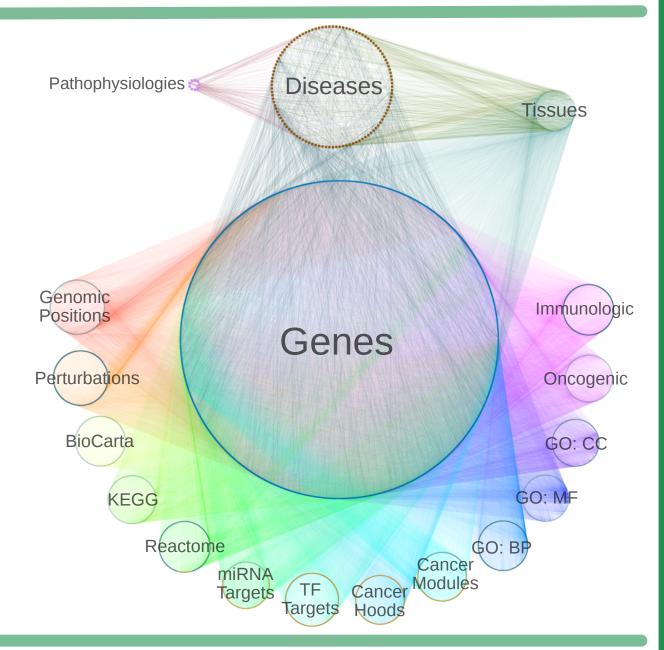
Daniel Himmelstein, Leo Brueggeman & Sergio Baranzini present Repurposing drugs on a heterogeneous network

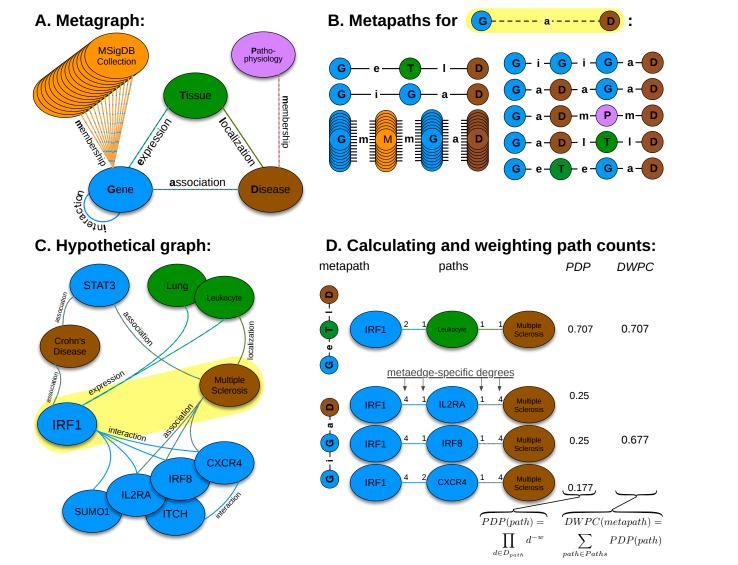
Last year, we introduced *heterogeneous network edge prediction* (HNEP) to <u>predict disease-associated genes</u>.

Heterogeneous networks contain multiple node and edge types.

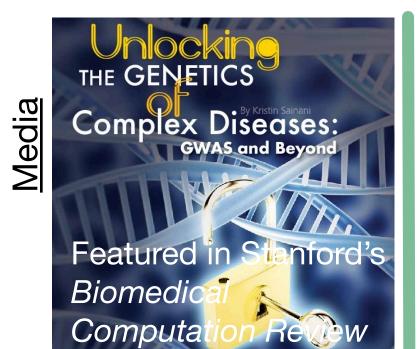
Our network contained 40,343 nodes (of 18 types) and 1,608,168 edges (of 19 types).

EP Method





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

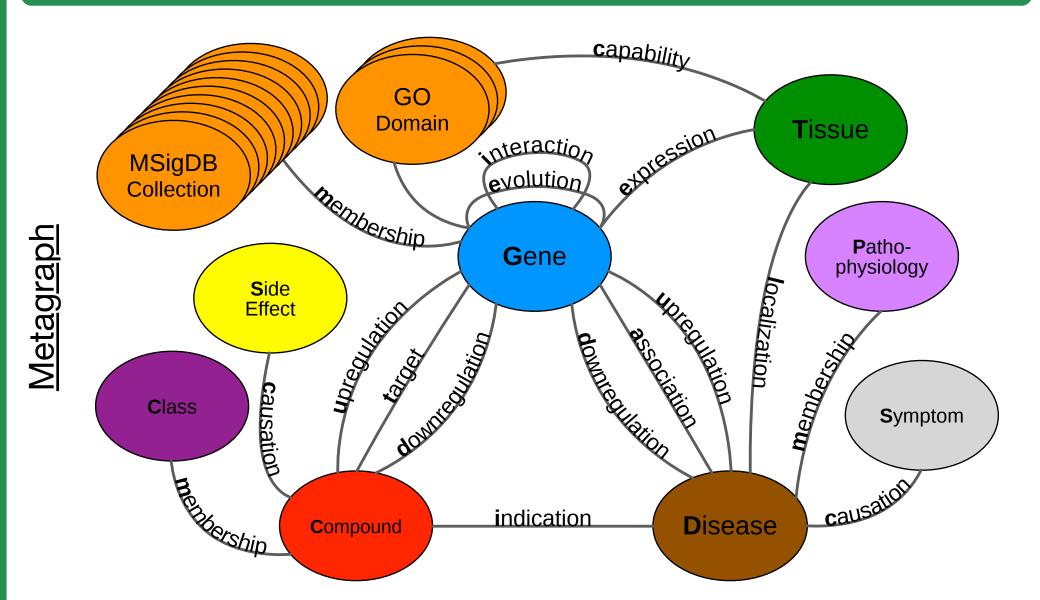


Predictions online at het.io

net.io Home	HNLP Disease Gen	i es → Media	Baranzini Lab ▼		
Prediction	s for <i>multipl</i>	e sclerc	sis		
	•	0 00/0/0	010		
Disease Onto	EFO EFO:0003885				
archart below show	ing each feature's contrib	ution to the over	all prediction. The existence	gene information. Clicking elsew of a GWAS-reported association	is indicated by status
where the levels indi-	cate whether the gene wa			r low-confidence (± LC-P) associ	
nnotation for a high				ns shows the number of disease	
nnotation for a high				ns shows the number of disease gene is showed by mean_predi	
nnotation for a high	ne is associated with. The				
nnotation for a high clerosis, that the ge	ne is associated with. The			gene is showed by mean_predi	
nnotation for a high sclerosis, that the ge	ne is associated with. The	average predic	tion across all diseases for a	gene is showed by mean_predi	ction.
nnotation for a high cclerosis, that the ge Show 10 : entri- gene_symbol	es \$\times \text{gene_code}\$	average predic	tion across all diseases for a	gene is showed by mean_predi Search: mean_prediction	ction.
annotation for a high cclerosis, that the ge Show 10 : entri- gene_symbol IL12A	es gene_code HGNC:5969	average predic	tion across all diseases for a to other_associations	gene is showed by mean_prediction \$\text{mean_prediction} \\ 4.312\%	oction .
annotation for a high scierosis, that the ge Show 10 : entri gene_symbol IL12A STAT4	es personal description of the second of th	average predic	other_associations 5	gene is showed by mean_prediction ### mean_prediction ### 4.312% ### 12.537%	⇒ prediction42.771%42.602%
nnotation for a high notation for a high clerosis, that the ge Show 10 : entri gene_symbol IL12A STAT4 IL12B	es \$\phi\$ gene_code HGNC:5969 HGNC:11365 HGNC:5970	average predic	other_associations 5 5	gene is showed by mean_predi Search:	prediction 42.771% 42.602% 41.675%
nnotation for a high clerosis, that the ge Show 10 : entri gene_symbol IL12A STAT4 IL12B PTPN2	es gene_code HGNC:5969 HGNC:11365 HGNC:5970 HGNC:9650	average predic	tion across all diseases for a the other_associations 2 5 5 4	gene is showed by mean_predi Search: \$ mean_prediction 4.312% 12.537% 24.642% 4.699%	prediction 42.771% 42.602% 41.675% 38.571%
unnotation for a high colerosis, that the ge Show 10 : entri gene_symbol IL12A STAT4 IL12B PTPN2 IL10	es continue to the continue t	status ± LC-P - + HC-P -	tion across all diseases for a the other_associations 2 5 5 4 4	gene is showed by mean_prediction \$ mean_prediction 4.312% 12.537% 24.642% 4.699% 8.114%	ction. prediction 42.771% 42.602% 41.675% 38.571% 38.345%
Innotation for a high colerosis, that the ge Show 10 : entri gene_symbol IL12A STAT4 IL12B PTPN2 IL10 IL2RA	es gene_code HGNC:5969 HGNC:5970 HGNC:5962 HGNC:6008	status ± LC-P - + HC-P -	tion across all diseases for a to other_associations 2 5 4 4 5	gene is showed by mean_prediction Search:	\$\text{prediction}\$ \$\perp \text{prediction}\$ 42.771% 42.602% 41.675% 38.571% 38.345% 36.025%
Innotation for a high clerosis, that the ge Show 10 : entri gene_symbol IL12A STAT4 IL12B PPTPN2 IL10 IL2RA IRF5	es	status ± LC-P - + HC-P -	other_associations other_associations cut of ther_associations cut of ther_associations cut of ther_associations cut of ther_associations cut of therase of the thera	gene is showed by mean_prediction 4.312% 12.537% 24.642% 4.699% 8.114% 18.681% 8.140%	prediction 42.771% 42.602% 41.675% 38.571% 36.025% 31.477%

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

Planning the Network Construction



	Туре	Resource	
	Compound	DrugBank	
	Disease	Disease Ontology	
9 S	Gene	Entrez Gene	
Nodes	Tissue	Uberon	
\mathbf{z}	Gene Set	MSigDB	
	Side Effect	UMLS	
	Pathophysiology	Manual	
	Symptom	MeSH	

Standardized terminologies:

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

Source	Target	Туре	Resource	
Compound	Disease	Indication	MEDI	
Compound	Disease	Indication	LabeledIn	Ideal resources are:
Compound	Gene	Expression	LINCS	 high-throughput
Compound	Side Effect	Causation	SIDER 2	 systematic
Compound	Side Effect	Causation	OFFSIDES	 unbiased
Disease	Gene	Target	ChEMBL	 diverse in aggregate
Disease	Gene	Association	GWAS Catalog	
Disease	Gene	Expression	STAR-GEO	
Disease	Pathophysiology	Membership	Manual	
Disease	Symptom	Causation	Human symptom	sdisease network
Gene	Gene	Interaction	Human Interacto	me Project
Gene	Gene	Interaction	The Incomplete I	nteractome
Gene	Gene	Evolution	Evolutionary Rate	e Covariation
Gene	Gene Set	Membership	MSigDB	
Gene	Tissue	Expression	GNF Gene Expre	ession Atlas

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

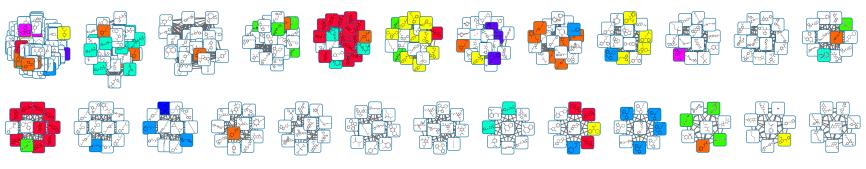


ThinkLab is:

- 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).

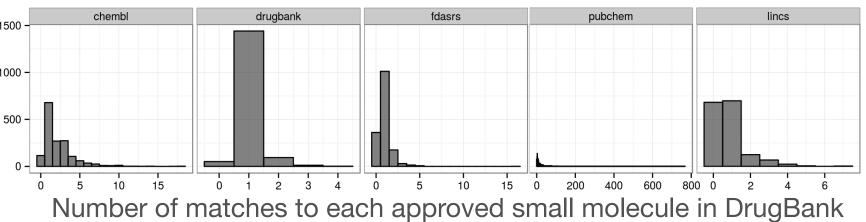
git.dhimmel.com/SIDER2

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

We mapped compound vocabularies to DrugBank using **UniChem** to enable fuzzy matching.



Number of matches to each approved small molecule in DrugBan git.dhimmel.com/drugbank/unichem-map.html

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.