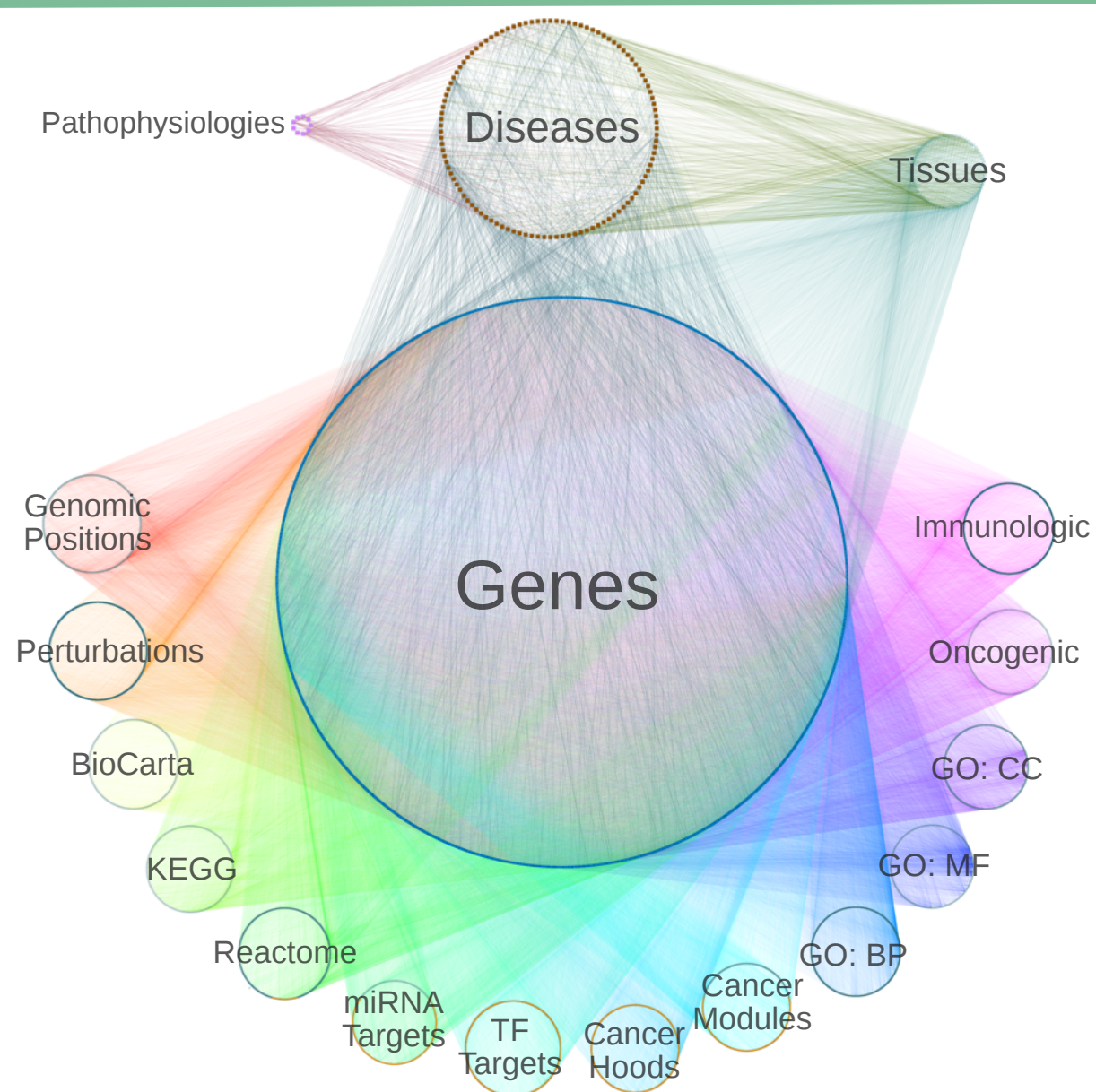


Daniel Himmelstein, Leo Brueggeman & Sergio Baranzini present ***Repurposing drugs on a heterogeneous network***

Last year, we introduced *heterogeneous network edge prediction* (HNEP) to predict disease-associated genes.

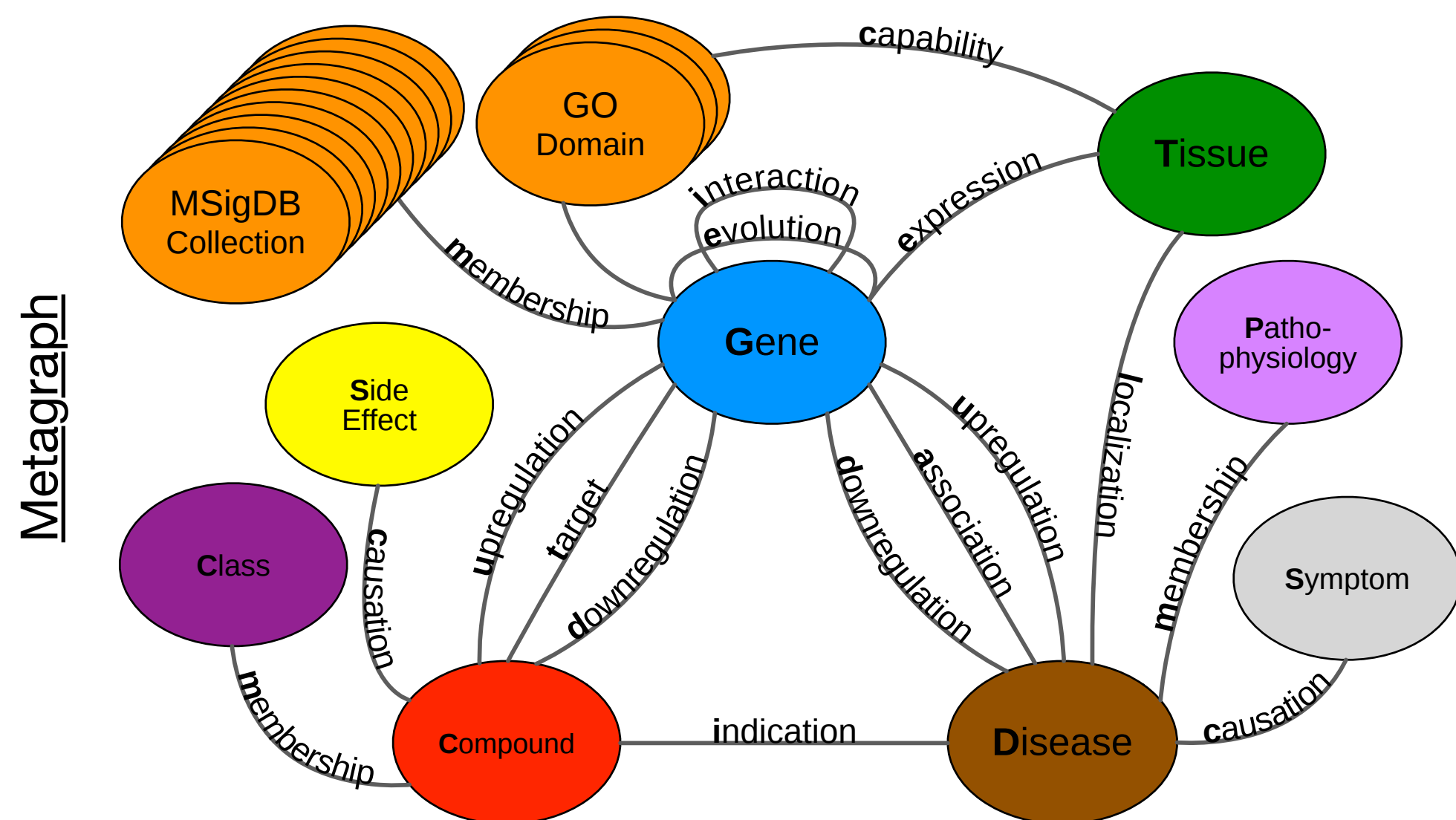
Heterogeneous networks contain multiple node and edge types.

Our network contained 40,343 nodes (of 18 types) and 1,608,168 edges (of 19 types).



Now in 2015, we will use this data integration approach to repurpose drugs on a heterogeneous network.

Planning the Network Construction



And you can follow in realtime and get paid to participate.

ThinkLab

thinklab.com/p/rephetio
doi:10.15363/thinklab.4

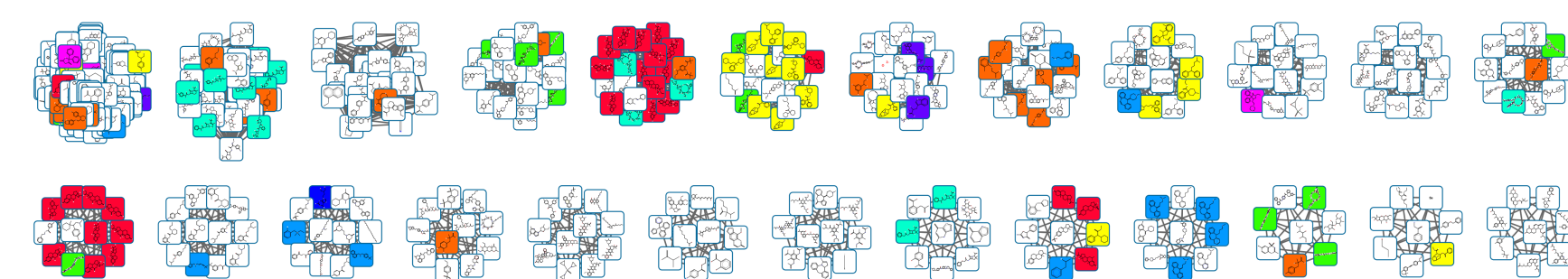


ThinkLab is:

- massively collaborative — all are welcome
- open science — all content is CC-BY
- incentivized — contributions are rewarded
- productive — scientific markdown editor
- efficient — code and results public upon commit

Results (as of March 2015)

We analyzed **SIDER 2** and investigated its strengths and weaknesses as well as pharmacological utility.



Side-effect similarity modules were concordant with structural similarity modules (colored).

git.dhimmel.com/SIDER2

We created a user-friendly service to retrieve **Gene Ontology annotations** with optional propagation.

Propagated	Unpropagated
Entrez	Symbol
All Genes	Protein-coding Genes

git.dhimmel.com/gene-ontology

Forthcoming in *PLOS Computational Biology*
preprint on *bioRxiv* [doi:10.1101/011569]



Predictions online at het.io

[Home](#)
[HPLP](#)
[Disease Genes](#)
[Media](#)
[Balance Lab](#)

Predictions for *multiple sclerosis*

Disease Ontology
EFO:0003885
UMLS Concepts
C0026759

The table shows the predicted probability that multiple sclerosis is associated with each protein-coding gene. Clicking the [gene symbol](#) will take you to the gene page. Clicking the [gene code](#) will take you to the external list with additional gene information. Clicking elsewhere on the row displays a breakdown below showing each feature's contribution to the overall prediction. The existence of a GWAS-reported association is indicated by [asterisks](#) where the levels indicate if there was a primary association for a high (> H2 P) or low confidence (< L2 P) association, or a secondary association for a high (> H2 S) or low confidence (< L2 S) association. [Click here](#) to view the number of diseases, including multiple sclerosis, that the gene is associated with. The average prediction across all diseases for a gene is shown by [clicking the gene symbol](#).

Show 10 entries				Search:	
gene_symbol	gene_code	status	other_associators	mean_prediction	prediction +
L3L3A	HGN03989	L2LP - 2	4	0.31251	42.7171%
L3R5X	HGN01395	- 5	2	0.25377	43.846%
L12H	HGN0370	H2CP - 5	5	26.6442	41.673%
PTF1B	HGN03650	- 4	4	0.6991	38.571%
L12H2	HGN03982	- 4	4	0.6114	36.848%
L2H4	HGN0308	H2CP - 5	5	0.6811	36.025%
H2FS	HGN01810	- 5	2	0.8475	31.477%
O05G55	HGN01787	- 3	3	1.86104	29.422%
H2L42C4H	HGN0360	H2HB - 2	2	4.2697	28.311%
TAF11P3	HGN011596	H2HS - 4	4	7.0977	25.666%

Showing 1 to 10 of 151,116 entries

Page
Previous
1
2
3
4
Next
Last

Nodes

Type	Resource
Compound	DrugBank
Disease	Disease Ontology
Gene	Entrez Gene
Tissue	Uberon
Gene Set	MSigDB
Side Effect	UMLS
Pathophysiology	Manual
Symptom	MeSH

Standardized terminologies:

- provide a scalable framework for data integration
- prevent redundancy
- enable semantic data

Edges

Source	Target	Type	Resource
Compound	Disease	Indication	MEDI
Compound	Disease	Indication	LabeledIn
Compound	Gene	Expression	LINCS
Compound	Side Effect	Causation	SIDER 2
Compound	Side Effect	Causation	OFFSIDES
Disease	Gene	Target	ChEMBL
Disease	Gene	Association	GWAS Catalog
Disease	Gene	Expression	STAR-GEO
Disease	Pathophysiology	Membership	Manual
Disease	Symptom	Causation	Human symptoms--disease network
Gene	Gene	Interaction	Human Interactome Project
Gene	Gene	Interaction	The Incomplete Interactome
Gene	Gene	Evolution	Evolutionary Rate Covariation
Gene	Gene Set	Membership	MSigDB
Gene	Tissue	Expression	GNF Gene Expression Atlas

Ideal resources are:

- high-throughput
- systematic
- unbiased
- aggregately diverse

Acknowledgements

We would like to thank our ThinkLab contributors (thinklab.com/p/rephetio/leaderboard) and Alex Pico for the SIDER visualization. This material is based upon work supported by the National Science Foundation under Grant No. 1144247 to DSH. SEB is a Harry Weaver Neuroscience fellow from the National Multiple Sclerosis Society.