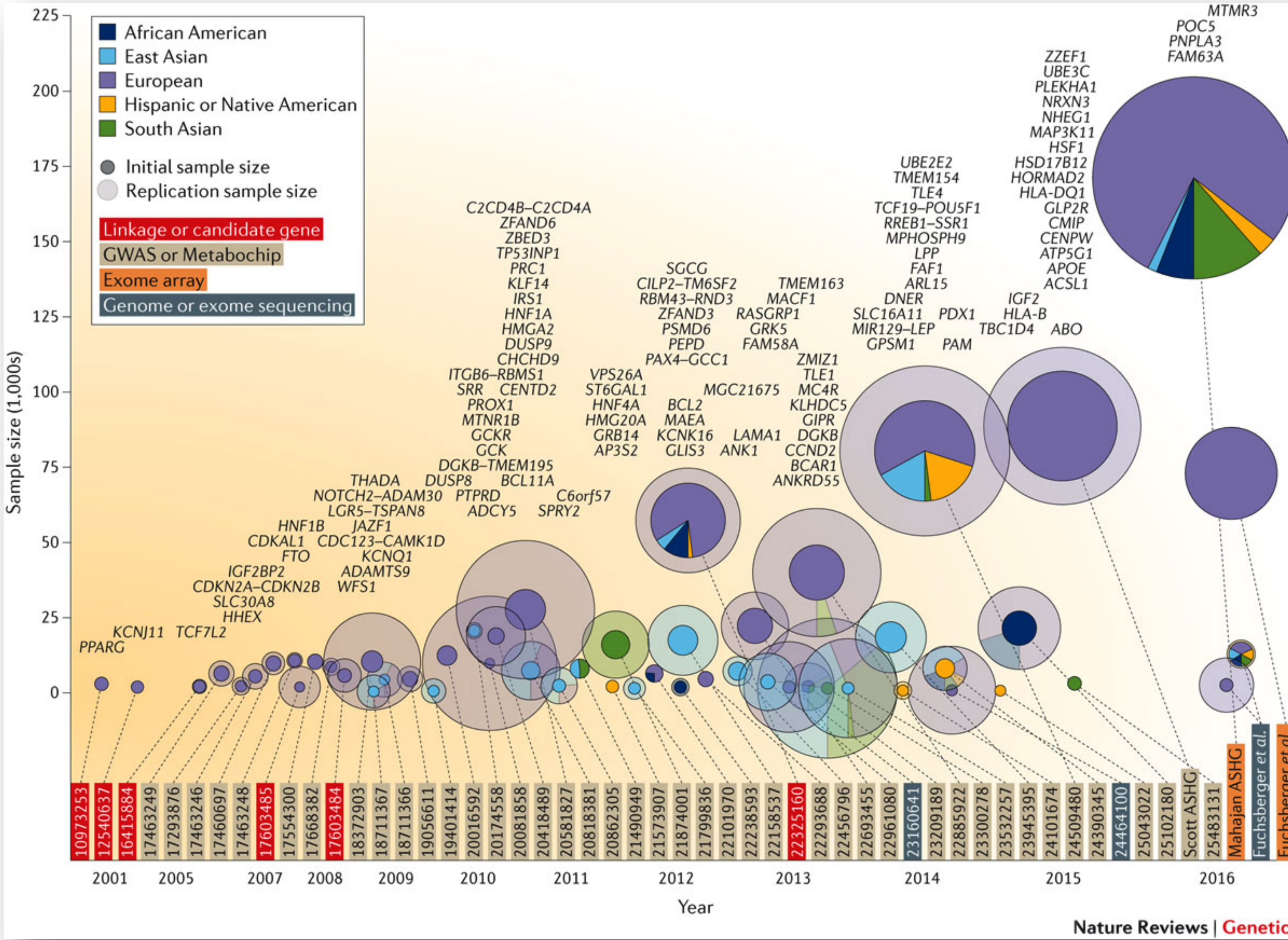
The DisGeNET knowledge management platform for disease genomics

Laura I. Furlong

Research Programme on Biomedical Informatics (GRIB)
Hospital del Mar Medical Research Institute (IMIM)
Pompeu Fabra University (UPF)
ELIXIR-ES



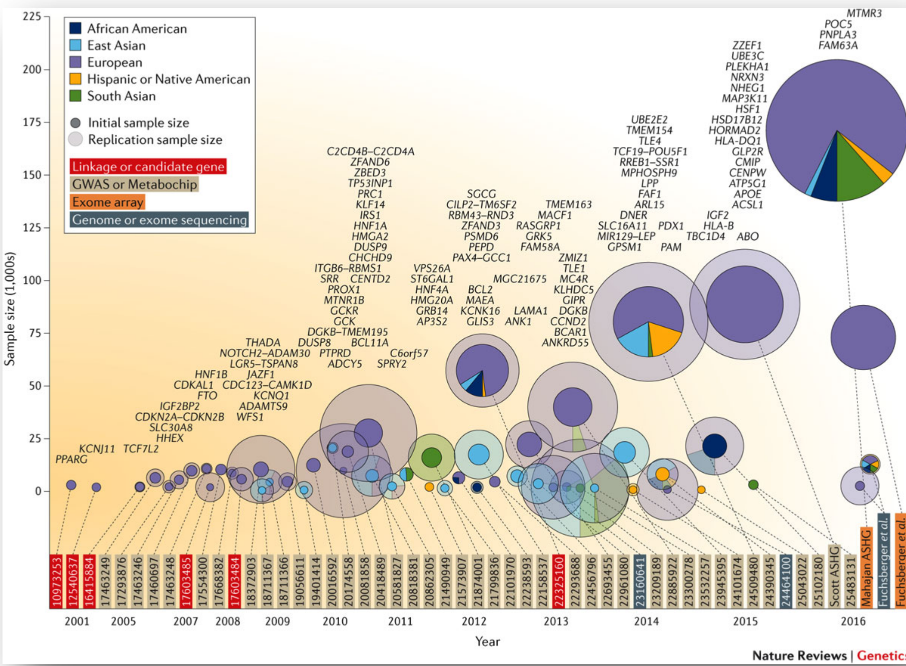
Flannick, J., & Florez, J. C. (2016). Type 2 diabetes: genetic data sharing to advance complex disease research. *Nature Reviews Genetics*, *17*(9), 535.

DIAGRAM Consortium

900K individuals
27M SNPs

Inventory of T2D variants

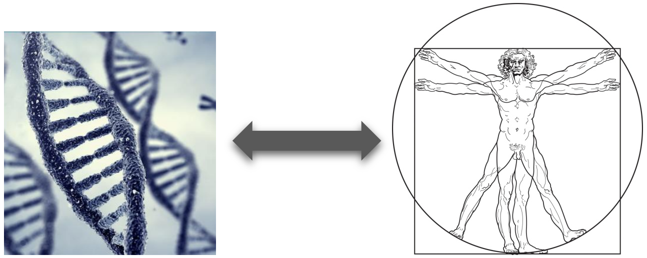
243 loci at genome-wide significance, including 135 new loci for type 2 diabetes



Flannick, J., & Florez, J. C. (2016). Type 2 diabetes: genetic data sharing to advance complex disease research. *Nature Reviews Genetics*, *17*(9), 535.

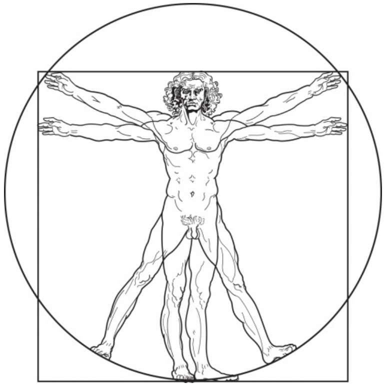
http://www.diagram-consortium.org/

2018



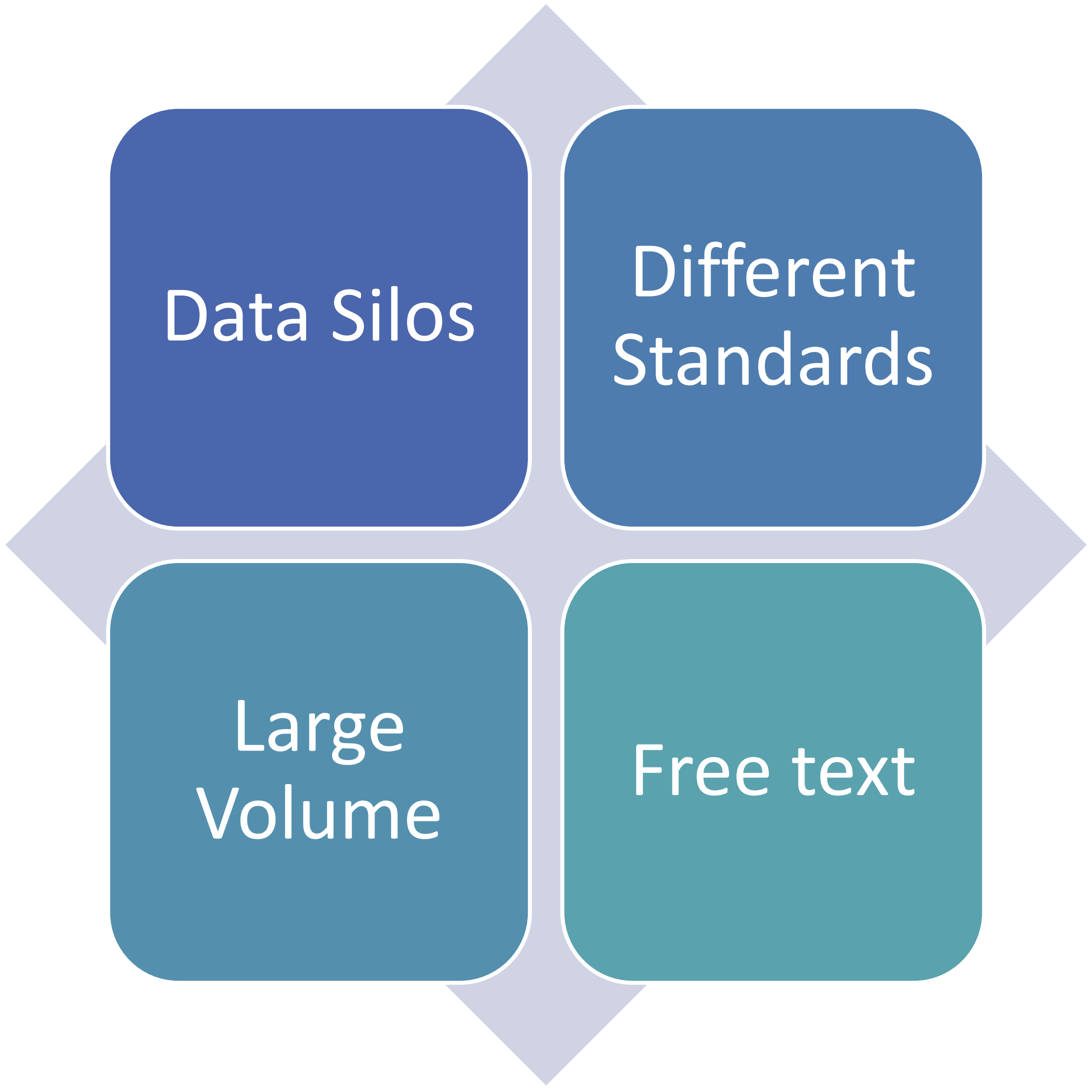
Genomics studies are generating a vast volume of data, claiming for solutions for **data management**, **data interoperability** and **knowledge extraction** for **genotype-phenotype** data.

The accumulation of large-scale data requires the development of **computational tools** able to explore and mine the vast amount **of biological knowledge** they contain.



genotype

phenotype


Data Silos
Different Standards
Large Volume
Free text

From genotype to phenotype: data silos

#277900

WILSON DISEASE

OMIM
Online Mendelian Inheritance in Man

Alternative titles; symbols

WND; WD

HEPATOLENTICULAR DEGENERATION

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance (in progress)	Phenotype mapping key	Gene/Locus
13q14.3	Wilson disease	277900	AR	3	ATP7B

orphanet

Rare diseases

- Search
- Search by sign
- Classifications
- Genes
- Disability
- Encyclopaedia for patients
- Encyclopaedia for professionals
- Emergency guidelines

Hepatolenticular Degeneration

Basics | Chemical–Gene Interactions | Chemicals | Genes

ctd™

1–50 of 11,475 results.

First | Previous | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Next | La

	Gene	Disease	Direct Evidence
1.	CP	Hepatolenticular Degeneration	M
2.	ATP7B	Hepatolenticular Degeneration	M
3.	PRNP	Hepatolenticular Degeneration	M

Genes related to Wilson Disease

Genes related to Wilson Disease (1 elite genes):

★ - Elite gene

id	Symbol ★	Description
1	ATP7B ★	ATPase, Cu++ transporting, beta polypeptide
2	CP	ceruloplasmin (ferroxidase)
3	ATP7A	ATPase, Cu++ transporting, alpha polypeptide
4	COMMD1	copper metabolism (Murr1) domain containing 1
5	ARSA	arylsulfatase A
6	HFE	hemochromatosis
7	SLC31A1	solute carrier family 31 (copper transporter), member 1

MalaCards
HUMAN DISEASE DATABASE

GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

Search the catalog

Examples: breast cancer, rs7329174, Yang, 2q37.1, HBS1L, 6:16000000-25000000

TATATCT ACCTCAC ClinVar

	Gene(s)	Condition(s)
ATP7B, 1-BP DEL, 2511A	ATP7B	Wilson disease
ATP7B, 3-BP DEL, 3892GTC	ATP7B	Wilson disease
ATP7B, 15-BP DEL, NT-441	ATP7B	Wilson disease

From genotype to phenotype: standards

GENE

- ✓ **Lipocalin 2**
- ✓ 24p3
- ✓ 25 KDa Alpha-2-Microglobulin-Related Subunit Of MMP-9
- ✓ HNL
- ✓ lipocalin 2 (oncogene 24p3)
- ✓ Lipocalin-2
- ✓ Migration-Stimulating Factor Inhibitor
- ✓ MSFI
- ✓ neutrophil gelatinase-associated lipocalin
- ✓ NGAL
- ✓ oncogene 24p3
- ✓ P25
- ✓ Siderocalin

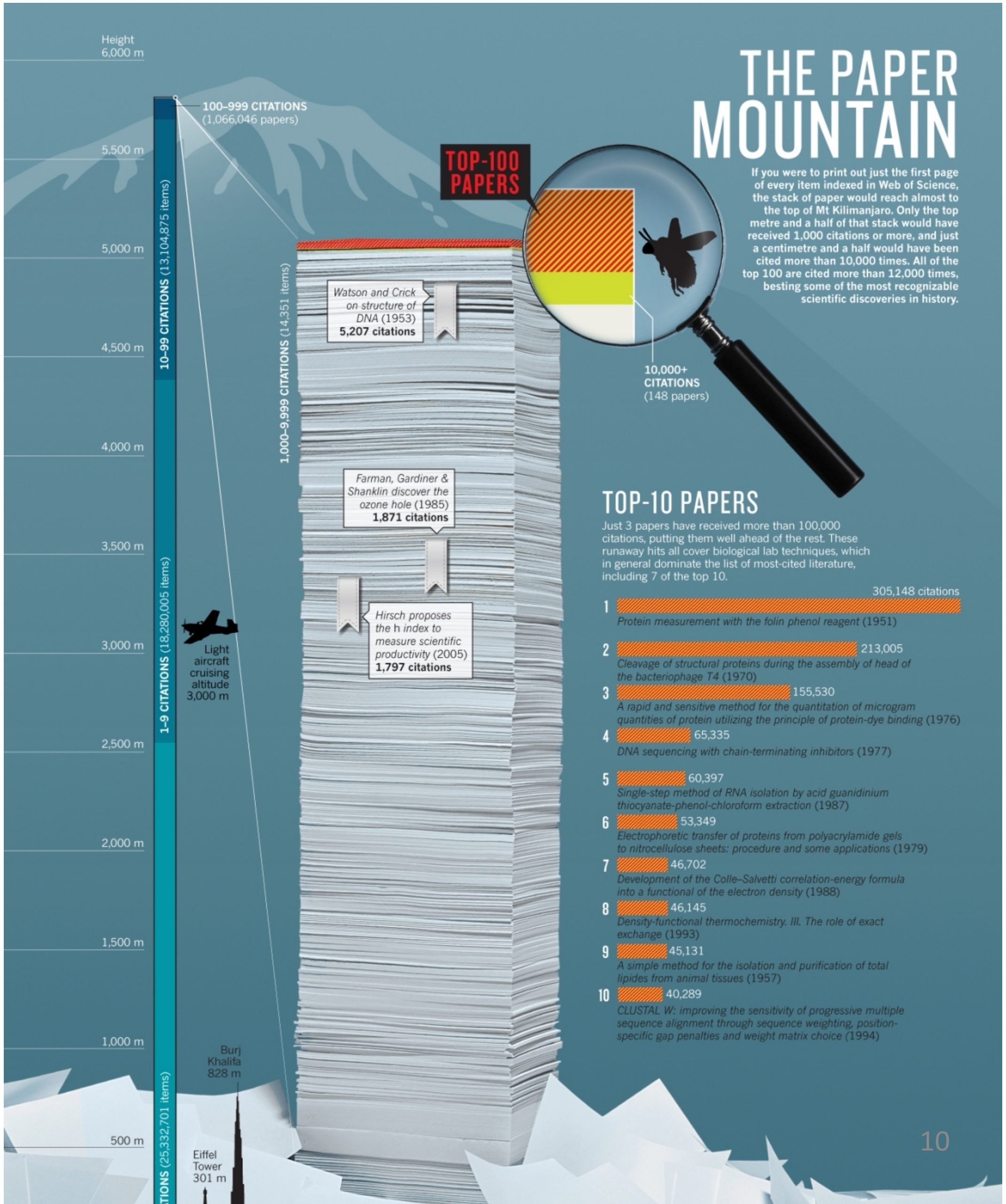
DISEASE

- ✓ **Wilson's disease**
- ✓ Cerebral Pseudosclerosis
- ✓ Copper Storage Disease
- ✓ Hepatic Form of Wilson Disease
- ✓ Hepato-Neurologic Wilson Disease
- ✓ Hepatocerebral Degeneration
- ✓ Hepatolenticular degeneration
- ✓ Kinnier-Wilson Disease
- ✓ Neurohepatic Degeneration
- ✓ Progressive Lenticular Degeneration
- ✓ Pseudosclerosis
- ✓ WD
- ✓ Westphal-Strumpell Syndrome
- ✓ Wilson Disease
- ✓ Wilson Disease, Hepatic Form

From genotype to phenotype: data volume

From genotype to phenotype: Free text

- ✓ 25,000 peer-reviewed journals
- ✓ 2.5M articles published per year
- ✓ 2 papers/minute in life sciences
- ✓ 1 article/hour about diseases and genes



Van Noorden, et al. 2014 doi:10.1038/514550a
Burger, et al (2014).

Data Silos

Different Standards

DisGeNET

Large Volume

Free text

RB1 Protein | Bladder cancer

UniProt

RB is overexpressed in bladder cancer samples as measured by….

PMID 1234567

Provenance

Data extraction and standardization

RB1

Carcinoma of bladder

AlteredExpression

Gene associated Disease

DisGeNET

Gene-Disease Association (GDA)

GENE ⟷ DISEASE

Variant-Disease Association (VDA)

VARIANT ⟷ DISEASE

- ✓ Standards
- ✓ Provenance
- ✓ Tools

genotype ⟷ **phenotype**

- Large in scale and growing rapidly (NGS)
- Large studies on genetics of disease available
- HGVS standard for sequence variation nomenclature
- Standards for data exchange
- UniProt, NCBI, Ensembl
- VarioML, VariO

- Phenotype data spans a wide spectrum of possible observations about an individual
- More difficult to capture and to standardize
- Human Phenotype Ontology, Disease Ontology
- Broad phenotype categories used in many studies

Standards in DisGeNET

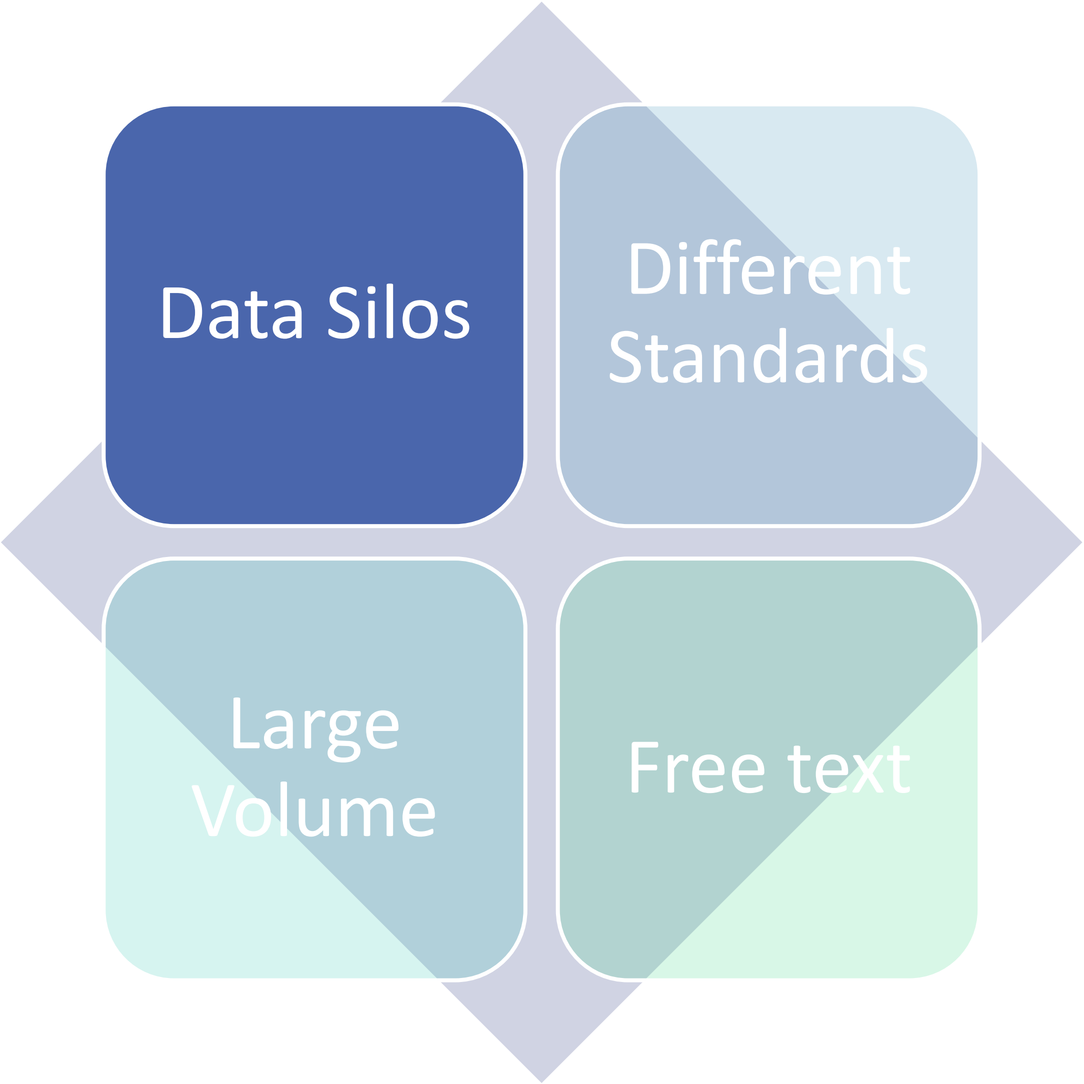
genotype ⟷ phenotype

- Gene, protein, SNPs
- Official Gene symbol
- NCBI Gene Id
- Uniprot accession
- dbSNP identifier for variants

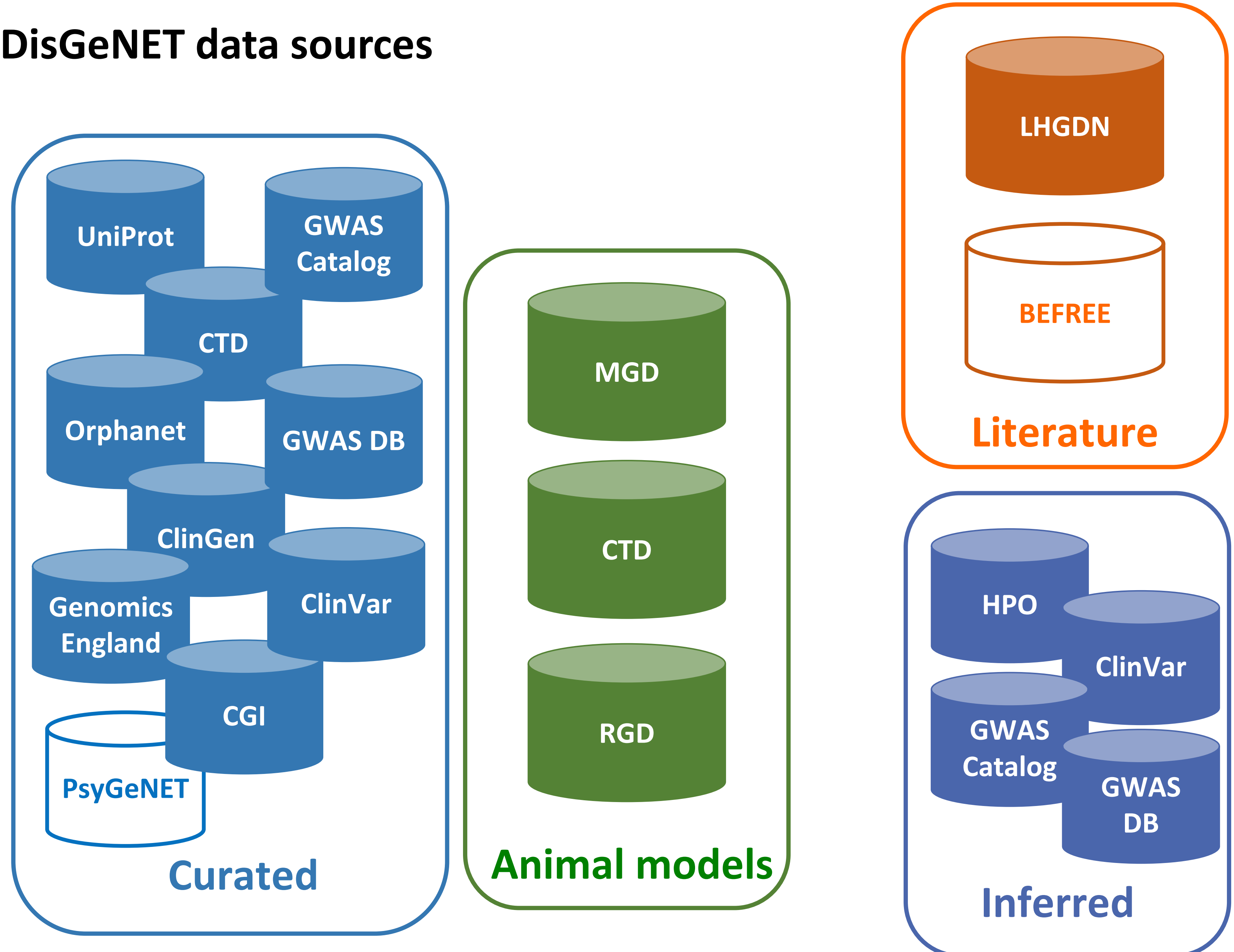
- Diseases, symptoms, phenotypes
- UMLS CUIs
- UMLS semantic types
- Disease Ontology
- Mappings to a variety of phenotype vocabularies and ontologies

DisGeNET association type ontology

DisGeNET data sources

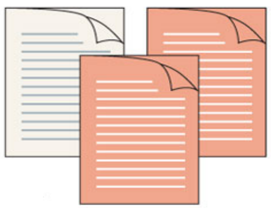


DisGeNET data sources



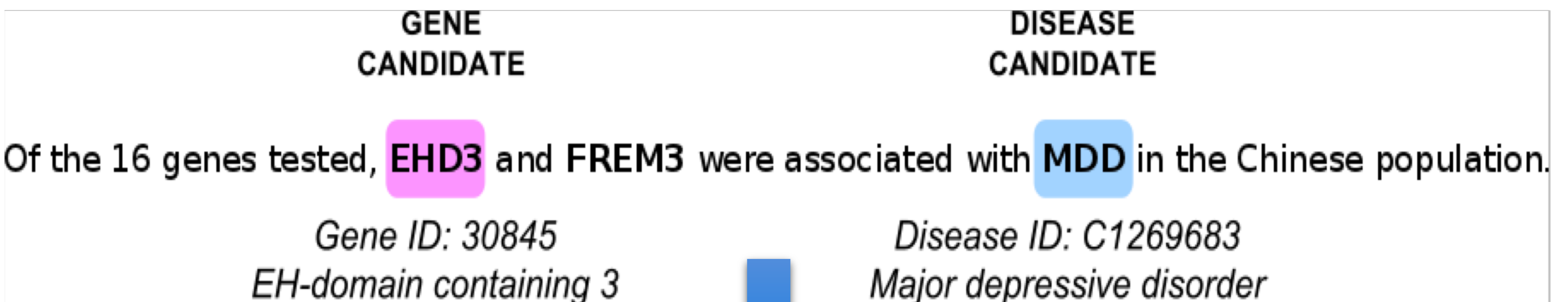
Text mining of GDAs and VDAs

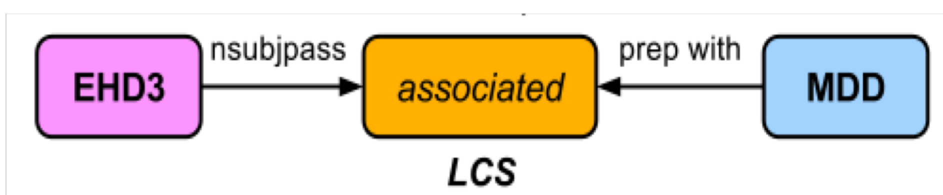
MEDLINE articles



GENE ⟷ DISEASE

VARIANT ⟷ DISEASE





Association types classified according to the DisGeNET ontology

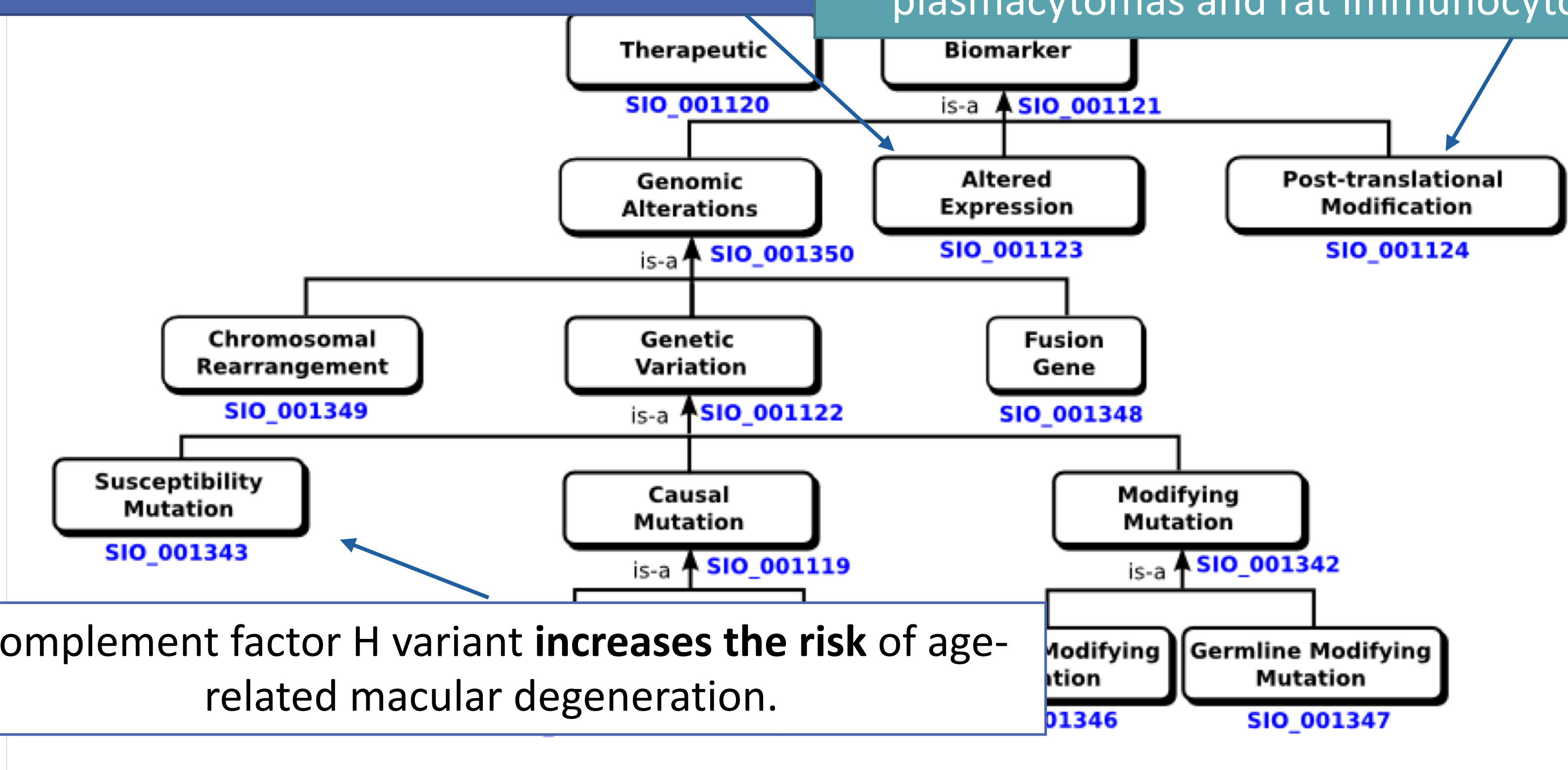
DisGeNET association type on

Gpc3 expression correlates with the phenotype of the Simpson-Golabi-Behmel syndrome.

phosphorylation state of Ser-129 in human alpha-synuclein determines neurodegeneration in a rat model of Parkinson disease

Unbalanced GLA mRNAs ratio quantified by real-time PCR in Fabry patients' fibroblasts results in Fabry disease.

The amino-terminal **phosphorylation sites** of C-MYC are frequently mutated in Burkitt's lymphoma lines but not in mouse plasmacytomas and rat immunocytomas.



Complement factor H variant **increases the risk** of age-related macular degeneration.

DisGeNET statistics

GENE ⟷ DISEASE

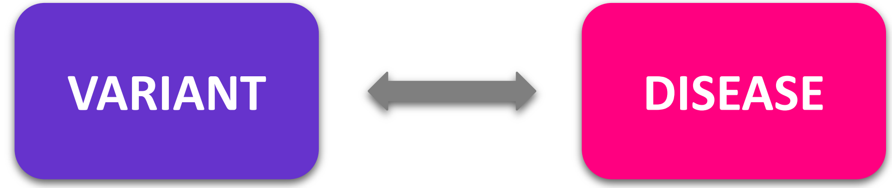
Gene-Disease Associations (GDAs)

Source	Genes	Diseases*	Associations
Curated	9413	10370	81746
Animal Models	2795	2789	11517
Inferred	8700	13176	163626
Literature	15283	12418	415583
All	**17549**	**24163**	**628685**

*diseases, traits, symptoms, disease groups

66 % are GDAs exclusively provided by BeFree

DisGeNET statistics

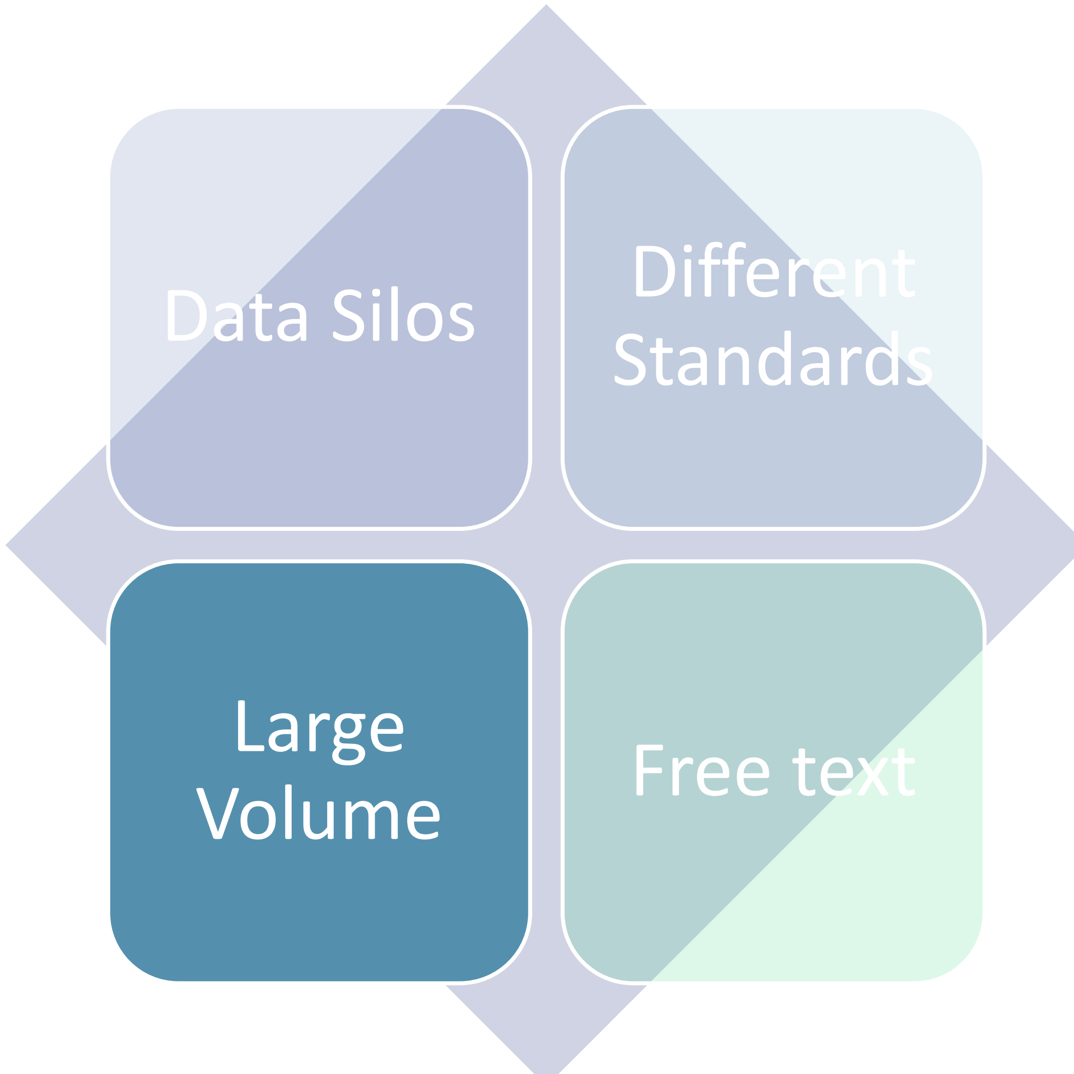


Variant-Disease Associations (VDAs)

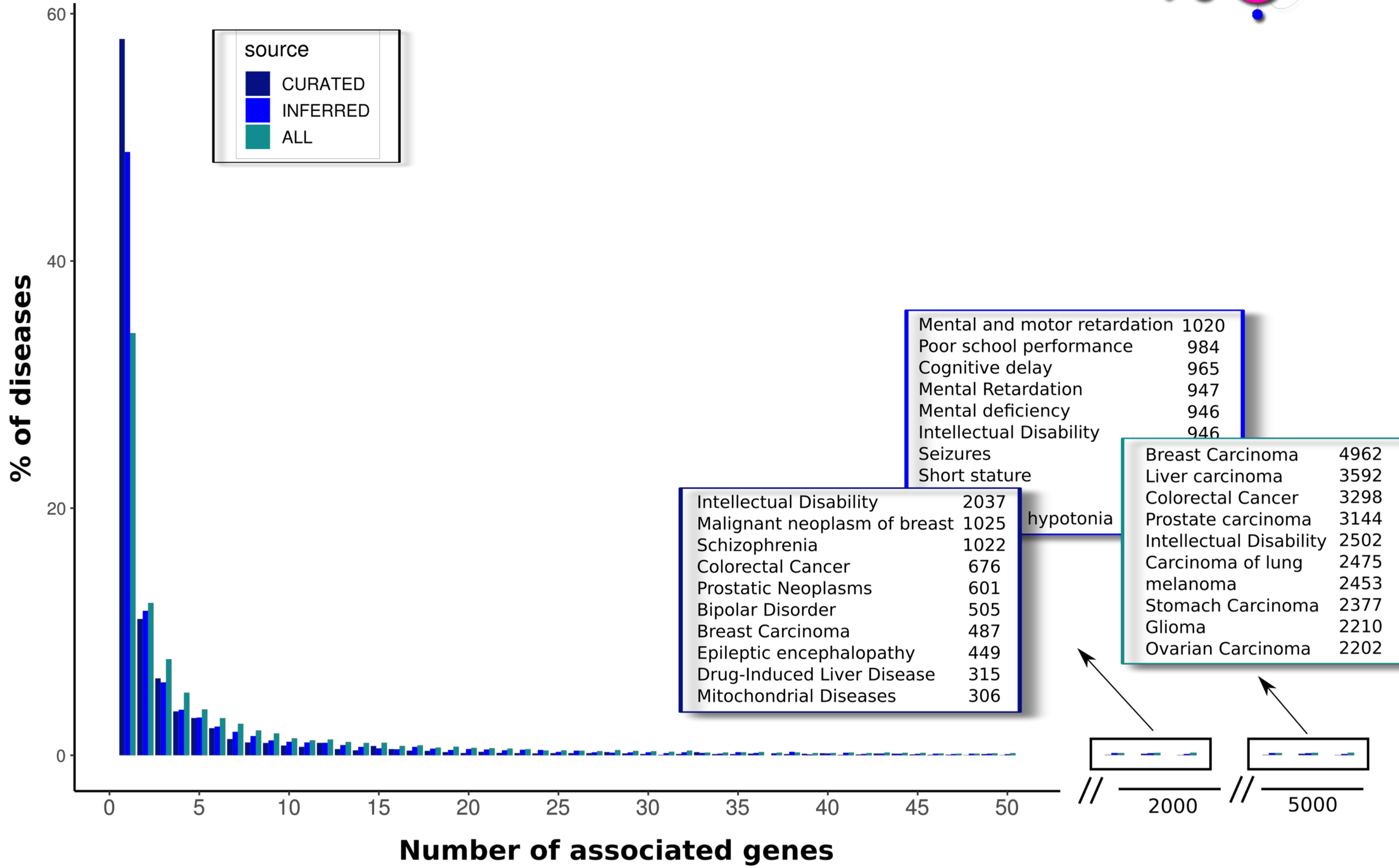
Source	Variants	Diseases*	Associations
Curated	104653	7954	165354
Literature	19407	4228	48998
All	**117337**	**10358**	**210498**

*diseases, traits, symptoms, disease groups

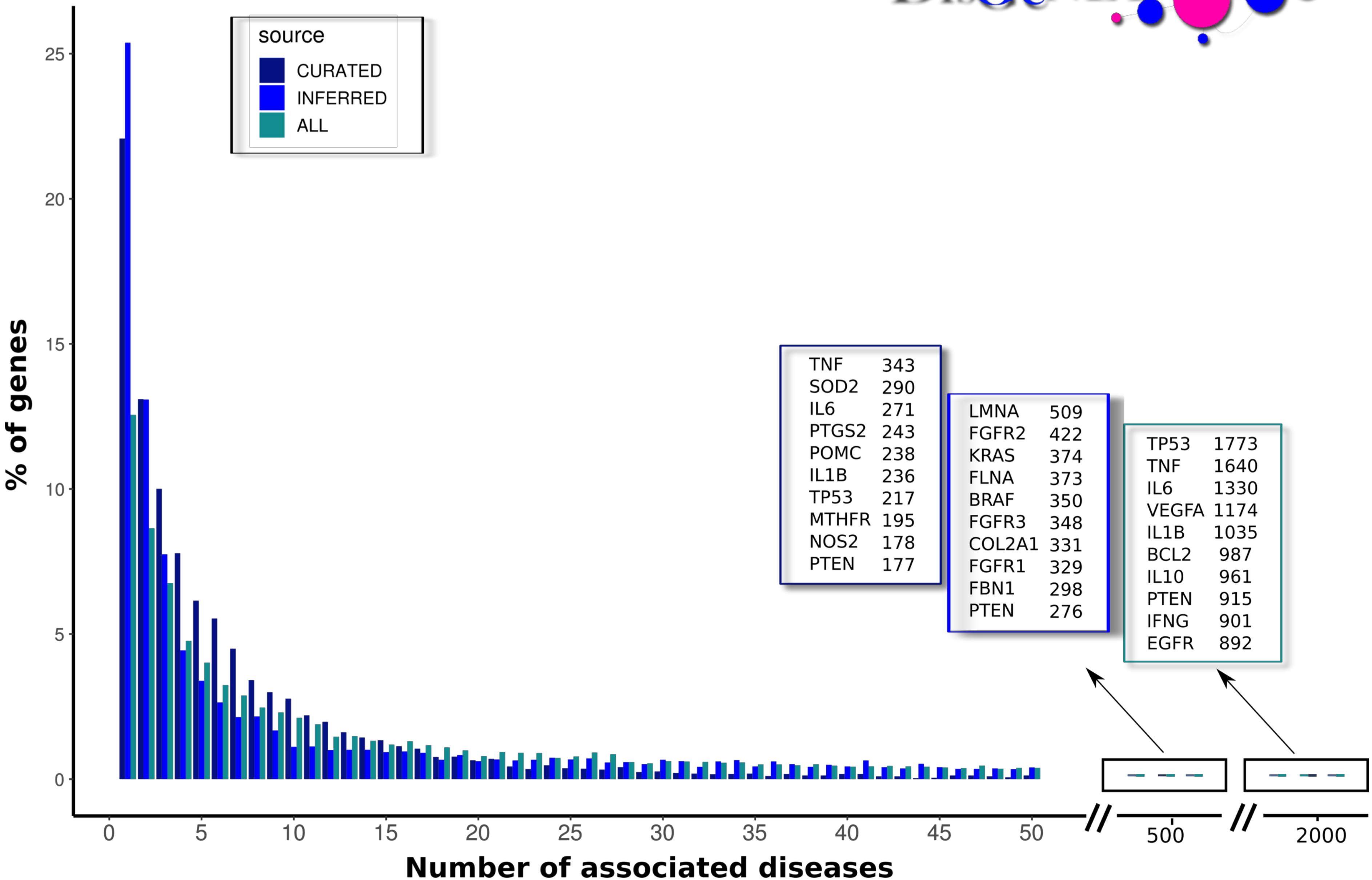
DisGeNET prioritization tools



DisGeNET



% of diseases
60
40
20
0
source
CURATED
INFERRED
ALL
Mental and motor retardation 1020
Poor school performance 984
Cognitive delay 965
Mental Retardation 947
Mental deficiency 946
Intellectual Disability 946
Seizures
Short stature
hypotonia
Intellectual Disability 2037
Malignant neoplasm of breast 1025
Schizophrenia 1022
Colorectal Cancer 676
Prostatic Neoplasms 601
Bipolar Disorder 505
Breast Carcinoma 487
Epileptic encephalopathy 449
Drug-Induced Liver Disease 315
Mitochondrial Diseases 306
Breast Carcinoma 4962
Liver carcinoma 3592
Colorectal Cancer 3298
Prostate carcinoma 3144
Intellectual Disability 2502
Carcinoma of lung 2475
melanoma 2453
Stomach Carcinoma 2377
Glioma 2210
Ovarian Carcinoma 2202
0 5 10 15 20 25 30 35 40 45 50
2000
5000
Number of associated genes

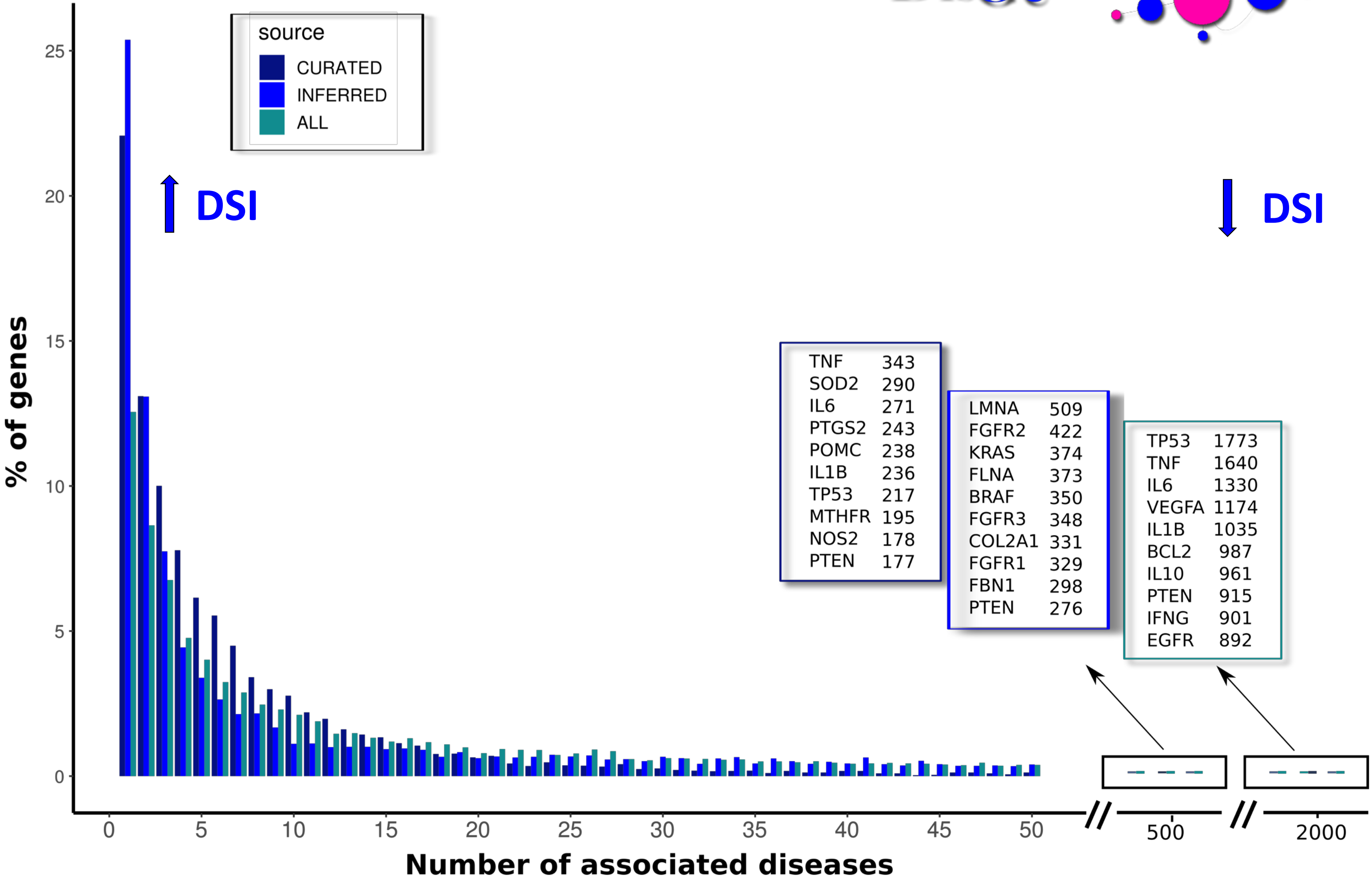

source
CURATED
INFERRED
ALL
% of genes
25
20
15
10
5
0
Number of associated diseases
0
5
10
15
20
25
30
35
40
45
50
500
2000
TNF 343
SOD2 290
IL6 271
PTGS2 243
POMC 238
IL1B 236
TP53 217
MTHFR 195
NOS2 178
PTEN 177
LMNA 509
FGFR2 422
KRAS 374
FLNA 373
BRAF 350
FGFR3 348
COL2A1 331
FGFR1 329
FBN1 298
PTEN 276
TP53 1773
TNF 1640
IL6 1330
VEGFA 1174
IL1B 1035
BCL2 987
IL10 961
PTEN 915
IFNG 901
EGFR 892

VARIANT GENE

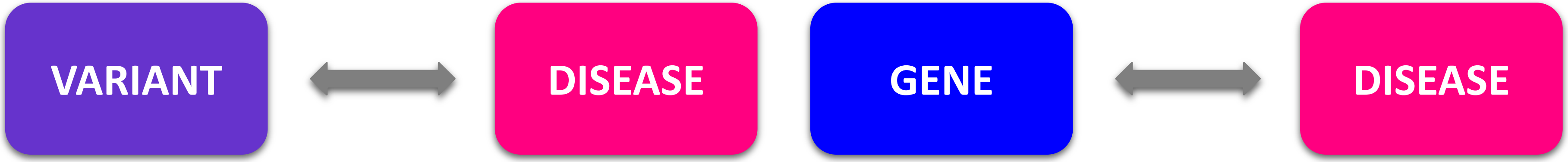
Tools for prioritization

- ✓ Protein functional classification
- ✓ Tolerance of genes to LoF variation
- ✓ Allele frequency, variant consequence type
- ✓ Disease Specificity Index (DSI)

  A gene/variant is more specific if it is associated to a small number of diseases (DSI closer to 1)



source
CURATED
INFERRED
ALL
DSI
DSI
% of genes
25
20
15
10
5
0
0
5
10
15
20
25
30
35
40
45
50
500
2000
Number of associated diseases
TNF 343
SOD2 290
IL6 271
PTGS2 243
POMC 238
IL1B 236
TP53 217
MTHFR 195
NOS2 178
PTEN 177
LMNA 509
FGFR2 422
KRAS 374
FLNA 373
BRAF 350
FGFR3 348
COL2A1 331
FGFR1 329
FBN1 298
PTEN 276
TP53 1773
TNF 1640
IL6 1330
VEGFA 1174
IL1B 1035
BCL2 987
IL10 961
PTEN 915
IFNG 901
EGFR 892



Tools for prioritization

- ✓ **DisGeNET association score:** popularity/novelty

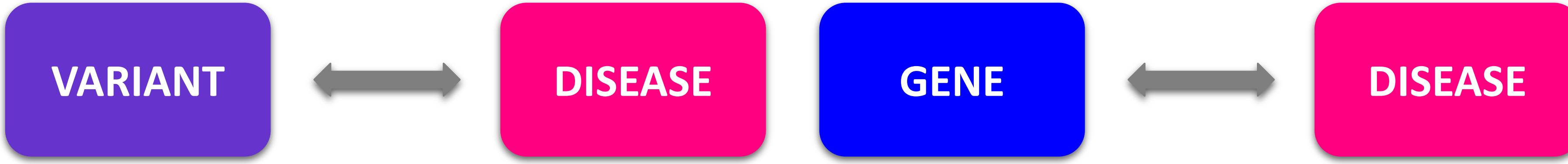
DisGeNET score

GDA score: Indicates **popularity** of a **gene-disease association (GDA)** across all data sources giving higher weight to curated sources vs. animal models GDAs, and to animal models vs. text-mining.

DisGeNET score = $S_{CURATED} + S_{MODELS} + S_{INFERRED} + S_{LITERATURE}$

VDA score: Indicates **popularity** of a **variant-disease association (VDA)** across all data sources giving higher weight to curated vs. text-mining VDAs.

DisGeNET score = $S_{CURATED} + S_{LITERATURE}$



Tools for prioritization

- ✓ **DisGeNET association score:** popularity/novelty
- ✓ **DisGeNET association type:** insight on biology
- ✓ **Evidence level:** confidence of the association
- ✓ **Evidence Index:** controversial field of research
- ✓ **Number of publications**

What is the advantage of data integration & standardization?

- Human genetics to support drug discovery
- Rare diseases research
- Annotation of NGS and variation data
- Disease comorbidity
- Insight on disease mechanisms and drug mode of action

Genomic data analysis identified 3,000 potentially “druggable” proteins in the human genome.

Only 10% of these potential targets have an FDA approved drug.

Santos, Rita, et al. "A comprehensive map of molecular drug targets." Nature Reviews Drug Discovery (2016).

NIH Pharos initiative:

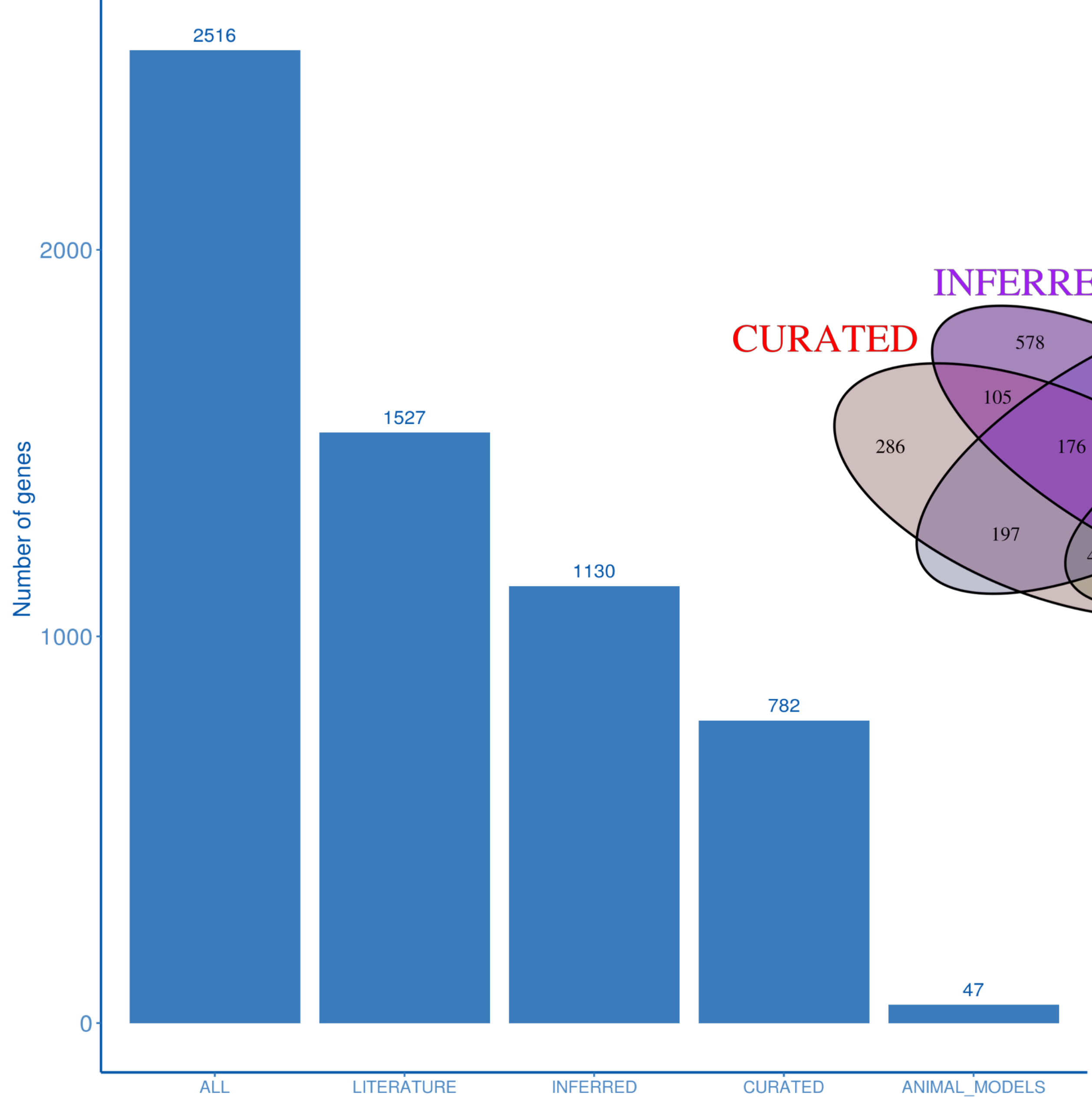
- To shed light on **poorly characterized proteins** that can potentially be modulated using small molecules or biologics

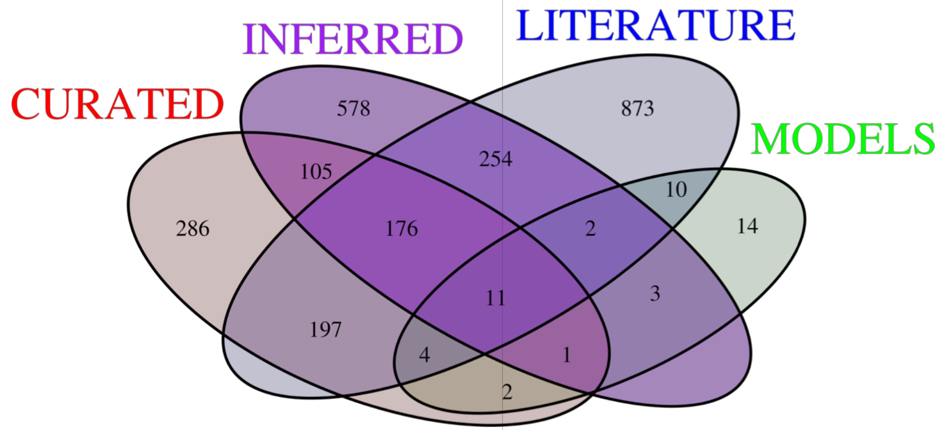
Target “dark” proteins

PHAROS

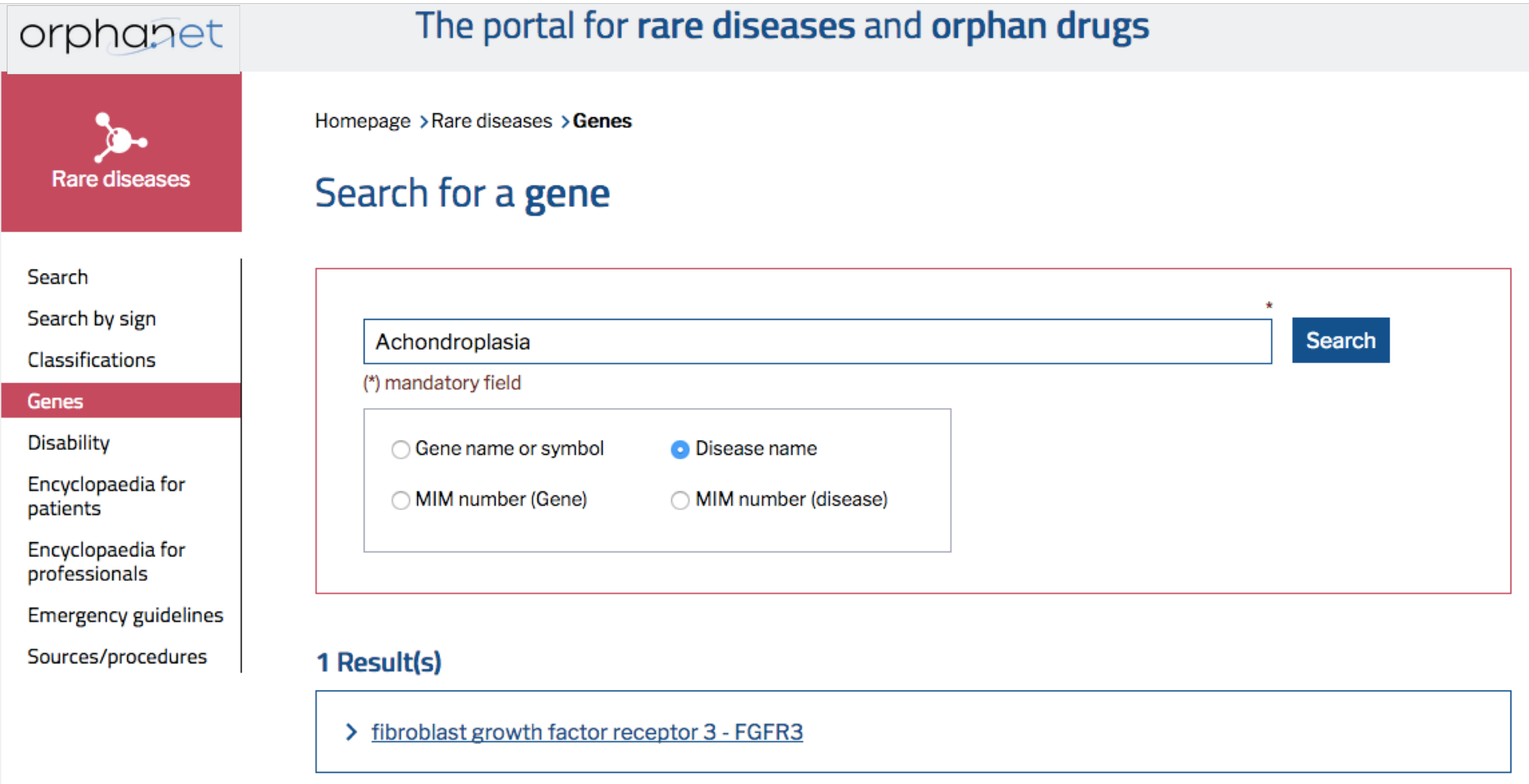
https://pharos.nih.gov/idg/index

DisGeNET annotates 40 % of Pharos Tdark proteins





Information on genetic basis of rare diseases



6850 GDAs in Orphanet involving 3496 genes and 3520 diseases

DisGeNET provides annotations to variants for 3455 Orphanet diseases (54 %)

6372 diseases

3520 diseases annotated to 3496 genes

2998 diseases with no gene annotation

DisGeNET provides additional annotation to the diseases

DisGeNET provides annotations for 1467 diseases (49 %)

Achondroplasia in DisGeNET

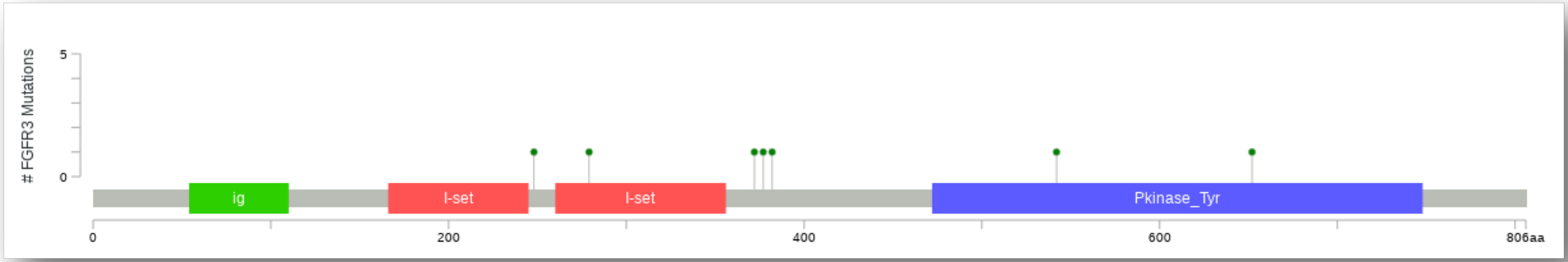
Top 10 genes

Gene	Gene name	DisGeNET score	N. PMIDs	N. SNPs
FGFR3	fibroblast growth factor receptor 3	1	133	7
SPRED2	sprouty related EVH1 domain containing 2	0.21	1	0
NPR2	natriuretic peptide receptor 2	0.21	3	0
PTHLH	parathyroid hormone like hormone	0.21	2	0
GH1	growth hormone 1	0.03	3	0
FGF1	fibroblast growth factor 1	0.02	2	0
PTH	parathyroid hormone	0.02	2	0
FGF2	fibroblast growth factor 2	0.02	2	0
NPPC	natriuretic peptide C	0.01	1	0
PTRH1	peptidyl-tRNA hydrolase 1 homolog	0.01	1	0

Variants in FGFR3 gene annotated to Achondroplasia

Variants in the FGFR gene associated to achondroplasia

variantid	score	npmids	most_severe_consequence	ref_alt	Protein_position	Amino_acids
rs121913105	0.70	0	missense variant	A/C,T	652	K/T
rs121913114	0.70	2	missense variant	A/G,T	279	S/G
rs121913479	0.01	1	missense variant	G/A,T	372	G/S
rs121913482	0.70	1	missense variant	C/T	248	R/C
rs28931614	0.90	55	missense variant	G/A,C	382	G/R
rs28933068	0.74	5	stop gained	C/A,G,T	542	N/K
rs75790268	0.85	7	missense variant	G/T	377	G/C

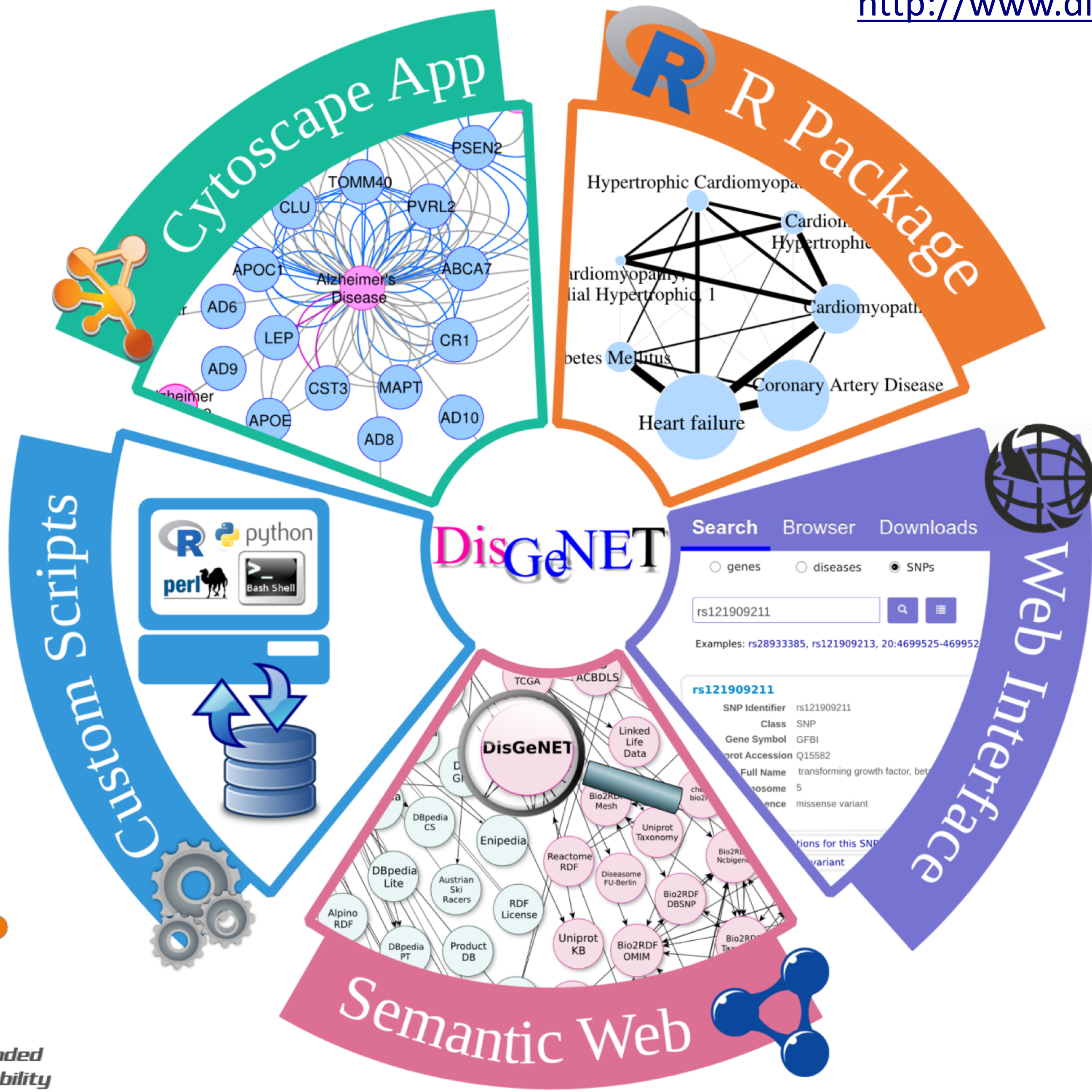


DisGeNET

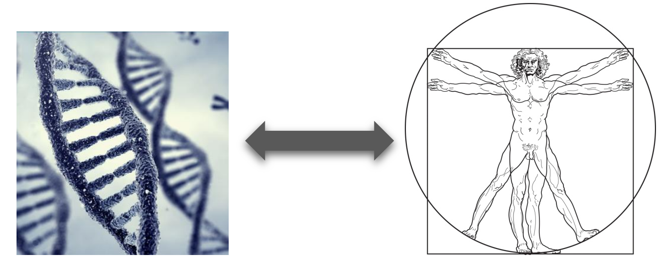
- One of the most comprehensive catalogs of genes and variants associated to human diseases and phenotypes publicly available

- Developed by integration of different public resources, including information extracted from the literature by text mining

- Provides different prioritization metrics and can be accessed with different tools http://www.disgenet.org/

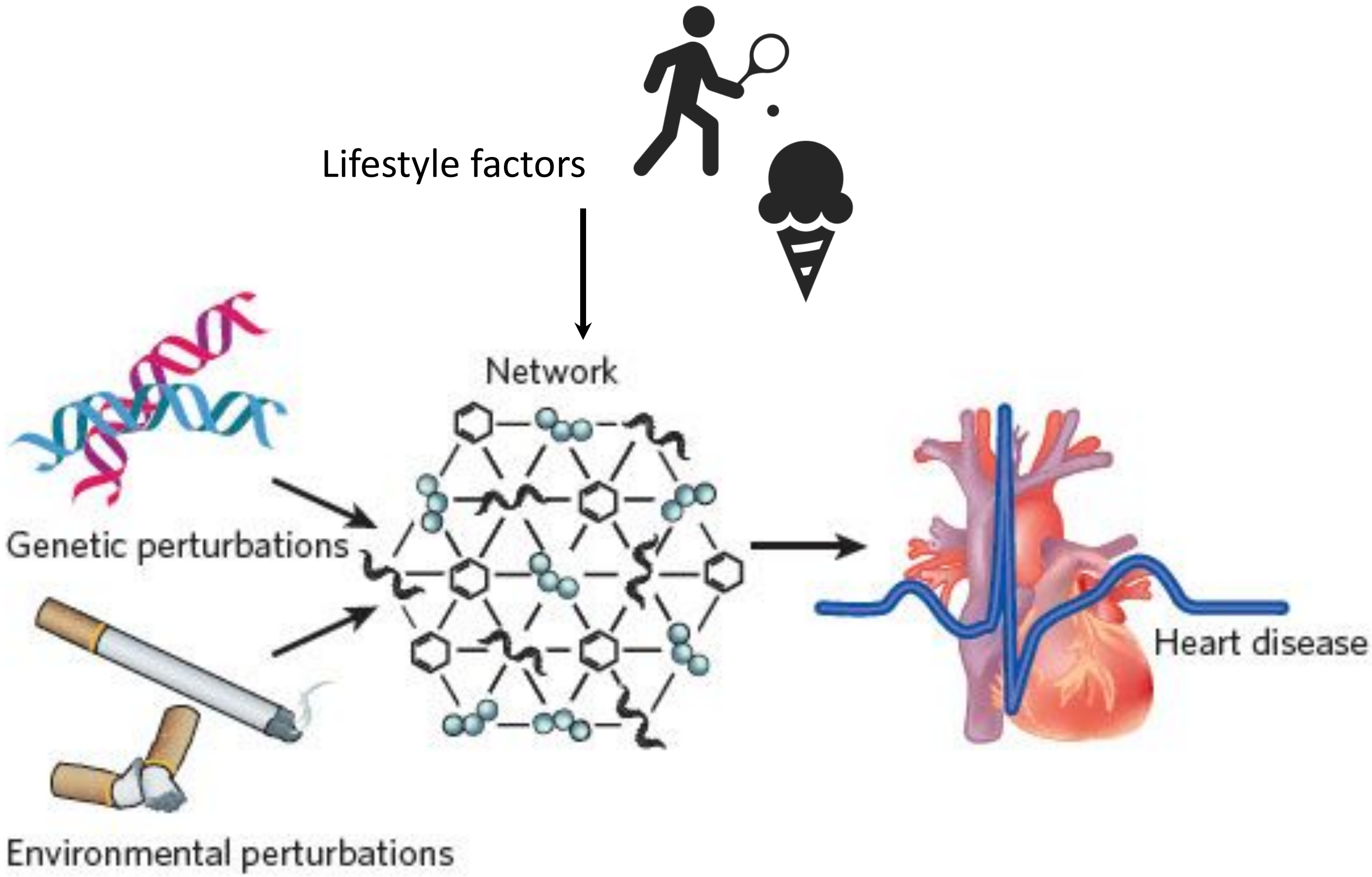


Recommended
Interoperability
Resources



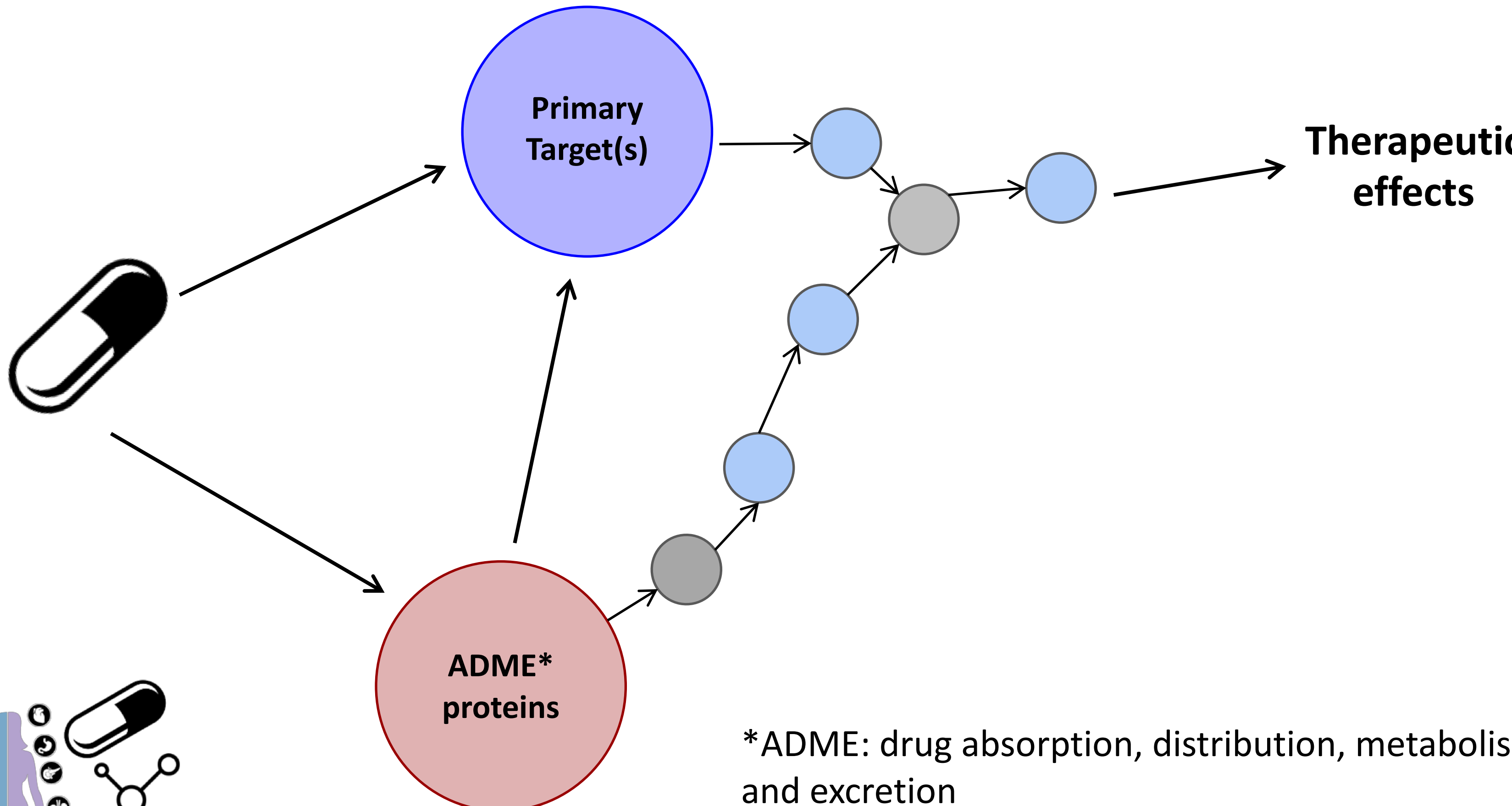
The accumulation of large-scale data requires the development of **computational tools** able to explore and mine the vast amount **of biological knowledge** they contain.

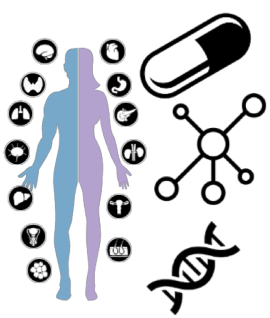
Network medicine to study human diseases



Modified from Schadt, E. E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature*, *461*(7261), 218.

“Classic” view of drug mode of action

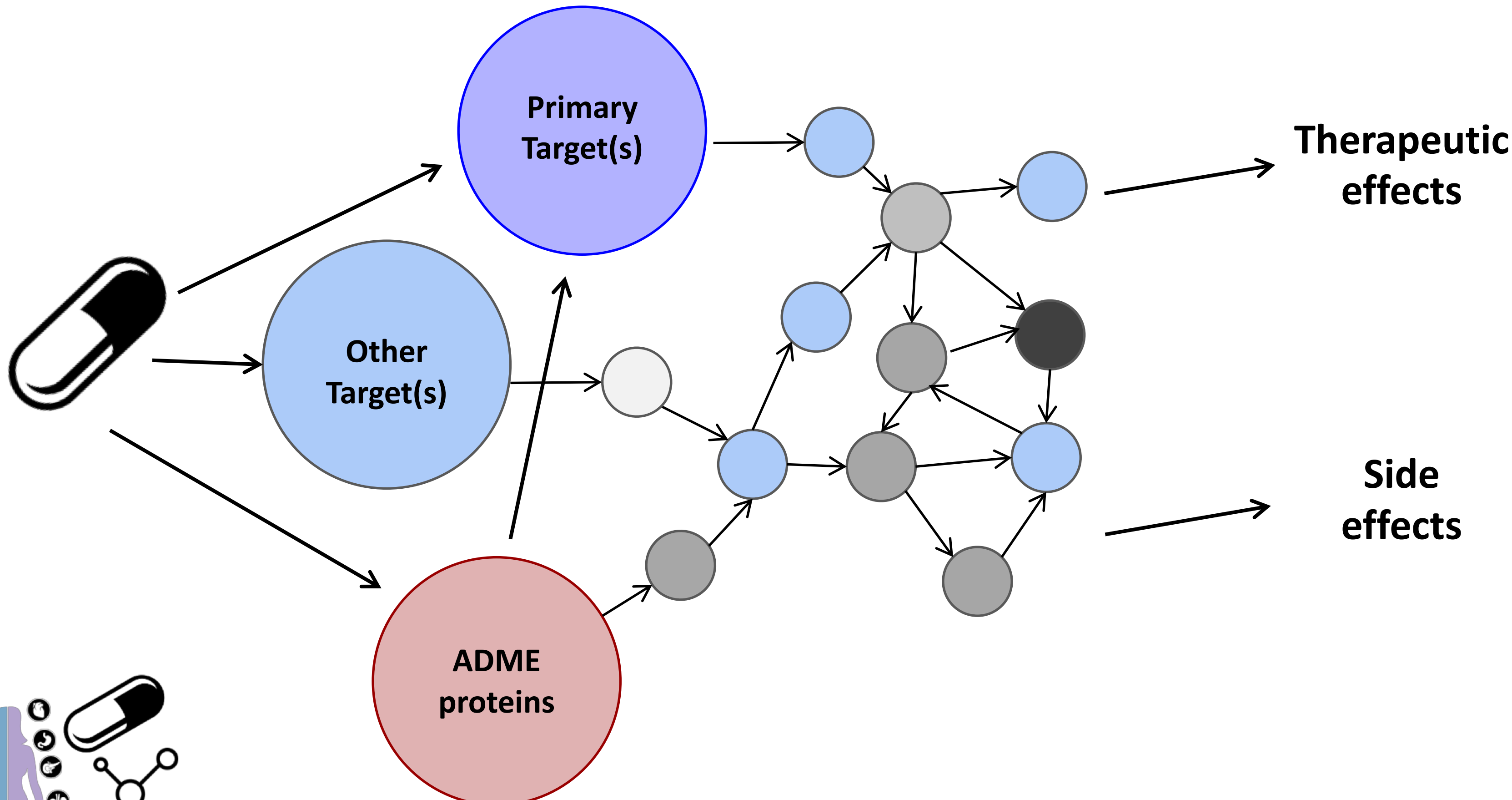


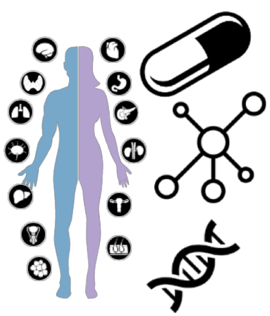


*ADME: drug absorption, distribution, metabolism, and excretion

Adapted from Berger and Iyengar 2009, Bioinformatics

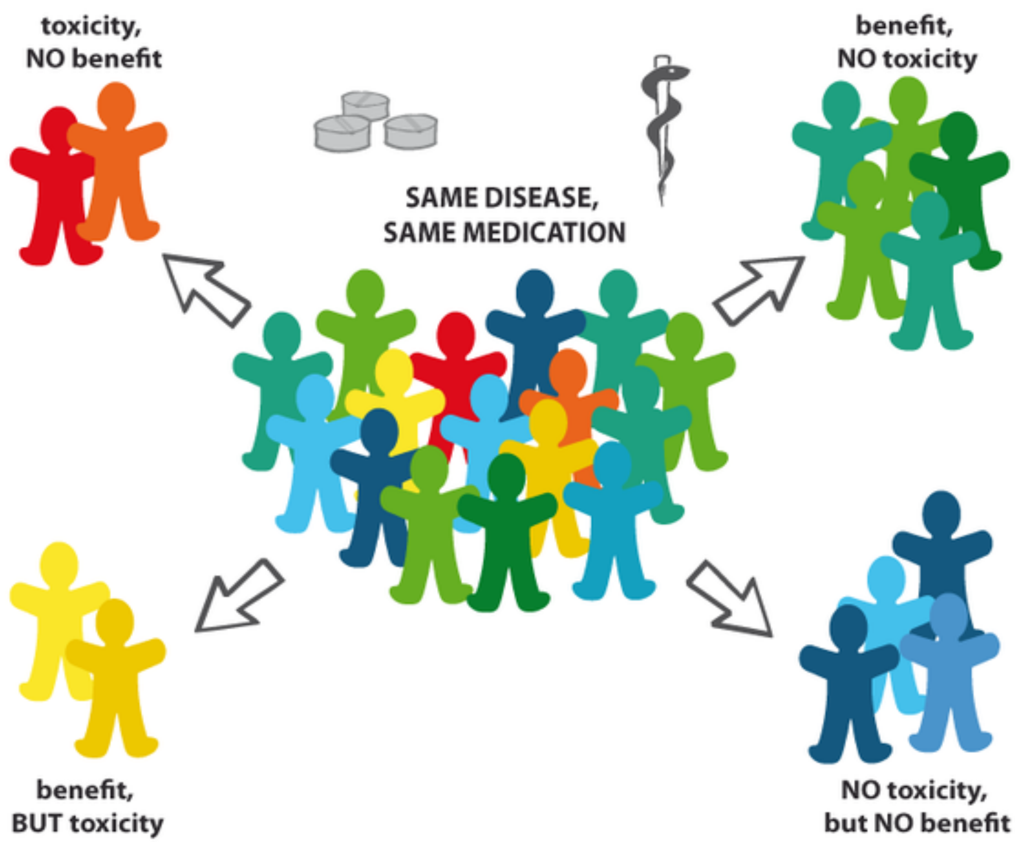
“Systems” view of drug mode of action





Adapted from Berger and Iyengar 2009, Bioinformatics

Variability of drug response

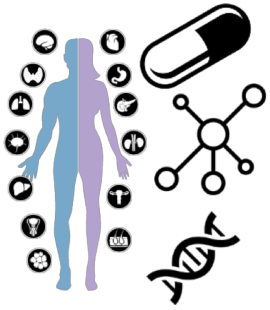


Variability of drug response

Drug	Gene	Effect
Pharmacokinetics		
Codeine	*CYP2D6* (34)	Increase in the amount of active drug by variants
Clopidogrel	*CYP2C19* (80)	Increase in the amount of active drug by variants
Warfarin	*CYP2C9* (81)	Changes in drug levels in blood by variants

Adapted from Zhao S, Iyengar R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annual review of pharmacology and toxicology. 2012 Feb 10;52:505-21.

Genes relevant to drug response have different transcriptomic, genomic and network properties



Genes relevant to drug response

ChEMBL
DrugCentral
DRUGBANK
DGIdb

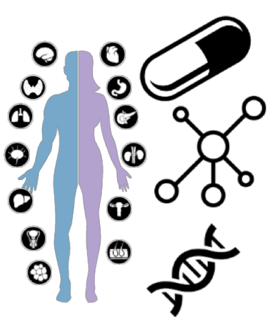
TARGETS

1934 drug targets

DRUGBANK

METAB

470 ADME proteins



AEOLUS
OffSides
OrganDB

DisGeNET

TOXPROT

4160 ADR proteins

Genes relevant to drug response

ChEMBL
DrugCentral
DRUGBANK
DGIdb

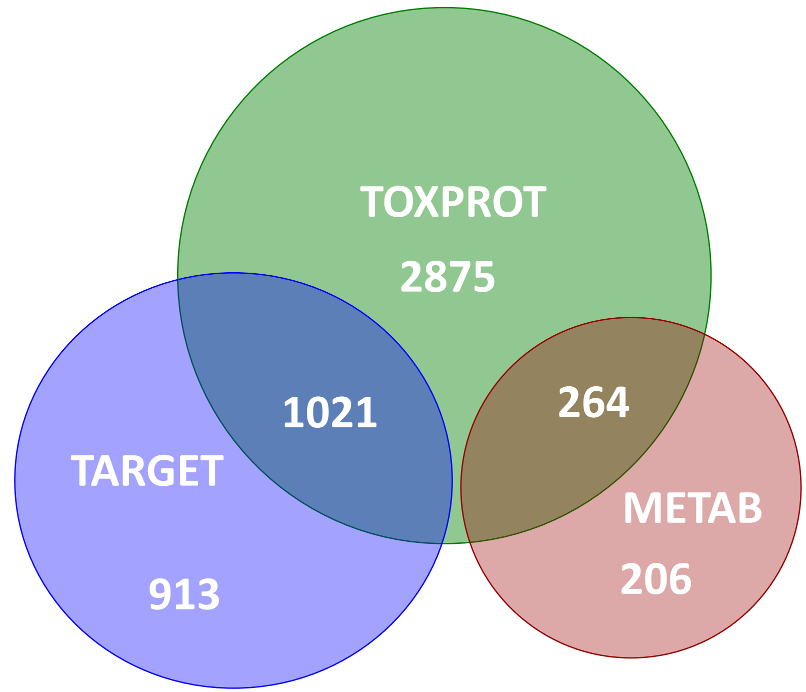
TARGETS

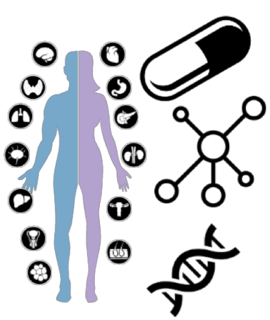
1934 drug targets

DRUGBANK

METAB

470 ADME proteins





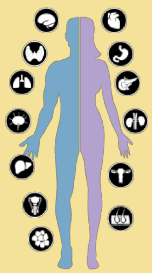
AEOLUS
OffSides
OrganDB

DisGeNET

TOXPROT

4160 ADR proteins

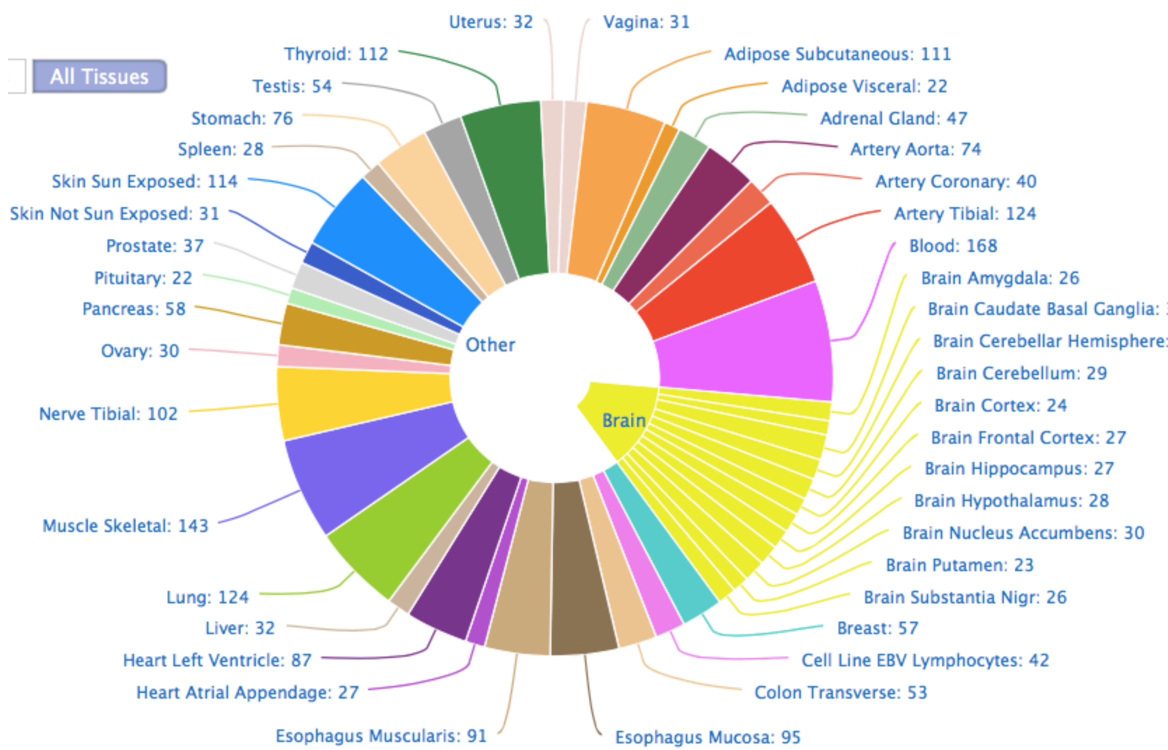
Transcriptomic analysis

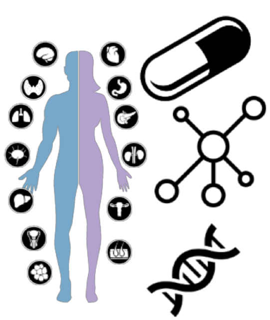


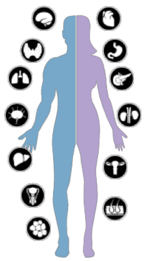
Transcriptomic data

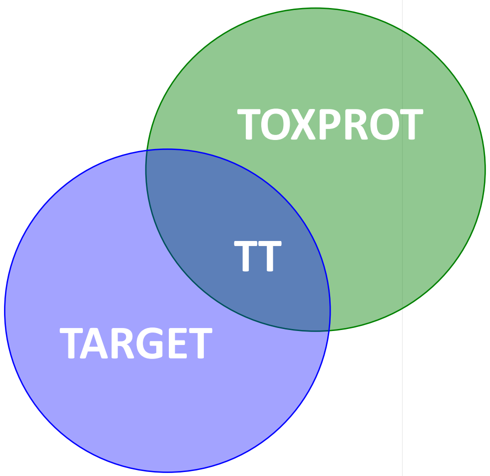
GTEx Portal

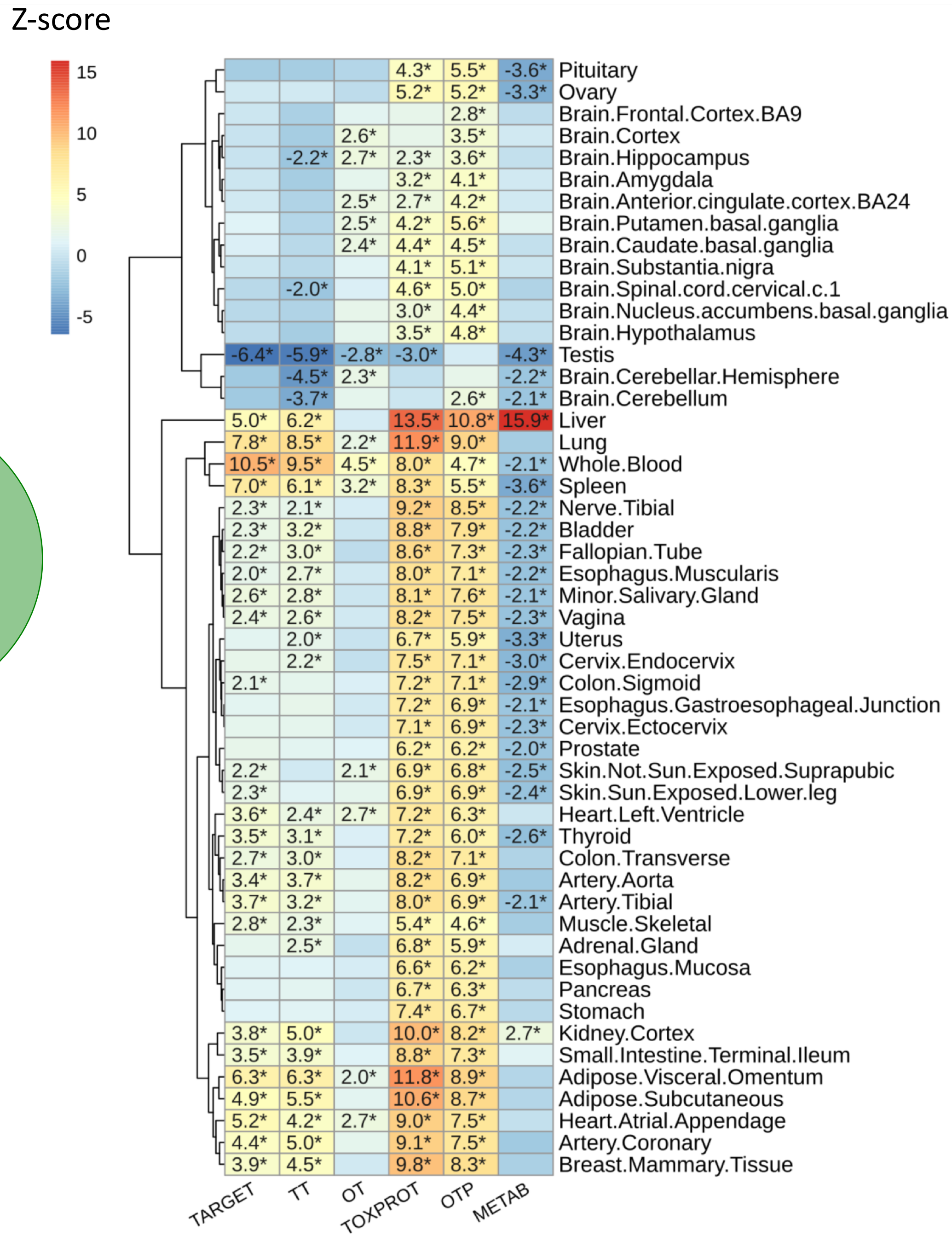
53 tissues
(TPM >=1)



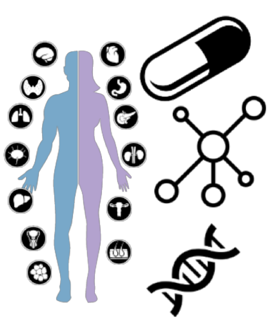





TOXPROT
TT
TARGET


Z-score
15
10
5
0
-5

	TARGET	TT	OT	TOXPROT	OTP	METAB
Pituitary				4.3*	5.5*	-3.6*
Ovary				5.2*	5.2*	-3.3*
Brain.Frontal.Cortex.BA9					2.8*	
Brain.Cortex			2.6*		3.5*	
Brain.Hippocampus		-2.2*	2.7*	2.3*	3.6*	
Brain.Amygdala				3.2*	4.1*	
Brain.Anterior.cingulate.cortex.BA24			2.5*	2.7*	4.2*	
Brain.Putamen.basal.ganglia			2.5*	4.2*	5.6*	
Brain.Caudate.basal.ganglia			2.4*	4.4*	4.5*	
Brain.Substantia.nigra				4.1*	5.1*	
Brain.Spinal.cord.cervical.c.1		-2.0*		4.6*	5.0*	
Brain.Nucleus.accumbens.basal.ganglia				3.0*	4.4*	
Brain.Hypothalamus				3.5*	4.8*	
Testis	-6.4*	-5.9*	-2.8*	-3.0*		-4.3*
Brain.Cerebellar.Hemisphere		-4.5*	2.3*			-2.2*
Brain.Cerebellum		-3.7*			2.6*	-2.1*
Liver	5.0*	6.2*		13.5*	10.8*	15.9*
Lung	7.8*	8.5*	2.2*	11.9*	9.0*	
Whole.Blood	10.5*	9.5*	4.5*	8.0*	4.7*	-2.1*
Spleen	7.0*	6.1*	3.2*	8.3*	5.5*	-3.6*
Nerve.Tibial	2.3*	2.1*		9.2*	8.5*	-2.2*
Bladder	2.3*	3.2*		8.8*	7.9*	-2.2*
Fallopian.Tube	2.2*	3.0*		8.6*	7.3*	-2.3*
Esophagus.Muscularis	2.0*	2.7*		8.0*	7.1*	-2.2*
Minor.Salivary.Gland	2.6*	2.8*		8.1*	7.6*	-2.1*
Vagina	2.4*	2.6*		8.2*	7.5*	-2.3*
Uterus		2.0*		6.7*	5.9*	-3.3*
Cervix.Endocervix		2.2*		7.5*	7.1*	-3.0*
Colon.Sigmoid	2.1*			7.2*	7.1*	-2.9*
Esophagus.Gastroesophageal.Junction				7.2*	6.9*	-2.1*
Cervix.Ectocervix				7.1*	6.9*	-2.3*
Prostate				6.2*	6.2*	-2.0*
Skin.Not.Sun.Exposed.Suprapubic	2.2*		2.1*	6.9*	6.8*	-2.5*
Skin.Sun.Exposed.Lower.leg	2.3*			6.9*	6.9*	-2.4*
Heart.Left.Ventricle	3.6*	2.4*	2.7*	7.2*	6.3*	
Thyroid	3.5*	3.1*		7.2*	6.0*	-2.6*
Colon.Transverse	2.7*	3.0*		8.2*	7.1*	
Artery.Aorta	3.4*	3.7*		8.2*	6.9*	
Artery.Tibial	3.7*	3.2*		8.0*	6.9*	-2.1*
Muscle.Skeletal	2.8*	2.3*		5.4*	4.6*	
Adrenal.Gland		2.5*		6.8*	5.9*	
Esophagus.Mucosa				6.6*	6.2*	
Pancreas				6.7*	6.3*	
Stomach				7.4*	6.7*	
Kidney.Cortex	3.8*	5.0*		10.0*	8.2*	2.7*
Small.Intestine.Terminal.Ileum	3.5*	3.9*		8.8*	7.3*	
Adipose.Visceral.Omentum	6.3*	6.3*	2.0*	11.8*	8.9*	
Adipose.Subcutaneous	4.9*	5.5*		10.6*	8.7*	
Heart.Atrial.Appendage	5.2*	4.2*	2.7*	9.0*	7.5*	
Artery.Coronary	4.4*	5.0*		9.1*	7.5*	
Breast.Mammary.Tissue	3.9*	4.5*		9.8*	8.3*	



Genomic analysis

Genomic data

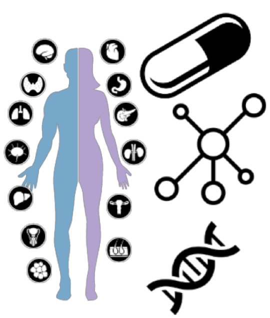
Exome Aggregation Consortium (ExAC)

Germline variants detected across 60k exomes

- ✓ pLI: the probability of a gene to be intolerant to heterozygous Loss of Function (LoF) mutations

  LoF intolerant genes: pLI ≥ 0.9
  LoF tolerant genes: pLI ≤ 0.1

- ✓ pNull: the probability of a gene to be tolerant to both heterozygous and homozygous LoF variation.

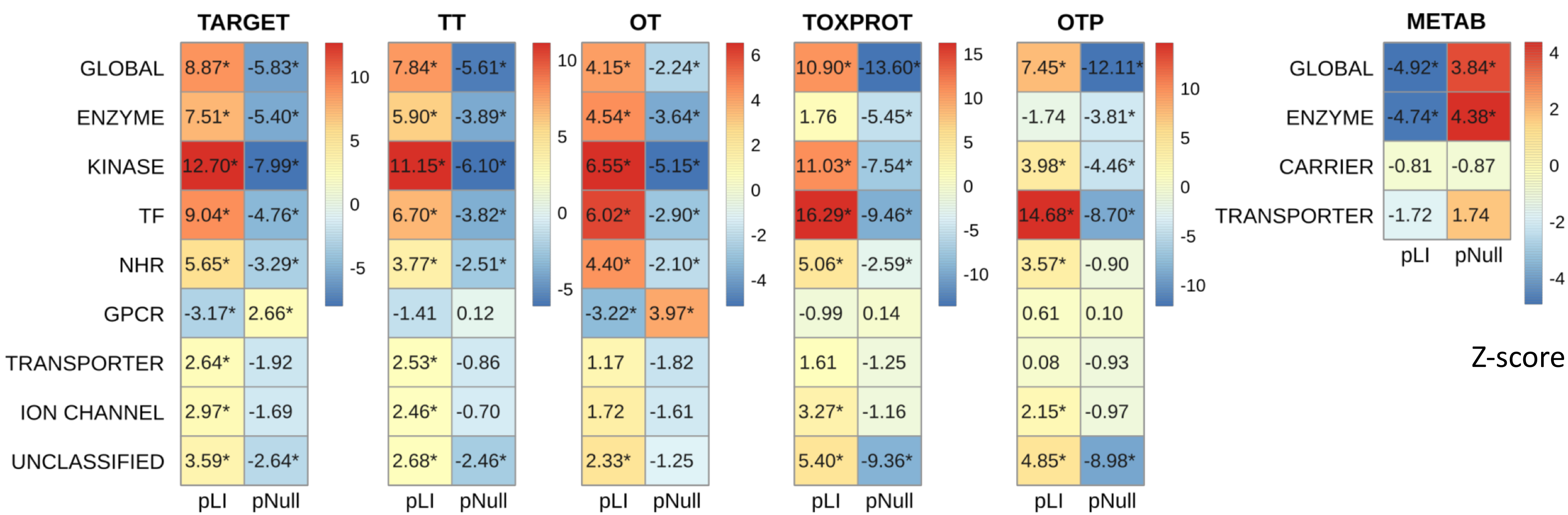


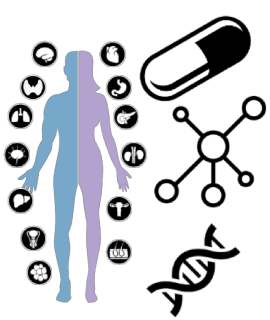
Gene constraint metrics for drug relevant genes

Gene Set	pLI			pNull		
	pLI	z-score	p-value	pNull	z-score	p-value
TARGET	0.380	8.87	7.31E-19	0.167	-5.829	5.58E-09
TOXPROT	0.365	10.9	1.15E-27	0.146	-13.604	3.79E-42
METAB	0.214	-4.92	8.65E-07	0.260	3.843	1.22E-04

- ✓ pLI: the probability of a gene to be intolerant to heterozygous LoF mutations
- ✓ pNull: the probability of a gene to be tolerant to both heterozygous and homozygous LoF variation.

Gene constraint metrics for drug relevant genes

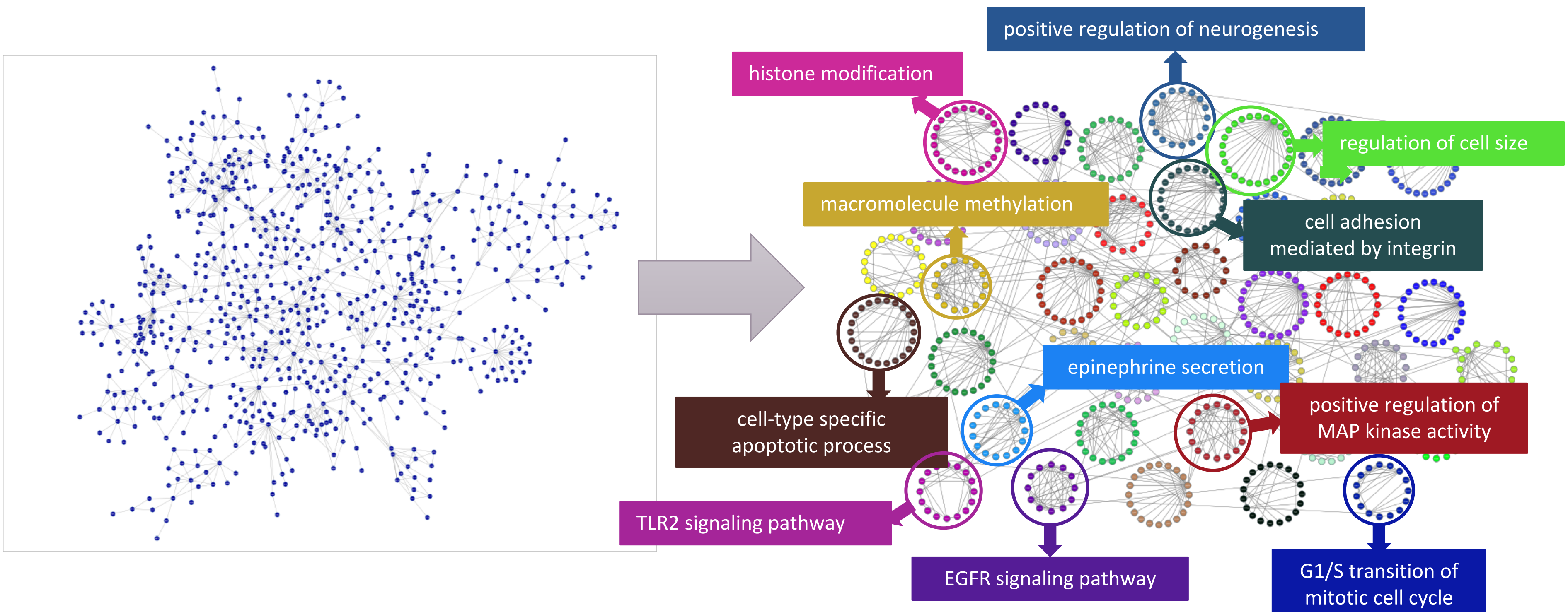


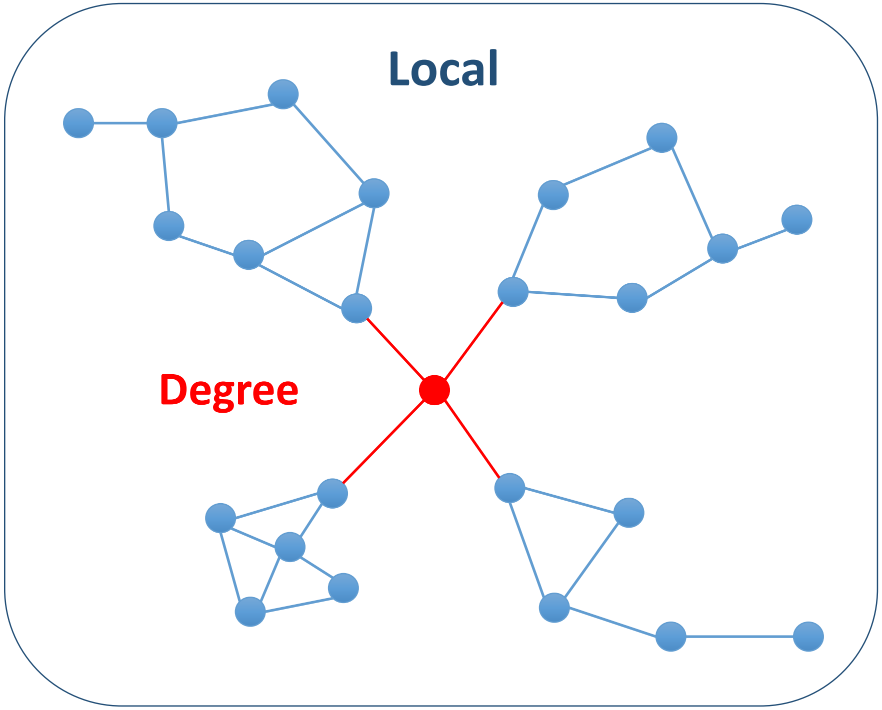


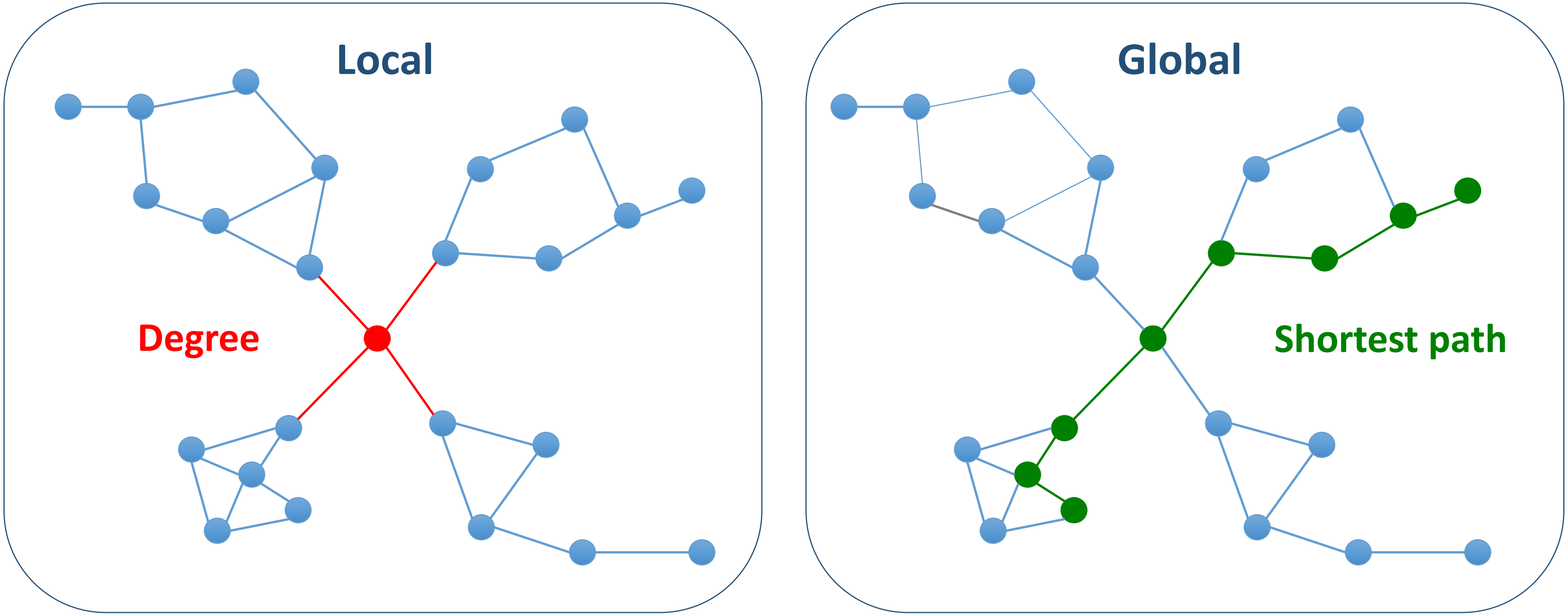
- ✓ pLI: the probability of a gene to be intolerant to heterozygous LoF mutations
- ✓ pNull: the probability of a gene to be tolerant to both heterozygous and homozygous LoF variation.

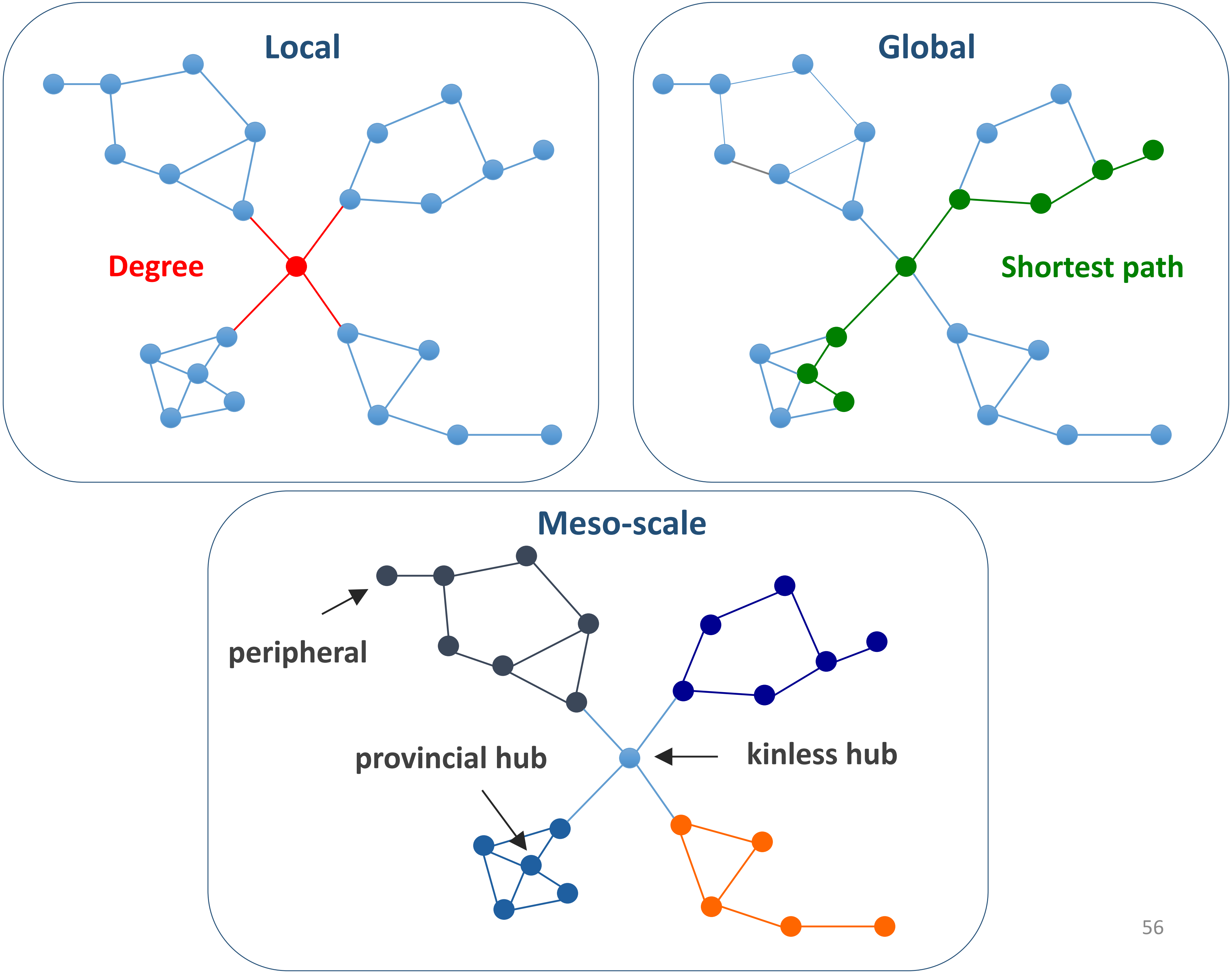
Network analysis

To connect **network structure** to **function**

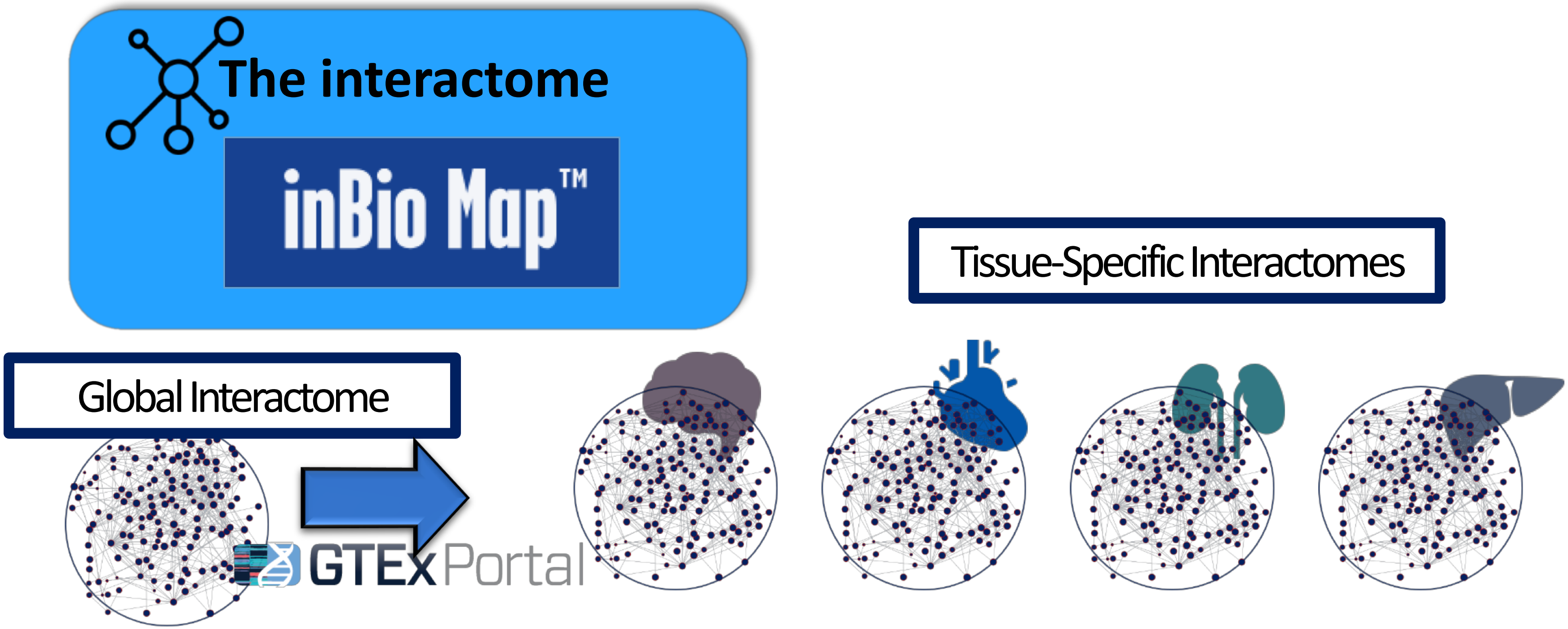


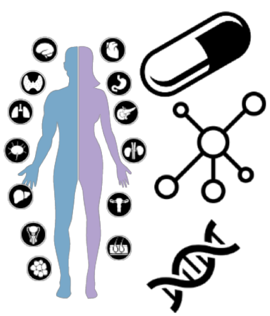

Local
Degree


Local
Degree
Global
Shortest path

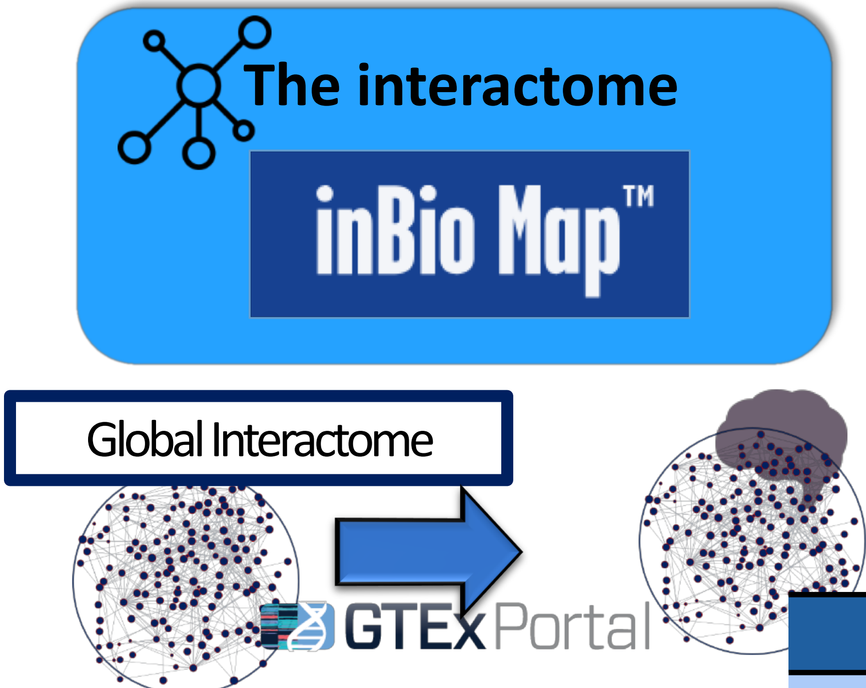

Local
Degree
Global
Shortest path
Meso-scale
peripheral
provincial hub
kinless hub

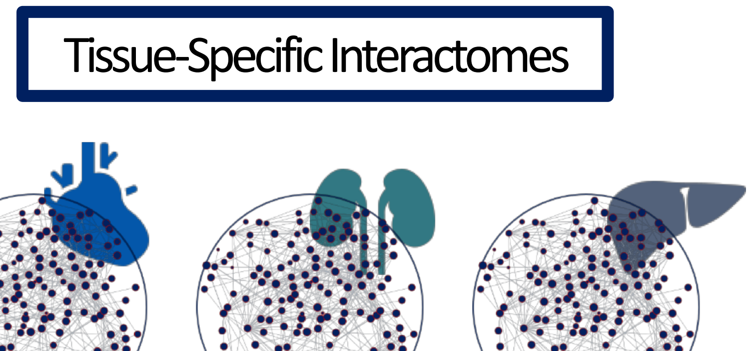
Network analysis

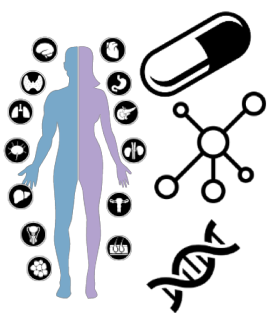




Network analysis

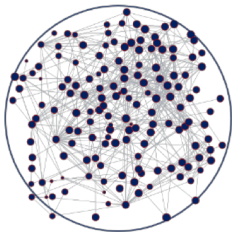




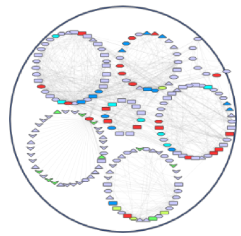


Network properties	
Local	Degree
	Clustering coefficient
Global	Betweennes centrality
Meso	Within-module degree Z
	Participation coefficient P
	Cartographic roles

Meso-scale network analysis



Network clustering



For each node, compute ***Z*** and ***P***

Assignment of one of seven cartographic roles to nodes

(Guimerà & Amaral, Nature 2005)

Participation Coefficient

$$P = 1 - \sum_{s=1}^{N_M} \left(\frac{K_{is}}{K_i} \right)^2$$

K_{is} **is the # of links of nodes i in module s**

K_i **is the degree of node i**

Within-Module Degree

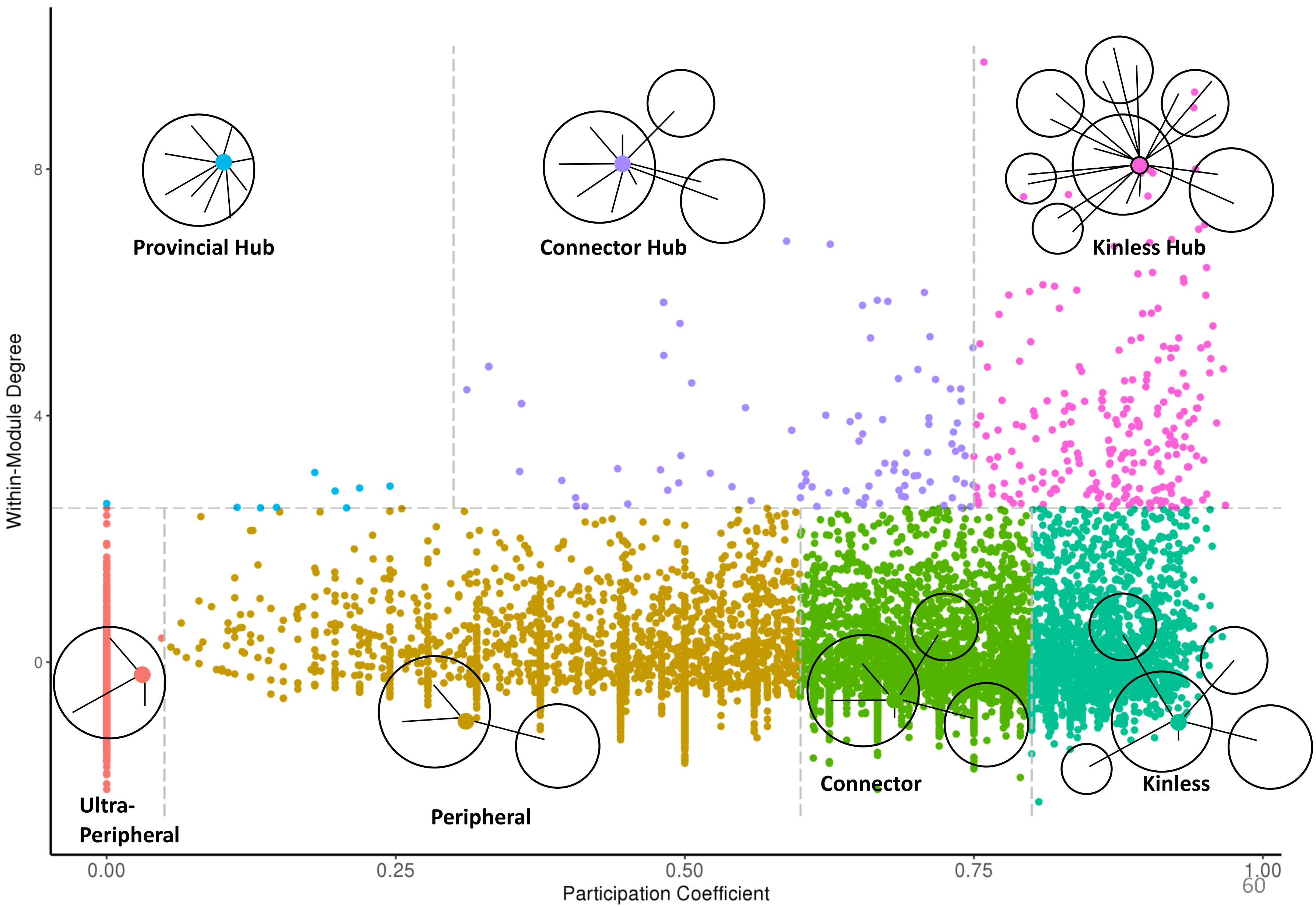
$$z_i = \frac{K_i - \bar{K}_{s_i}}{\sigma K_{s_i}}$$

$\bar{K}_{s_i}$ **he mean degree of nodes in module** s_i

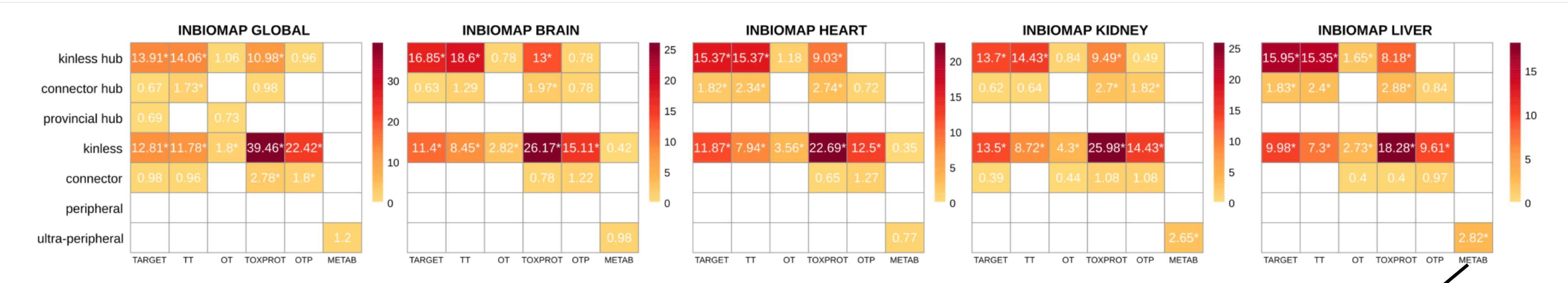
σK_{s_i} **the standard deviation of degree in** s_i

Piñero et al. *Scientific Reports* 6:24570 (2016)doi:10.1038/srep24570.

Cartographic representation of the interactome

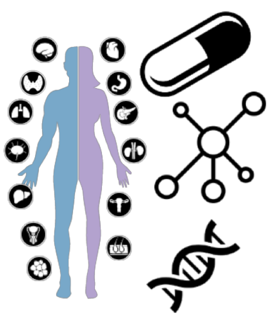


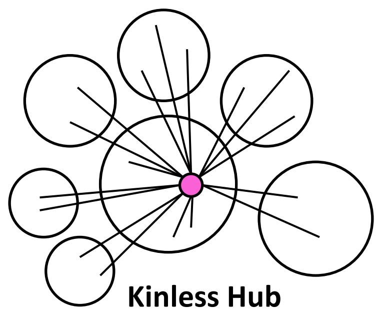
Meso-scale network properties



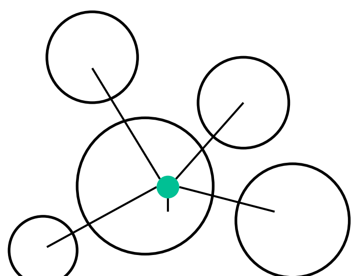
TARGETS

TOXPROT



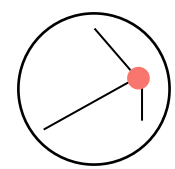


Kinless Hub



Kinless

METAB



Ultra-Peripheral

Network properties at different scales

Z-score

HIPPIE GLOBAL

	degree	P	Z	BET	CC
TARGET	15.3*	6.8*	12.8*	15.0*	-3.7*
TT	15.0*	6.0*	13.1*	15.7*	-4.2*
OT	5.8*	3.1*	4.1*	4.1*	-0.8
TOXPROT	11.9*	11.4*	13.3*	11.1*	-4.7*
OTP	4.0*	9.1*	6.8*	2.4*	-2.7*
METAB	-4.5*	-3.4*	-1.7	-1.5	-4.5*

HIPPIE BRAIN

	degree	P	Z	BET	CC
TARGET	15.9*	8.3*	12.3*	13.9*	-3.0*
TT	15.0*	7.2*	12.5*	14.4*	-3.3*
OT	6.3*	4.2*	3.9*	4.0*	-0.6
TOXPROT	11.9*	11.0*	12.1*	9.8*	-4.4*
OTP	3.9*	7.7*	5.6*	2.0*	-2.8*
METAB	-3.7*	-1.2	-2.1*	-1.4	-4.4*

HIPPIE HEART

	degree	P	Z	BET	CC
TARGET	16.3*	8.4*	12.1*	13.9*	-3.2*
TT	15.4*	6.9*	11.7*	14.7*	-3.9*
OT	6.5*	4.5*	4.2*	3.9*	-0.3
TOXPROT	11.4*	9.4*	11.2*	9.5*	-4.9*
OTP	3.2*	6.3*	5.1*	1.6	-3.0*
METAB	-3.9*	-1.4	-1.8	-1.8	-4.4*

HIPPIE KIDNEY

	degree	P	Z	BET	CC
TARGET	16.6*	9.1*	11.8*	14.1*	-3.3*
TT	15.1*	6.6*	11.5*	14.2*	-3.5*
OT	7.1*	5.5*	4.2*	4.3*	-0.9
TOXPROT	11.7*	9.4*	11.1*	9.6*	-4.9*
OTP	3.4*	6.4*	5.2*	1.7	-3.3*
METAB	-4.2*	-2.8*	-2.0*	-1.5	-5.3*

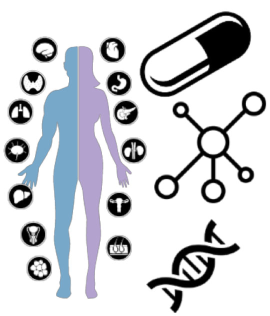
HIPPIE LIVER

	degree	P	Z	BET	CC
TARGET	16.2*	6.6*	11.8*	13.9*	-2.6*
TT	15.2*	5.2*	11.1*	14.3*	-2.9*
OT	6.4*	3.8*	4.6*	3.9*	-0.6
TOXPROT	11.3*	9.0*	10.9*	9.4*	-5.1*
OTP	3.0*	6.7*	5.1*	1.5	-3.8*
METAB	-4.2*	-3.8*	-1.2	-1.3	-4.5*

TARGETS

TOXPROT

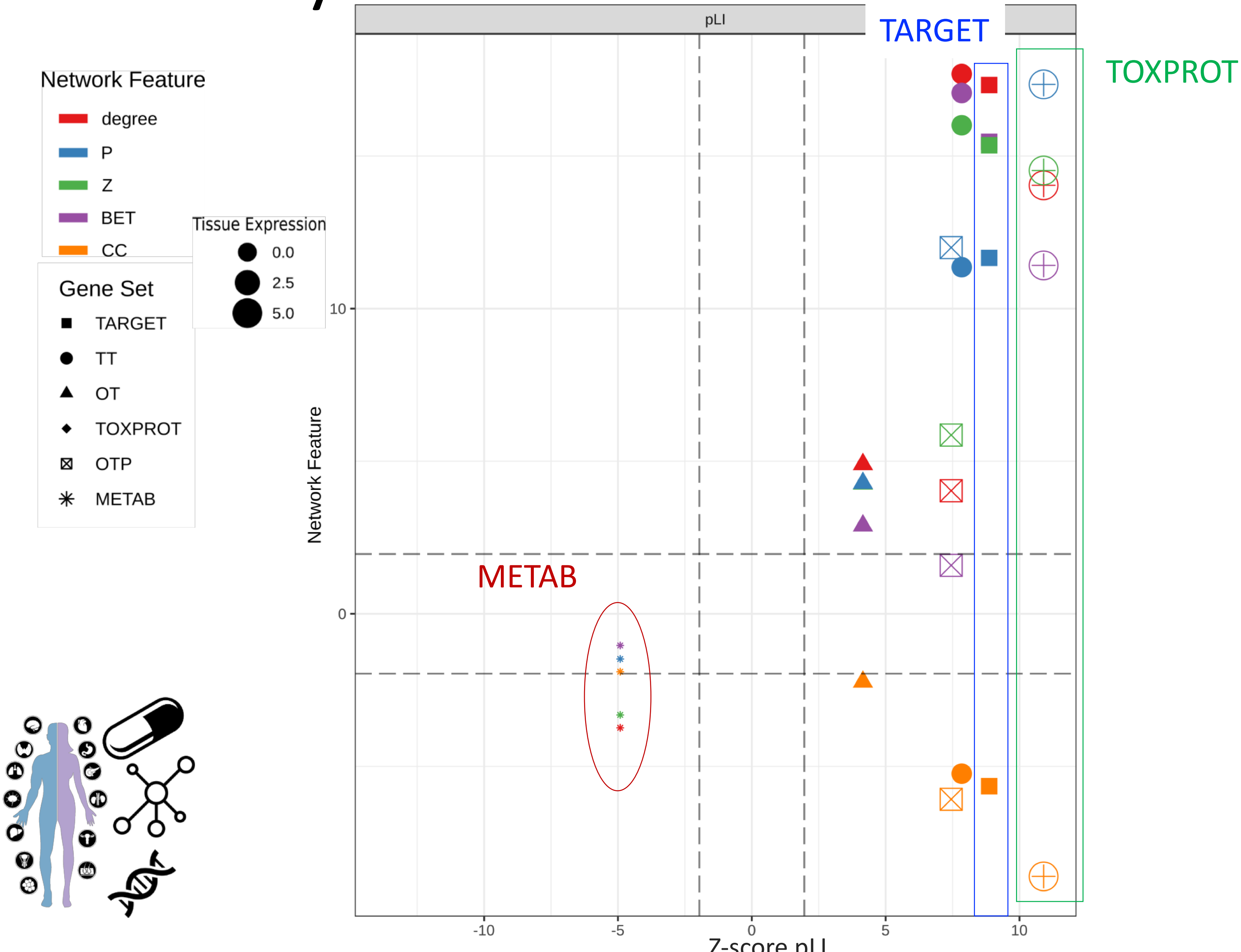
- ✓ higher degree, participation coefficient, within-module degree, and betweenness
- ✓ lower clustering coefficient



METAB

- ✓ lower degree, participation coefficient, within-module degree

Summary



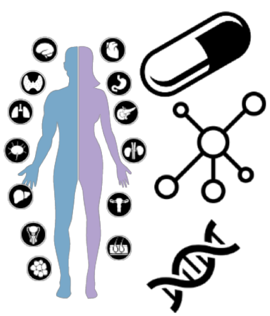
pLI
TARGET
TOXPROT
Network Feature
degree
P
Z
BET
CC
Tissue Expression
0.0
2.5
5.0
Gene Set
TARGET
TT
OT
TOXPROT
OTP
METAB
METAB
Network Feature
10
0
-10
-5
0
5
10
Z-score pLI

Take home messages

- ✓ Drug targets that mediate side effects are more central in cellular networks, more intolerant to LoF variation, and show a wider breadth of tissue expression than targets not mediating side effects.
- ✓ Among drug targets, GPCRs are tolerant to LoF variation and not central in the network
- ✓ Drug metabolizing enzymes are less central in the interactome, more tolerant to deleterious variants, and are more constrained in their tissue expression pattern.

Take home messages

The integrated analysis of *omics* and clinical data reveals distinct features of proteins associated to drug response, which could be applied to prioritize drugs with fewer probabilities of causing side effects.



Integrative Biomedical Informatics Group

RESEARCH PROGRAMME ON BIOMEDICAL INFORMATICS

Josep Saüch Pitarch

INB Spanish National Bioinformatics Institute

MedBioinformatics

transQST

elixir SPAIN

Instituto de Salud Carlos III

PRB²

eTRANSAFE

excelerate