



The 16th KJC Bioinformatics Symposium
Sokendai, Japan, August 29-31

Building a Knowledgebase for Precision Medicine

Lei Liu

Institute of Biomedical Sciences

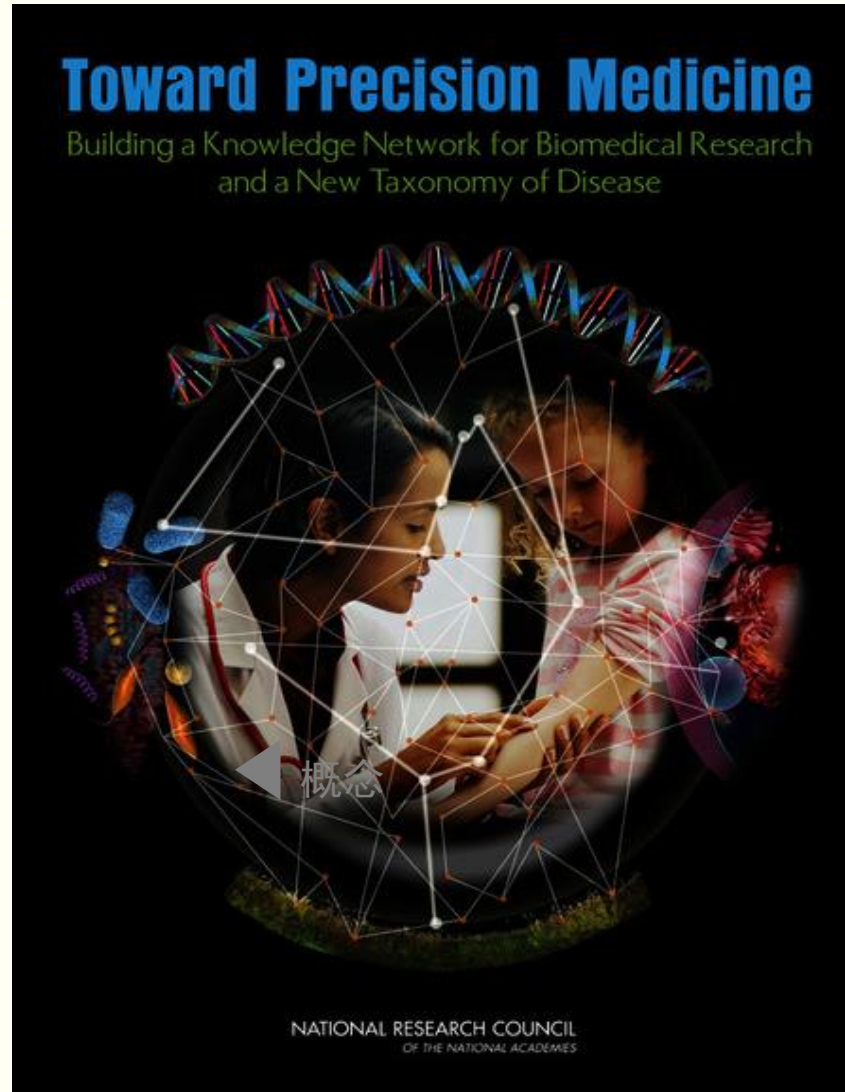
Fudan University

2018.8.30

Precision Medicine



NRC: 《Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease》 November, 2011



Toward Precision Medicine

- 10 types of cancer
- Type 2 diabetes
- Rare diseases

Biospecimen

- Cohorts
- Clinical trials
- Biobank

Precision Medicine:

- Diagnosis
- Prognosis
- Therapy
- Monitoring
- Prevention

Phenotype

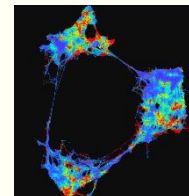
- EHR/EMR
- Clinical information
- Survival data
- Human Phenomics Project

Omics

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics

Big Data

- Data quality
- Standardization
- Data mining
- SOPs



China Precision Medicine Initiative

Cohort Studies



- Genotypes
- Phenotypes
- Environment
- Life Styles
- ...

Omics Analysis



- Genomics
- Proteomics
- Metabolomics
- ...

Large Samples, Big Data

Knowledgebase

Heterogeneous
Huge Data Sets



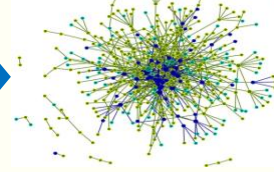
Ontologies

Literatures

Databases

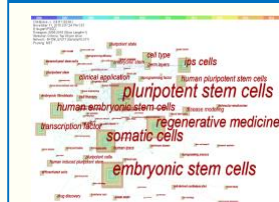
Data Curation

Knowledge Graph



Applications in Precision Medicine

Research



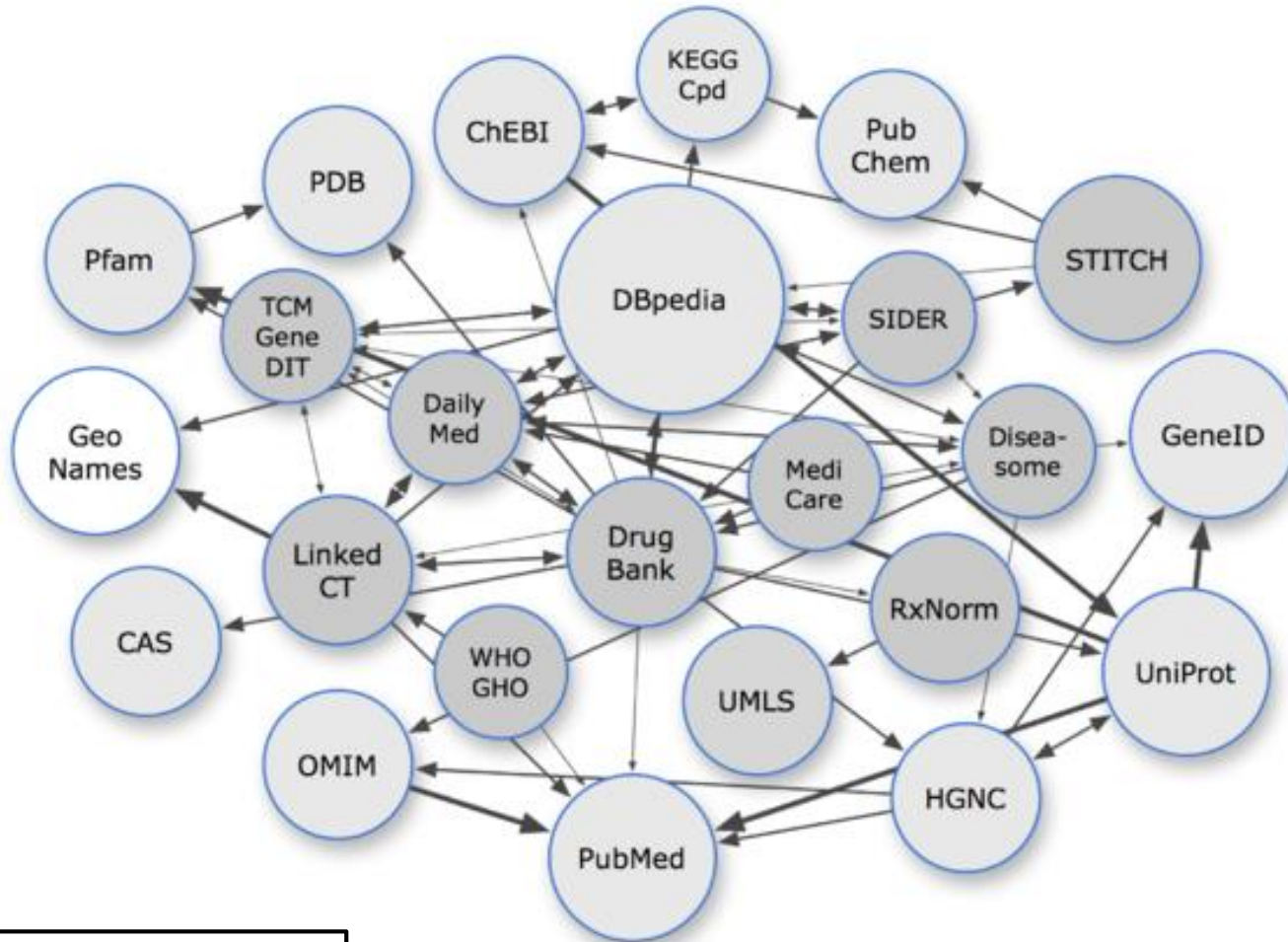
- Search Engine
- Data Analysis
- Discovery
- ...

Clinical



- Diagnosis
- Prognosis
- Treatment
- ...

Linked Open Data in Biomedicine



Source: <http://www.w3.org>



National Key R&D Plan Project:

Knowledgebase of Precision Medicine for Disease Studies

Lead Institute: Fudan University

Lead PI: Lei Liu

Participants:

Duration: 2016–2020

Institute of Medical Informatics, Chinese Academy of Medicine

Chinese Military Academy of Medicine

National Center for Protein Sciences

Shanghai Institute of Life Sciences, Chinese Academy of Science

Beijing Institute of Genomics, Chinese Academy of Science

Harbin Institute of Technology

Zhejiang University

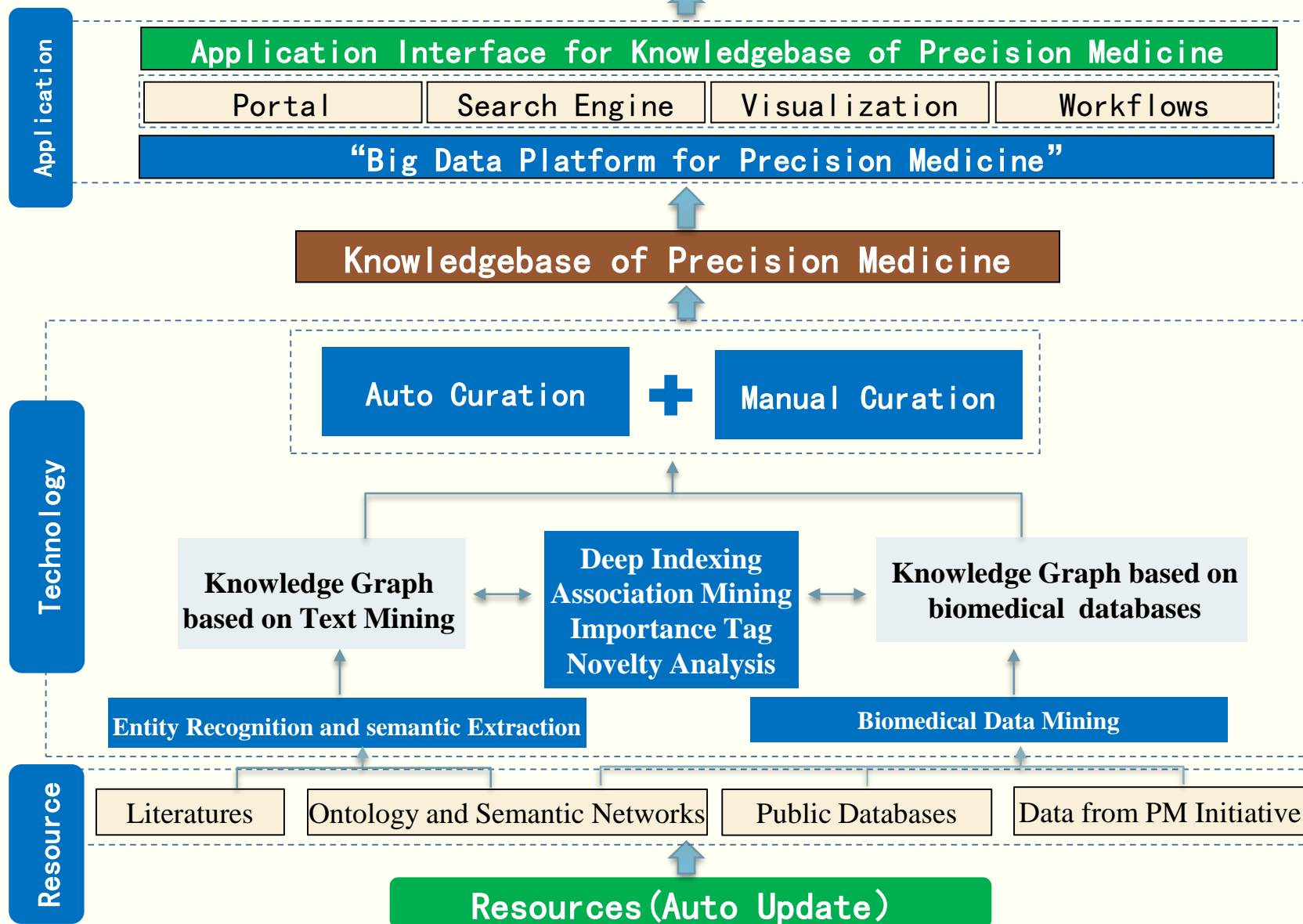
Dalian University of Technology

Shanghai Center for Bioinformation Technology

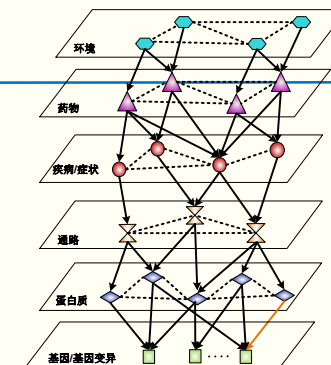
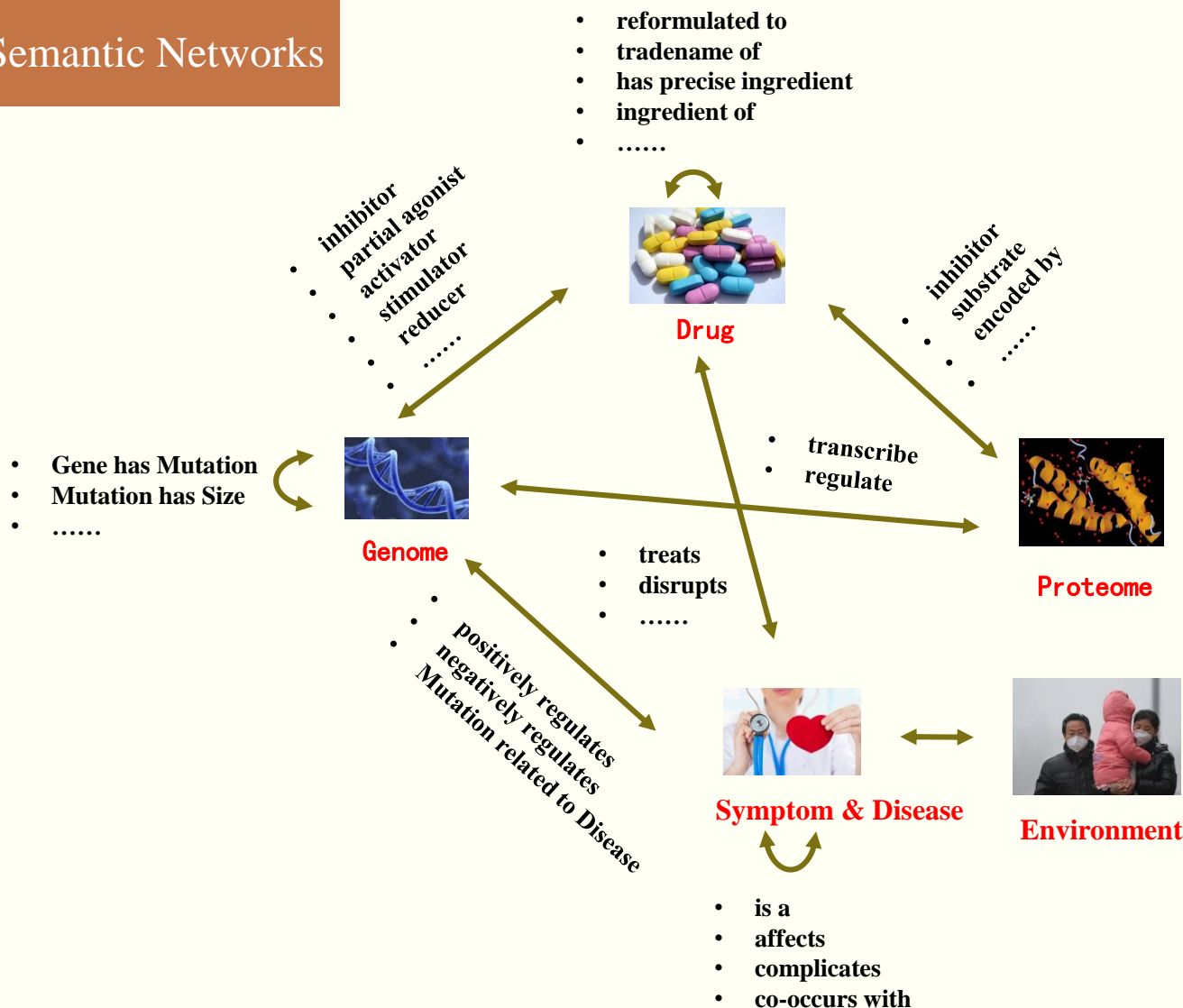
Implementation Schema



USERS (Research and Clinical Application)



PM Ontology and Semantic Networks



面向恶性肿瘤、代谢系统疾病、呼吸系统疾病、心脑血管疾病等重大疾病

Ontologies at Different Levels

Biology

Technology

Length Metric

10^0

10^{-1}

10^{-2}

10^{-3}

10^{-4}

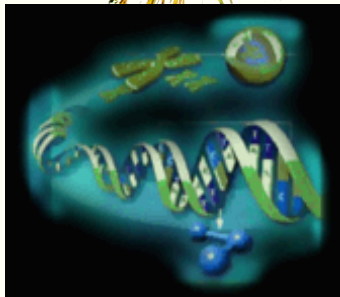
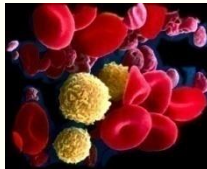
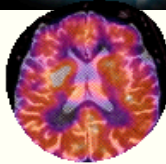
10^{-5}

10^{-6}

10^{-7}

10^{-8}

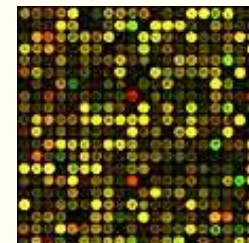
10^{-9}



ICD10(疾病)
SNOMED(疾病)
NCBI 分类系统
SNOMED (器官)

哺乳动物的表型
SNOMED (形态学)
ATCC (细胞株)

Cell Ontology 细胞本体 (细胞类型)
Gene Ontology 基因本体 (亚细胞)
Gene Nomenclature 基因命名
Quaternary code 四进制码



Population

Individual

Organ

Tissue

cell

subcellular

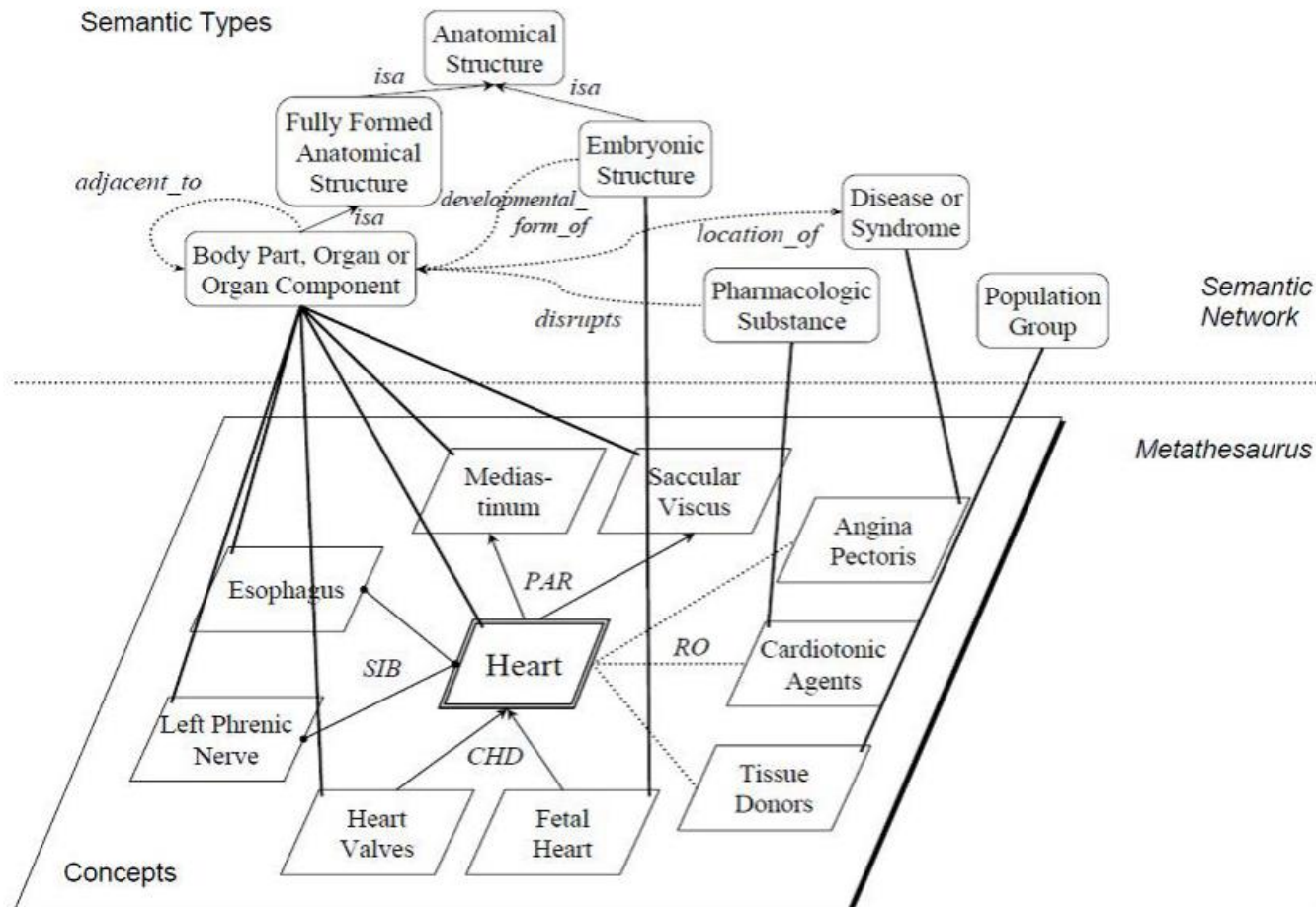
Virus

DNA

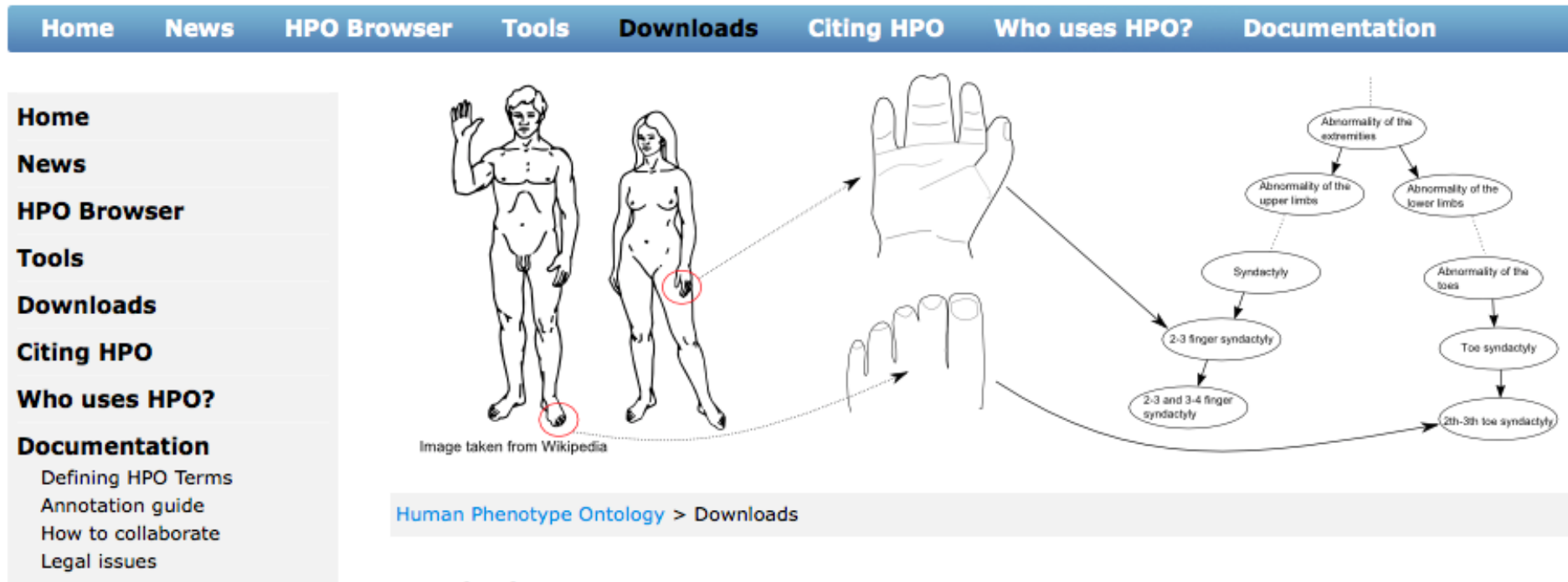
Bases

UMLS Semantic Network

- **UMLS** = Unified Medical Language System (NLM)
- Composed of:
 - Metathesaurus
 - Semantic Network
 - Lexicon
- Contains approximately 5 million codes representing 1 million concepts derived from 100 source terminologies




HPO: human phenotype ontology



Structure between phenotype
Relationship between phenotype and disease

DO: Disease ontology

[Home](#) [Tutorial](#) [Resources](#) [Downloads](#) [FAQ](#) [About](#) [Contact Us](#)



[Go »](#) | [Advanced Search »](#)

Navigation

- Open new metadata panel
- disease
 - disease by infectious agent
 - disease of anatomical entity
 - disease of cellular proliferation
 - disease of mental health
 - disease of metabolism
 - genetic disease
 - physical disorder
 - syndrome

Welcome

The **Disease Ontology** has been developed as a standardized ontology for human disease with the purpose of providing the biomedical community with consistent, reusable and sustainable descriptions of human disease terms, phenotype characteristics and related medical vocabulary disease concepts through collaborative efforts of researchers at Northwestern University, Center for Genetic Medicine and the University of Maryland School of Medicine, Institute for Genome Sciences.

The Disease Ontology semantically integrates disease and medical vocabularies through extensive cross mapping of DO terms to MeSH, ICD, NCI's thesaurus, SNOMED and OMIM.

To get started please visit the [tutorial page](#).

Database Updated

Posted on 2015-04-18

The Disease Ontology database has been updated to the latest ontology as of 2015-04-18. DO now includes two cancer slim files: DO_cancer_slim and TopNodes_DOcancerslim, published in Database.

You can view it [here](#).

Citation: Wu, T.-J., Schriml, L.M., Chen, Q.-R., Colbert, M., Crichton, D. J., Finney, R., Hu, Y., Kibbe, W. A., Kincaid, H., Meerzaman, D., Mitraka, E., Pan, Y., Smith, K. M., Srivastava, S., Ward, S., Yan, C. and Mazumder, R. (2015) Generating a focused view of disease ontology cancer terms for pan-cancer data integration and analysis. Database, 2015, bav032.

Database Updated

Posted on 2015-03-20

The Disease Ontology database has been updated to the latest ontology as of 2015-03-20. DO now includes a DO_MGI_slim.

Disease Ontology paper published in NAR

Posted on 2014-11-20

The Disease Ontology paper has been published in the NAR Database Issue 2015. You can

© 2015 - Institute for Genome Sciences - University of Maryland School of Medicine | [Home](#) | [Tutorial](#) | [Resources](#) | [Downloads](#) | [FAQ](#) | [About](#) | [Contact Us](#) |

Gene ontology



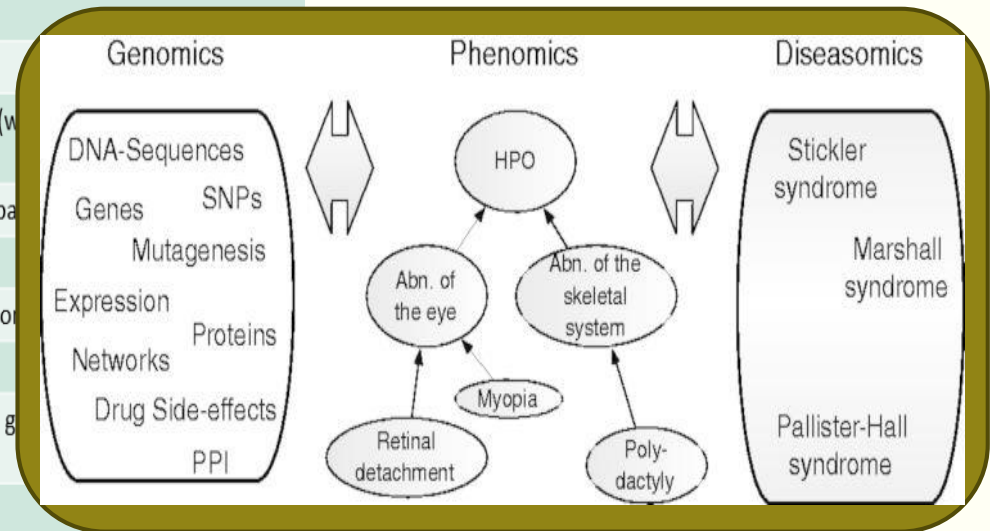
Human phenotype ontology



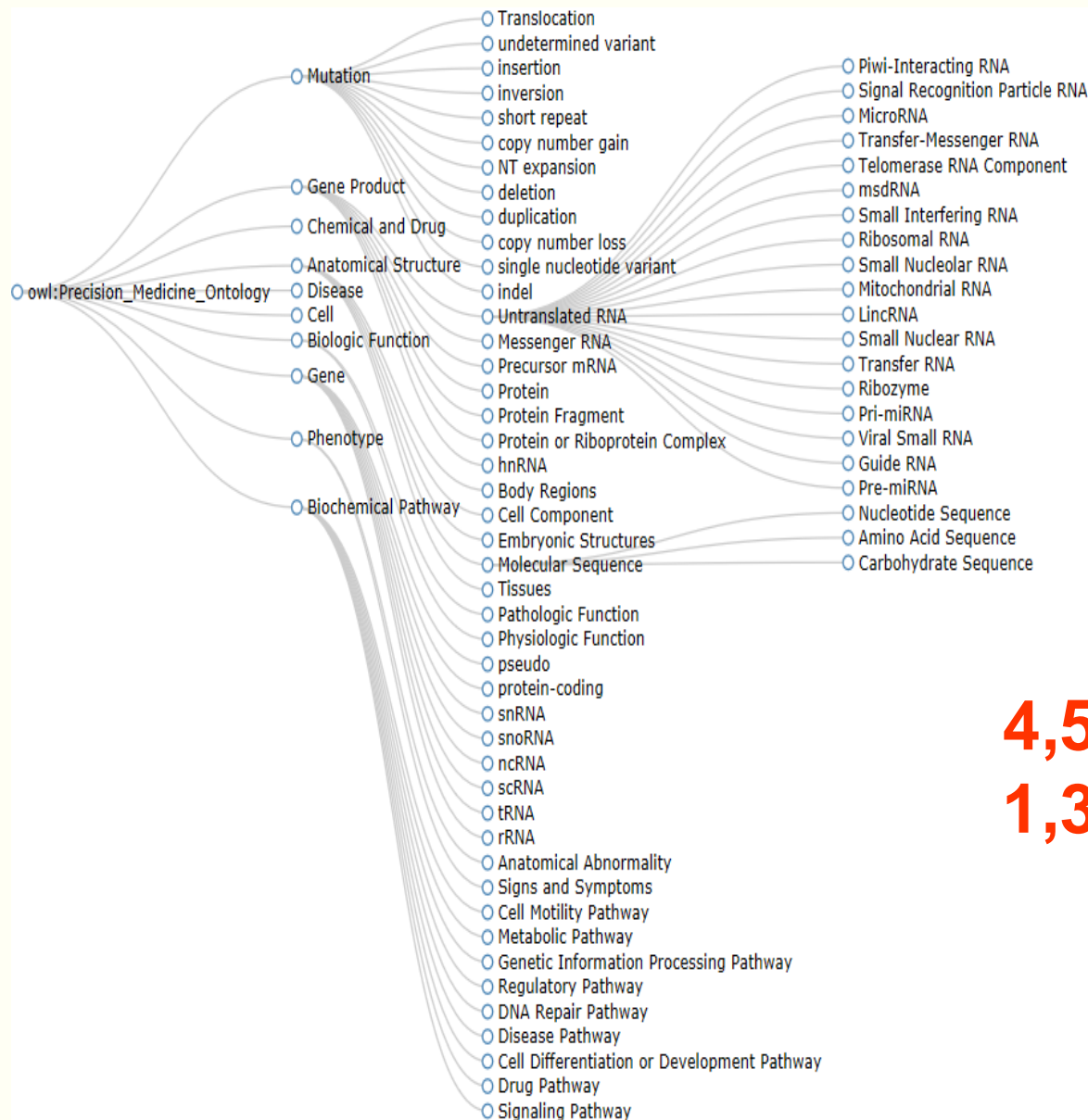
Disease ontology



Relationship	Definition
Gene has Mutation	A mutation occurs in or near a gene, usually at a given position.
Patient/Cohort has Mutation	A patient or cohort has a specific genetic variation.
Mutation related to Disease	A mutation is associated with (or causes) a disease.
Mutation has Size	Indicates the number or frequency of mutations.
Disease related to Gene	A disease is associated with a gene—that is, a gene (w causes a disease.
Disease related to Body Part	A disease may occur in a body part, or have a body pa
Patient has Age	A patient has a given age
Cohort has Age	A summary age for a cohort. Often listed as a mean o
Patient/Cohort has Gender	A patient or cohort is male or female.
Patient/Cohort has Geographic Location	A patient or cohort has a given ethnicity or lives in a g
Patient/Cohort has Disease	A patient or cohort has a disease.
Cohort has Size	The size of a cohort group



PMOV2.0—2018.05



58 ontologies, thesaurus,
controlled vocabularies

SNOMED CT

ICD-10 Version:2016

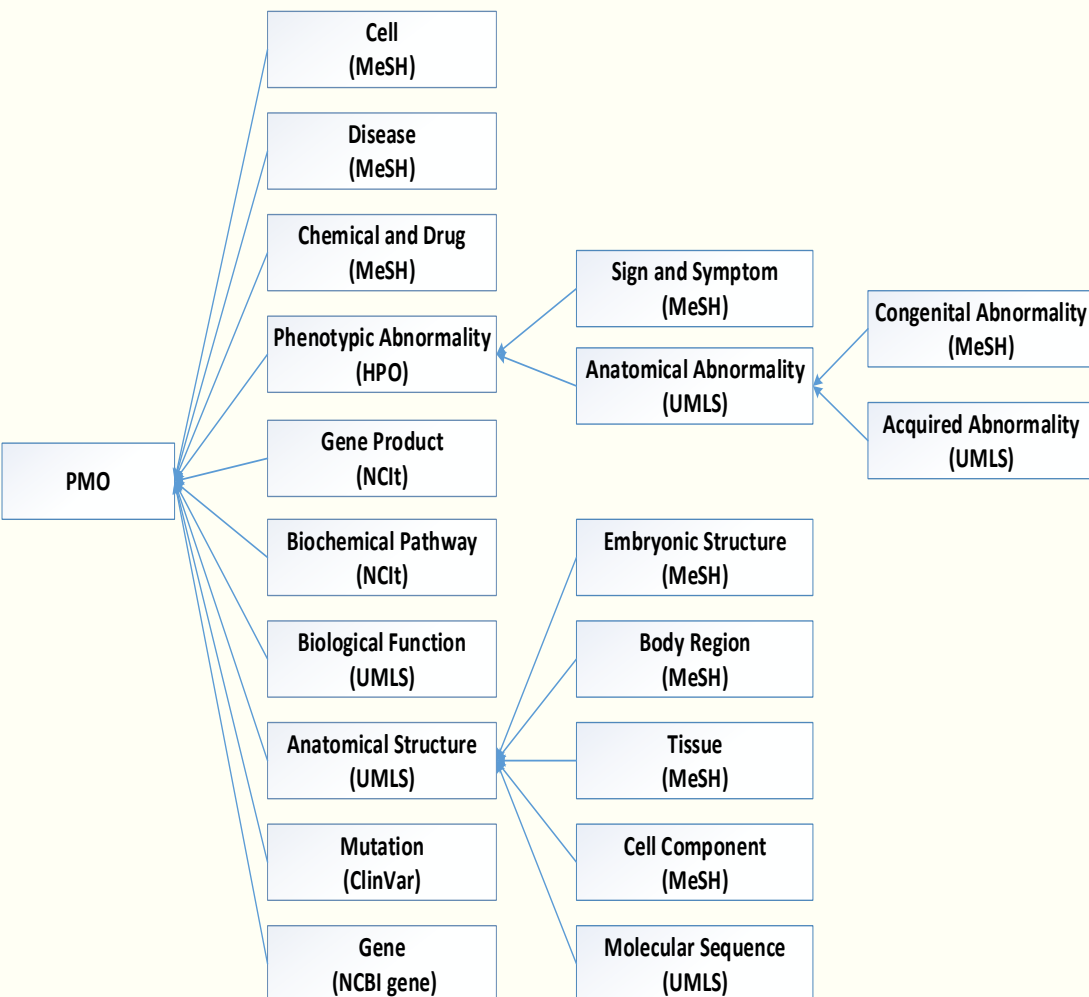
NCI thesaurus



Medical Subject Headings

4,567,208 terms;
1,372,967 concepts

PMO: sources and statistics



NUMBER OF CLASSES	15621
NUMBER OF INSTANCES	707892
NUMBER OF PROPERTIES	93
MAXIMUM DEPTH	13
MAXIMUM NUMBER OF CHILDREN	102
AVERAGE NUMBER OF CHILDREN	4
CLASSES WITH A SINGLE CHILD	1702
CLASSES WITH MORE THAN 25 CHILDREN	83

Concept types related to genetic mutations

Concept	Definition
Gene	A segment of DNA that codes for a protein
Mutation	A mutation is an alteration (deletion, insertion, substitution) of nucleotides (DNA, RNA) or amino acids (Protein)
Body part	An organ or anatomical location in a person.
Disease	An abnormal condition affecting the body of an organism.
Patient	An individual with a disease
Cohort	A group of people; specifically any group or population of people that may be assigned a disease or characteristic.
Size	A number indicating the number of people in a <i>cohort</i> , or the number/frequency of a <i>mutation</i> .
Age	A number or range indicating how old a person/group of people is.
Gender	Terms indicating whether someone is male or female
Geographical location	Terms indicating where a person/group of people comes from, either based on ethnic origin or where they live.
.....

- Verspoor, K., A. Jimeno Yepes, L. Cavedon, T. McIntosh, A. Herten-Crabb, Z. Thomas and J. P. Plazzer (2013). "Annotating the biomedical literature for the human variome." Database (Oxford) **2013**: bat019.

Relationship	Definition
<i>Gene has Mutation</i>	A mutation occurs in or near a gene, usually at a given position.
<i>Patient/Cohort has Mutation</i>	A patient or cohort has a specific genetic variation.
<i>Mutation related to Disease</i>	A mutation is associated with (or causes) a disease.
<i>Mutation has Size</i>	Indicates the number or frequency of mutations.
<i>Disease related to Gene</i>	A disease is associated with a gene—that is, a gene (when mutated) is linked to, or causes a disease.
<i>Disease related to Body Part</i>	A disease may occur in a body part, or have a body part in its name.
<i>Patient has Age</i>	A patient has a given age
<i>Cohort has Age</i>	A summary age for a cohort. Often listed as a mean or an age limit.
<i>Patient/Cohort has Gender</i>	A patient or cohort is male or female.
<i>Patient/Cohort has Geographic Location</i>	A patient or cohort has a given ethnicity or lives in a given place.
<i>Patient/Cohort has Disease</i>	A patient or cohort has a disease.
<i>Cohort has Size</i>	The size of a cohort group
.....

Semantic Relationships in PMO

Properties Tree

Create Delete

- physical interaction
 - physical interaction-protein-microorganism
 - physical interaction-protein-nucleic acid
 - physical interaction-protein-protein
 - physical interaction-protein-small molecule
 - physical interaction-RNA-RNA
- play functional roles in
 - play functional roles-cellular processes in
 - play functional roles-molecular processes in
 - play functional roles-organismal processes in
- play role in biological process**
- play role in cell
- protein functions/functional domains
- regulated by
- regulates
- regulating toxicology
 - regulating toxicology-hepatotoxicity
 - regulating toxicology-nephrotoxicity
- related disease
- secondary to
- targeted by
- targeting drug
 - antagonized by
 - binds by
 - inhibited by
- translocation
- transport
- treated by
- treats

Object property description for play role in biological process

Display name

IRI

Annotations

Domain

Range

Precision Medicine Ontology Collaborative Construction Platform

Precision Medicine Ontology Collaborative Construction Platform

pmo

Home

Class Management

Property Management

Version Management

User Management

You are working on the 1.0.0 version of pmo_Beta

Search

Disease

owl:Thing

+ Anatomical Structure

+ Cell

+ Phenotype

+ Biochemical Pathway

+ Biologic Function

+ Disease

+ Chemical and Drug

+ Gene

+ Gene Product

+ Mutation

Name : Disease

Annotations :

Select

Input Value...

Select

Save

Select

database_cross_reference

Source of Example

Synonym

Annotation properties

Tree Number

dcterms:description

MRID

Definition

Example

dcterms:creator

dcterms:issued

dcterms:created

dcterms:modified

MCID

is_a

dterm:title

rdfs:label

owl:versionInfo

CCPSS:0007490

Select

Save

Delete

CHV:0000004014

Select

Save

Delete

CSP:0944-4756

Select

Save

Delete

ICPC2P:A99001

Select

Save

Delete

LCH:U001423

Select

Save

Delete

LCH:U006403

Select

Save

Delete

LCH_NW:sh85038411

Select

Save

Delete

database_cross_refe

LNC:LP21006-9

Select

Save

Delete

database_cross_refe

MEDCIN:39448

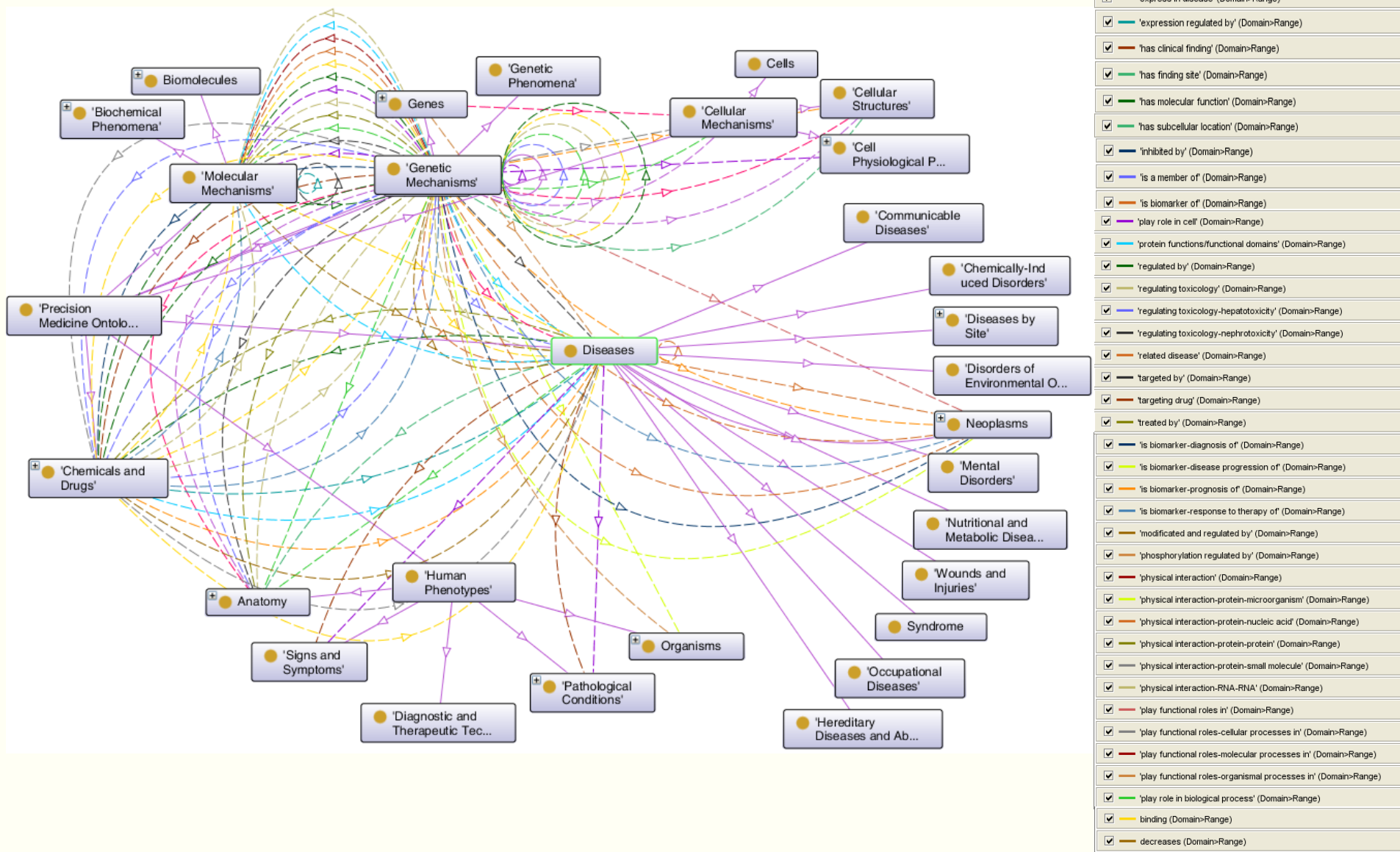
Select

Save

Delete

19

Complex Semantic Relationship



complex semantic relationships embedded in text

6 The presence or absence of MSI was determined for 1,022 colorectal cancers obtained from nine large regional hospitals in southeastern Finland [9].

7 There were 895 (87.6%) MSI- cancers and 127 (12.4%) MSI+ cancers.

8 The MSI+ cancers were further classified as sporadic (N = 98 or 9.6% of all cancers) or HNPCC (N = 29 or 2.9% of all cancers) based on germline M.

10 Ages at cancer can be used to estimate likely numbers of oncogenic mutations required before transformation [3-6,11].

11 Average ages for sporadic MSI+, MSI-, and HNPCC cancers were respectively 71.5, 67.5, and 50.3 years (Figure 1A).

15 We genotyped p.Lys618Ala in 1034 individuals (373 sporadic colorectal cancer [CRC] patients, 250 index subjects from families suspected of having LS [revised Bethesda guidelines] and 411 controls).

16 Three well-characterized LS families that fulfilled the Amsterdam II Criteria and consisted of members with the p.Lys618Ala variant were included to assess co-occurrence and co-segregation.

17 A subset of colorectal tumour DNA samples from 17 patients carrying the p.Lys618Ala variant was screened for microsatellite instability using five mononucleotide markers.

19 Results

21 Twenty-seven individuals were heterozygous for the p.Lys618Ala variant; nine had sporadic CRC (2.41%), seven were suspected of having hereditary CRC (2.8%) and 11 were controls (2.68%).

22 There were no significant associations in the case-control and case-case studies.

23 The p.Lys618Ala variant was co-existent with pathogenic mutations in two unrelated LS families.

Search

Text Entity Event Relation

Type - Any -

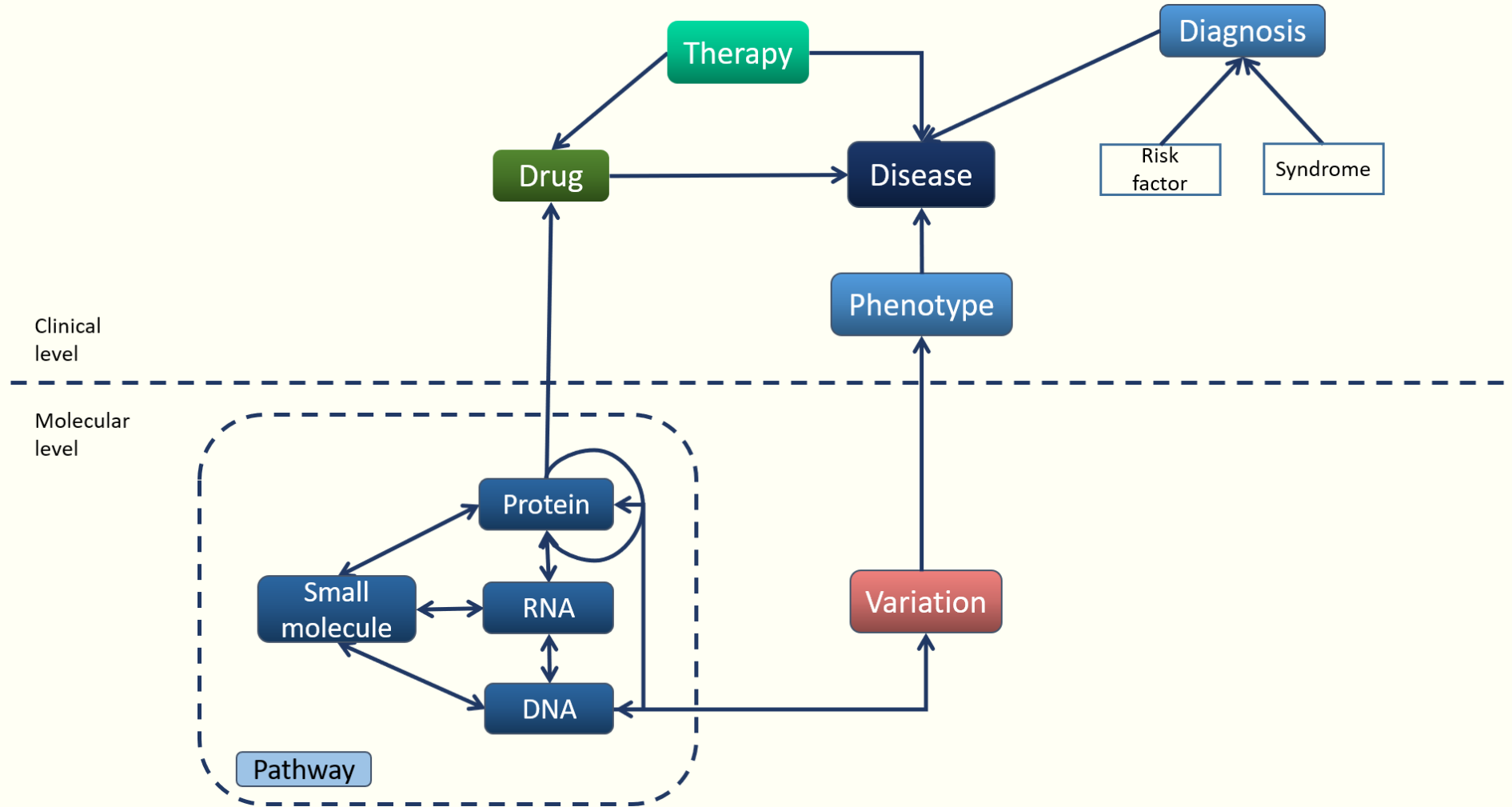
Trigger - Any -

- An v

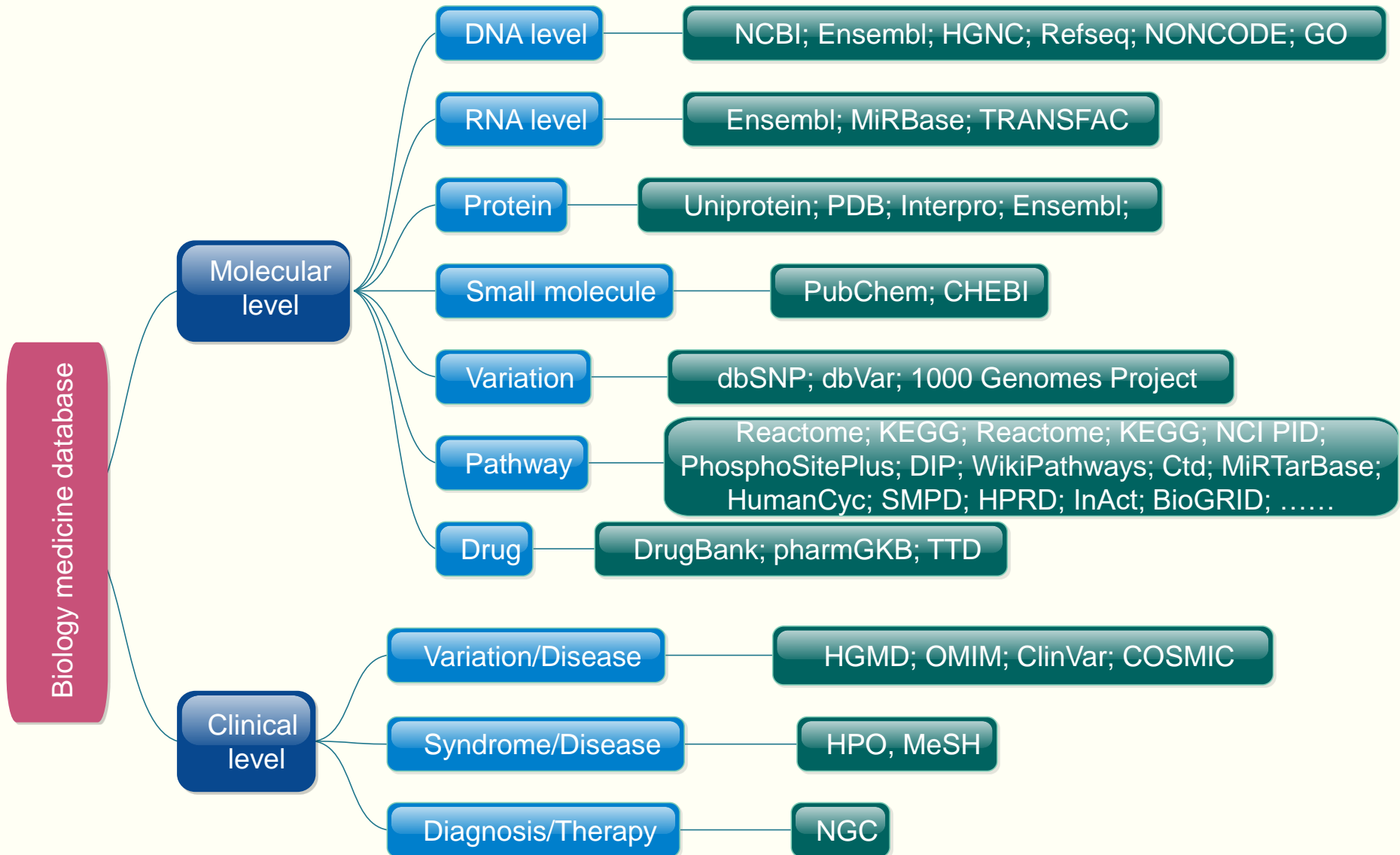
- DNA methylation
- DNA demethylation
- Protein_modification_process
- Acylation
- Acetylation
- Palmitoylation
- Alkylation
- Methylation
- Glycosylation
- Hydroxylation
- Phosphorylation
- Lipidation
- Prenylation
- Protein_modification_by_small_protein_conjugation
- Neddylation
- Sumoylation
- Ubiquitination
- Deacylation

<http://www.opennicta.com.au/home/health/variome>

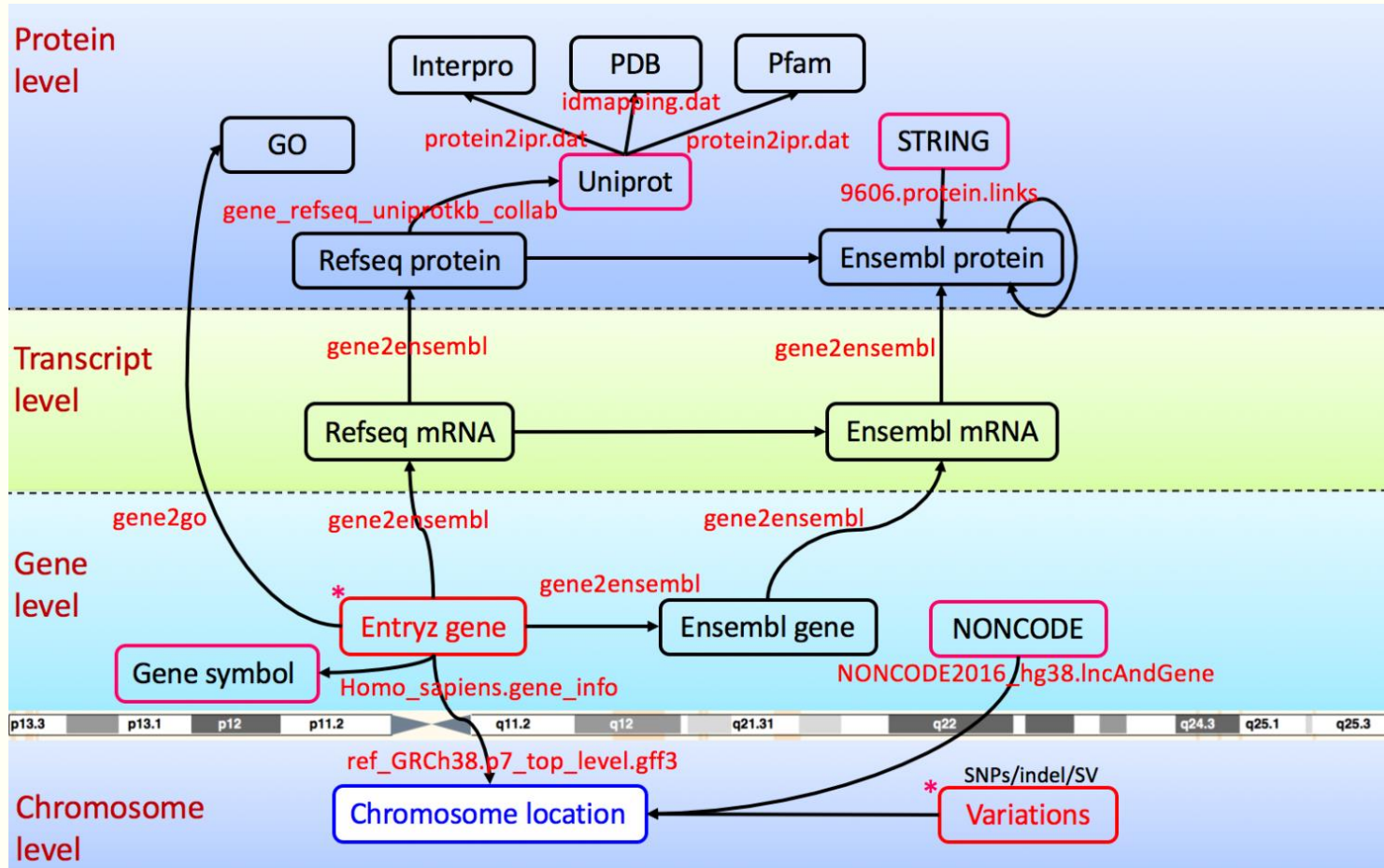
Top Design of Knowledge Integration



Integration of Databases



Genomics Data Integration



gene	Protein coding	20,656
	Protein no-coding	38,943

Variation Integration

Variation ↔ Disease



- 10,725 Diseases
- 18,022 genes

Variation



Mutation → 38,836 Genes

dbSNP

dbSNP

Short Genetic Variations

dbVar

ClinVar

GaP

PubMed

Nucleotide

Protein

Search small variations in dbSNP or large structural variations in dbVar

Search Entrez dbSNP for Go

Have a question about dbSNP? Try searching the SNP FAQ Archive!

Go

ANNOUNCEMENT

Interested in structural variations? Visit NCBI [dbVar](#)

GENERAL

RSS Feed

Contact Us

Site Map

dbSNP Homepage

NCBI Variation

Resources

Announcements

dbSNP Summary

FTP Download

HUMAN VARIATION

SNP SUBMISSION

DOCUMENTATION

SEARCH

RELATED SITES

Search by IDs on All Assemblies

Note: **rs#** and **ss#** must be prefixed with "rs" or "ss", respectively (i.e. rs25, ss25)

ID: Reference cluster ID(rs#)

Submission Information

- By Submitter
- New Submitted Batches
- Method
- Population
- Publication

Batch

- Enter List
 - NCBI Assay ID(ss)
 - Reference SNP ID(rs)
 - Local SNP ID
- Upload List
 - NCBI Assay ID(ss)
 - Reference SNP ID(rs)
 - Local SNP ID

[Batch Query Help](#)

NCBI: [Disclaimer](#) [Privacy statement](#)

NCBI Resources How To

dbVar Genome Browser

Homo sapiens: GRCh37 (GCF_000001405.13) Chr 2 (NC_000002.11): 1.299M - 1.299M

[Reset All](#) [Share this page](#) [Help](#) [Back to Setup Page](#)

View on Genome

Select Assembly

GRCh37

Select an assembly to change view

Search

Name	Location
rs9677798	Chr2 1,298,915

Your Data

Region Summary

Data in view

Click (-) to remove track

Study ID	Variant Calls
Study ID	Variant Calls

Data available for region

Click (+) to add track

Study ID	Variant Calls
estd203	1 (+)

dbSNP build 141: 61,060,456

<http://www.ncbi.nlm.nih.gov/SNP/>

 [Resources](#)  [How To](#)  [Sign in to NCBI](#)

ClinVar

ClinVar

Search ClinVar for gene symbols, HGVS expressions, conditor

Search

AdvancedHelp

Home

About

Data use and maintenance

Using the website

How to submit

Statistics

FTP site

ClinVar submissions

This page summarizes the number of genes and distinct variant locations currently represented in ClinVar from independent submissions. A gene is reported if a variant in ClinVar is either found within, or includes, that gene. Thus the number of genes is as high as it is because of the structural variants in the database that span many genes.

This page lists all submitters and the summary of their contributions. We acknowledge their support.

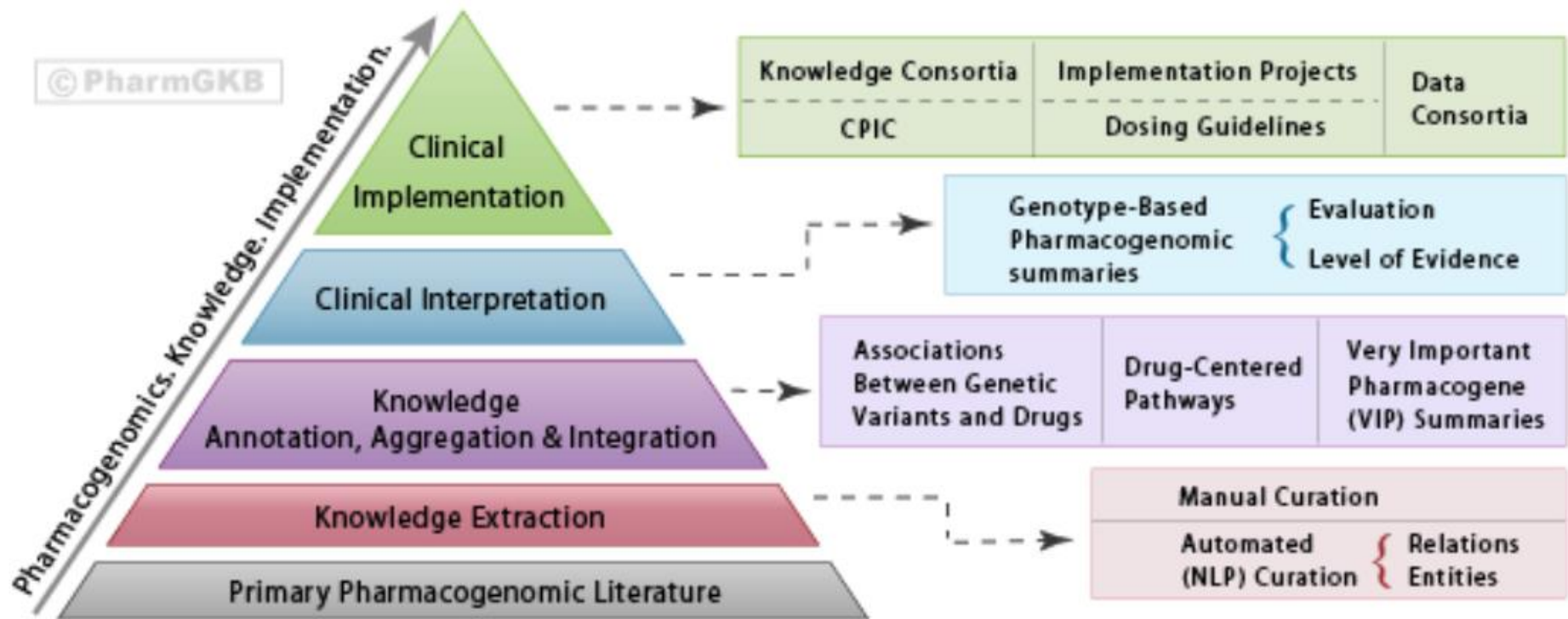
Submission overview

Category of analysis	Current total (Jun 08, 2015)
Total accessioned submissions	175033
Total genes represented	26422
Total genes, in submissions with assertions, with variants in one gene	7468
Total genes, in submissions with assertions, with variants in multiple genes	26235
Total variations represented	147215
Total variations, in submissions with assertions	120309
Total submitters	332

Submitters and their submissions

Submitter	Total submissions	Total submissions with assertions	Total Genes	Last updated
OMIM: Johns Hopkins University	25662	25659	3798	Jun 05, 2015
Precancer Genomics; Leeds Institute of Molecular Medicine	22512	22512	3477	Jun 20, 2012

The PharmGKB Knowledge Pyramid



PharmGKB提供以下信息:

VA: Variant Annotations

PW: Drug-Centered Pathway

VIP: Very Important Pharmacogene Summaries

CA: Clinical Annotations

DG: Pharmacogenomics-Based Drug-Dosing Guidelines

DL: Drug Labels with Pharmacogenomic Information

文章发表情况:

Total Publications: 284

- [Pathway Publications](#): 38
- [Guideline Publications](#): 23
- [VIP Publications](#): 32
- [Other Publications](#): 191

Downloads

In addition to the PharmGKB website, we are pleased to make PharmGKB data and knowledge available as downloads. We have found that it is often critical to check with our curators at feedback@pharmgkb.org before embarking on a large project using these data, to be sure that the files and data we make available are being interpreted correctly. PharmGKB generally does NOT need to be a co-author on such analyses; we just want to make sure that there is a correct understanding of our data before lots of resources are spent.

[Examples of papers that have been written by others using PharmGKB information](#)

Primary PharmGKB Data

[Versions of external data sources used in these files](#)

- Genes: [genes.zip](#) (3 MB)
- RSID mapping: [rsid.zip](#) (151 MB)
- Drugs: [drugs.zip](#) (456 KB)
- Diseases: [diseases.zip](#) (312 KB)
- Pathways in [BioPax](#) format: [pathways-biopax.zip](#) (484 KB)
- Pathways in [tsv](#) format: [pathways-tsv.zip](#) (67 KB)
- Dosing Guidelines in [JSON](#) format: [dosingGuidelines.json.zip](#) (414 KB)
- Drug Label summary in [tsv](#) format: [drugLabels.zip](#) (6 KB)

Variant and Clinical Annotations Data

To access PharmGKB's variant and clinical annotations data please submit a [Variant and Clinical Annotations Data Request](#).

Variant, Gene and Drug Relationship Data

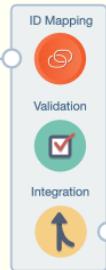
To access PharmGKB's variant, gene and drug relationship data please submit a [Variant, Gene and Drug Relationship Data Request](#).

Network Integration

Data Sources

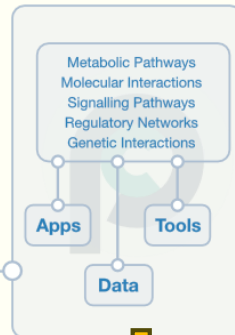
Pathway & interaction databases

BIND
BioGRID
CORUM
CTD
DIP
DrugBank
HPRD
HumanCyc
IntAct (Complex)
INOH
KEGG
miRTarBase
MSigDB
NetPath
PANTHER Pathway
NCI - PID
PhosphoSitePlus
Reactome
Recon X
SMPD
WikiPathways



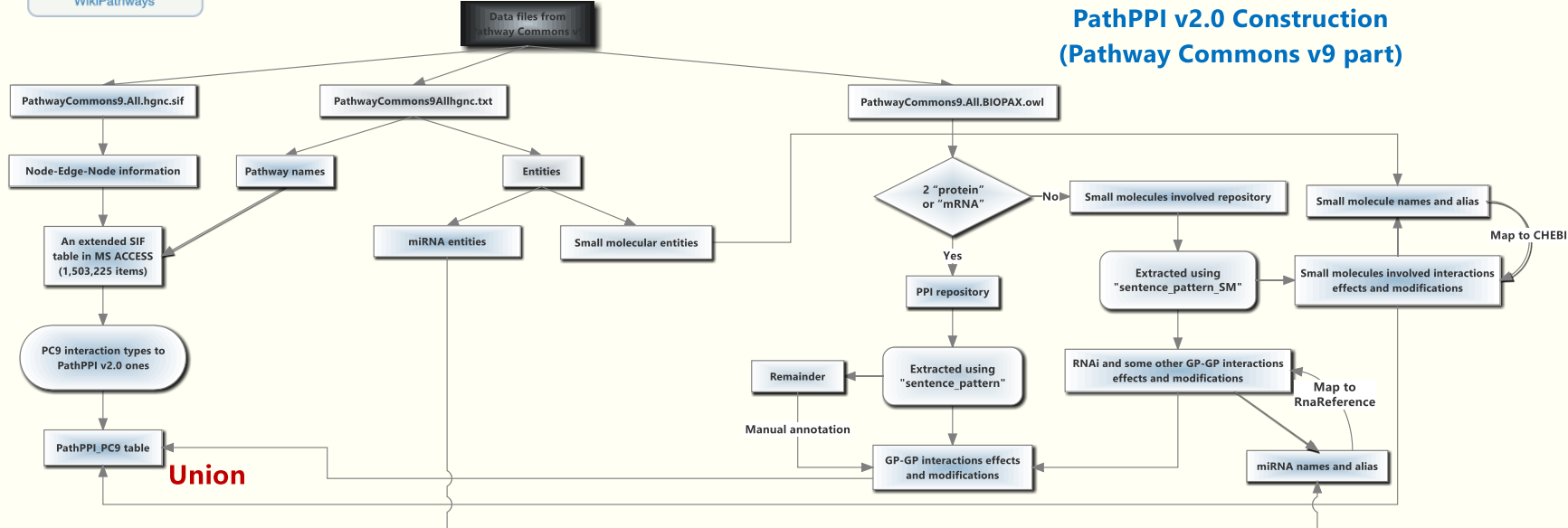
Pathway Commons

Over 4 000 pathways
1.3 million interactions



**Pathway Commons v9, BioPAX L3 →
PathPPI v2.0**

PathPPI v2.0 Construction (Pathway Commons v9 part)



Network Integration

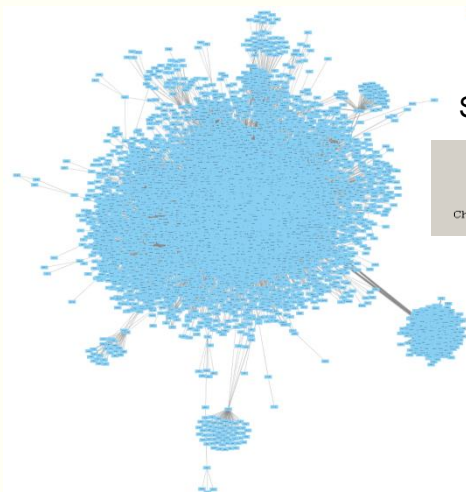
Pathway Type	Edge Type	Directionality	Edge No.
Signaling Pathway	SR: S ignaling R egulation	Directed	172,765
	ER: E xpression R egulation	Directed	122,786
	CAI: C omplex A ssembly I nteraction	Undirected	177,227
	TR: T ransport R egulation	Directed	7,296
	TRc: T ransport R egulation c hemical	Directed	3,285
	ca: x c hemical a ffects P	Directed	469,519
	RNAi: RNAi nterference	Directed	317,556
TechPPI	TechPPI: T echnical P rotein- P rotein I nteraction	Undirected	316,437
Metabolic Pathway	sp: metabolic reaction $s \xrightarrow{E} p \xrightarrow{F}$	Directed	14,428
	sE: metabolic reaction $s \xrightarrow{E} p \xrightarrow{F}$	Directed	22,480
	Ep: metabolic reaction $s \xrightarrow{E} p \xrightarrow{F}$	Directed	21,334
	EE: metabolic reaction $s \xrightarrow{E} p \xrightarrow{F}$	Directed	154,975
	rw: x r eacts w ith y	Undirected	3,912

Network

Source Database	PathPPI v1.0	PathPPI v2.0
Reactome	2012.03.14	v56
NCI PID	2012.03.17	2015.17.27
NetPath	2012.04.14	2012.04.14
INOH	2010.01.31	2011.03.22 (v4.0)
KEGG	2009.10.12	2011.07
BioCarta	2010.08.13	
SPIKE	2011.03.22	
PhosphoSitePlus		2016.03.15
HumanCyc		v19.5
PANTHER Pathway		v3.4
TRANSFAC		v7.4
miRTarBase		v4.5
DrugBank		v4.3
Recon X		v2.03
Ctd		2015.06.03
SMPD		v2.0
WikiPathways		2015.09.29
HPRD	R9	R9
IntAct	2013.10.22	2016.02.16
BioGRID	R3.2.105	R3.44.135
DIP	2013.07.07	2016.01.14
MINT	2013.03.26	
BIND		2010.12.15
CORUM		2012.02.17

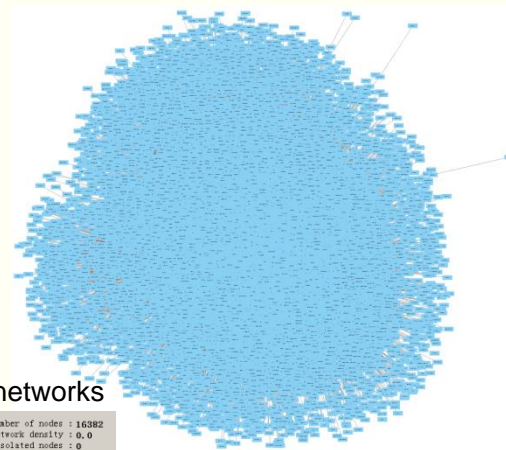
26 Modification

Modification	v1.0	v2.0	Modification	v1.0	v2.0
Acetylation	●	●	ADP-ribosylation		●
Botinylation	●		Acylation		●
Decanoylation	●		Alkylation		●
Dimethylation	●		Amination		●
Farnesylation	●	●	Carbamoylation		●
Fucosylation	●		Carboxylation		●
Galactosylation	●		Ethylation		●
Geranylgeranylation	●		Geranoylation		●
Glucosylation	●		Glucuronidation		●
Glycosylation	●	●	Glutathionylation		●
Glycylation	●		Glycation		●
Hydroxylation	●	●	N-glycosylation		●
Lipoylation	●	●	O-glycosylation		●
Methylation	●	●	Nitrosation		●
Myristoylation	●		Prenylation		●
Octanoylation	●		Ribosylation		●
Palmitoylation	●		Sulfation		●
Phosphopantetheine	●				
Phosphorylation	●	●			
Sumoylation	●	●			
Trimethylation	●				
Ubiquitination	●	●			



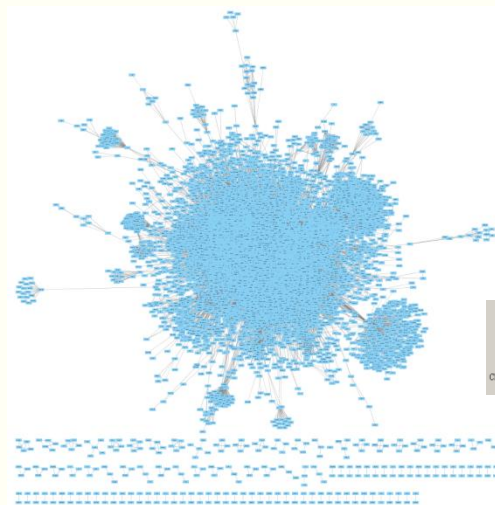
Signaling networks

Clustering coefficient : 0.105	Number of nodes : 8911
Connected components : 13	Network density : 0.0
Network diameter : 10	Isolated nodes : 0
Network radius : 1	Number of self-loops : 0
Shortest paths : 31576996 (39%)	Multi-edge node pairs : 7855
Characteristic path length : 3.860	Analysis time (sec) : 13.015
Avg. number of neighbors : 37.013	



Expression regulatory networks

Clustering coefficient : 0.201	Number of nodes : 16382
Connected components : 1	Network density : 0.0
Network diameter : 10	Isolated nodes : 0
Network radius : 1	Number of self-loops : 0
Shortest paths : 19916671 (7%)	Multi-edge node pairs : 524
Characteristic path length : 5.632	Analysis time (sec) : 12.774
Avg. number of neighbors : 14.926	

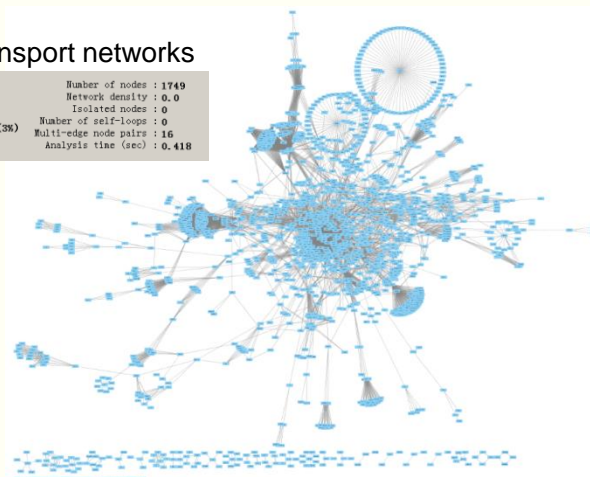


Complex assembling networks


Clustering coefficient : 0.539	Number of nodes : 8831
Connected components : 174	Network density : 0.005
Network diameter : 11	Network heterogeneity : 1.580
Network radius : 1	Isolated nodes : 0
Network centralization : 0.097	Number of self-loops : 0
Shortest paths : 64281890 (88%)	Multi-edge node pairs : 0
Characteristic path length : 3.743	Analysis time (sec) : 19.294
Avg. number of neighbors : 41.549	

Transport networks

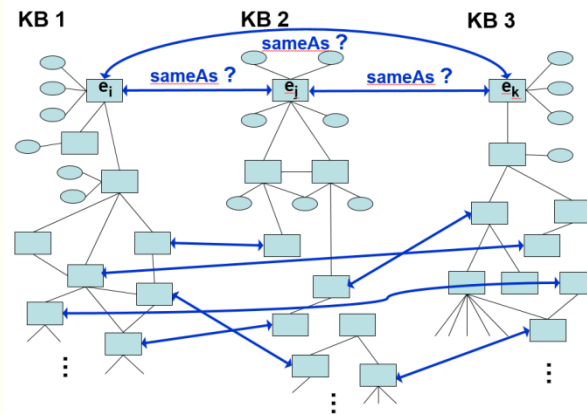
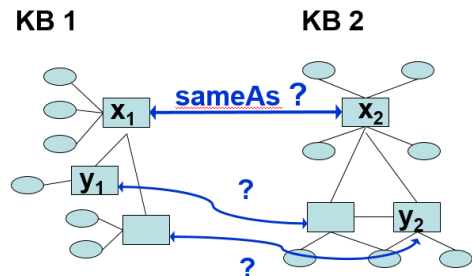
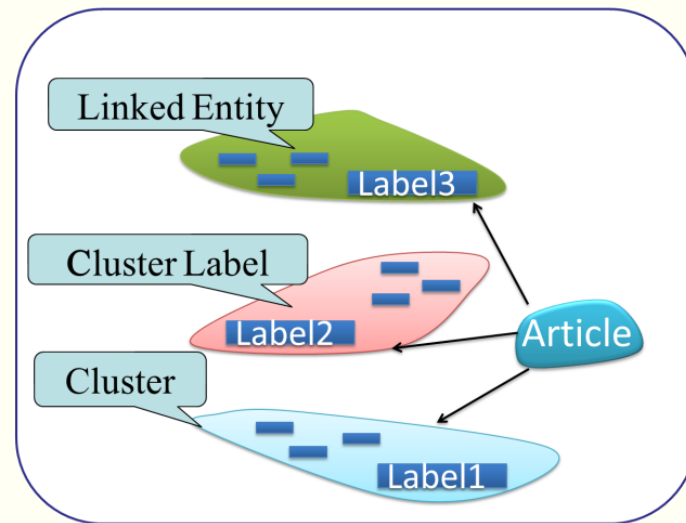
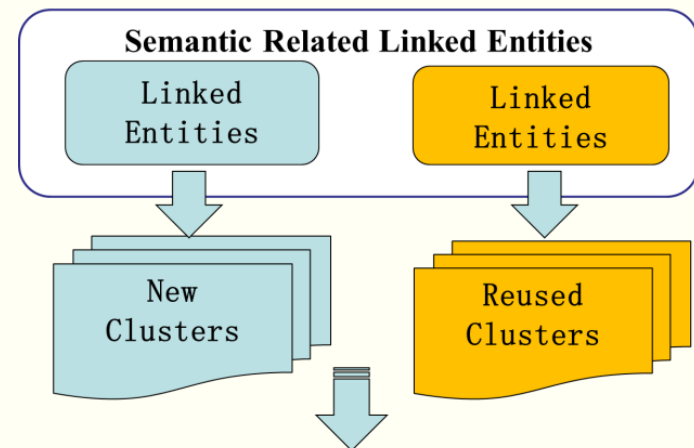
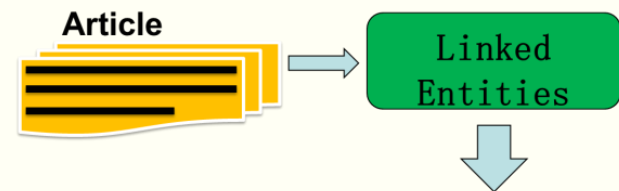
Clustering coefficient : 0.025	Number of nodes : 1749
Connected components : 34	Network density : 0.0
Network diameter : 12	Isolated nodes : 0
Network radius : 1	Number of self-loops : 0
Shortest paths : 93773 (3%)	Multi-edge node pairs : 16
Characteristic path length : 4.626	Analysis time (sec) : 0.418
Avg. number of neighbors : 8.325	



Knowledge Graph

Gene Name	Gene Id	Alternative Transcripts	OMIM	 COSMIC <small>Catalogue of somatic mutations in cancer</small> Record
TP53	COSG501	TP53_ENST0000054585, TP53_ENST0000026930	191170	

Subject	Predicate	Object	RDF
COSG501	Gene name	TP53	
COSG501	Alternative Transcripts	TP53_ENST0000054585	
COSG501	Alternative Transcripts	TP53_ENST0000026930	
COSG501	OMIM	OMIM_TP53	

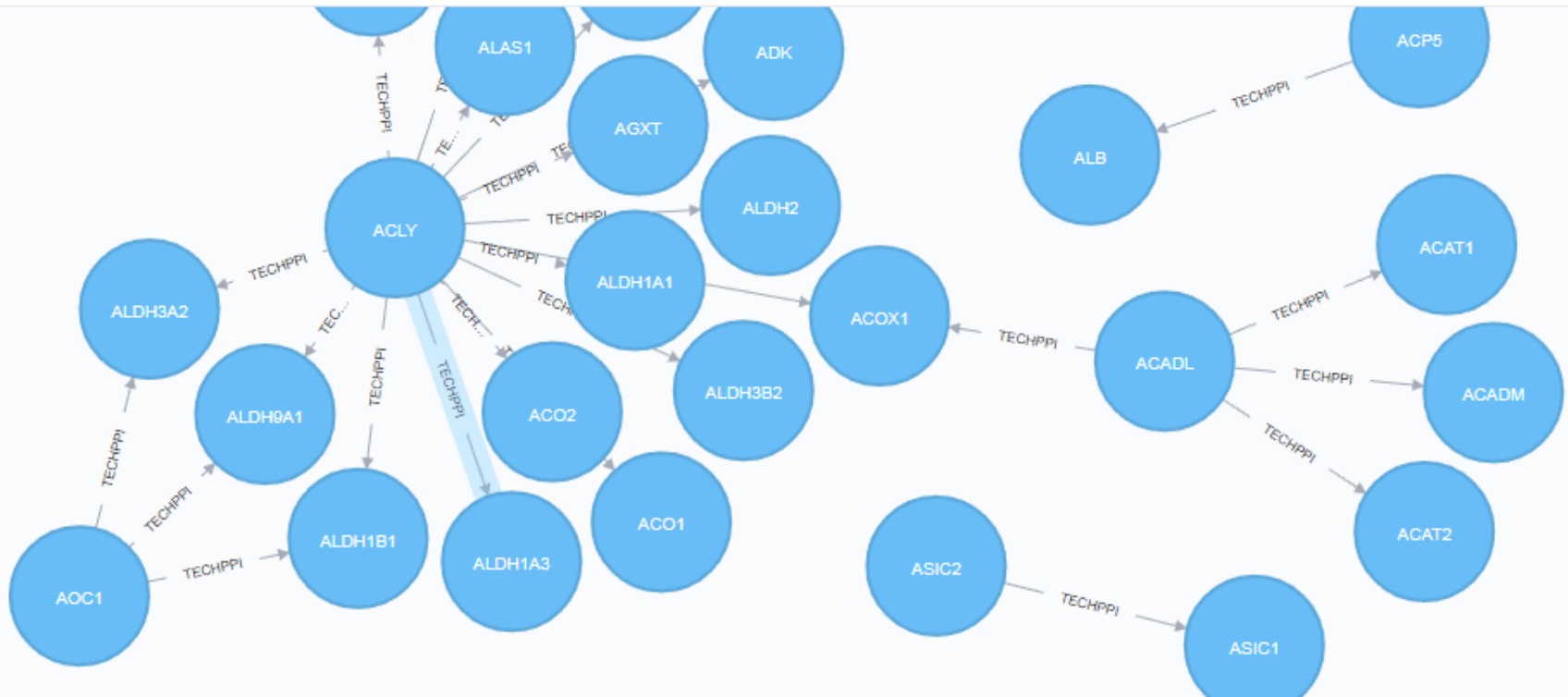


Visualization of Knowledge Graph

```
$ match(a)-[r:TECHPPI]->(b) return a,b limit 25
```

*(26)

GENE(26)



TECHPPI <id>: 41647 sourceDB: STRING

Tools for the Knowledgebase: PMap

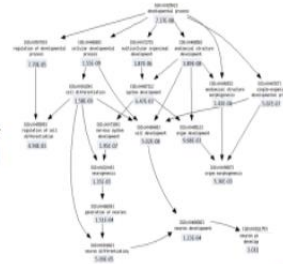
PMap

Home Retrieve Analysis Explore Help Contract Us

PMap : a Precision Medicine knowledgebase Platform

PMap is a disease Knowledge for Precision Medicine, which includes the information from biological molecular, human disease, phenotype, drugs, etc. PMap supports functions of retrieval by various biomedical terms, workflow analyses for omics datasets and also the intelligent knowledge discovery.

PMap is supported by the Chinese Program of Precision Medicine (Construction of Precision Medicine Knowledgebase for Disease Research, 2016YFC0901905).



KB>>Pathway

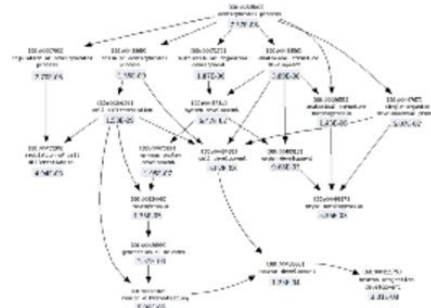
TP53

Advanced

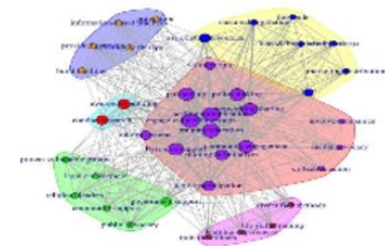
SEARCH



Literature



Ontology



PMap-ODAS

Search Engine

PM

Search

GENE:EGF

Display

AllNone

- ☒ Function
- ☒ Protein
- ☒ Interaction
- ☒ Drugs
- ☒ Disease

Gene | **EGF**1950

Description | **epidermal growth factor**

Location | *Homo Sapiens (Human)*4q25 [genome browser](#)

Function

GO - Molecular function

- Ras guanyl-nucleotide exchange factor activity
- phosphatidylinositol-4,5-bisphosphate 3-kinase activity
- growth factor activity
- calcium ion binding
- transmembrane receptor protein tyrosine kinase activator activity
- protein tyrosine kinase activity
- epidermal growth factor receptor binding
- Wnt-activated receptor activity
- protein binding
- Wnt-protein binding

GO - Biological process

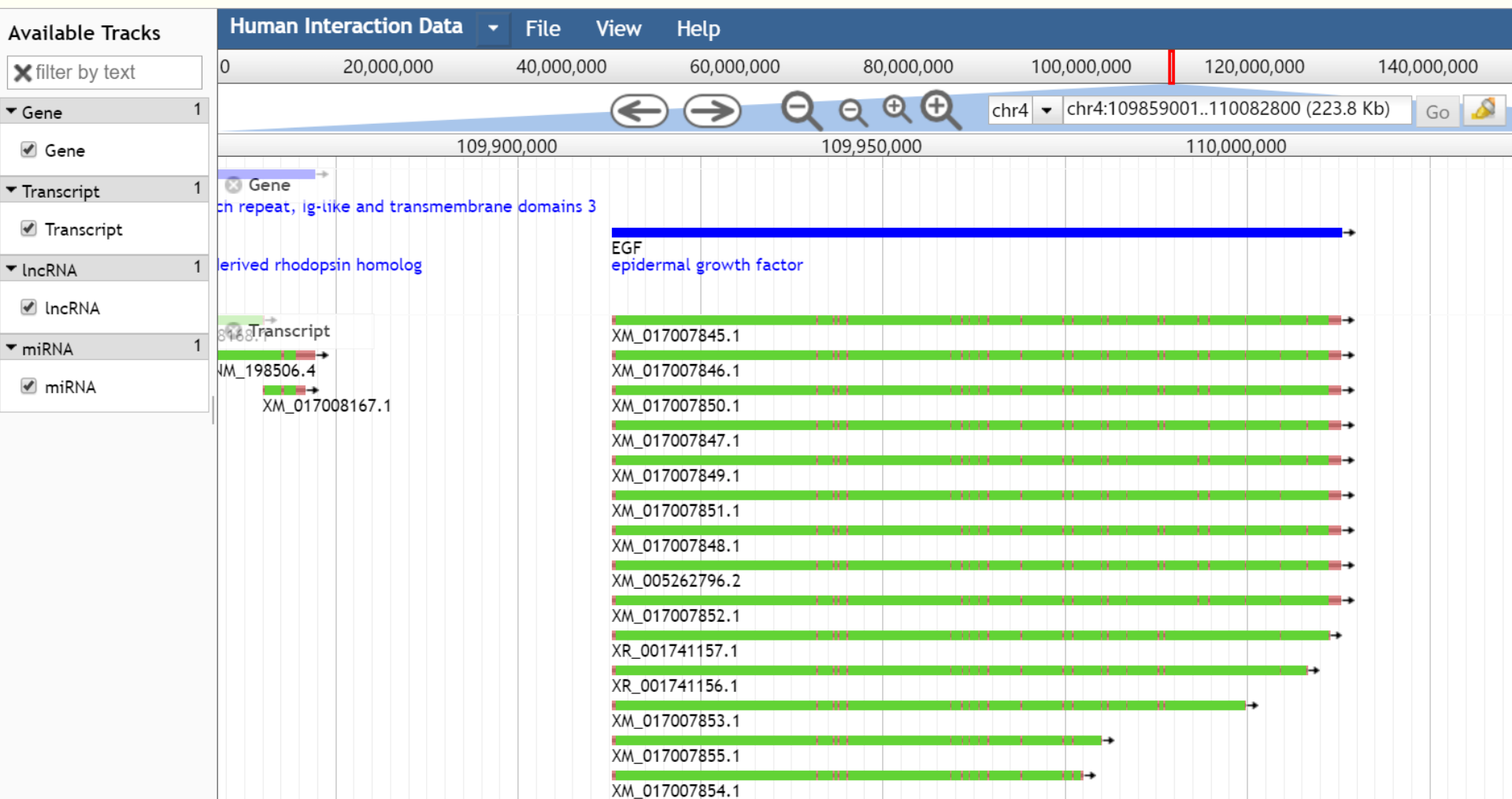
- activation of MAPK activity
- activation of MAPKK activity
- negative regulation of epidermal growth factor receptor signaling pathway

Proteins

Uniprot

Uniprot	Description	PDB
P01133	Pro-epidermal growth factor	1IVO 1JL9 1NQL 1P9J 2KV4 3NJP

Genome browser: EGF



EGF Related Drugs

Drugs

- **cetuximab** Source:PharmGKB
- bevacizumab Source:PharmGKB
- cholecalciferol Source:PharmGKB
- gefitinib Source:PharmGKB
- oxaliplatin Source:PharmGKB
- fluorouracil Source:PharmGKB
- irinotecan Source:PharmGKB
- ethanol Source:PharmGKB
- capecitabine Source:PharmGKB
- dexamethasone Source:PharmGKB
- docetaxel Source:PharmGKB
- panitumumab Source:PharmGKB
- Fatty acid derivatives Source:PharmGKB
- Insulins And Analogues Source:PharmGKB
- estrogens Source:PharmGKB
- glucocorticoids Source:PharmGKB

Display

AllNone

<input checked="" type="checkbox"/>	Function
<input checked="" type="checkbox"/>	Protein
<input checked="" type="checkbox"/>	Interaction
<input checked="" type="checkbox"/>	Drugs
<input checked="" type="checkbox"/>	Disease

Disease

OMIM

OMIM	Hypomagnesemia 4, renal, 611718 (3)
------	-------------------------------------

COSMIC

Show 10 entries

Drug-Drug Interaction

PM

Search

DRUG:Cetuximab

Display

AllNone

- ☒ Interaction
- ☒ Disease
- ☒ Genes

Drug | **Cetuximab** DrugBank DB00002

Formular | **C6484H10042N1732O2023S36**

UNII | PQX0D8J21J

PharmGKB | PA10040

Interaction

Drugs

Drug	Description
Acetyldigitoxin	Acetyldigitoxin may decrease the cardiotoxic activities of Cetuximab.
Belimumab	The risk or severity of adverse effects can be increased when Cetuximab is combined with Belimumab.
Bevacizumab	Bevacizumab may increase the cardiotoxic activities of Cetuximab.
Cabazitaxel	The risk or severity of adverse effects can be increased when Cabazitaxel is combined with Cetuximab.
Cyclophosphamide	Cyclophosphamide may increase the cardiotoxic activities of Cetuximab.
Deslanoside	Deslanoside may decrease the cardiotoxic activities of Cetuximab.
Digitoxin	Digitoxin may decrease the cardiotoxic activities of Cetuximab.

Cetuximab Targets and Treatments

Trastuzumab

Trastuzumab may increase the cardiotoxic activities of Cetuximab.

Foods

Diseases

- Cetuximab, used in combination with irinotecan, is indicated for the treatment of EGFR-expressing, metastatic colorectal carcinoma in patients who are refractory to irinotecan-based chemotherapy. Cetuximab administered as a single agent is indicated for t [Source:drugbank](#)
- Colorectal cancer [Source:TTD](#)

Genes

- Epidermal growth factor receptor [Source:drugbank](#)
- Low affinity immunoglobulin gamma Fc region receptor III-B [Source:drugbank](#)
- Complement C1r subcomponent [Source:drugbank](#)
- Complement C1q subcomponent subunit A [Source:drugbank](#)
- Complement C1q subcomponent subunit B [Source:drugbank](#)
- Complement C1q subcomponent subunit C [Source:drugbank](#)
- Low affinity immunoglobulin gamma Fc region receptor III-A [Source:drugbank](#)
- Complement C1s subcomponent [Source:drugbank](#)
- High affinity immunoglobulin gamma Fc receptor I [Source:drugbank](#)
- Low affinity immunoglobulin gamma Fc region receptor II-a [Source:drugbank](#)
- Low affinity immunoglobulin gamma Fc region receptor II-b [Source:drugbank](#)
- Low affinity immunoglobulin gamma Fc region receptor II-c [Source:drugbank](#)
- cyclin-dependent kinase 2 [Source:PharmGKB](#)
- cyclin-dependent kinase 4 [Source:PharmGKB](#)

Display

AllNone

- ☒ Interaction
- ☒ Disease
- ☒ Genes

History

- ▼ Project_20170810
- Sample
- ▼ Workflow Analysis
- ✓ R:Differential gene analysis
- + Add analysis
- ▼ Genelist1
- ✓ GO
- ✓ KEGG
- + Add anotation
- ▼ Genelist2
- ✓ GO
- Network
- + Add annotation
- + Add genelist
- ▼ Project_20170814
- Sample
- ▼ Workflow Anaysis
- ✓ R:Differential gene analysis
- + Add analysis
- + Add genelist
- + New Project

Project_20170810

R:Differential Gene Analysis

Data

Chart

Basic Information

Name:	GeneList	Create time:	20170210 16:57
Gene Num:	750	Update time:	20170210 17:09

Choose Gene List Annotation:

Ontology Annotations

☒ Gene Ontology (GO)

☐ Disease Ontology (DO)

☐ Human Phenotype Ontology (HPO)

☐

Network Annotation

Molecular network

☐ Protein network

☐ Transcription factor regulatory network

Next

Showing 1 to 10 of 750 rows

Search

Ontology Annotations

- ☒ Gene Ontology (GO)
- ☐ Disease Ontology (DO)
- ☐ Human Phenotype Ontology (HPO)
- ☐

Network Annotation

Molecular network

- ☐ Protein network
- ☐ Transcription factor regulatory network
- ☐ Phosphorylation network
- ☐ Ubiquitination network
- ☐

Pathway

- ☐ KEGG analysis
- ☐ Reactome Analysis

Knowledge network

- ☐

History



▼ Project_20170810



▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

▼ Genelist1

✓ GO

✓ KEGG

+ Add anotation

▼ Genelist2

✓ GO

Network

+ Add annotation

+ Add genelist

▼ Project_20170814



▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

+ Add genelist

+ New Project

Project_20170810

R: GeneOntology

Basic Information

Name: GOAnalysisResult Create time: 20170210 16:57

GO Version: *****版本 Update time: 20170210 17:09

Input data: GeneList(750) Size on disk: 5 kb

Table View

Tree View

Download

Mapped Gene Num.: 500

Sig. GO terms: 50

Category Display Display GO Full Hierachy ⓘ

PValue >= 0.05

GO terms view: table ⓘ

Targets >= 2

Selected 30 GO terms, [Export to tree view.](#)

Sig. Depleted Depleted Enriched Sig. Enriched

GO term ID	GeneRatio	BgRatio	PValue	p.adjust	Targets
GO:0005515	372(76.23%)	10893(64.14%)	3.92E-09	3.92E-06	372
GO:0008009	12(2.46%)	48(0.28%)	7.94E-09	3.92E-06	12
GO:0042379	12(2.46%)	58(0.34%)	7.43E-08	2.58E-05	12
GO:0003678	11(2.25%)	50(0.29%)	1.34E-07	3.48E-05	11
GO:0043138	6(1.23%)	12(0.07%)	4.36E-07	9.07E-05	6
GO:0042802	71(14.55%)	1376(8.10%)	9.10E-07	0.000152493	71
GO:0015631	25(5.12%)	289(1.70%)	1.03E-06	0.000152493	25
GO:0008017	21(4.30%)	217(1.28%)	1.26E-06	0.000164208	21
GO:0048248	4(0.82%)	5(0.03%)	3.39E-06	0.000356267	4
GO:0045236	6(1.23%)	16(0.09%)	3.42E-06	0.000356267	6
GO:0004386	16(3.28%)	150(0.88%)	6.73E-06	0.000586543	16
GO:0019901	35(7.17%)	541(3.19%)	7.03E-06	0.000586543	35
GO:0019900	38(7.79%)	612(3.60%)	7.32E-06	0.000586543	38
GO:0005524	71(14.55%)	1473(8.67%)	1.01E-05	0.000747546	71
GO:0003777	11(2.25%)	77(0.45%)	1.19E-05	0.000793709	11
GO:0032559	72(14.75%)	1509(8.89%)	1.22E-05	0.000793709	72
GO:0030554	72(14.75%)	1520(8.95%)	1.56E-05	0.000954452	72
GO:0003688	5(1.02%)	13(0.08%)	2.04E-05	0.001180507	5
GO:0003682	31(6.35%)	482(2.84%)	2.64E-05	0.001448643	31
GO:0043168	110(22.54%)	2666(15.70%)	3.63E-05	0.001888793	110
GO:0003690	41(8.40%)	738(4.35%)	4.36E-05	0.002162286	41
GO:0016787	105(21.52%)	2544(14.98%)	5.69E-05	0.002694465	105
GO:0048020	7(1.43%)	36(0.21%)	6.28E-05	0.002841317	7
GO:0097367	93(19.06%)	2203(12.97%)	7.42E-05	0.003218342	93
GO:0035173	5(1.02%)	17(0.10%)	8.92E-05	0.003696812	5
GO:0061575	3(0.61%)	4(0.02%)	9.23E-05	0.003696812	3
GO:0005488	439(89.96%)	14281(84.09%)	1.00E-04	0.003854447	439

Records per page: 10 ▾ Showing 1 to 10 of 1000 GO terms

First Previous 1 2 3 4 5 Next Last

History



▼ Project_20170810



▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

▼ Genelist1

✓ GO

✓ KEGG

+ Add anotation

▼ Genelist2

✓ GO

Network

+ Add annotation

+ Add genelist

▼ Project_20170814



▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

+ Add genelist

+ New Project

Project_20170810

R: GeneOntology

Basic Information

Name: GOAnalysisResult Create time: 20170210 16:57

GO Version: *****版本 Update time: 20170210 17:09

Input data: GeneList(750) Size on disk: 5 kb

Table View

Tree View

Download

Mapped Gene Num.: 500 Sig. GO terms: 50

Category Display Display GO Full Hierachy ⓘ PValue >= 0.05

GO terms view: ontology ⓘ Targets >= 2

Selected 30 GO terms, [Export to tree view.](#) ■ Sig. Depleted ■ Depleted ■ Enriched ■ Sig. Enriched

GO term ID		GeneRatio	BgRatio	PValue	p.adjust	Targets
Edit filters.		Edit filters. ▾	Edit filter: ▾	Edit filter: ▾	Edit filter: ▾	Edit filters: ▾
<input checked="" type="checkbox"/> GO:0003674	Molecular function	372(76.23%)	10893(64.14%)	No Filter	1.92E-06	372
<input checked="" type="checkbox"/> GO:0009888	Transcription Factor	12(2.46%)	48(0.28%)	Equal	1.92E-06	12
<input checked="" type="checkbox"/> GO:0001071	Activity, Protein Binding	12(2.46%)	58(0.34%)	Not Equal	1.58E-05	12
<input checked="" type="checkbox"/> GO:0003677	Nucleic Acid Binding	11(2.25%)	50(0.29%)	<input checked="" type="checkbox"/> Less	1.48E-05	11
<input checked="" type="checkbox"/> GO:0003723	Transcription Factor	6(1.23%)	12(0.07%)	Greater	1.07E-05	6
<input checked="" type="checkbox"/> GO:0003729	Activity	71(14.55%)	1376(8.10%)	1.03E-06	0.000152493	71
<input checked="" type="checkbox"/> GO:0003924	DNA Binding	25(5.12%)	289(1.70%)	1.26E-06	0.000152493	25
<input checked="" type="checkbox"/> GO:0004386	RNA Binding	21(4.30%)	217(1.28%)	3.29E-06	0.000164208	21
<input checked="" type="checkbox"/> GO:0004518	MRNA Binding	4(0.82%)	5(0.03%)	3.42E-06	0.000356267	4
<input checked="" type="checkbox"/> GO:0004871	MRNA Binding	6(1.23%)	16(0.09%)	6.73E-06	0.000356267	6
<input checked="" type="checkbox"/> GO:0005198	GTPase Activity	16(3.28%)	150(0.88%)	7.03E-06	0.000586543	16
<input checked="" type="checkbox"/> GO:0008092	Helicase Activity	35(7.17%)	541(3.19%)	7.32E-06	0.000586543	35
<input checked="" type="checkbox"/> GO:0008134	Helicase Activity	38(7.79%)	612(3.60%)	1.01E-05	0.000747546	38
<input checked="" type="checkbox"/> GO:0008168	Nuclease Activity	71(14.55%)	1473(8.67%)	1.19E-05	0.000747546	71
<input checked="" type="checkbox"/> GO:0008233	Signal Transducer Activ	11(2.25%)	77(0.45%)	1.22E-05	0.000793709	11
<input checked="" type="checkbox"/> GO:0008289	Structural Molecule	72(14.75%)	1509(8.89%)	1.56E-05	0.000793709	72
<input checked="" type="checkbox"/> GO:0008330	Activity	72(14.75%)	1520(8.95%)	2.04E-05	0.000954452	72
<input checked="" type="checkbox"/> GO:0008423	Cytoskeletal Protein	5(1.02%)	13(0.08%)	2.64E-05	0.001180507	5
<input checked="" type="checkbox"/> GO:0008545	Binding	31(6.35%)	482(2.84%)	3.63E-05	0.001180507	31
<input checked="" type="checkbox"/> GO:0008648	Binding	110(22.54%)	2666(15.70%)	4.36E-05	0.001448643	110
<input checked="" type="checkbox"/> GO:0008742	Transcription Factor	41(8.40%)	738(4.35%)	5.69E-05	0.001888793	41
<input checked="" type="checkbox"/> GO:0008839	Binding	105(21.52%)	2544(14.98%)	6.28E-05	0.002162286	105
<input checked="" type="checkbox"/> GO:0008955	Methyltransferase	7(1.43%)	36(0.21%)	7.42E-05	0.002694465	7
<input checked="" type="checkbox"/> GO:0009003	Activity	93(19.06%)	2203(12.97%)	8.92E-05	0.002841317	93
<input checked="" type="checkbox"/> GO:0009031	Peptidase Activity	5(1.02%)	17(0.10%)	9.23E-05	0.003218342	5
<input checked="" type="checkbox"/> GO:0009239	Lipid Binding	3(0.61%)	4(0.02%)	1.00E-04	0.003696812	3
<input checked="" type="checkbox"/> GO:0009408	Protein Transporter	439(89.96%)	1428(84.09%)		0.003854447	439

Records per page: 10 Showing 1 to 10 of 1000 GO terms

First Previous 1 2 3 4 5 Next Last

History



▼ Project_20170810



 Sample

Workflow Analysis

- ✓ R: Differential gene analysis

+ Add analysis

▼ Geneliste

✓ KEGG

+ Add anotation

▼ Genelist2

 Network

+ Add annotation

[+ Add genelists](#)

▼ Project_20170814



 Sample

▼ Workflow Analysis

- ✓ R: Differential gene analysis

+ Add analysis

[+ Add genelists](#)

[+ New Project](#)

Project_20170810

R: GeneOntology

Basic Information

Name: GOAnalysisResult



Create time: 20170210 16:57

GO Version: *****版本

Update time: 20170210 17:09

Input data: GeneList(750)

Size on disk: 5 kb

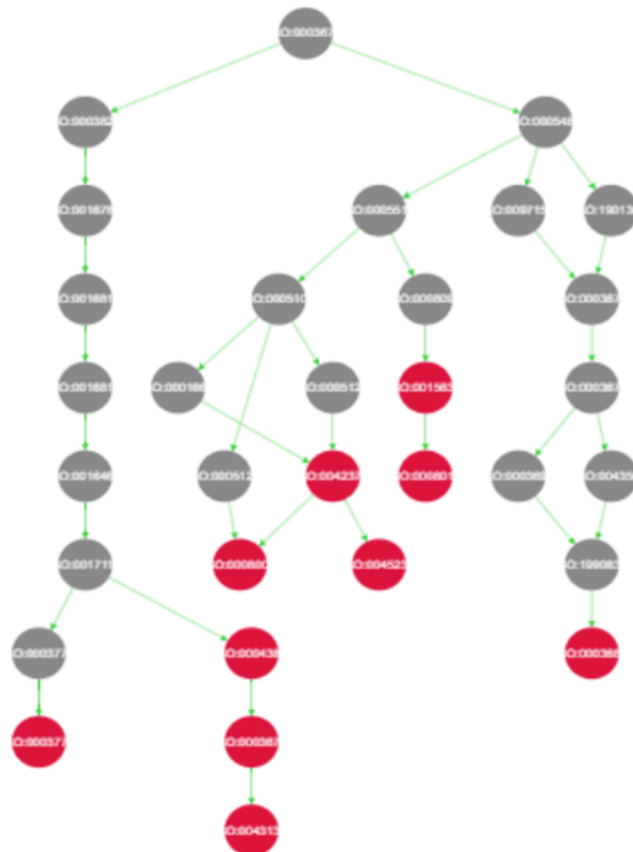
Table View

Tree View

 [Export picture](#) [Download](#)

[Download](#)

Tree View of the Selected GO Terms



History



▼ Project_20170810



Sample

▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

▼ Genelist1

✓ GO

✓ KEGG

+ Add annotation

▼ Genelist2

✓ GO

Network

+ Add annotation

+ Add genelist

▼ Project_20170814



Sample

▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

+ Add genelist

+ New Project

Project_20170810

R: ProteinNetwork

Basic Information

Name: NetworkAnalysisResult  Create time: 20170210 16:57

Sig. Networks: 50 Update time: 20170210 17:09

Input data: GeneList(750) Size on disk: 5 kb

Table View

Network View

 Download

Sig. Networks: 50 Score >= 20

Records per page: 10 700 Targets >= 2

Default

GO

Selected 30 GO terms, [Export to Network view.](#)

Network ID ▾	Molecules in Network ▾	Score ▾	Targets ▾
<input type="checkbox"/> 1	vANN, ^APOBR, ^ARPIN/C15orf38-AP3S2, ^CADM3, vCHCHD10, vENPP4, ^FOXA1, ^FOXA2, ^GATA3, vITGB4, vITGB7, ^RCID1, ^KLF2, vLHFP	40	35
<input type="checkbox"/> 2	vADAD2, ^RPI, ^APP, ^CCDC149, vTOX2, vENPP4, ^FOXA1, ^FOXA2, ^GATA3, vITGB4, vITGB7, ^RCID1, ^KLF2, vLHFP	39	34
<input type="checkbox"/> 3	vCCDC149, vTOX2, vENPP4, ^FOXA1, ^FOXA2, ^GATA3, vITGB4, vITGB7, ^RCID1, ^KLF2, vLHFP	35	35
<input type="checkbox"/> 4	^CADM3, vCHCHD10, vENPP4, ^FOXA1, vANN, ^APOBR	34	32
<input type="checkbox"/> 5	^CADM3, vCHCHD10, vENPP4, ^FOXA1	30	31

Records per page: 10 Showing 1 to 10 of 50 rows

First Previous 1 2 3 4 5 Next Last

Targets mapped to this term:

Code	GO ID	GO Description	Targeted ID	Targeted Name	Gene Description
1.4	GO:0008009	chemokine activity	3627	A3GALT2	Alpha-1,3-galactosyltransferase 2 (EC 2.4.1.87) (Isoglobotriaosylceramide synthase) (iGb3 synthase) (iGb3S) [A3GALT2P] [IGBS3S]
1.4	GO:0008009	chemokine activity	10563	AADACL3	Arylacetamide deacetylase-like 3 (EC 3.1.1.-)
1.4	GO:0008009	chemokine activity	6373	AADACL4	Arylacetamide deacetylase-like 4 (EC 3.1.1.-)
1.4	GO:0008009	chemokine activity	4283	ABCA4	Retinal-specific ATP-binding cassette transporter (ATP-binding cassette sub-family A member 4) (RIM ABC transporter) (RIM protein) (RmP) (Stargardt disease protein) [ABCR]
1.4	GO:0008009	chemokine activity	6362	ABCB10	ATP-binding cassette sub-family B member 10, mitochondrial precursor (ATP-binding cassette transporter 10) (ABC transporter 10 protein) (Mitochondrial ATP-binding cassette 2) (M-ABC2)
1.4	GO:0008009	chemokine activity	6355	ABCD3	ATP-binding cassette sub-family D member 3 (70 kDa peroxisomal membrane protein) (PMP70) [PMP70] [PXMP1]
1.4	GO:0008009	chemokine activity	2921	ABL2	Abelson tyrosine-protein kinase 2 (EC 2.7.10.2) (Abelson murine leukemia viral oncogene homolog 2) (Abelson-related gene protein) (Tyrosine-protein kinase ARG) [ABLL] [ARG]
1.4	GO:0008009	chemokine activity	6364	ACADM	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial precursor (EC 1.3.8.7) (MCAD)
1.4	GO:0008009	chemokine activity	3576	ACAP3	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 3 (Centaurin-beta-5) (Cnt-b5) [CENTB5]
1.4	GO:0008009	chemokine activity	6352	ACBD3	Golgi resident protein GCP60 (Acyl-CoA-binding domain-containing protein 3) (Golgi complex-associated protein 1) (GOCAP1) (Golgi phosphoprotein 1) (GOLPH1) (PBR- and PKA-associated protein 7) (Peripheral benzodiazepine receptor-associated protein PAP7) [Contains: Golgi resident protein GCP60, N-terminally processed] [GCP60]
1.4	GO:0008009	chemokine activity	6347	ACBD6	Acyl-CoA-binding domain-containing protein 6
1.4	GO:0008009	chemokine activity	6351	ACKR1	Atypical chemokine receptor 1 (Duffy antigen/chemokine receptor) (Fy glycoprotein) (GpFy) (Glycoprotein D) (Plasmodium vivax receptor) (CD234 antigen) [DARC] [FY] [GPD]

History



▼ Project_20170810



Sample

▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

▼ Genelist1

✓ GO

✓ KEGG

+ Add anotation

▼ Genelist2

✓ GO

Network

+ Add annotation

+ Add genelist

▼ Project_20170814



Sample

▼ Workflow Analysis

✓ R:Differential gene analysis

+ Add analysis

+ Add genelist

+ New Project

Project_20170810

R: ProteinNetwork

Basic Information

Name: NetworkAnalysisResult

Create time: 20170210 16:57

Sig. Networks: 50

Update time: 20170210 17:09

Input data: GeneList(750)

Size on disk: 5 kb

Table View

Network View

Export picture Download

Title: NetworkAnalysisChart

Color palettes: green-red

Layout: Force-Directed

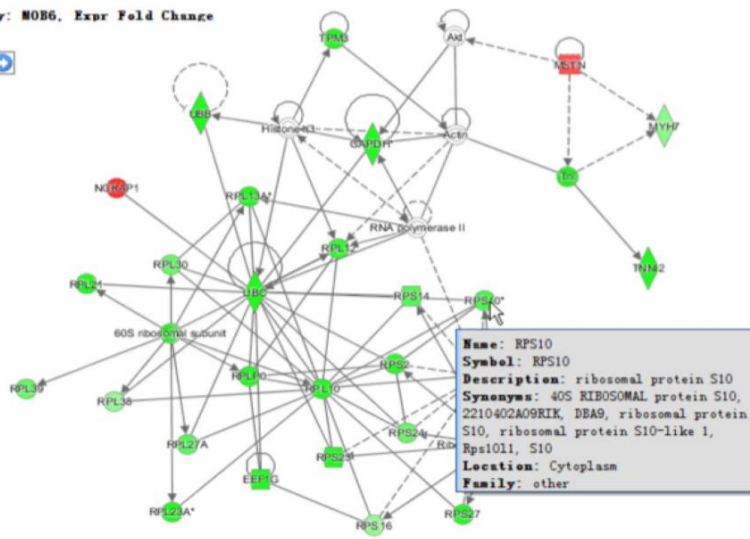
Border color: none

Update time: 2017-02-10 17:09

Default

GO

Overlay: NOB6, Expr Fold Change



国家重点研发计划“疾病研究精准医学知识库构建”项目研讨会

2017.04.14 杭州



Thank You!