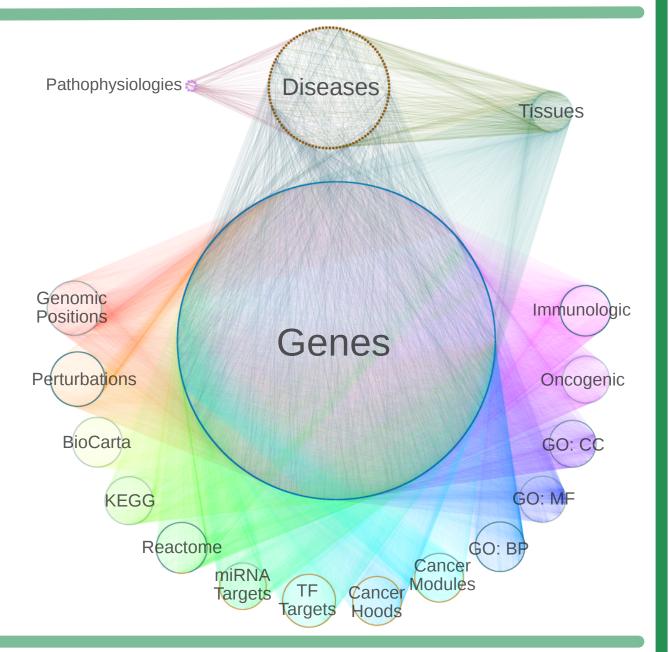
# Daniel Himmelstein, Leo Brueggeman & Sergio Baranzini present Repurposing drugs on a heterogeneous network

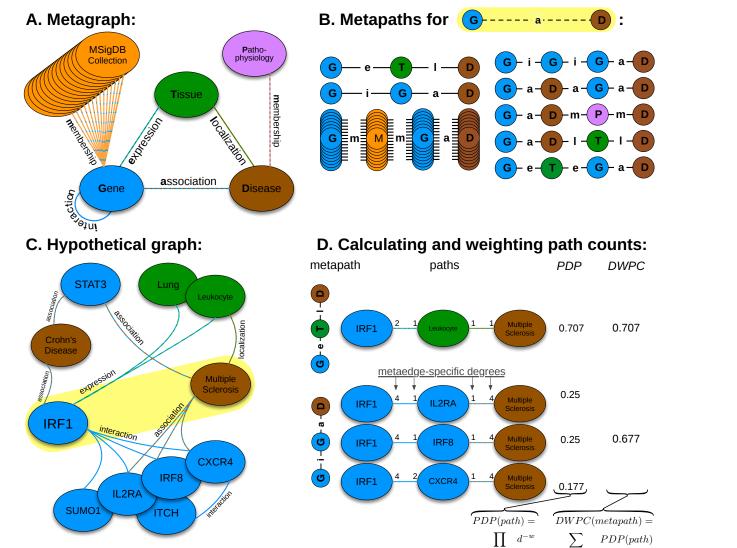
Last year, we introduced *heterogeneous network edge prediction* (HNEP) to <u>predict disease-associated genes</u>.

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]

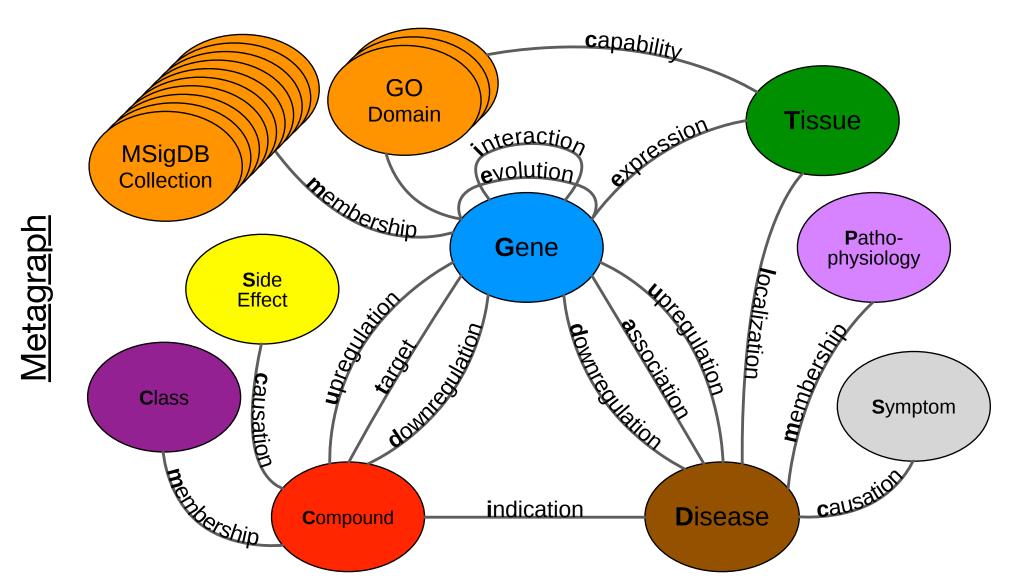
THE GENETICS
Complex Diseases:
GWAS and Beyond
Featured in Stanford's
Biomedical
Computation Review

### Predictions online at het.io

net.io Home	HNLP Disease Ger	nes ▼ Media	Baranzini Lab ▼			
Prediction	s for <i>multip</i>	le sclero	sis			
	•	10 001010	010			
Disease Onto	logy DOID:2377 EFO EFO:0003885					
UMLS Conc	epts C0026769					
where the levels indic	(+ HC-S) or low-confider	nce (+ LC-S) asso	ciation other association	shows the number of diseas	es excluding	multiple
annotation for a high sclerosis, that the ge	ne is associated with. The			s shows the number of diseas gene is showed by mean_pred Search:		multiple
annotation for a high	ne is associated with. The	e average predict		gene is showed by mean_pred		
annotation for a high sclerosis, that the gel	ne is associated with. The	e average predict	ion across all diseases for a	gene is showed by mean_pred	oction.	
annotation for a high sclerosis, that the gereshow 10 : entries	es  \$\phi\$ gene_code	e average predict	ion across all diseases for a	gene is showed by mean_pred  Search:  mean_prediction	⇒ prec 42.7	diction
annotation for a high sclerosis, that the get Show 10 : entrice gene_symbol	es  \$\text{gene_code}\$  HGNC:5969	e average predict	on across all diseases for a  other_associations	gene is showed by mean_pred  Search:  mean_prediction  4.312%	<ul><li>prec</li><li>42.7</li><li>42.6</li></ul>	diction
annotation for a high sclerosis, that the get Show 10 : entri gene_symbol IL12A STAT4	es  \$\phi\$ gene_code  HGNC:5969  HGNC:11365	e average predict	on across all diseases for a  other_associations 2 5	gene is showed by mean_pred  Search:  mean_prediction 4.312% 12.537%		diction 771% 602%
annotation for a high scierosis, that the get Show 10 : entrice gene_symbol IL12A STAT4 IL12B	es  \$\phi\$ gene_code  HGNC:5969  HGNC:11365  HGNC:5970	e average predict	on across all diseases for a  other_associations  2  5  5	gene is showed by mean_pred  Search:	<ul><li>prec</li><li>42.7</li><li>42.6</li><li>41.6</li><li>38.5</li></ul>	diction 771% 602%
annotation for a high scierosis, that the ger Show 10 : entrice gene_symbol IL12A STAT4 IL12B PTPN2	pe is associated with. The  gene_code  HGNC:5969  HGNC:11365  HGNC:5970  HGNC:9650	e average predict	ion across all diseases for a  the other associations  2  5  5  4	Search:  Search:  mean_prediction  4.312%  12.537%  24.642%  4.699%		diction 771% 602% 671%
annotation for a high scierosis, that the get Show 10 : entric gene_symbol IL12A STAT4 IL12B PTPN2 IL10	es  pene_code  HGNC:5969  HGNC:5970  HGNC:9650  HGNC:5962	a status b LC-P HC-P	ion across all diseases for a  the other_associations  2  5  5  4  4	Search:  Search:  mean_prediction  4.312%  12.537%  24.642%  4.699%  8.114%	\$\rightarrow\$ prec         42.7         42.6         41.6         38.5         38.3         36.0	diction 771% 602% 675% 671%
annotation for a high scolerosis, that the get Show 10 : entrice gene_symbol IL12A STAT4 IL12B PTPN2 IL10 IL2RA	es  gene_code  HGNC:5969  HGNC:5970  HGNC:9650  HGNC:5962  HGNC:6008	a status b LC-P HC-P	on across all diseases for a  the other_associations  continuous c	Search:  Search:  mean_prediction  4.312%  12.537%  24.642%  4.699%  8.114%  18.681%	\$\rightarrow\$ prec         42.7         42.6         41.6         38.5         38.3         36.0         31.4	diction 771% 602% 675% 671% 645%
annotation for a high sockerosis, that the geressis, that the geressis with the gene symbol ill. 12A STATA IL. 12B PTPN2 IL. 10 IL. 2RA IRF5	es	a status b LC-P HC-P	on across all diseases for a  other_associations  2  5  4  4  5  5	Search:  pean_prediction 4.312% 12.537% 24.642% 4.699% 8.114% 18.681% 8.140%	prec 42.7 42.6 41.6 38.5 38.3 36.0 31.4 29.4	diction 771% 602% 675% 671% 645% 625% 6477%

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	
SS	Gene	Entrez Gene	
Nodes	Tissue	Uberon	
$\geq$	Gene Set	MSigDB	
	Side Effect	UMLS	
	Pathophysiology	Manual	
	Symptom	MeSH	

Standardized terminologies:

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

Sou	ırce	Target	Туре	Resource	
	npound	Disease	Indication	MEDI	
	npound	Disease	Indication	LabeledIn	Ideal resources are:
Con	npound	Gene	Expression	LINCS	<ul> <li>high-throughput</li> </ul>
Con	npound	Side Effect	Causation	SIDER 2	systematic
Con	npound	Side Effect	Causation	OFFSIDES	<ul> <li>unbiased</li> </ul>
Dise	ease	Gene	Target	ChEMBL	<ul> <li>aggregately diverse</li> </ul>
Dise	ease	Gene	Association	GWAS Catalog	
Dise	ease	Gene	Expression	STAR-GEO	
Dise	ease	Pathophysiology	Membership	Manual	
Dise	ease	Symptom	Causation	Human symptom	nsdisease network
Gen	ne	Gene	Interaction	Human Interacto	me Project
Gen	ne	Gene	Interaction	The Incomplete I	nteractome
Gen	ne	Gene	Evolution	Evolutionary Rate	e Covariation
Gen	ne	Gene Set	Membership	MSigDB	
Gen	ne	Tissue	Expression	GNF Gene Expre	ession Atlas

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

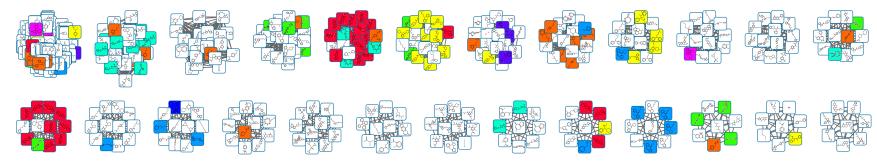


### ThinkLab is:

- 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).

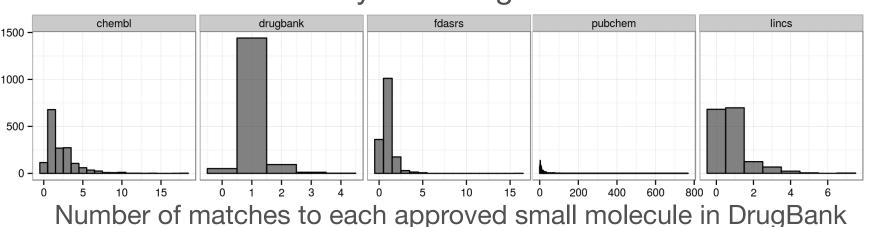
git.dhimmel.com/SIDER2

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.