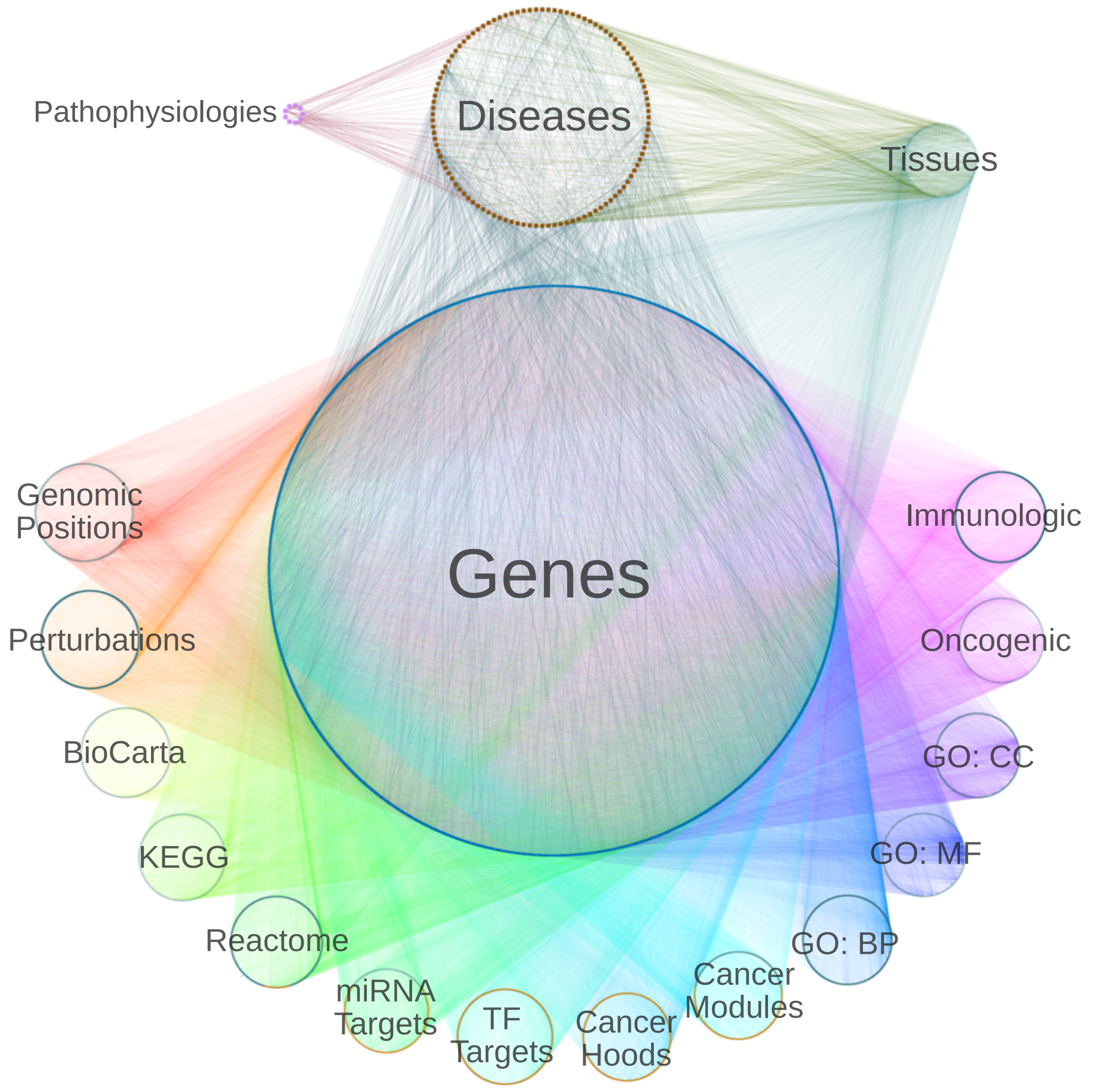
Daniel Himmelstein, Leo Brueggeman & Sergio Baranzini present *Repurposing drugs on a heterogeneous network*

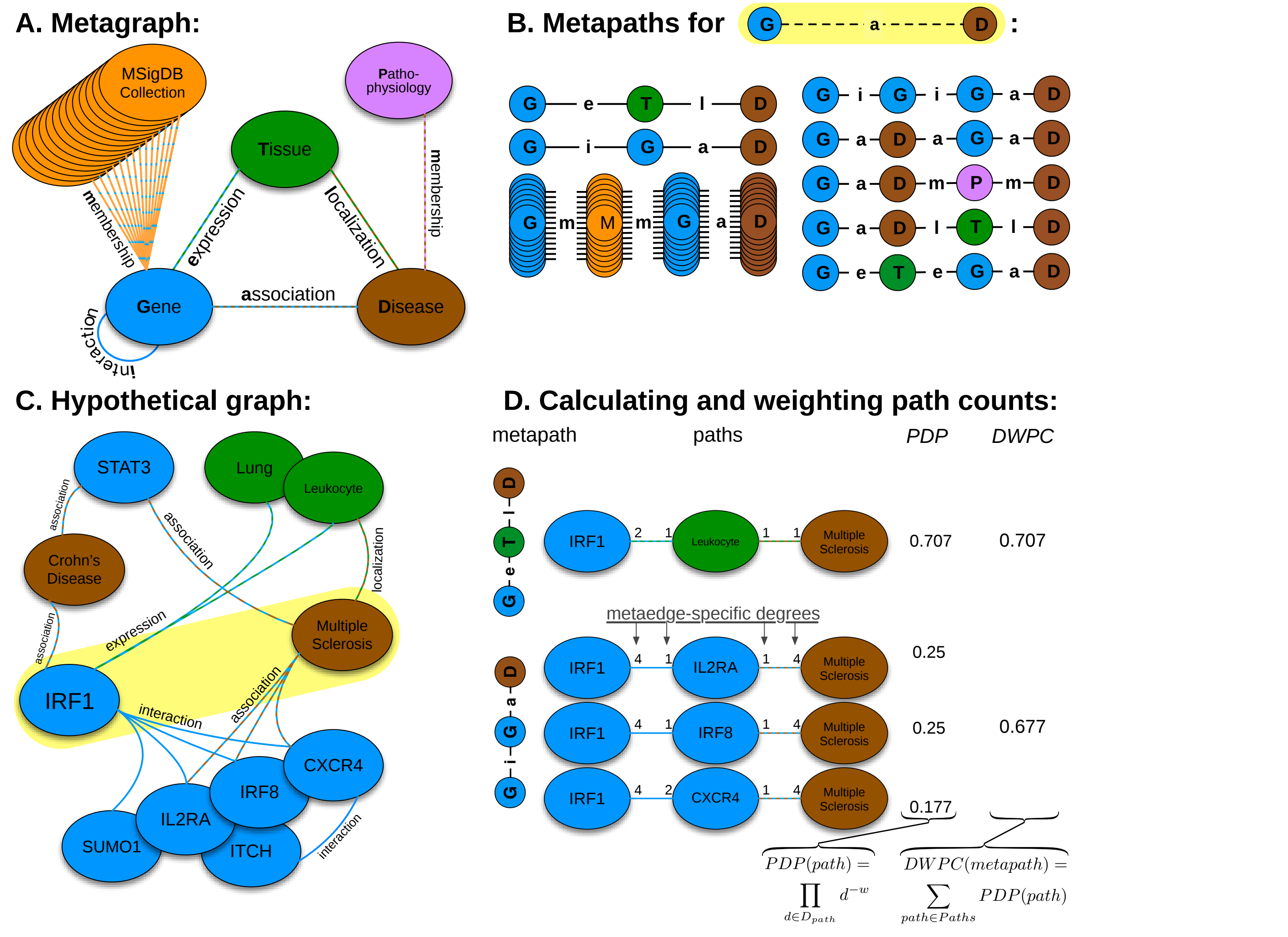
Last year, we introduced *heterogeneous network edge prediction* (HNEP) to predict disease-associated genes.

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).



HNEP Method



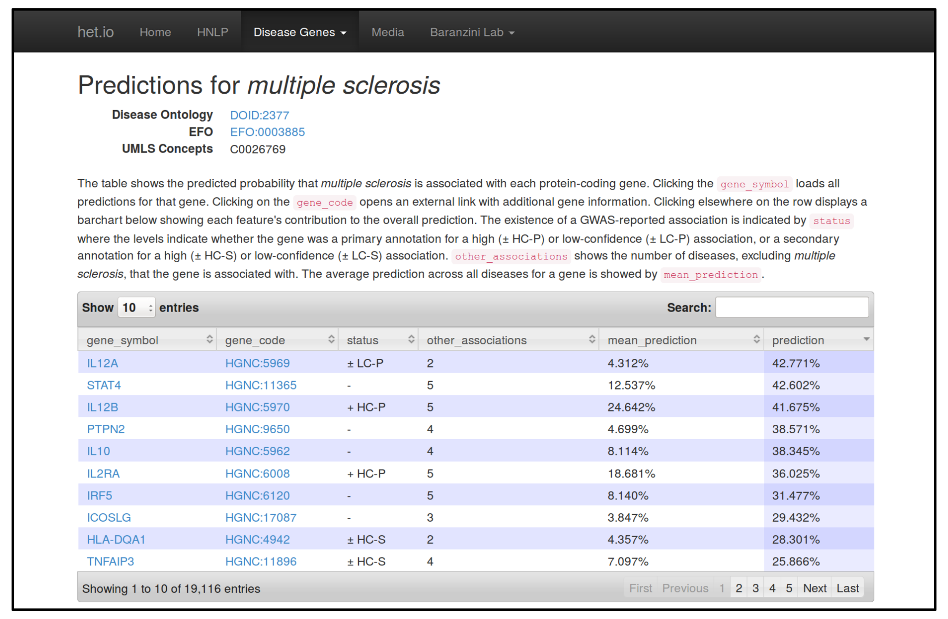
$$PDP(path) = \prod_{d \in D_{path}} d^{-w}$$

$$DWPC(metapath) = \sum_{path \in Paths} PDP(path)$$

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

Media

Predictions online at het.io



Predictions for *multiple sclerosis*

Disease Ontology DOID:2377
EFO EFO:0003885
UMLS Concepts C0026769

The table shows the predicted probability that *multiple sclerosis* is associated with each protein-coding gene. Clicking the `gene_symbol` loads all predictions for that gene. Clicking on the `gene_code` opens an external link with additional gene information. Clicking elsewhere on the row displays a barchart below showing each feature's contribution to the overall prediction. The existence of a GWAS-reported association is indicated by `status` where the levels indicate whether the gene was a primary annotation for a high (± HC-P) or low-confidence (± LC-P) association, or a secondary annotation for a high (± HC-S) or low-confidence (± LC-S) association. `other_associations` shows the number of diseases, excluding *multiple sclerosis*, that the gene is associated with. The average prediction across all diseases for a gene is showed by `mean_prediction`.

Show 10 entries Search:

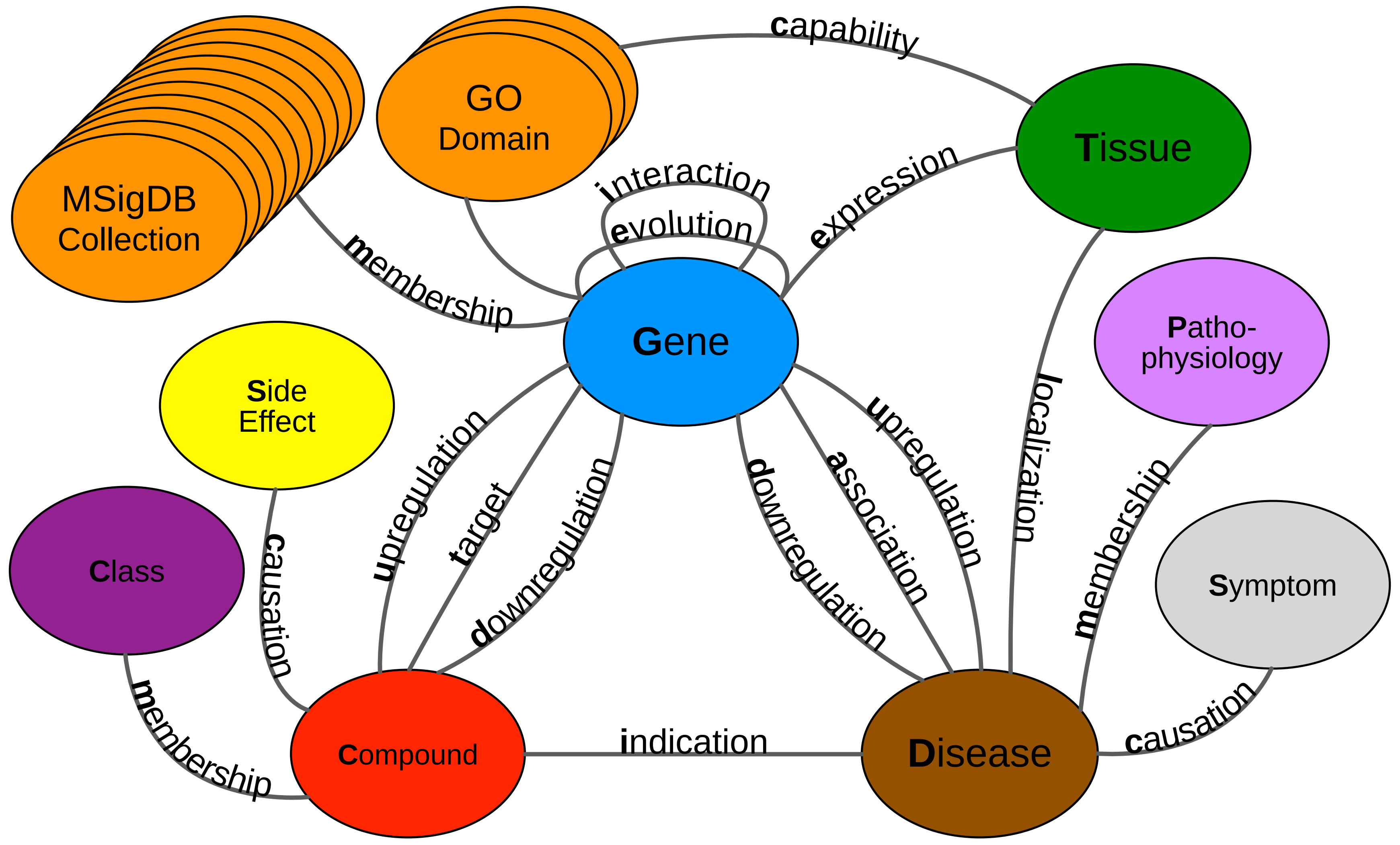
gene_symbol	gene_code	status	other_associations	mean_prediction	prediction
IL12A	HGNC:5969	± LC-P	2	4.312%	42.771%
STAT4	HGNC:11365	-	5	12.537%	42.602%
IL12B	HGNC:5970	+ HC-P	5	24.642%	41.675%
PTPN2	HGNC:9650	-	4	4.699%	38.571%
IL10	HGNC:5962	-	4	8.114%	38.345%
IL2RA	HGNC:6008	+ HC-P	5	18.681%	36.025%
IRF5	HGNC:6120	-	5	8.140%	31.477%
ICOSLG	HGNC:17087	-	3	3.847%	29.432%
HLA-DQA1	HGNC:4942	± HC-S	2	4.357%	28.301%
TNFAIP3	HGNC:11896	± HC-S	4	7.097%	25.866%

Showing 1 to 10 of 19,116 entries First Previous 1 2 3 4 5 Next Last

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

Planning the Network Construction

Metagraph



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

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

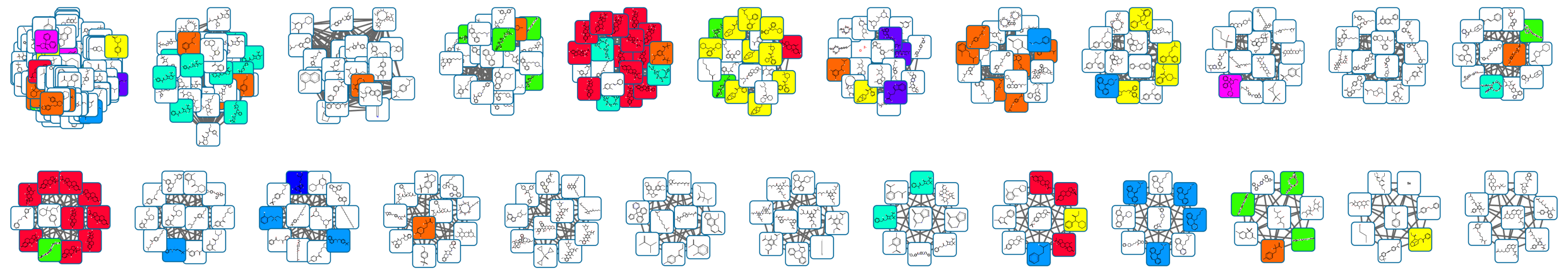
ThinkLab
thinklab.com/p/rephetio
doi:10.15363/thinklab.4

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)

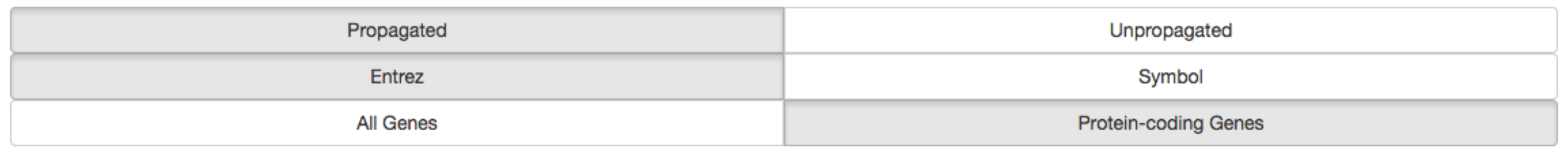
We analyzed **SIDER 2** and investigated its strengths and weaknesses as well as pharmacological utility.



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

git.dhimmel.com/SIDER2

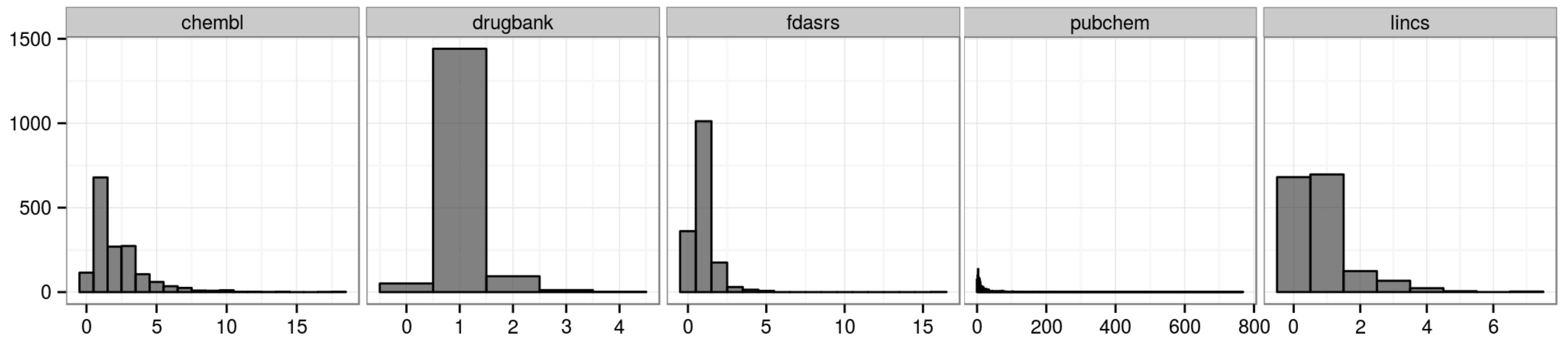
We created a user-friendly service to retrieve **Gene Ontology annotations** with optional propagation.



Propagated	Unpropagated
Entrez	Symbol
All Genes	Protein-coding Genes

git.dhimmel.com/gene-ontology

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



Number of matches to each approved small molecule in DrugBank

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.