experimentally validated. This website represents an regularly updated electronic review in that field. The aim is to help biologists who need to select the most promising genes from large gene lists by describing what are the current available options. We describe more than 20 strategies (see the tool section) and give more general information to support user decision (see the statistics section).

Reference

Tranchevent LC., Bonachela Capdevila F., Nitsch D., De Moor B., De Causmaecker P., Moreau Y. **A guide to web tools to prioritize candidate genes.** *Brief. Bioinform.* (2010)
link to paper

Summary:

33 tools
12 data sources
82 publications
22 validations

A random tool:

PGMapper
More details

Useful resource

Nature Reviews Genetics | AOP, published online 3 July 2012; doi:10.1038/nrg3253

REVIEWS

Computational tools for prioritizing candidate genes: boosting disease gene discovery

Yves Moreau and Léon-Charles Tranchevent

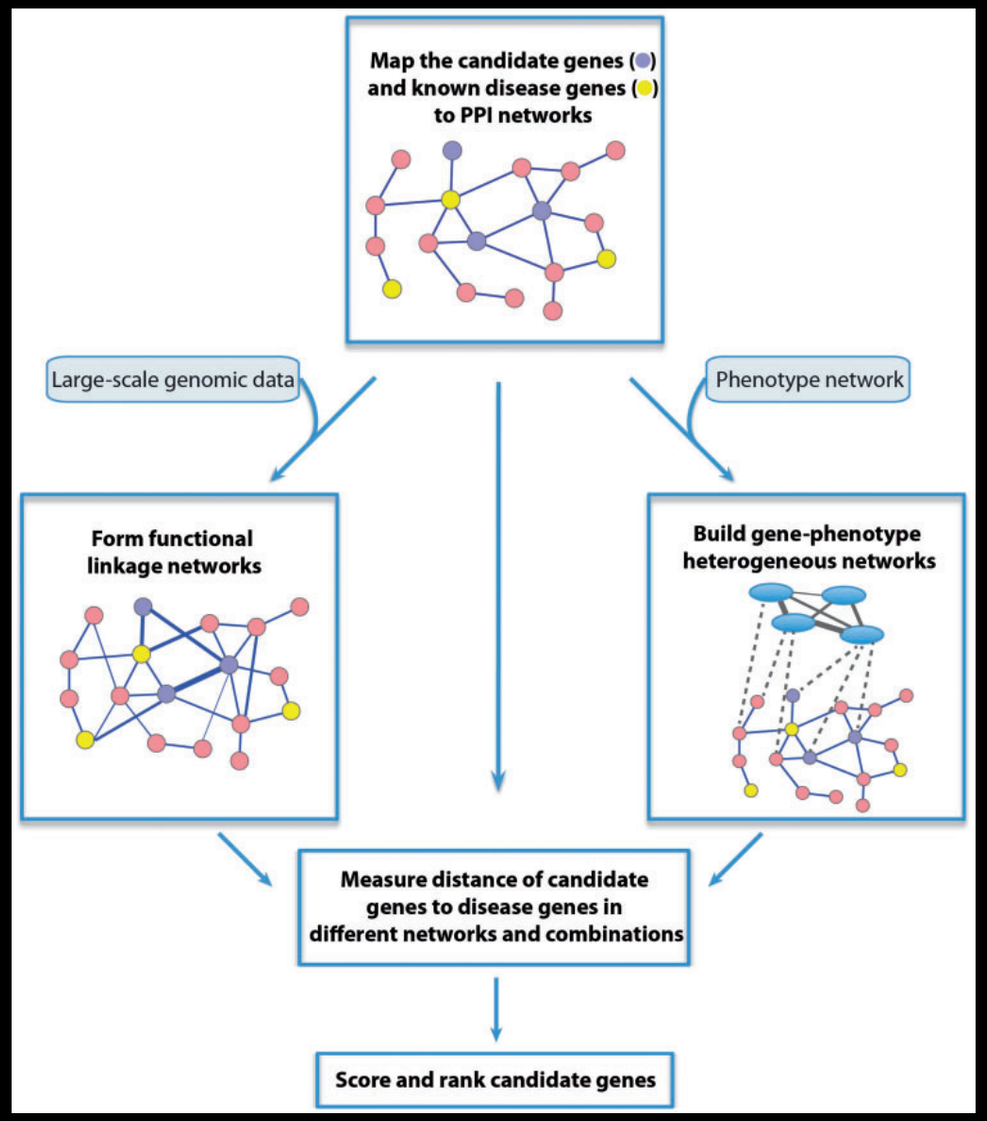
Abstract | At different stages of any research project, molecular biologists need to choose — often somewhat arbitrarily, even after careful statistical data analysis — which genes or proteins to investigate further experimentally and which to leave out because of limited resources. Computational methods that integrate complex, heterogeneous data sets — such as expression data, sequence information, functional annotation and the biomedical literature — allow prioritizing genes for future study in a more informed way. Such methods can substantially increase the yield of downstream studies and are becoming invaluable to researchers.

Network based method

- **Background**
  - Genes/proteins interact with each other in the network tend to have similar biological processes and functions
  - Identification of subnetworks containing a set of disease genes with novel candidate disease genes could help to improve the ability of disease gene discovery

Network based method

- **Two approaches**
  - Molecular networks
  - Integrated networks



Wang et al., Network-based methods for human disease gene discovery, Briefings in Functional Genomics and Proteomics 2011

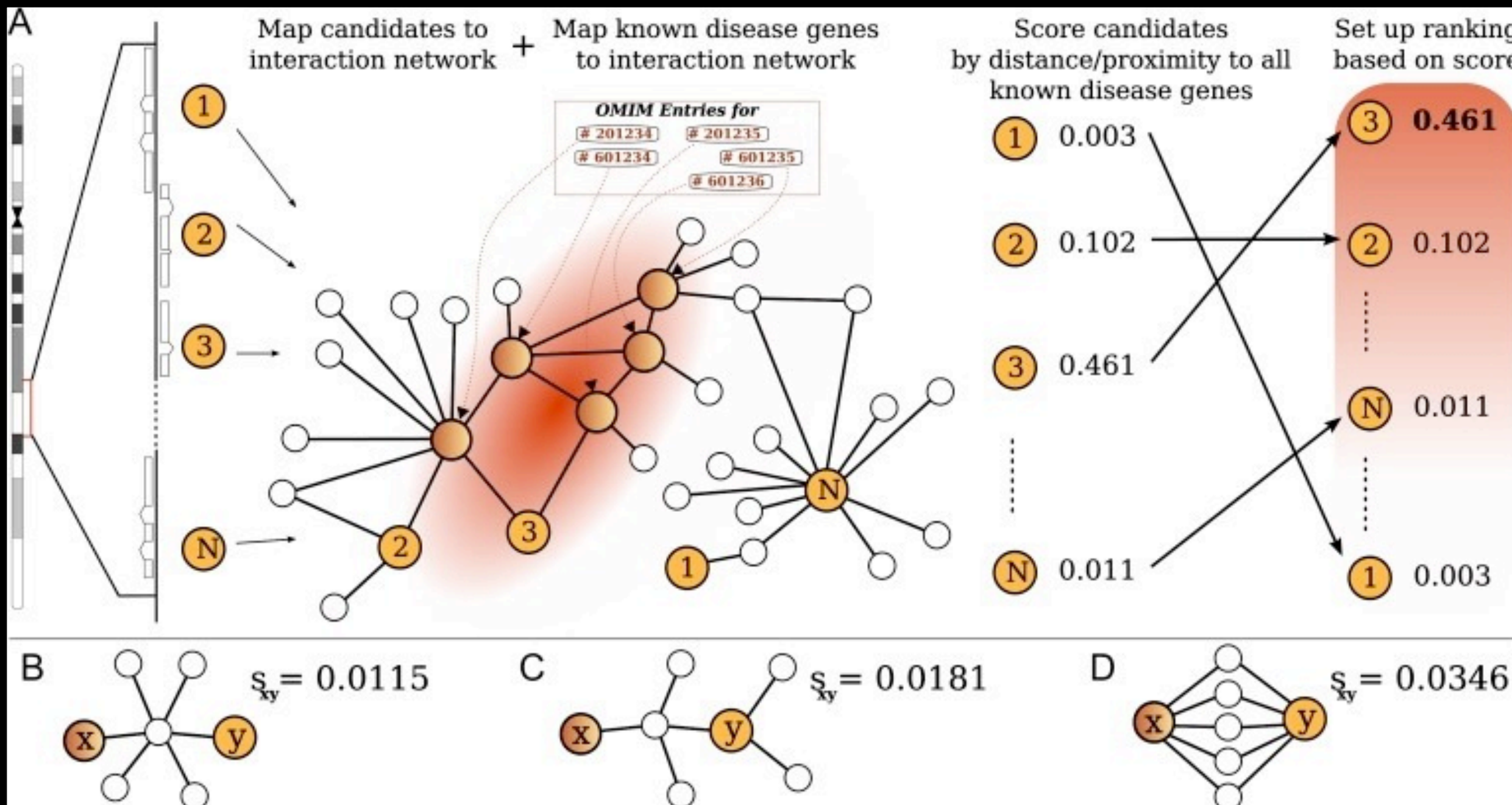
Network based method

- **Network-based methods for the use of molecular interaction networks**
  - Input: known genes (training), a set of candidate loci, molecular network
  - Output: a list of ranked candidate genes

Koller et al., Walking the interactome for prioritization of candidate disease genes, Am J Hum Genet. 2008 Apr;82(4):949-58. Epub 2008 Mar 27.

Network based method

- **Network-based methods for the use of molecular interaction networks**



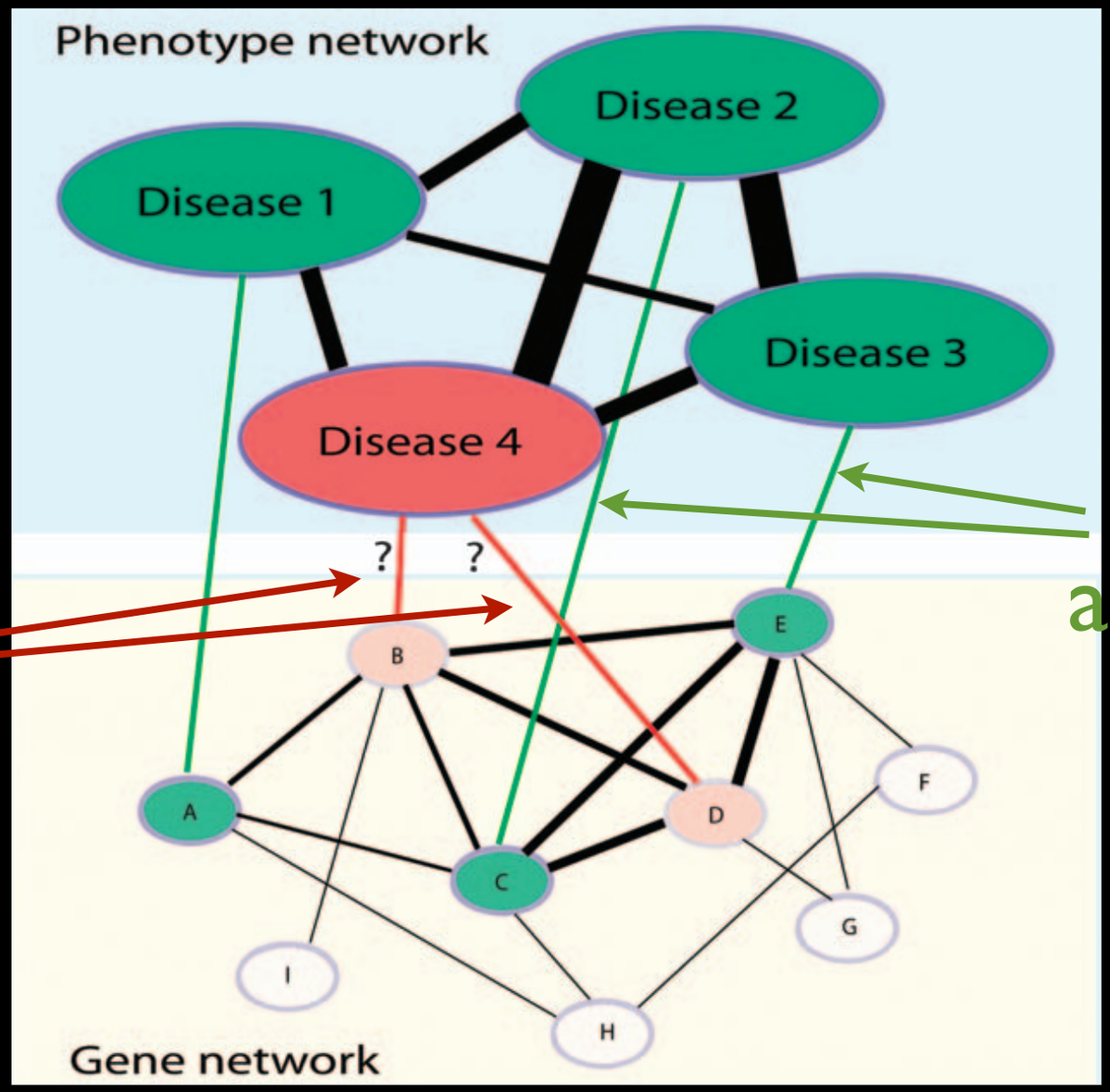
Koller et al., Walking the interactome for prioritization of candidate disease genes, Am J Hum Genet. 2008 Apr;82(4):949-58. Epub 2008 Mar 27.

Network based method

- **Network-based methods for the use of integrated networks (e.g., disease phenotype similarity networks, disease-gene association networks, gene-gene interaction networks)**
  - Input: a query disease phenotype
  - Output: a list of ranked candidate genes

Motivation

- **Modular view of disease and gene networks**
  - Phenotypically similar diseases are caused by functionally related genes



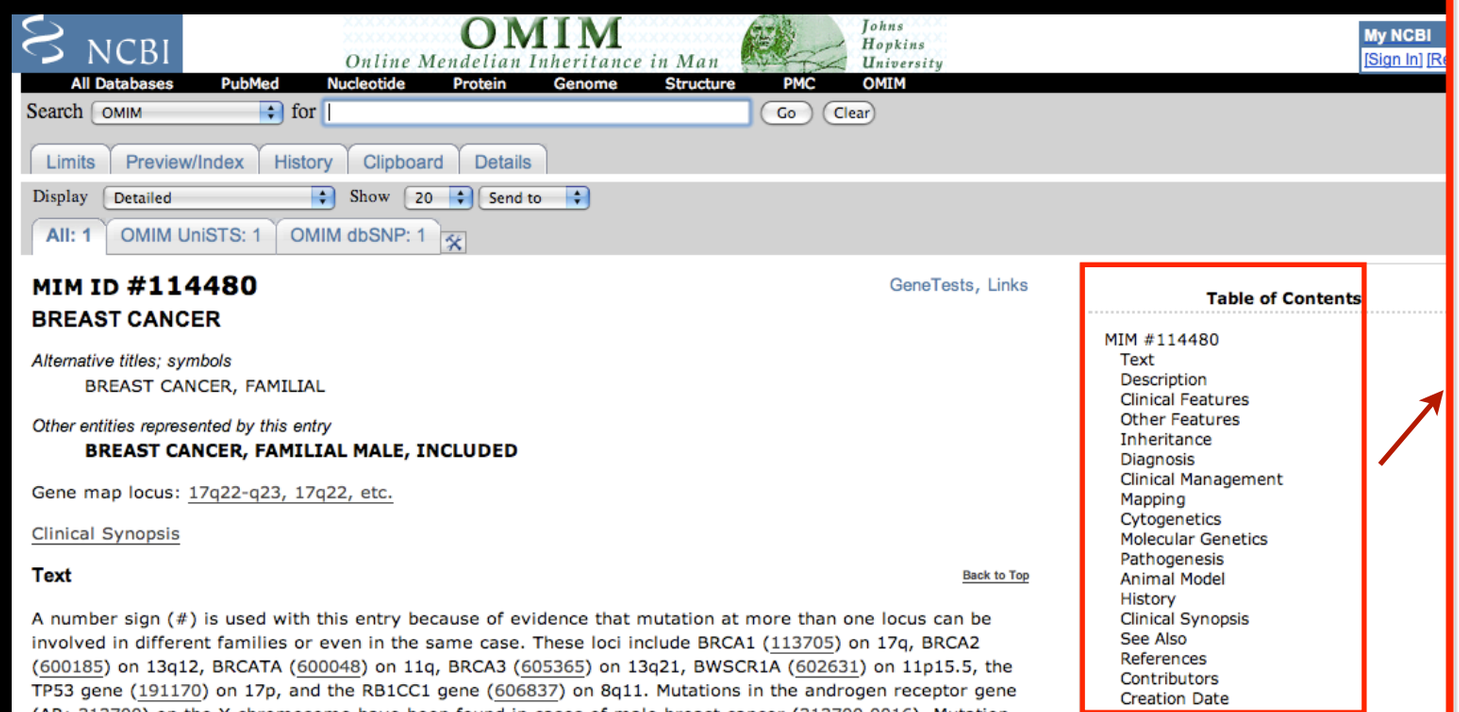
Missing association

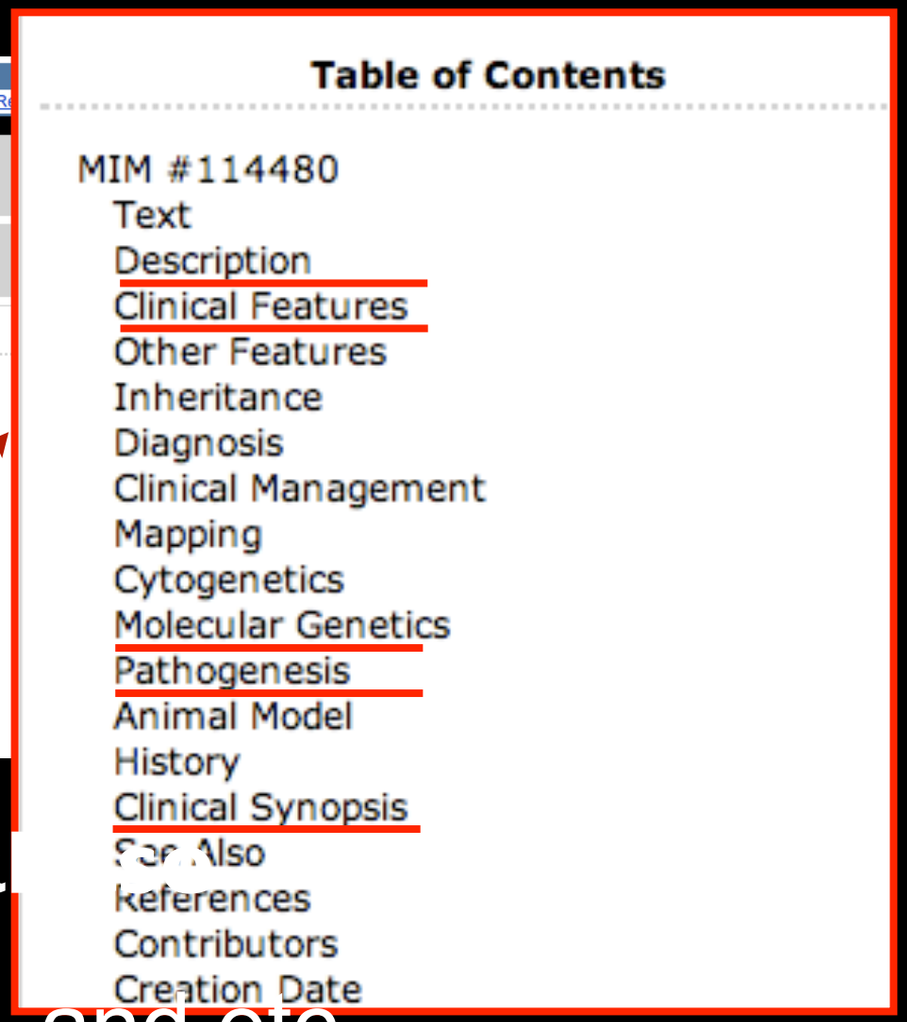
Known association

Oti M, Brunner HG, The modular nature of genetic disease, Clinical Genetics 2007

Public database

- **Disease phenotype database**
  - Online Mendelian Inheritance in Man (OMIM)

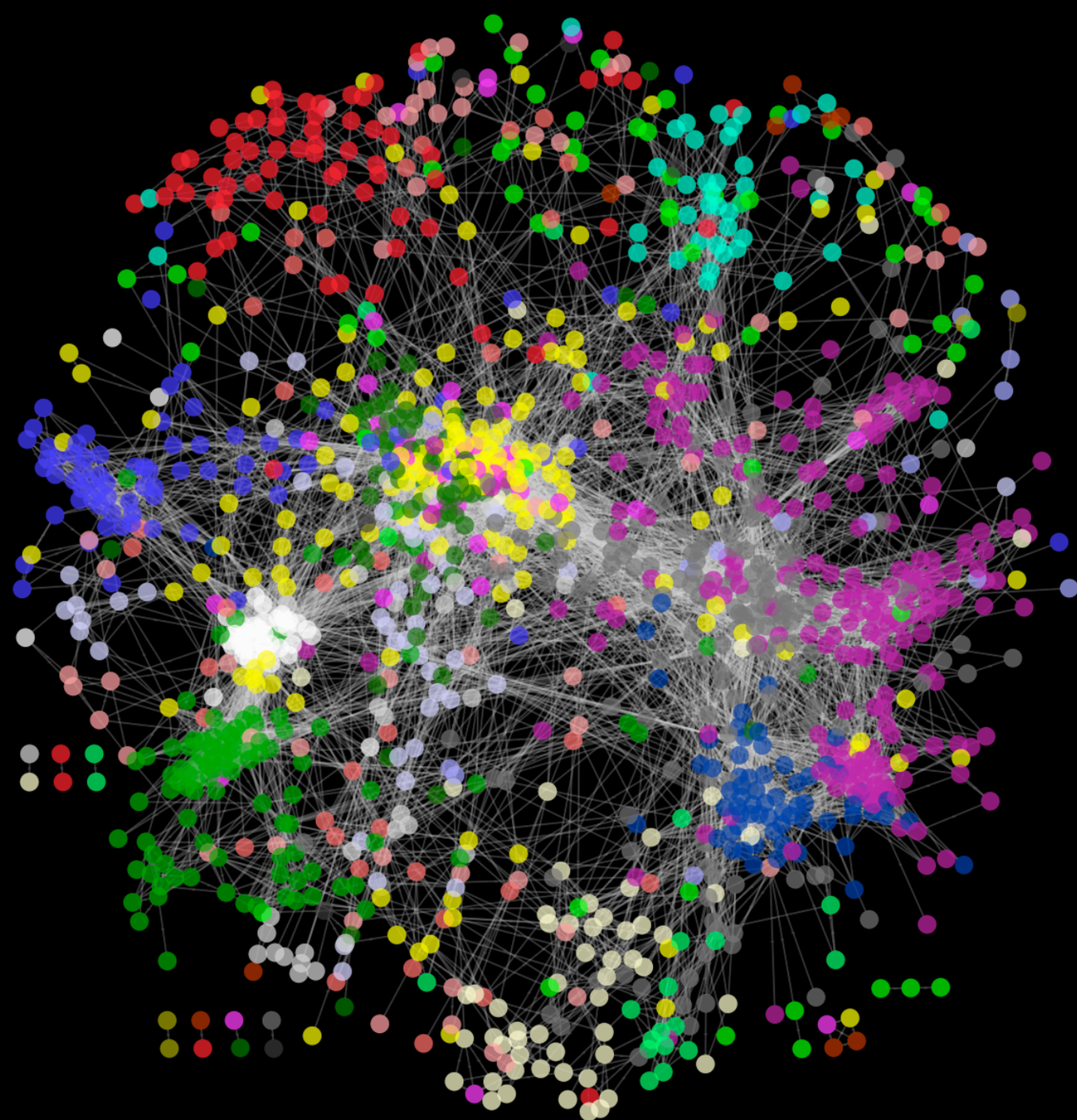


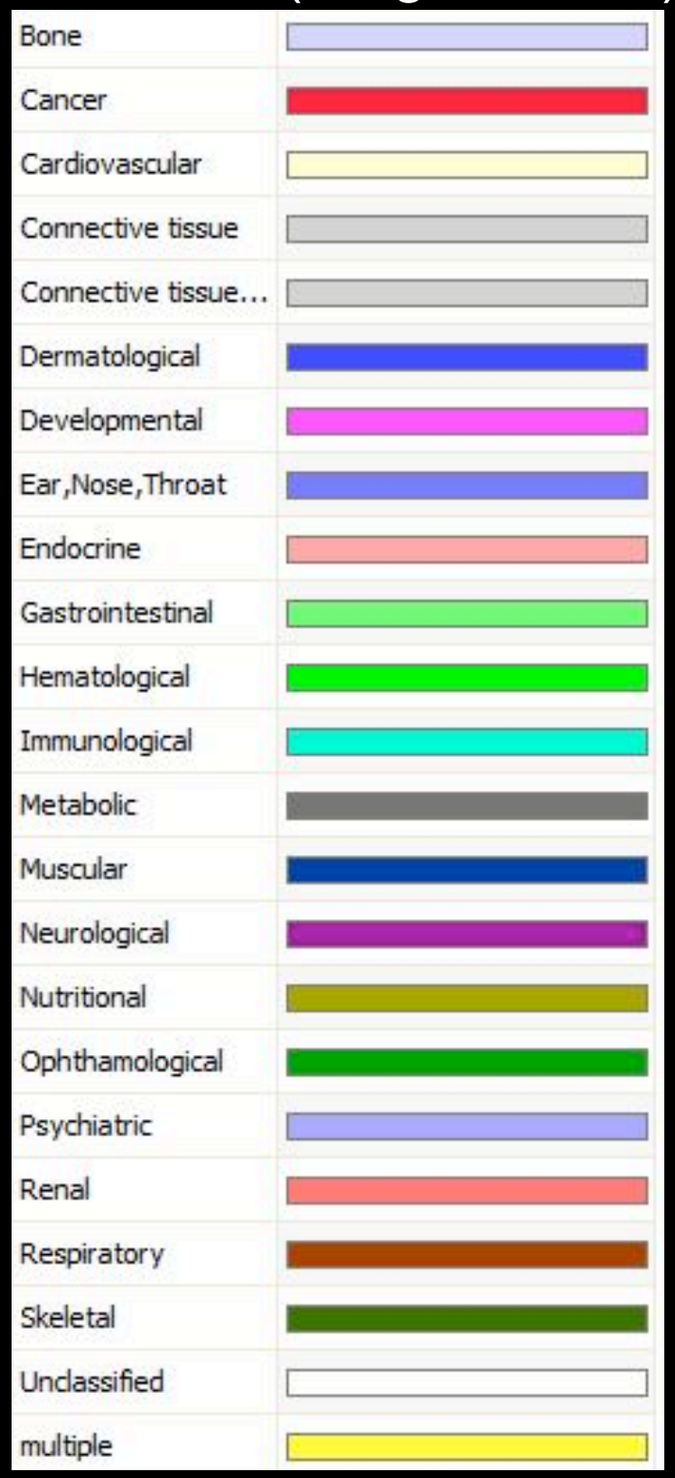


- **Disease-gene association data**
  - OMIM, dbSNP, GWAS, literature, and etc.
- **Gene-gene interaction networks**
  - Protein-interaction, co-expression, and etc.,

Disease network

- Node: disease phenotype in OMIM
- Edge: phenotypical similarity calculated by text mining* with OMIM database (weights > 0.4)



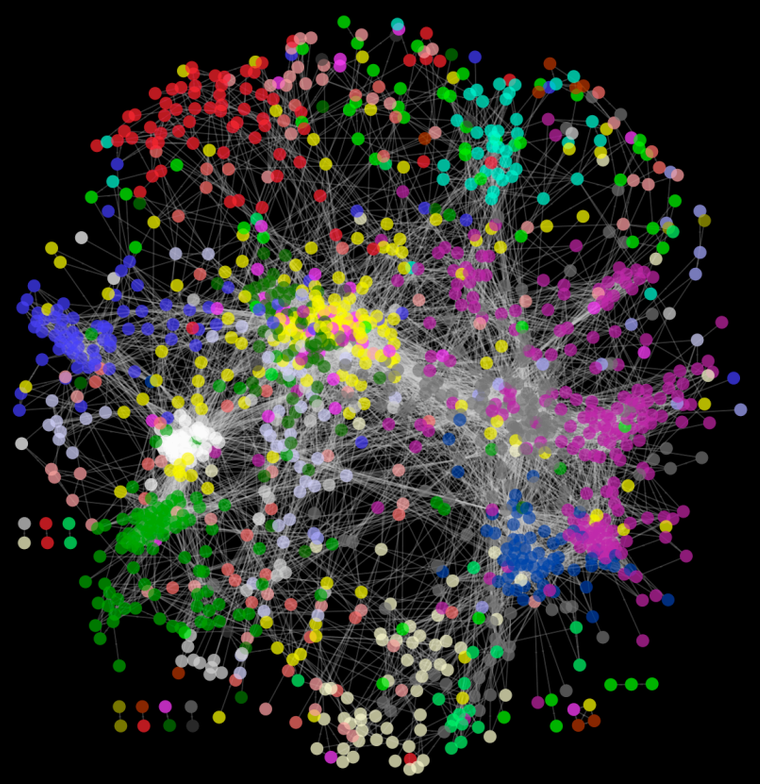


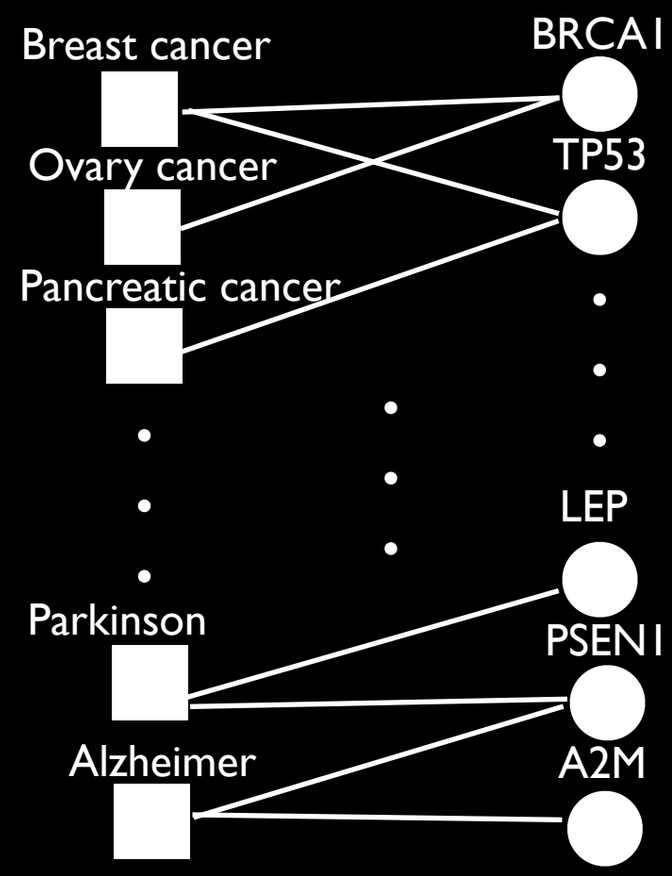
Disease class annotation from Goh et. al, PNAS 2007

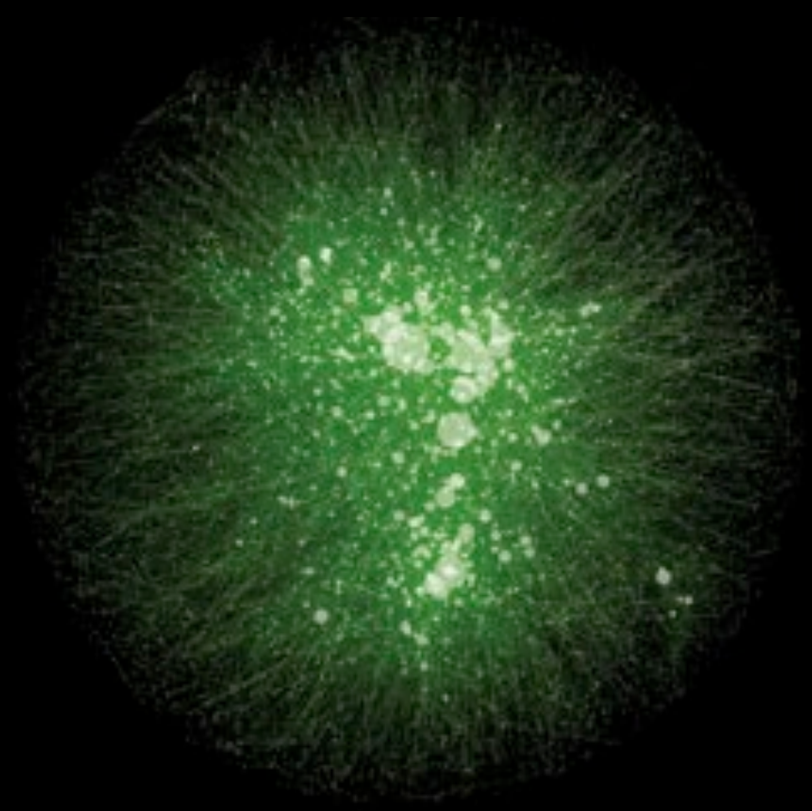
*Marc Driel, et al. "A text-mining analysis of the human phenome", European Journal of Human Genetics 2006

Challenge

"One challenge computationally is integrating heterogeneous data sets to build a network model" - Ilya Shmulevich, Institute for System Biology, Nature 2010







Disease network

text mining from OMIM
comordity from patients records
microarray gene expression

Disease-gene association network

OMIM
dbGaP
GWAS

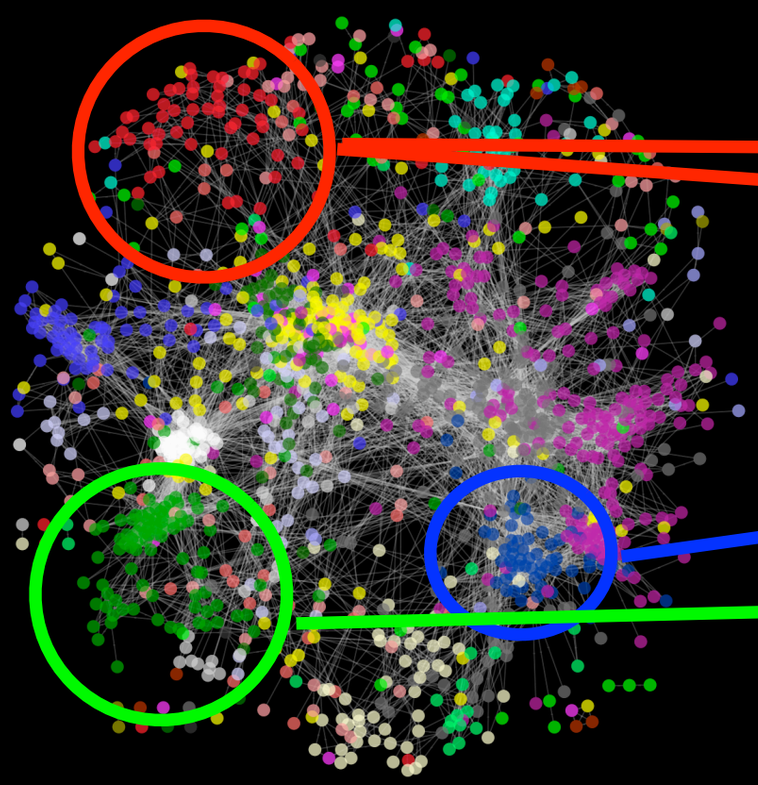
Gene interaction network

Protein interaction network
Co-expression
Genetic interaction network

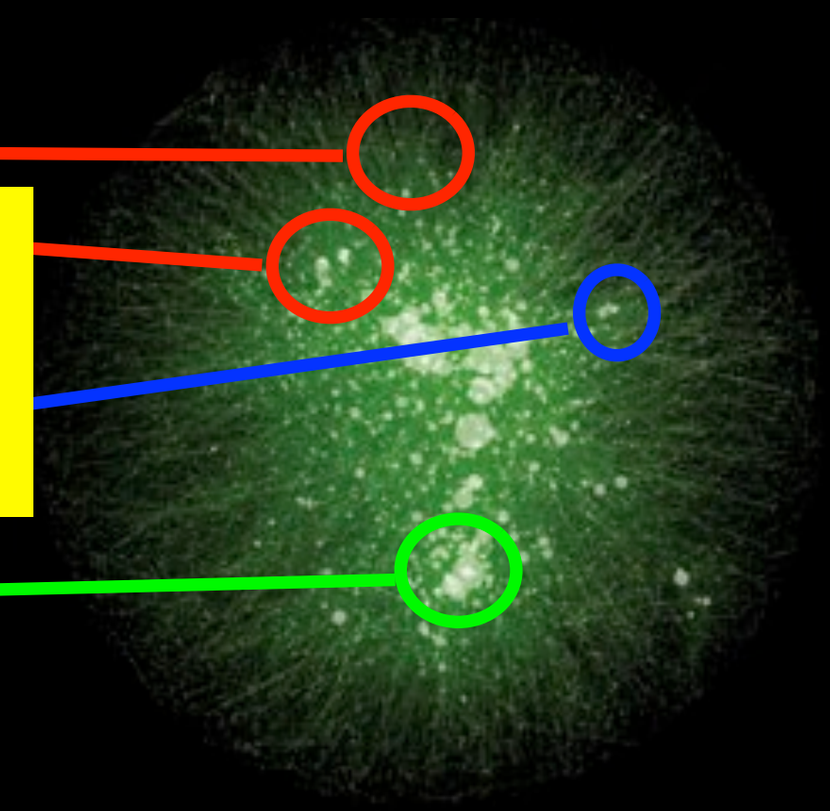
Challenge

- **Generalizability and scalability**
  - Efficient optimization
- **Provable theoretical guarantees**
  - Consistency, convergence rate, etc
- **Interpretability**
  - Biologically interpretable

?



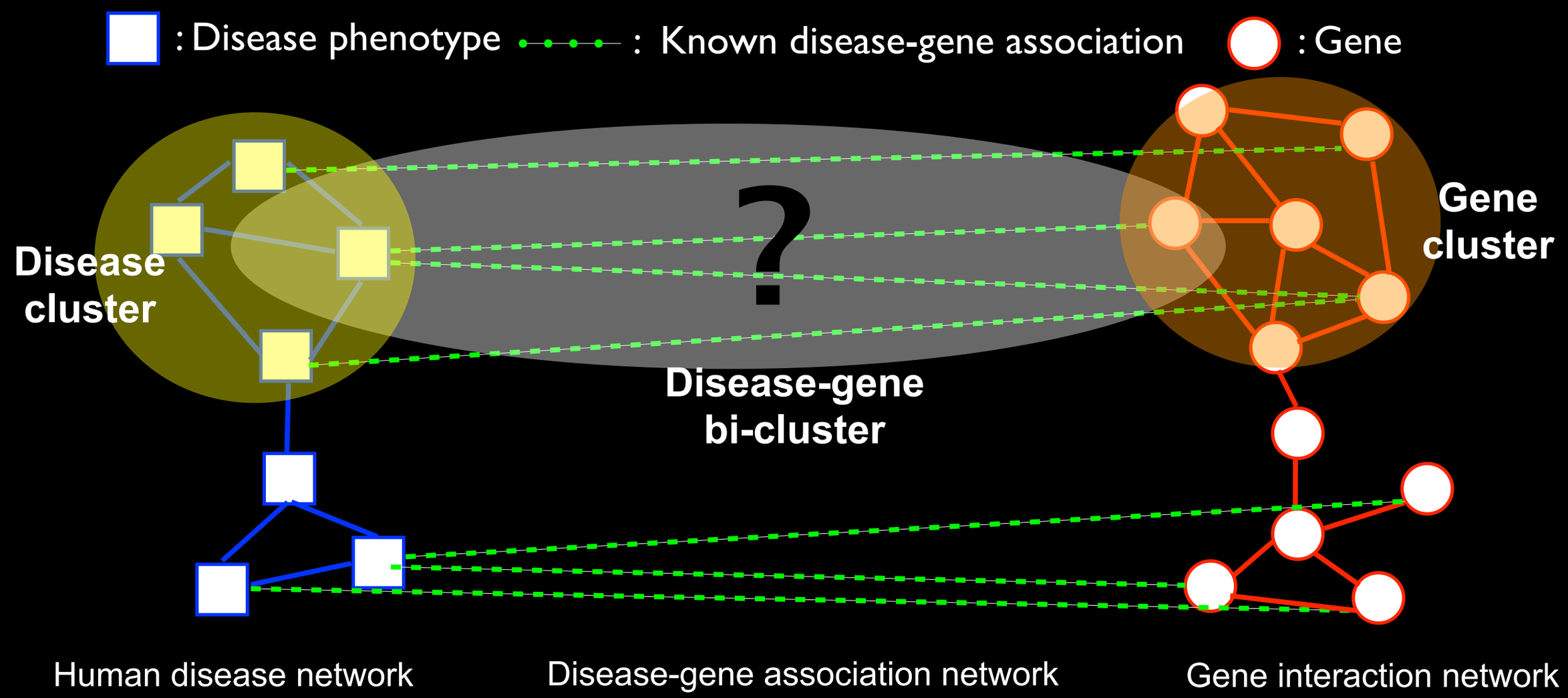
How to
find missing associations
between modules
in **networks**



Disease network	Disease-gene association network	Gene interaction network
text mining from OMIM	OMIM	Protein interaction network
comordity from patients records	dbGaP	Co-expression
microarray gene expression	GWAS	Genetic interaction network

Data integration

- Different network data could be combined as an integrated heterogeneous network
- Exploring cluster structures in each network independently



T. Hwang, and Rui Kuang, “A heterogeneous label propagation for disease gene discovery”. **SDM** 2010

Problem formulation

- Given: an integrated heterogeneous network and a query disease phenotype
- Task: predict candidate disease causative genes of the query disease phenotype
  - Input: initial activation values on the query node
  - Output: a ranked gene list based on final activation values

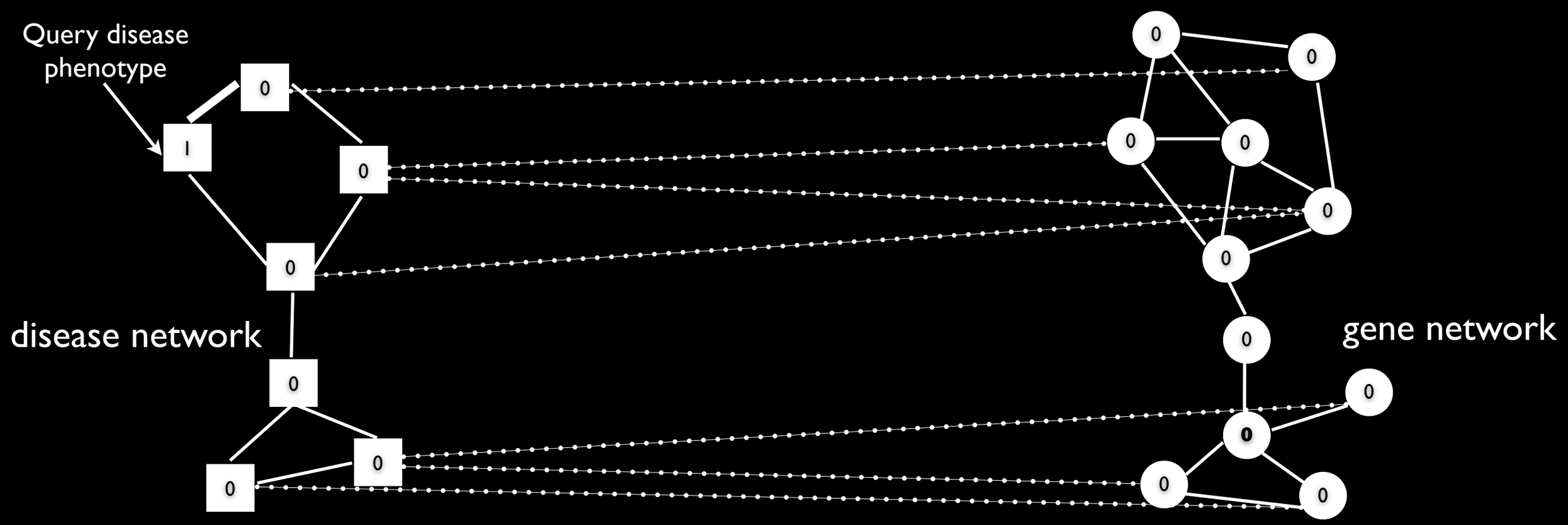
Q: Find candidate disease genes associated with a query disease phenotype

T. Hwang, and Rui Kuang, "A heterogeneous label propagation for disease gene discovery". **SDM** 2010

Working example (1/5)

- Q: Find candidate disease genes associated with a query disease phenotype

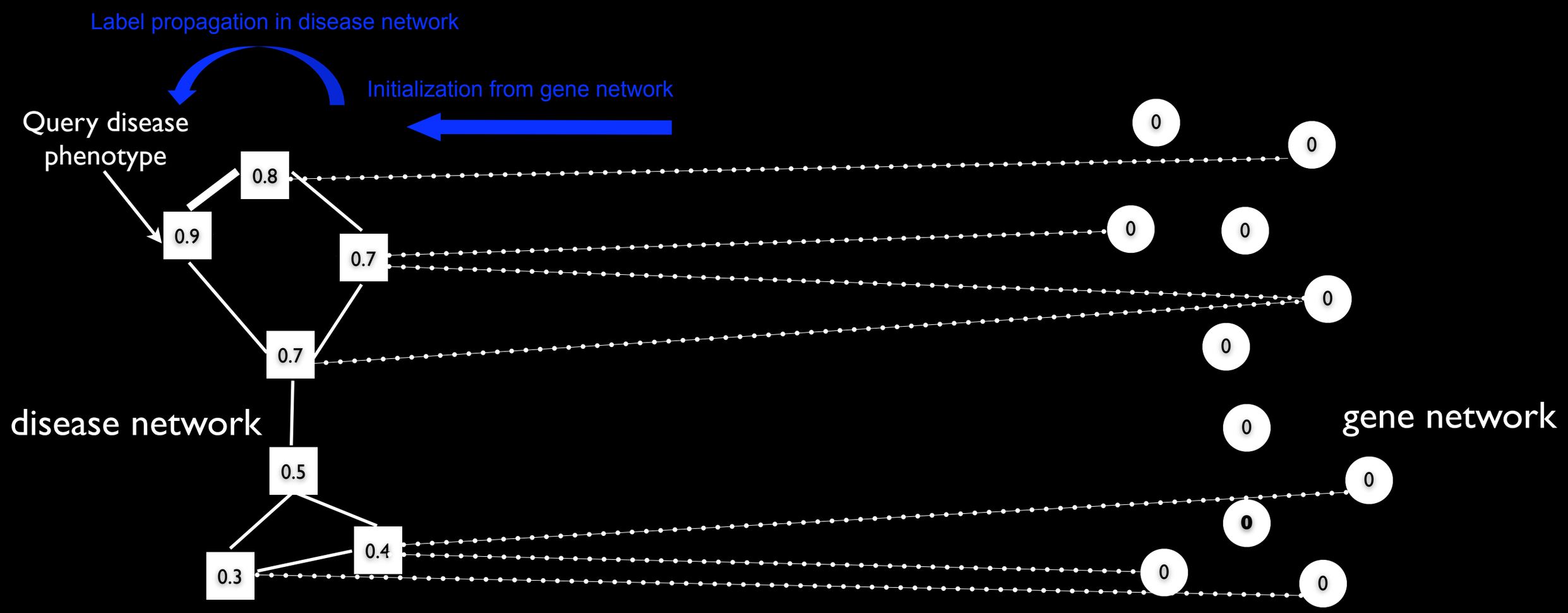
1. Initialize activation values on nodes (i.e. query node: 1 and others: 0)



Working example (2/5)

- Q: Find candidate disease genes associated with a query disease phenotype

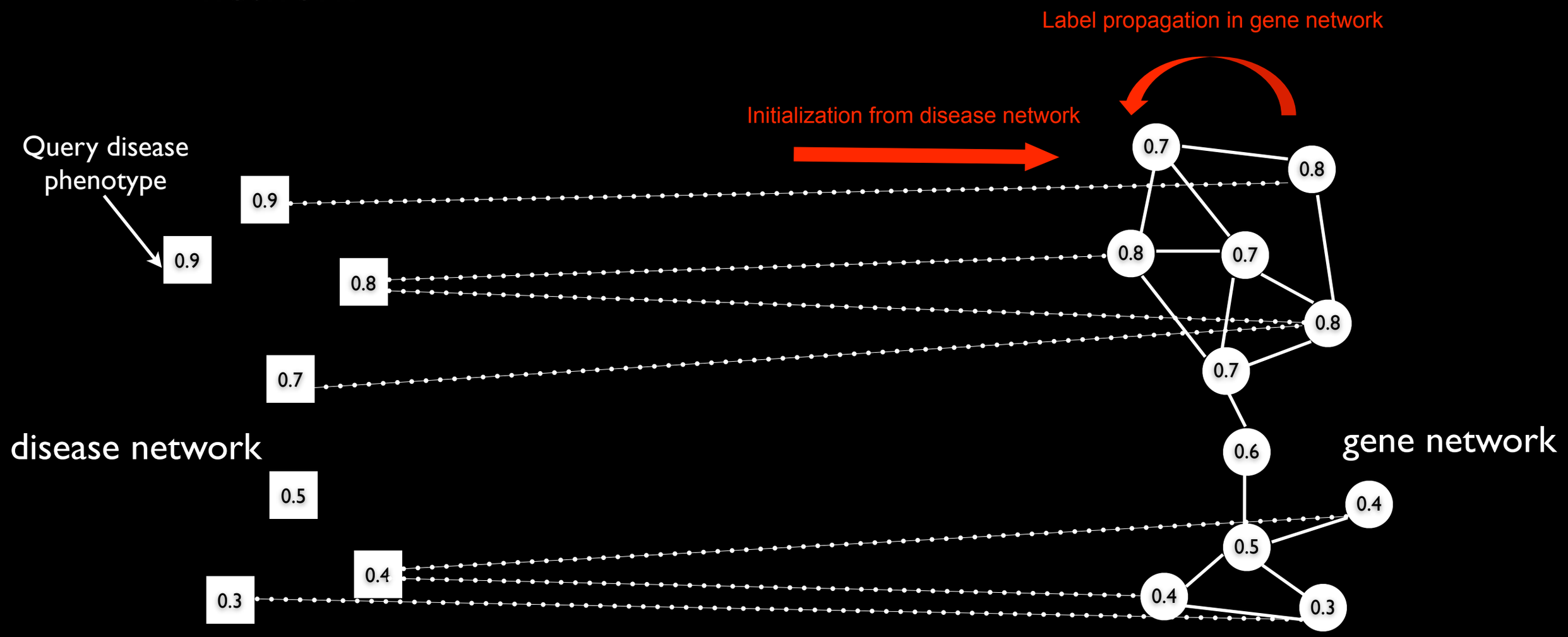
1. Initialize activation values on nodes (i.e. query node: 1 and others: 0)
2. Run label propagation on each network interactively
   - Run label propagation on disease network with initialization from gene network



Working example (3/5)

- Q: Find candidate disease genes associated with a query disease phenotype

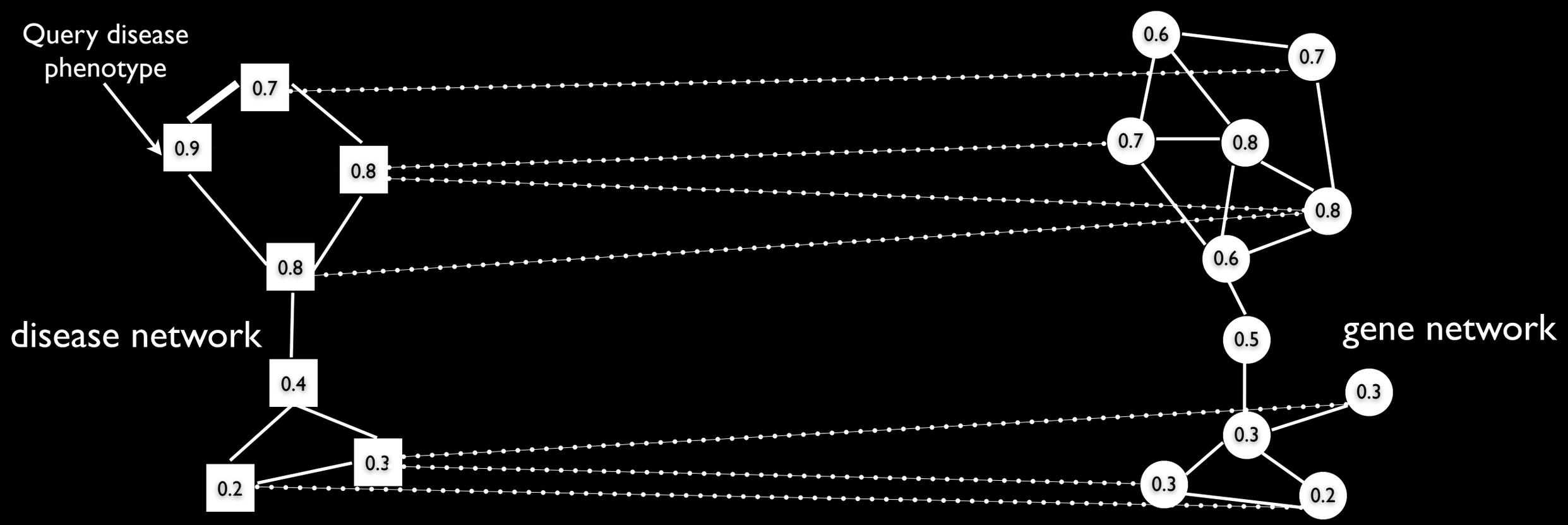
1. Initialize activation values on nodes (i.e. query node: 1 and others: 0)
2. Run label propagation on each network interactively
   - Run label propagation on gene network with initialization from disease network



Working example (4/5)

- Q: Find candidate disease genes associated with a query disease phenotype

1. Initialize activation values on nodes (i.e. query node: 1 and others: 0)
2. Run label propagation on each network interactively
   - Repeat until activation values on all nodes converge

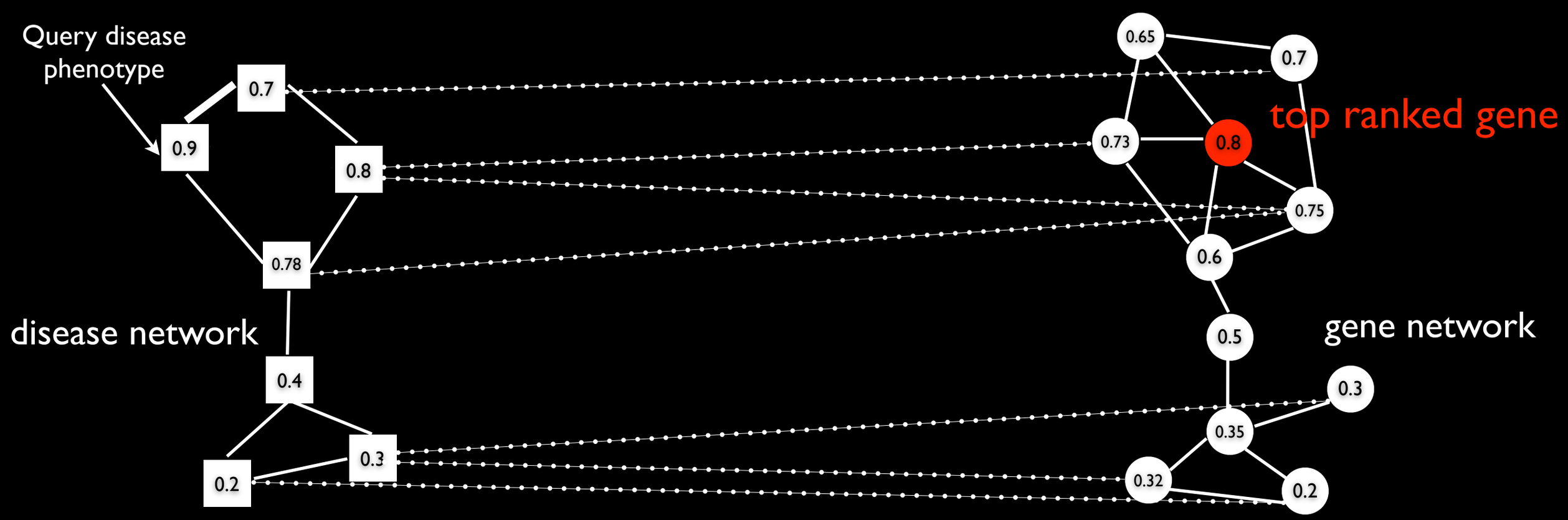


Working example (5/5)

- Q: Find candidate disease genes associated with a query disease phenotype

1. Initialize activation values on nodes (i.e. query node: 1 and others: 0)
2. Run label propagation on each network interactively
3. Analysis: Rank genes based on their final activation values

✓ **Highly ranked genes can be candidate disease genes of query disease**



Data preparation

1. Disease phenotype similarity network

- 5080 disease phenotypes
- Edges are weighted by pairwise disease similarities among 5080 disease phenotypes calculated by text mining techniques [Marc Driel, et al., European Journal of Human Genetics 2006]

2. Disease-gene association network [OMIM database., May 2007]

- an undirected bi-partite graph with disease and gene vertices
- 1126 disease-gene associations

3. Protein interaction networks [HPRD database., May 2007]

- 8919 proteins are mapped to human genes
- 34364 binary-valued undirected interactions between 8919 proteins
- Self-interactions are removed

Case study (1/2)

- Experimental setup
  - ✓ Use old-version of disease-gene associations (before May 2007) to predict new disease genes for disease phenotype
  - ✓ Compare prediction results with **recent association data (April 2010)**
    - 538 new associations
      - 404 associations between newly discovered disease genes and disease phenotypes
      - 134 associations between known disease genes and disease phenotypes

Case study (2/2)

- Our approach is capable to identify **true disease causative genes of disease phenotypes**

MIM#	Phenotype Name	HGNC symbol	Ranking MINProp	Ranking CIPHER SP	Ranking PRINCE	Status
601626	LEUKEMIA, ACUTE MYELOID	MLF1	**3**	4323	4323	new
		JAK2	**15**	354	280	new
		ETV6	**23**	769	769	new
		GMPS	245	4512	4512	new
300299	NEUTROPENIA	WAS	**1**	1656	30	known
171300	PHEOCHROMOCYTOMA	VHL	**1**	1105	1105	new
		GDNF	**16**	1400	1400	known
		KIF1B	228	512	512	new
607174	MENINGIOMA, FAMILIAL	NF2	**1**	1279	1279	known
		PTEN	**5**	1307	1307	known
166710	OSTEOPOROSIS	LRP5	**2**	1541	1541	known
		CALCR	**4**	7661	7661	new
		COL1A1	**5**	8086	8086	known
		VDR	42	1402	1402	known
202300	ADRENOCORTICAL CARCINOMA	TP53	**1**	1249	430	known
601367	STROKE, ISCHEMIC	PRKCH	**14**	448	448	new
		ALOX5AP	154	7892	7892	known

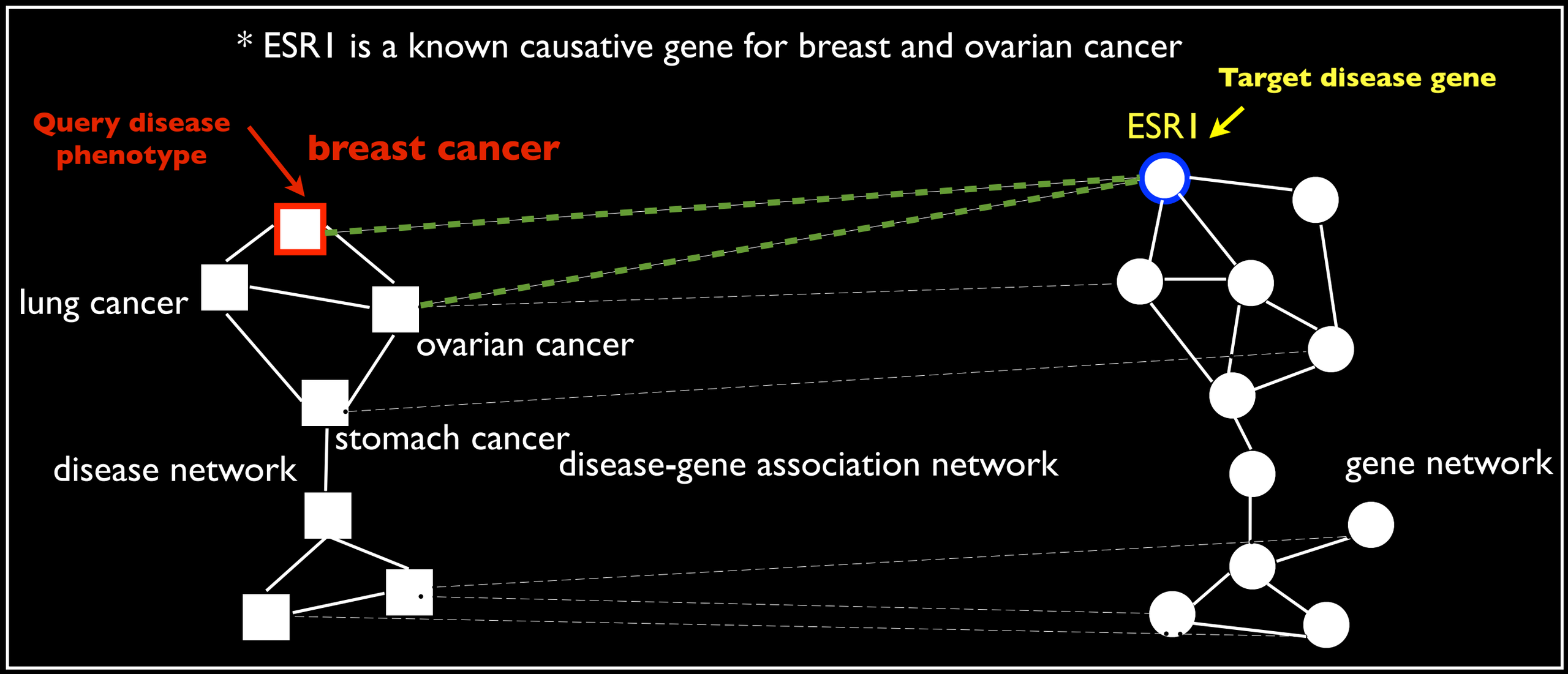
Leave-one out cross validation

1. Uncovering associations with known disease genes

   Ex) **Remove the direct association** btw ESR1 and breast cancer (keep the association btw ESR1 and ovarian cancer)

2. Discovering associations with unknown disease genes

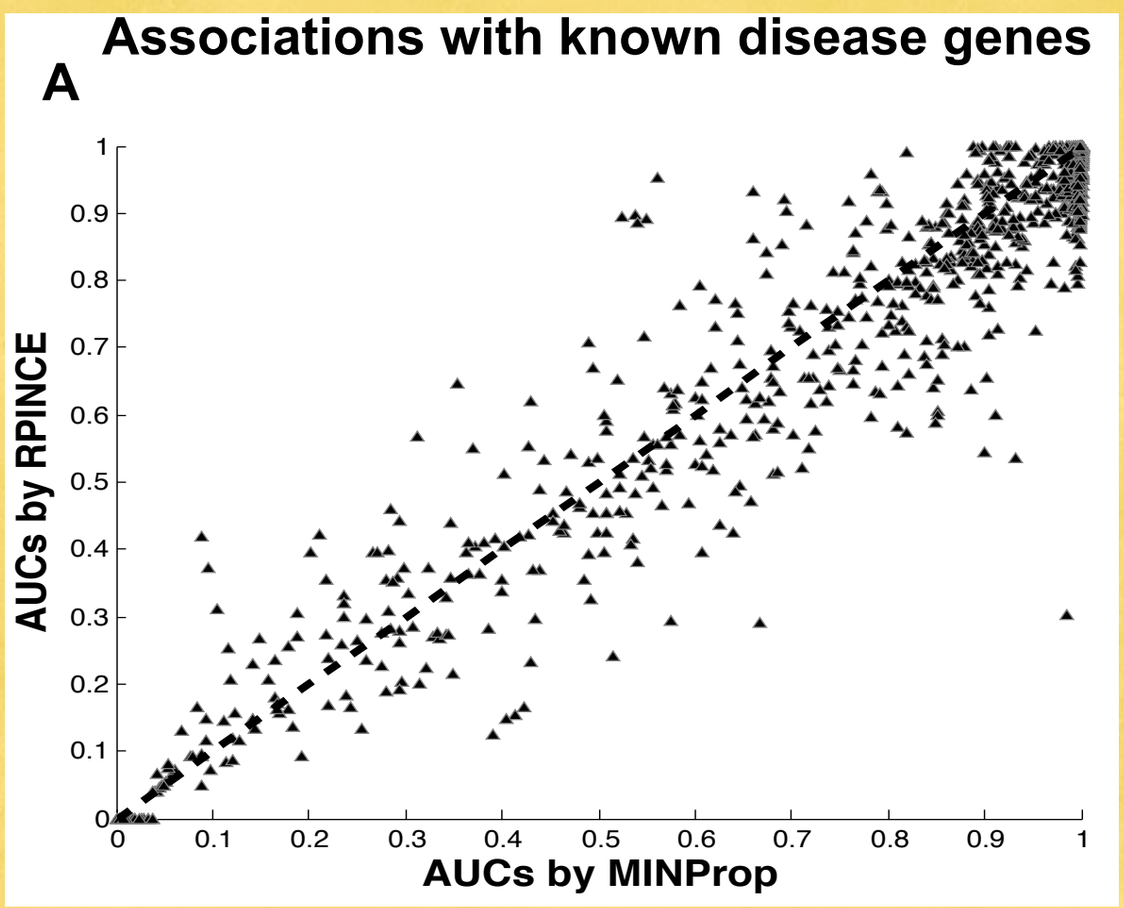
   Ex) **Remove all association** btw ESR1 and other disease

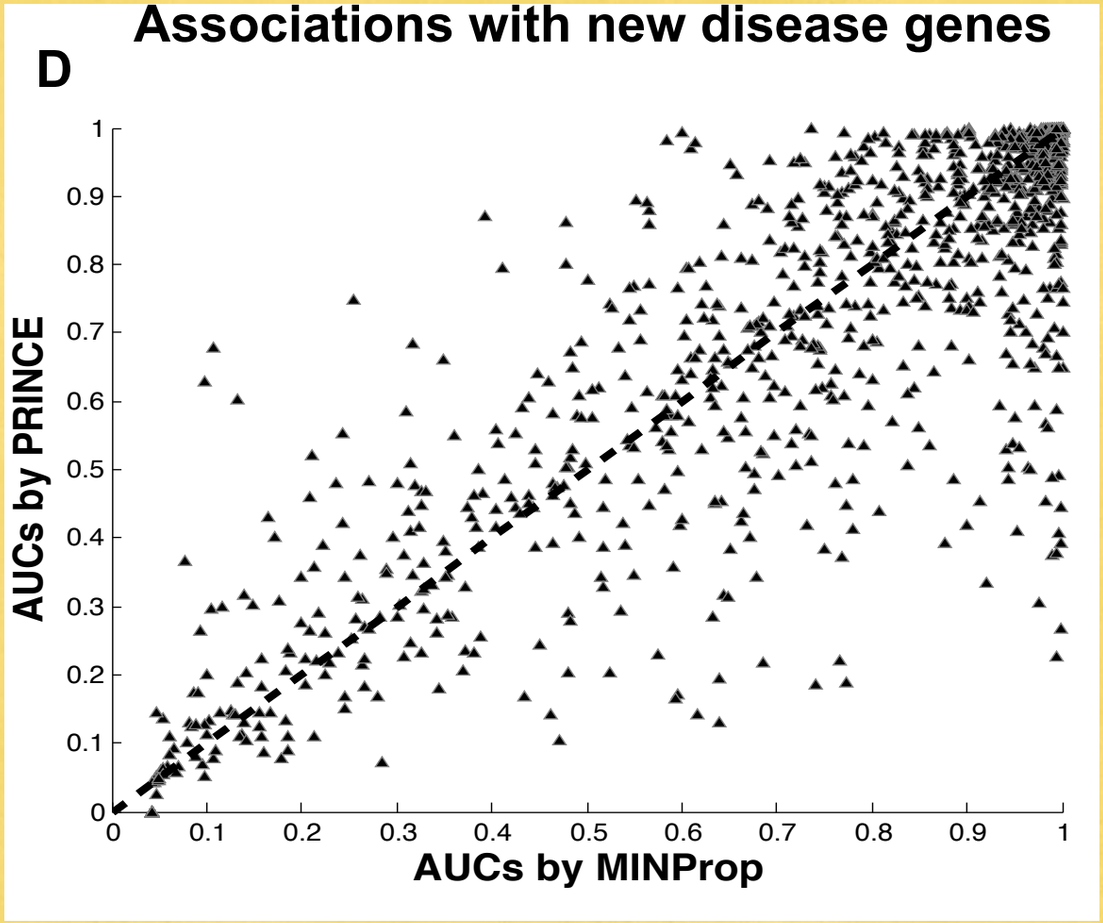


Ranking diseaes genes

- Overall, **MINProp achieved best performances** in leave one out cross validation for two experiments set-up

Methods	Known disease genes Avg. AUC *(win/draw/loss)*	New disease genes Avg. AUC *(win/draw/loss)*
MINProp vs. PRINCE	**0.805** vs. 0.785 (796/24/306)	**0.728** vs. 0.703 (642/8/476)
MINProp vs. Random Walk	**0.805** vs. 0.797 (738/75/313)	**0.728** vs. 0.648 (1045/2/79)
MINProp vs. CIPHER-DN	**0.863** vs. 0.738 (565/5/288)	**0.821** vs. 0.738 (515/11/332)
MINProp vs. CIPHER-SP	**0.805** vs. 0.734 (678/8/440)	0.728 vs. 0.729 (538/54/534)





Exploring modularity of genes

- How well the method could explore modular structures (i.e., cluster or subnetwork) of genes?
  - ✓ In most of most cases that disease genes of query phenotypes have higher clustering coefficients, MINProp performs better that that of baselines
  - ✓ Hybrid case shows better performances against MINProp

CC	MINProp vs. PRINCE Avg. AUC	MINProp vs. C-DN Avg. AUC	MINProp vs. C-SP Avg. AUC	Hybrid vs. MINProp Avg. AUC
$[0.1, 1]$	**0.875** vs. 0.854	**0.889** vs. 0.855	**0.875** vs. 0.813	**0.886** vs. 0.875
$[0.01, 0.1)$	**0.902** vs. 0.886	**0.906** vs. 0.799	**0.902** vs. 0.801	**0.911** vs. 0.902
$[0, 0.1)$	**0.653** vs. 0.626	**0.770** vs. 0.688	0.654 vs. **0.693**	**0.692** vs. 0.654
Total	**0.728** vs. 0.703	**0.821** vs. 0.738	0.728 vs. 0.729	**0.756** vs. 0.727

* Higher average clustering coefficients of disease genes indicate strong modularity of genes in the protein interaction network

Today's topic

- **Disease phenotype-gene association study**
  - Identify genetic variations affecting the phenotypic changes on a genome-scale
- **Applications**

1. Disease gene prediction
   - Predict candidate disease genes associated with a query disease phenotype
2. Predicting phenotypic/functional impact of candidate disease genes
   - Give a gene (or a set of genes), predict its target disease phenotypes/functions

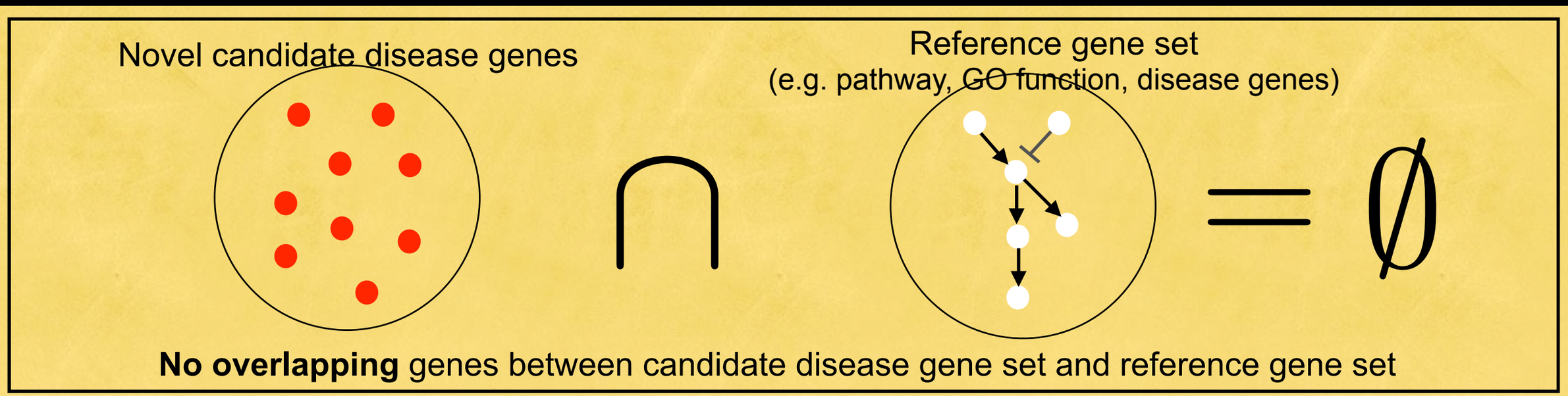
Inferring disease and gene set association

- **Background**
  - Numerous genome-scale disease studies are conducted to discover candidate disease causing genes
  - Overrepresentation based gene set enrichment analysis widely used for validation for their findings
    - ✓ GSEA (Broad), DAVID (NIH), and etc.

T. Hwang et. al., “Inferring disease and gene set association using rank coherent”. **Bioinformatics** 2011

Challenge

- **Current knowledge for gene function, pathway, and disease genes are still incomplete**
- **Novel disease susceptibility genes are often not well characterized and studied (e.g. unknown for their functions, pathways and associations with disease)**
  - **Ex) Only less** than **one-quarter** of significantly altered copy number regions contain previously validated cancer-causing genes. [*Beroukhim et al.*, Nature 2010]



No overlapping = No association ?

T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

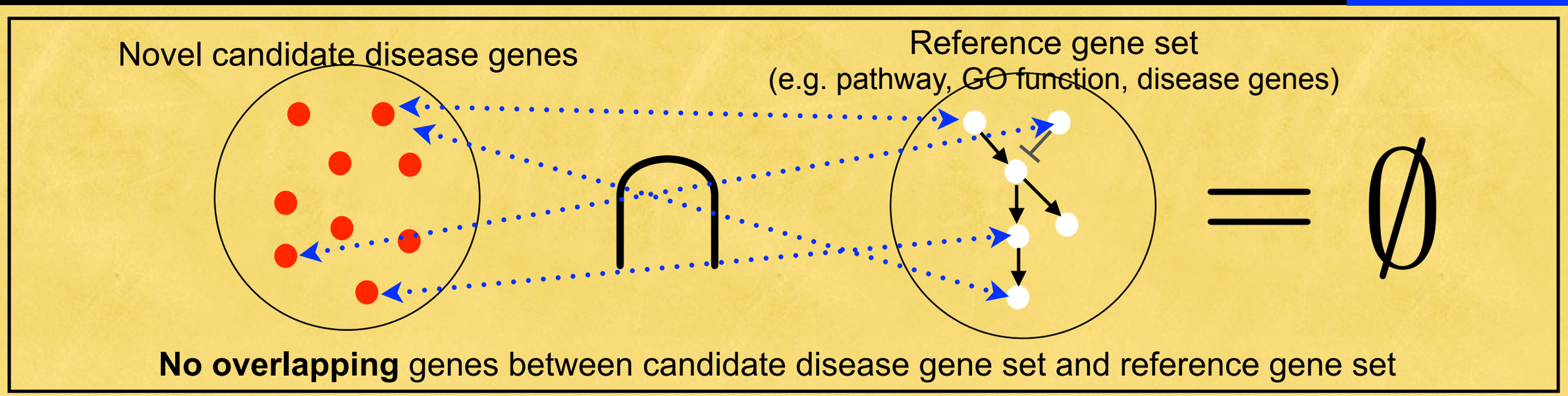
Challenge

- **Current knowledge for gene function, pathway, and disease genes are still incomplete**
- **Novel disease susceptibility genes are often not well characterized and studied (e.g. unknown for their functions, pathways and associations with disease)**

What if **candidate disease genes interact with genes** in the reference gene set?

et al., Nature 2010]

gene interaction



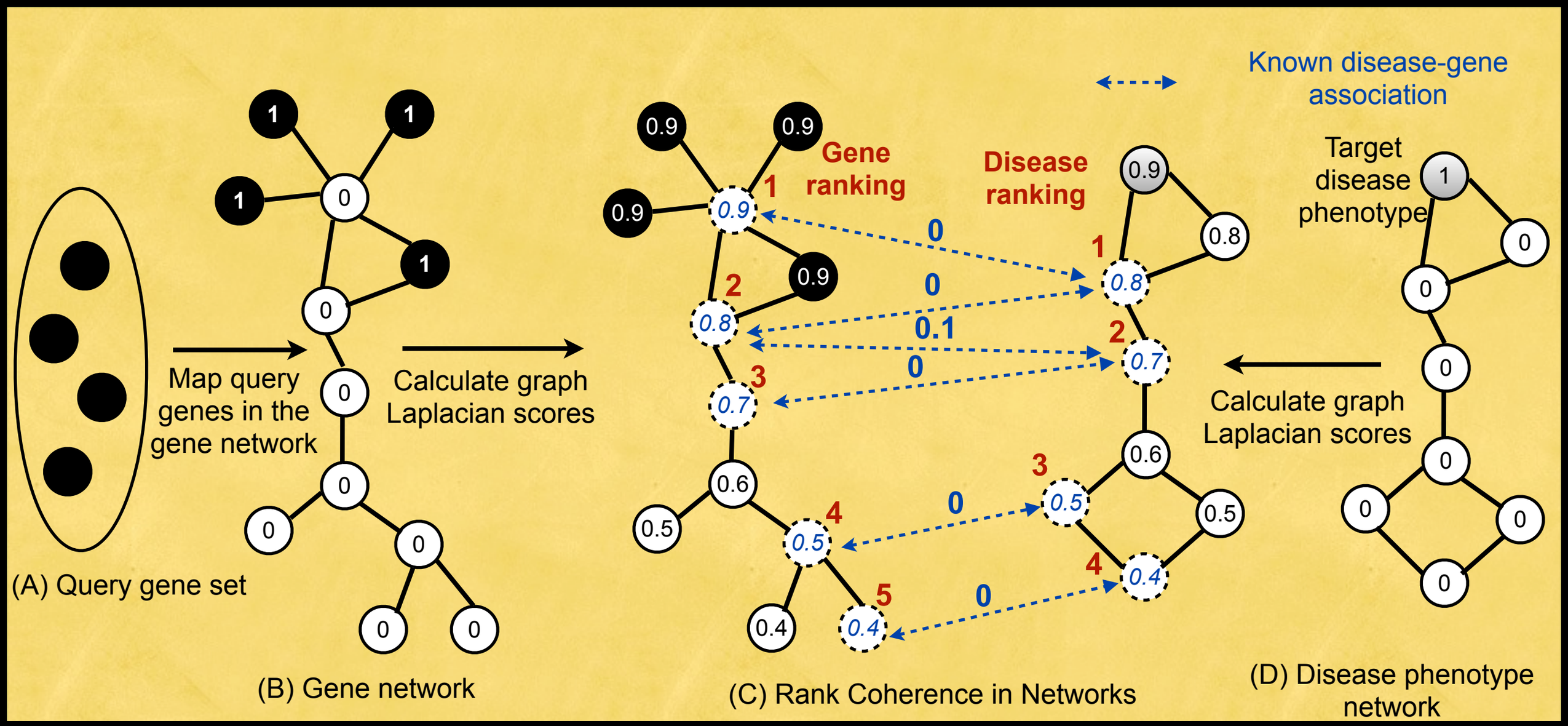
No overlapping genes between candidate disease gene set and reference gene set

No overlapping = No association ? No!

T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

Network based method

- **By querying the networks with a given gene set, we want to retrieve a list of disease phenotypes with the highest predicted association with the gene set.**

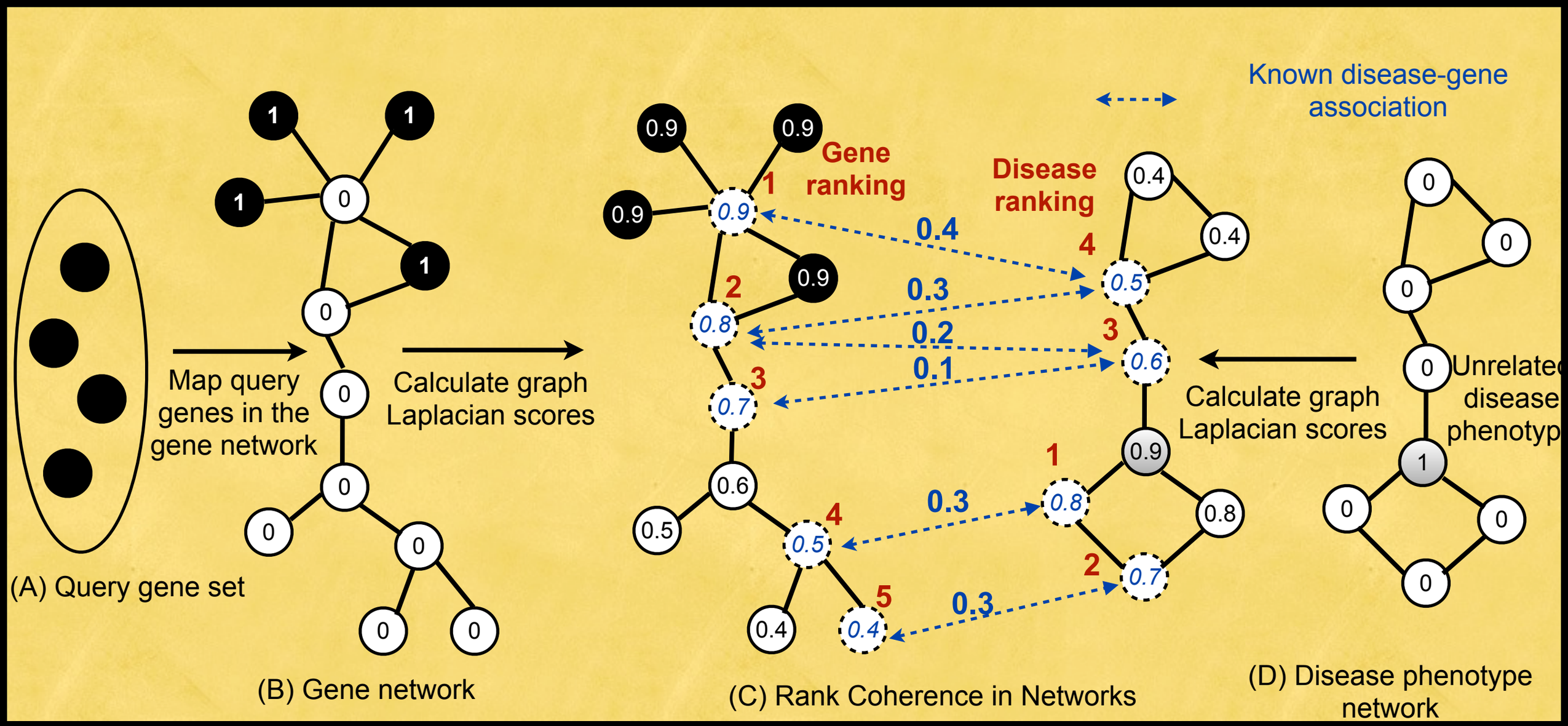


Two rankings between disease gene and its target disease are **coherent!**

T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

Network based method

- **By querying the networks with a given gene set, we want to retrieve a list of disease phenotypes with the highest predicted association with the gene set.**



Two rankings between disease gene and its target disease are **coherent!**

T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

Network based method

- **Objective: Given a query gene set, find a disease phenotype maximizing coherence between rankings of disease gene, and its target disease**

1. A ridge regression model

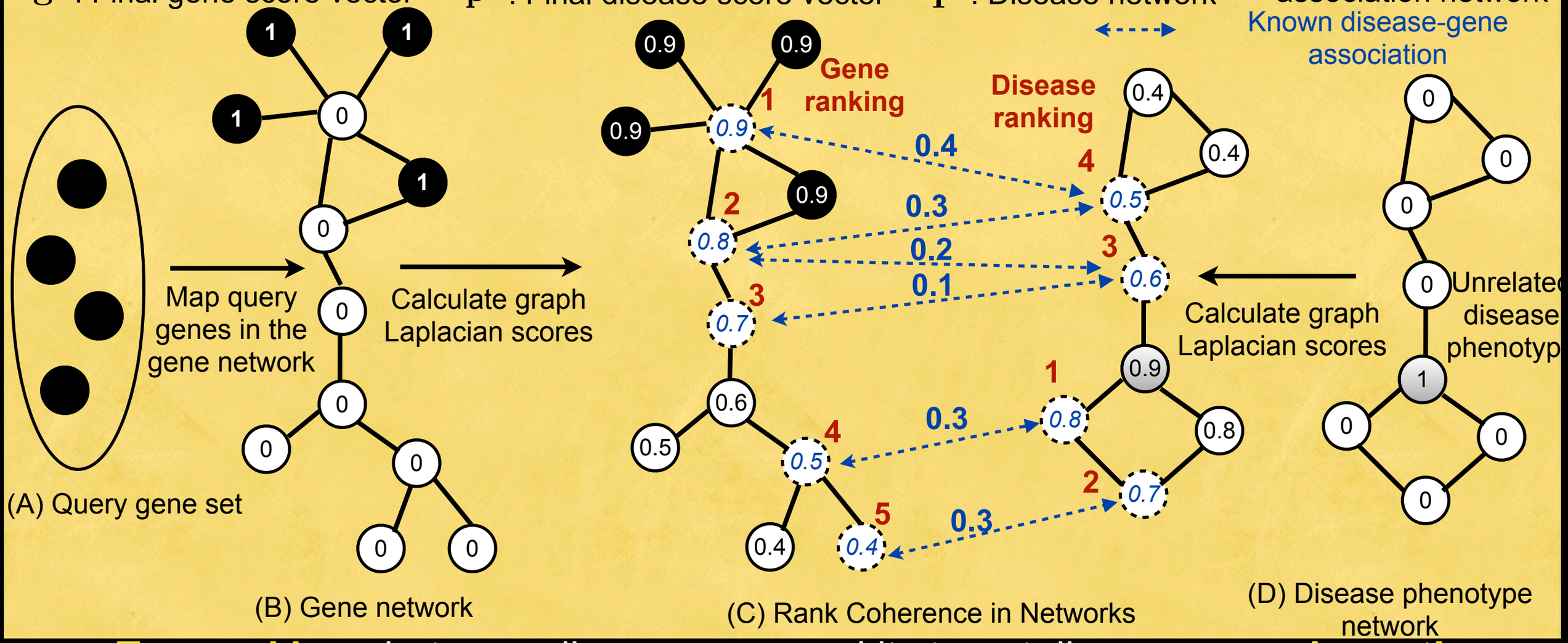
$$\Omega = ||\mathbf{A}\tilde{\mathbf{p}} - \tilde{\mathbf{g}}||^2 + \kappa||\mathbf{p}||^2$$

2. Enumeration methods

$$\mathbf{rcNet_{corr}}(\tilde{\mathbf{g}}, \tilde{\mathbf{p}}, \mathbf{A}) = \mathbf{corr}(\mathbf{A}\tilde{\mathbf{p}}, \tilde{\mathbf{g}})$$

$$\mathbf{rcNet_{lap}}(\tilde{\mathbf{g}}, \tilde{\mathbf{p}}, \mathbf{A}) = -\sum_{\mathbf{i,j}} \mathbf{A_{i,j}}(\tilde{\mathbf{p}}_\mathbf{i} - \tilde{\mathbf{g}}_\mathbf{j})^2$$

$\mathbf{g}$: Initial gene score vector
$\tilde{\mathbf{g}}$: Final gene score vector
$\mathbf{p}$: Initial disease score vector
$\tilde{\mathbf{p}}$: Final disease score vector
$\mathbf{G}$: Gene network
$\mathbf{P}$: Disease network
$\mathbf{A}$: Disease-gene association network



Two rankings between disease gene and its target disease are **coherent!**

T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

Network based method

- **Objective: Given a query gene set, find a disease phenotype maximizing coherence between rankings of disease gene, and its target disease**

1. A ridge regression model

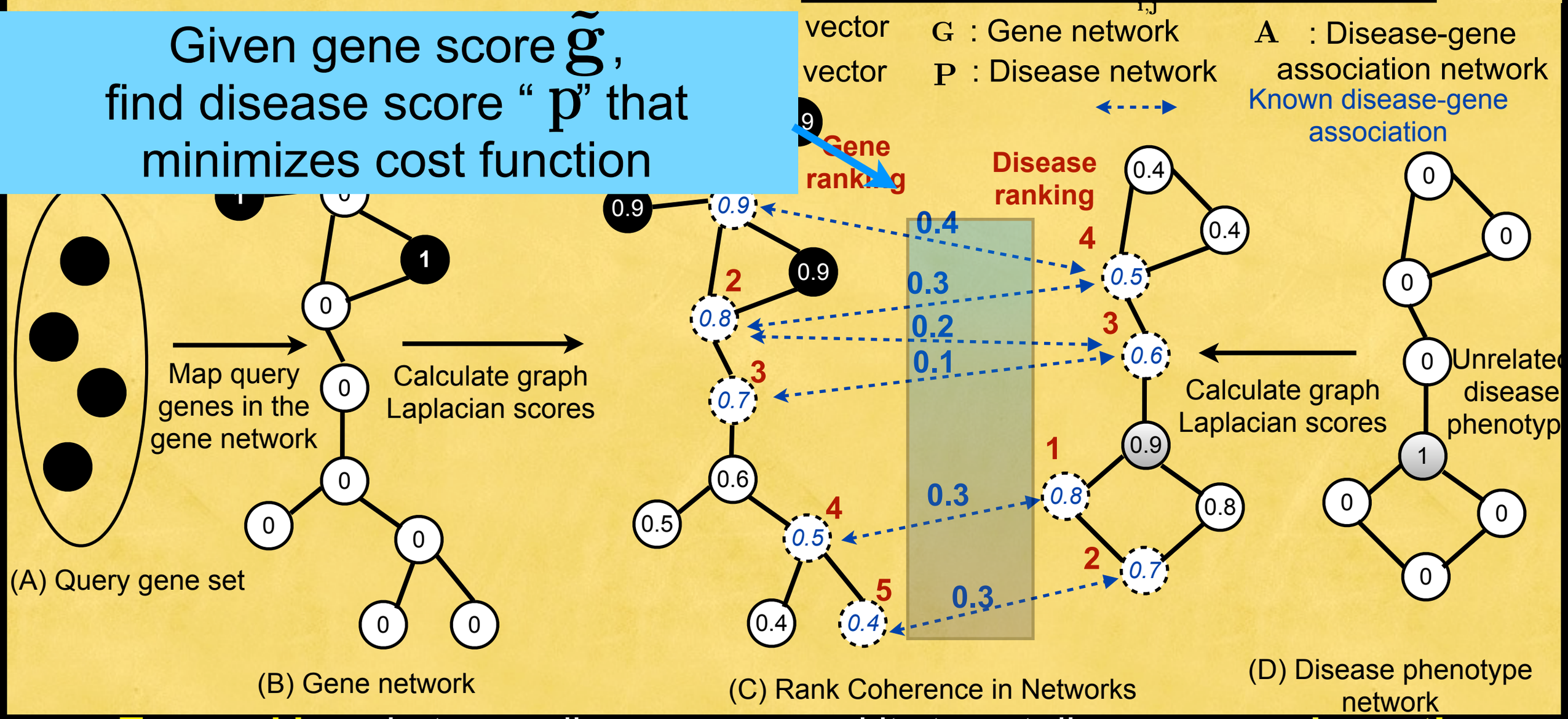
$$\Omega = ||\mathbf{A}\tilde{\mathbf{p}} - \tilde{\mathbf{g}}||^2 + \kappa||\mathbf{p}||^2$$

2. Enumeration methods

$$\mathbf{rcNet_{corr}}(\tilde{\mathbf{g}}, \tilde{\mathbf{p}}, \mathbf{A}) = \mathbf{corr}(\mathbf{A}\tilde{\mathbf{p}}, \tilde{\mathbf{g}})$$

$$\mathbf{rcNet_{lap}}(\tilde{\mathbf{g}}, \tilde{\mathbf{p}}, \mathbf{A}) = -\sum_{\mathbf{i,j}} \mathbf{A_{i,j}}(\tilde{\mathbf{p}}_\mathbf{i} - \tilde{\mathbf{g}}_\mathbf{j})^2$$

Given gene score $\tilde{g}$, find disease score " p" that minimizes cost function



Two rankings between disease gene and its target disease are **coherent!**

T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

Network based method

- **Objective: Given a query gene set, find a disease phenotype maximizing coherence between rankings of disease gene, and its target disease**

1. A ridge regression model

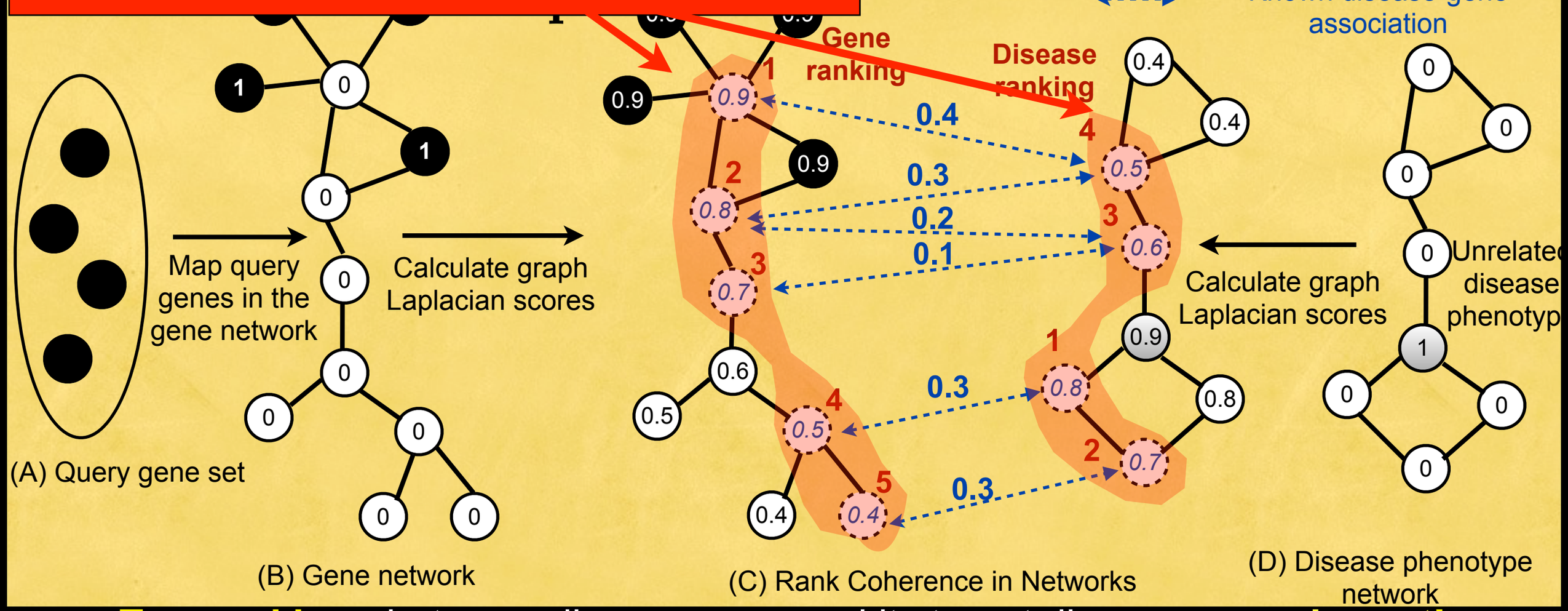
$$\Omega = ||\mathbf{A}\tilde{\mathbf{p}} - \tilde{\mathbf{g}}||^2 + \kappa||\mathbf{p}||^2$$

2. Enumeration methods

$$\mathbf{rcNet_{corr}}(\tilde{\mathbf{g}}, \tilde{\mathbf{p}}, \mathbf{A}) = \mathbf{corr}(\mathbf{A}\tilde{\mathbf{p}}, \tilde{\mathbf{g}})$$

$$\mathbf{Net_{lap}}(\tilde{\mathbf{g}}, \tilde{\mathbf{p}}, \mathbf{A}) = -\sum_{\mathbf{i,j}} \mathbf{A_{i,j}}(\tilde{\mathbf{p}}_\mathbf{i} - \tilde{\mathbf{g}}_\mathbf{j})^2$$

Given gene and diseae score , find diseae score "P" that maximize scores

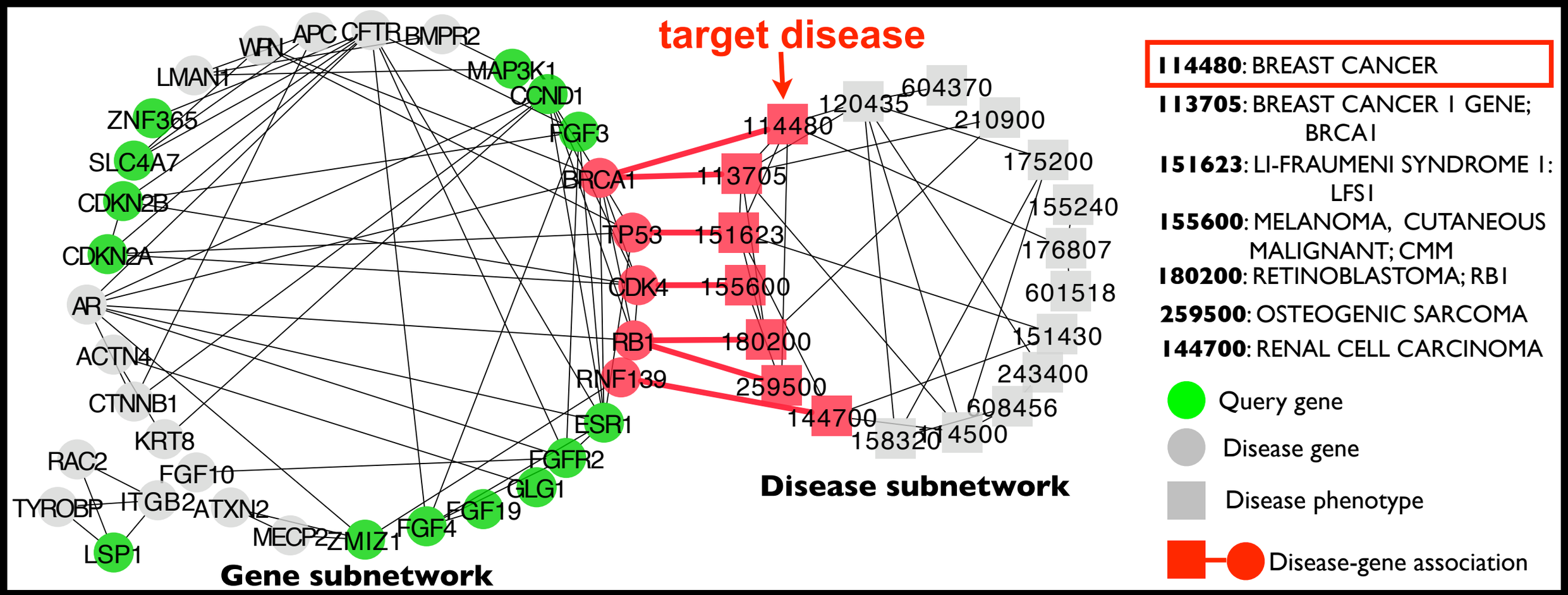


Two rankings between disease gene and its target disease are **coherent!**

T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

Real example

- Query **novel breast cancer susceptibility genes** from recent GWAS to predict target disease phenotype (breast cancer)
  - Genes in the query gene set are not known for any associations with any disease phenotypes



T. Hwang et. al., "Inferring disease and gene set association using rank coherent". **Bioinformatics** 2011

Data preparation

1. Disease phenotype similarity network

- 5080 disease phenotypes
- Edges are weighted by pairwise disease similarities among 5080 disease phenotypes calculated by text mining techniques [Marc Driel, et al., European Journal of Human Genetics 2006]

2. Disease-gene association network [OMIM database., May 2007]

- an undirected bi-partite graph with disease and gene vertices
- 1126 disease-gene associations

3. Protein interaction networks [HPRD database., May 2007]

- 8919 proteins are mapped to human genes
- 34364 binary-valued undirected interactions between 8919 proteins
- Self-interactions are removed

4. Functional linkage networks [Huttenhower et al., 2009]

- 24,433 genes in the network
- 60 million weighted (undirected) interactions between 24,433 genes
- Self-interactions are removed

Experiments

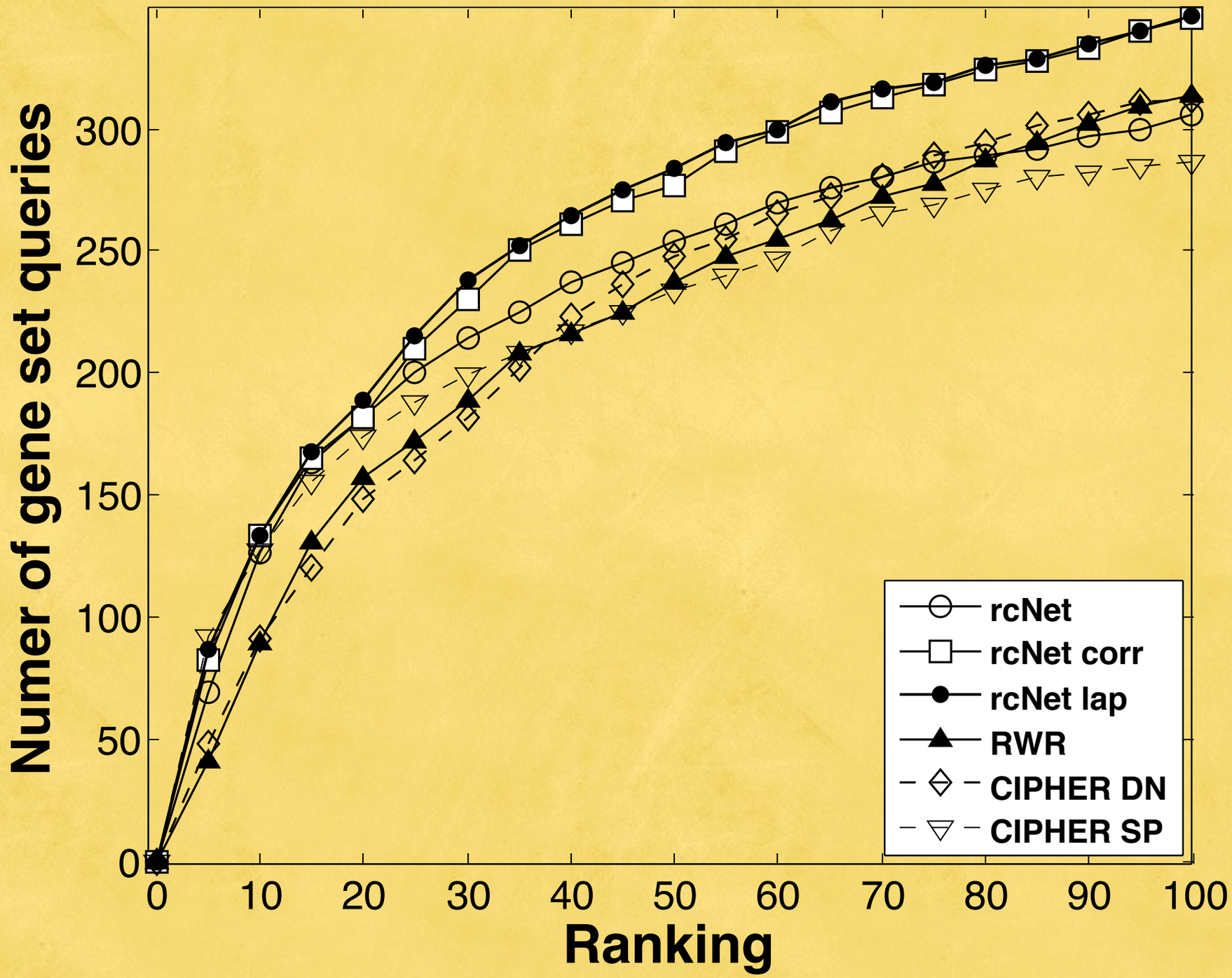
- Baselines
  - Cipher [Wu et al, Molecular System Biology 2009]
  - Random walk restart [Y Li, Bioinformatics 2010]
- Task
  - Given a query disease gene set, find target disease phenotype
    - rank candidate target diseases of the query gene set

- Validations
  - Leave-one-out cross-validations
  - Case study
    - recent OMIM
    - GWAS
    - Copy number data
    - Gene expression data

***The classification performance of all methods are evaluated using area under the receiver operating characteristics (ROC) score.**

T. Hwang et. al., “Inferring diseae and gene set association using rank coherent”. **Bioinformatics** 2012

Leave one out cross validation



T. Hwang et. al., “Inferring diseae and gene set association using rank coherent”. **Bioinformatics** 2012

GWAS experiments

- Query novel disease susceptibility genes from recent GWAS to predict the disease phenotype
  - Q) How a set of novel candidate disease genes from GWAS affects to disease phenotypes?

Table 2. Ranking the target disease phenotype of the disease susceptibility genes identified from GWAS. The disease categories in the first column are based on the definition in Goh *et al.* (2007). In the third column, the PubMed IDs marked with '*' denote multiple GWASs for a disease/trait. Refer to supplementary Table for the results of the full list of the GWAS cases.

Category	Disease/Trait	PubMed Index	OMIM Index	Gene Set Size	Rank by rcNet	Rank by $\mathbf{rcNet_{corr}}$	Rank by $\mathbf{rcNet_{lap}}$
Cancer	Prostate cancer	20676098*	176807	15	2 (0.03%)	2 (0.03%)	2 (0.03%)
	Breast cancer	20872241*	113705	26	7 (0.1%)	51 (1%)	43 (0.8%)
	Basal cell carcinoma (cutaneous)	18849993	605462	5	7 (0.1%)	189 (3.7%)	228 (4.5%)
	Basal cell carcinoma (cutaneous)	18849993	604451	5	90 (2%)	202 (4%)	256 (5%)
	Urinary bladder cancer	18794855	109800	1	14 (0.2%)	48 (0.9%)	60 (1.1%)
	Acute lymphoblastic leukemia (childhood)	20670164*	159555	3	19 (0.04%)	51 (1.0%)	45 (0.8%)
	Lung cancer	20304703*	211980	12	22 (0.4%)	587 (12%)	1610 (32%)
	Lung adenocarcinoma	20871597*	211980	6	52 (1%)	838 (16%)	1815 (36%)
	Chronic lymphocytic leukemia	20062064*	151430	14	57 (1%)	318 (6.3%)	306 (6%)
	Neuroblastoma (high-risk)	19412175	600613	1	143 (3%)	110 (2%)	138 (3%)
Immunological	Systemic lupus erythematosus	20169177*	152700	10	46 (0.9%)	178 (4%)	161 (3%)
	Leprosy	20018961	246300	4	78 (1.5%)	62 (1.2%)	64 (1.3%)
	Leprosy	20018961	607572	4	272 (5%)	54 (1%)	55 (1%)
Endocrine	Type 2 diabetes	20862305*	125853	9	97 (2%)	718 (14%)	1912 (38%)
	Type 1 diabetes	19966805*	222100	26	331 (7%)	690 (13%)	191 (3.8%)
Gastrointestinal	Crohns disease	17684544	266600	2	60 (1.2%)	1396 (27%)	3012 (59%)

T. Hwang et. al.,, "Inferring diseae and gene set association using rank coherent". **Bioinformatics** 2012

Copy number experiments

- Query disease susceptibility genes in significantly altered copy number
  - Q) How a set of significantly altered genes in copy number affects to disease phenotypes?

Table 3. Ranking the target disease phenotypes of the candidate disease genes with copy number changes. This experiment includes 13 human cancer copy number studies from (Beroukhim *et al.*, 2010).

Disease/Trait	Rank by rcNet	Rank by $\mathbf{rcNet_{corr}}$	Rank by $\mathbf{rcNet_{lap}}$
Neuroblastoma	5	13	126
Colorectal cancer	14	20	613
Renal cancer	22	14	33
Non small cell lung cancer	34	48	558
Breast cancer	68	136	521
Medulloblastoma	77	826	2007
Prostate cancer	129	127	2447
Ovarian cancer	322	73	1108
Small cell lung cancer	759	53	909
Mesothelioma	959	21	54
Gastrointestinal stromal tumor	1169	787	1679
Hepatocellular carcinoma	4241	952	1295
Glioma	4705	787	951

T. Hwang et. al.,, “Inferring diseae and gene set association using rank coherent”. **Bioinformatics** 2012

Gene expression experiments

- Query differentially expressed genes to predict target disease phenotype

Table 4. Ranking the target disease of differentially expressed genes. The first column represents the target disease of a microarray gene expression study, and the second column gives the GEO number of the dataset.

Disease/Trait	GEO Num.	Rank by rcNet	Rank by $\mathbf{rcNet_{corr}}$	Rank by $\mathbf{rcNet_{lap}}$
AML	GSE9476	576	316	359
Breast cancer	GSE7390	14	49	51
	GSE2034	40	130	146
	GSE6532	129	151	182
	GSE1456	138	102	109
	GSE3494	161	709	1313
Gastric cancer	GSE13911	248	298	362
Lung cancer	GSE10072	206	755	2219
	E-MEXP-231	318	608	1115
	GSE7670	379	1330	4002
Ovarian cancer	GSE6008	414	1494	2283
Prostate cancer	E-MEXP-1327	271	1446	2057
	GSE8218	900	1214	2498

T. Hwang et. al.,, “Inferring diseae and gene set association using rank coherent”. **Bioinformatics** 2012

rcNet web tool

http://compbio.cs.umn.edu/dgsa_rcNet

Computational Biology Lab

Department of Computer Science and Engineering, University of Minnesota - Twin Cities

Introduction | User Guidance | Reference | Credits

rcNet: A Web Tool for Inferring Disease and Gene Set Association

Gene Set Query | Phenotype Query

Enter your genes into the the text area:
(One gene in one line)

C2orf43
CTBP2
EHBP1
FOXP4
GPRC6A
GSPT2
JAZF1
KLK3
LMTK2
MAGED1

Search gene:

Add

Select a ranking method:

rcNet_lap(Laplacian Score)

The number of phenotypes to display:

10

Clear | Search

View query gene set:

TNRC6B | SLC22A3 | LMTK2 | MAGED1 | JAZF1 | GPRC6A | FOXP4 | KLK3 | MSMB | CTBP2 | EHBP1 | NUDT10 | NUDT11 | GSPT2 LESS

Phenotype ranking:

Ranking	Phenotype Name	Relavance Score
1	TETRALOGY OF FALLOT	1273.4128
2	PROSTATE CANCER	1439.6707
3	IMMUNODYSREGULATION, POLYENDOCRINOPATHY, AND ENTEROPATHY, X-LINKED; IPEX	1494.3404
4	PROGEROID FACIAL APPEARANCE WITH HAND ANOMALIES	1556.6287
5	MESOAXIAL HEXADACTYLY AND CARDIAC MALFORMATION	1614.4134
6	ECTRODACTYLY OF LOWER LIMBS, CONGENITAL HEART DEFECT, AND MICROGNATHIA	1629.6243
7	SPINOCEREBELLAR ATAXIA 2; SCA2	1640.2059
8	MICROCEPHALY-CARDIOMYOPATHY	1652.7193
9	OVARIAN GERM CELL CANCER	1654.8965
10	BRACHYMORPHISM-ONYCHODYSPLASIA-DYSPHALANGISM SYNDROME	1688.1062

About Us | Site Map | Privacy Policy | Contact Us | ©2011 Computational Biology Lab, UMN

rcNet algorithm

$\mathbf{dgsa\_rcNet}(\mathbf{g}, \bar{\mathbf{G}}, \bar{\mathbf{P}}, \mathbf{A}, \alpha, \beta)$

1 $\mathbf{p} = \mathbf{0}$

2 $\tilde{\mathbf{g}} = (\mathbf{1} - \alpha)(\mathbf{I} - \alpha\bar{\mathbf{G}})^{-\mathbf{1}}\mathbf{g}$ (equation (3)).

3 $\bar{\mathbf{A}} = (\mathbf{1} - \beta)\mathbf{A}(\mathbf{I} - \beta\bar{\mathbf{P}})^{-\mathbf{1}}$

4 $\mathbf{p}^* = (\bar{\mathbf{A}}^{\mathbf{T}}\bar{\mathbf{A}} + \kappa\mathbf{I})^{-\mathbf{1}}\bar{\mathbf{A}}^{\mathbf{T}}\tilde{\mathbf{g}}$

5 $\mathbf{p}(\mathbf{p}^* > \mathbf{a}) = \mathbf{1}$ *(target selection with threshold* $\mathbf{a}$*)*

6 **return** $(\mathbf{p})$

Fig. 2. rcNet Algorithm - Rank Coherence in Networks.

rcNet algorithm (corr and lap)

$\mathbf{dgsa\_rcNet\_enu}(\mathbf{g}, \bar{\mathbf{G}}, \bar{\mathbf{P}}, \mathbf{A}, \alpha, \beta)$

$\tilde{\mathbf{g}} = (\mathbf{1} - \alpha)(\mathbf{I} - \alpha\bar{\mathbf{G}})^{-\mathbf{1}}\mathbf{g}$
$\mathbf{p} = \mathbf{0}, \mathbf{s} = \mathbf{0}$
for $\mathbf{i} = \mathbf{1}$ **to** $\mathbf{n}$
    $\mathbf{p_i} = \mathbf{1}$
    $\tilde{\mathbf{p}} = (\mathbf{1} - \beta)(\mathbf{I} - \beta\bar{\mathbf{P}})^{-\mathbf{1}}\mathbf{p}.$
    $\mathbf{s_i} = \mathbf{corr}(\mathbf{A}\tilde{\mathbf{p}}, \tilde{\mathbf{g}})$ or $-\sum_{\mathbf{i,j}} \mathbf{A_{i,j}}(\tilde{\mathbf{p}}_\mathbf{i} - \tilde{\mathbf{g}}_\mathbf{j})^\mathbf{2}$
    $\mathbf{p_i} = \mathbf{0}$
$\mathbf{j} = \mathbf{argmax_i}\ \mathbf{s_i}$
$\mathbf{p_j} = \mathbf{1}$
return $(\mathbf{p})$

Fig. 3. $\text{rcNet}_{\text{corr}}$ and $\text{rcNet}_{\text{lap}}$ Algorithms - Rank Coherence in Networks by Enumeration.

Regularization framework

$$\Omega(f) = \sum_{i=1}^{k} \underbrace{\left(f_i^T \Delta^{(i)} f_i + \mu_i \parallel f_i - y_i \parallel^2\right)}_{\text{label propagation in the homo-subnetwork}}$$

$$+ \frac{1}{2} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \underbrace{\mu_{ij} [f_i^T f_j^T] \Sigma^{(i,j)} \begin{bmatrix} f_i \\ f_j \end{bmatrix}}_{\text{label propagation in the hetero-subnetwork}},$$

- f : predicted label
- y : initial label
- Δ : graph laplacian of homo - subnetwork
- k : number of subnetwork
- μ_i and μ_{ij} : positive constants
- Σ : graph laplacian of hetero - subnetwork

T. Hwang, and Rui Kuang, “A heterogeneous label propagation for disease gene discovery”. **SDM** 2010

Algorithm

Algorithm 1 MINProp

Input

k: number of homo-subnetworks

σ: convergence threshold

$y_1, y_2, ..., y_k$: vectors of initial label values

$\alpha_1, \alpha_2, ..., \alpha_k$: diffusion parameters

$S^{(1)}, S^{(2)}, ..., S^{(k)}$: homo-subnetwork matrices

$S^{(1,2)}, ..., S^{(k-1,k)}$: hetero-subnetwork matrices

Output

$f_1, f_2, ..., f_k$: vectors of final label values

$f_i = 0$ for $i = 1...k$;
do
  $f_i^{old} = f_i$ for $i = 1...k$;
  for $i = 1...k$
    $t = 0, f_i^0 = 0$;
    $y' = \frac{1-k\alpha_i}{1-\alpha_i} y_i + \frac{\alpha_i}{1-\alpha_i} \sum_{j \neq i} S^{(i,j)} f_j$;
    do
      $t = t + 1$;
      $f_i^t = (1 - \alpha_i) y' + \alpha_i S^{(i)} f_i^{t-1}$;
    while$(\| f_i^t - f_i^{t-1} \| > \sigma)$;
    $f_i = f_i^t$;
  end for
while $(\exists i$ **s.t.** $\| f_i - f_i^{old} \| > \sigma)$;
return $f_1, f_2, ..., f_k$;

T. Hwang, and Rui Kuang, “A heterogeneous label propagation for disease gene discovery”. **SDM** 2010

rcNet algorithm (corr and lap)

$\mathbf{dgsa\_rcNet\_enu}(\mathbf{g}, \bar{\mathbf{G}}, \bar{\mathbf{P}}, \mathbf{A}, \alpha, \beta)$

$\tilde{\mathbf{g}} = (\mathbf{1} - \alpha)(\mathbf{I} - \alpha\bar{\mathbf{G}})^{-\mathbf{1}}\mathbf{g}$
$\mathbf{p} = \mathbf{0}, \mathbf{s} = \mathbf{0}$
for $\mathbf{i} = \mathbf{1}$ **to** $\mathbf{n}$
    $\mathbf{p_i} = \mathbf{1}$
    $\tilde{\mathbf{p}} = (\mathbf{1} - \beta)(\mathbf{I} - \beta\bar{\mathbf{P}})^{-\mathbf{1}}\mathbf{p}.$
    $\mathbf{s_i} = \mathbf{corr}(\mathbf{A}\tilde{\mathbf{p}}, \tilde{\mathbf{g}})$ or $-\sum_{\mathbf{i,j}} \mathbf{A_{i,j}}(\tilde{\mathbf{p}}_\mathbf{i} - \tilde{\mathbf{g}}_\mathbf{j})^\mathbf{2}$
    $\mathbf{p_i} = \mathbf{0}$
$\mathbf{j} = \mathbf{argmax_i}\ \mathbf{s_i}$
$\mathbf{p_j} = \mathbf{1}$
return $(\mathbf{p})$

Fig. 3. $\text{rcNet}_{\text{corr}}$ and $\text{rcNet}_{\text{lap}}$ Algorithms - Rank Coherence in Networks by Enumeration.