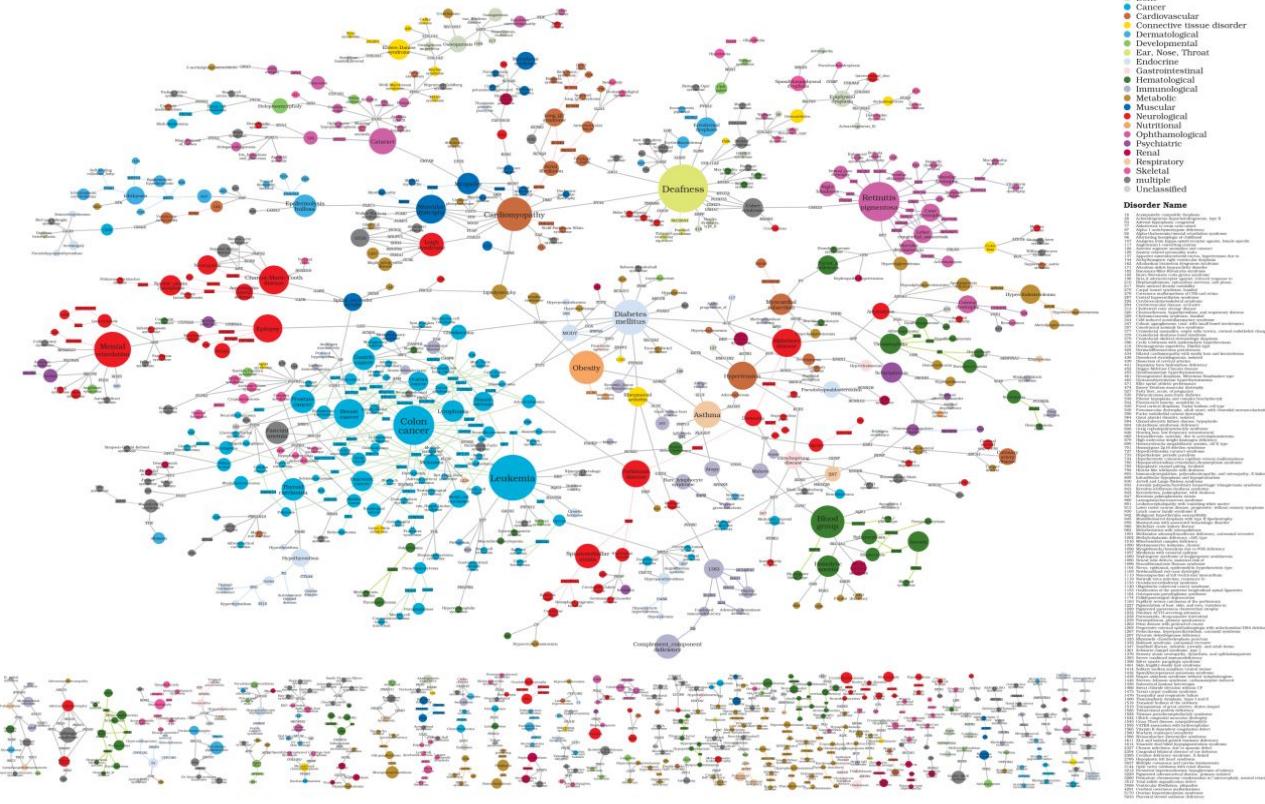


Databases for Drug Repurposing

08.09.2025 **Andreas Maier (PhD student & junior group leader at CoSyBio in Hamburg)**

Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L (2007) Proc Natl Acad Sci USA 104:8685-8690



Network medicine

Network medicine is the representation of molecular biological or medical entities like

- Diseases
- Genes
- Proteins
- Drugs
- etc.

as networks

Human disease network:
Disease-Disease network based on
shared gene associations

Relevant Databases

Sources for network-based drug repurposing

General Info & Terminology

Data sources:

- **primary**: Scientific publications, raw experimental data, ...
 - **secondary**: Build directly on knowledge extracted/aggregated from primary sources
 - **tertiary**: Aggregated information from secondary sources
- Network databases **are usually secondary or tertiary data sources**

Networks consist of:

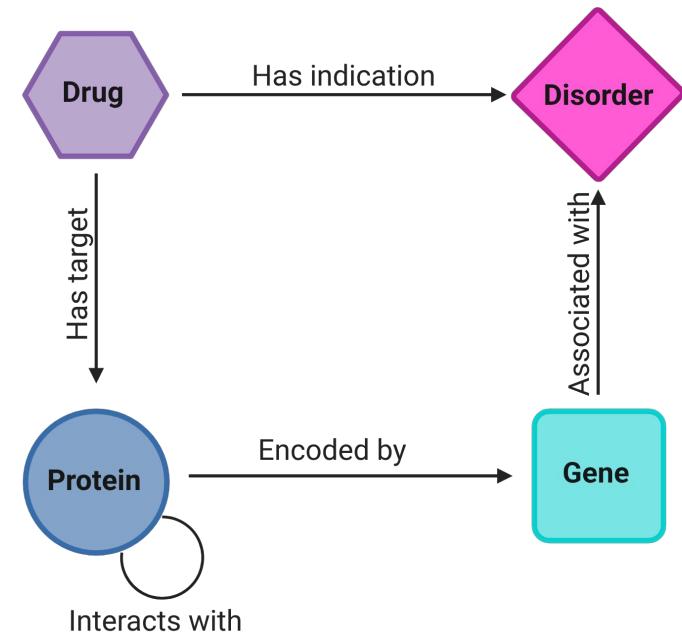
- **Nodes**: Single entities representing *drugs, genes, proteins, diseases*, ...
- **Edges**: Links, relationships, or connections between two nodes

Types of networks:

- **homogeneous**: Contain only one type of node and edge
- **heterogeneous (nonhomogeneous)**: Nodes and/or edges can be partitioned into distinct types/classes

Most relevant information layers (for DR and today)

- **Protein-Protein Interactions (PPI):** Interactome, highlighting relationships between proteins (homogeneous)
- **Disorder-gene association:** Records if there is any association between a gene and a disorder
- **Drug target interactions:** Known mechanistic targets of drugs in the human system
- **Drug-disorder indications:** Any indications about a drug being effective/in use for a disorder
- **Gene-Protein encoding:** Information about which genes encode which protein.



Do the information sources matter?

Example (non-biological):

*You want to create a network of social interactions based on online social networks.
Does it matter which social network (e.g. LinkedIn, Facebook, Instagram) you use?*

Of course! Different sources have different meanings!

Node-Level: Social network profiles: *Same person multiple personas or even accounts*

Edge-Level: Social network links:

- *What is a link in LinkedIn?* → Mix of peers and business connections
- *What is a link in Facebook?* → Mostly known/private connections
- *What is a link in Instagram?* → Mix of peers and foreign connections

→ **Source of information affects meaning nodes and edges**

Databases for Nodes

Protein - Databases and information sources

UniProt:

- **Description:** A comprehensive, high-quality resource for protein sequence and functional information.
- **URL:** <https://www.uniprot.org/>
- **Published as:** UniProt Consortium, 2025
- **Noteworthy attributes:**
 - “protein review status”



Number of Entries

 Reviewed (Swiss-Prot)
573,661 entries

 Unreviewed (TrEMBL)
253,061,696 entries

[Explore the 2025_03 release »](#)

Downloads

Reviewed (Swiss-Prot)
[fasta](#) [text](#) [xml](#)

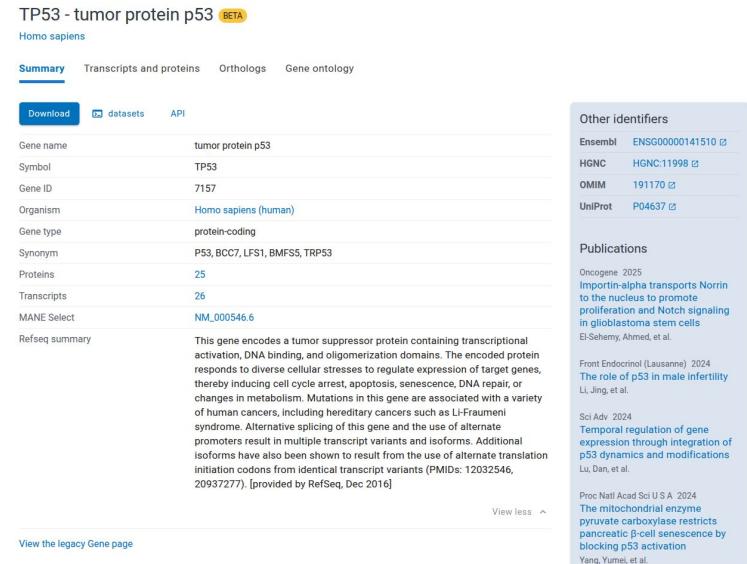
Unreviewed (TrEMBL)
[fasta](#) [text](#) [xml](#)

[Explore more in FTP](#)

Gene - Databases and information sources

NCBI:

- **Description:** Comprehensive public repository for gene-specific information, including sequence, function, and variation data.
- **URL:** <https://www.ncbi.nlm.nih.gov/gene/>
- **Published:** Sayers et al., 2025



TP53 - tumor protein p53 BETA

Homo sapiens

Summary Transcripts and proteins Orthologs Gene ontology

Download datasets API

Gene name	tumor protein p53
Symbol	TP53
Gene ID	7157
Organism	Homo sapiens (human)
Gene type	protein-coding
Synonym	TP53, BCC7, LFS1, BMFS5, TRP53
Proteins	25
Transcripts	26
MANE Select	NM_000546.6
Refseq summary	This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. Mutations in this gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome. Alternative splicing of this gene and the use of alternate promoters result in multiple transcript variants and isoforms. Additional isoforms have also been shown to result from the use of alternate translation initiation codons from identical transcript variants (PMIDs: 12032546, 20937277). [provided by RefSeq, Dec 2016]

View less ^

Other identifiers

Ensembl	ENSG00000141510
HGNC	HGNC:11998
OMIM	191170
UniProt	P04637

Publications

- Oncogene 2025 Importin-alpha transports Notch to the nucleus to promote proliferation and Notch signaling in glioblastoma stem cells El-Shermy, Ahmed, et al.
- Front Endocrinol (Lausanne) 2024 The role of p53 in male infertility Li, Jing, et al.
- Sci Adv 2024 Temporal regulation of gene expression through integration of p53 dynamics and modifications Lu, Dan, et al.
- Proc Natl Acad Sci U S A 2024 The mitochondrial enzyme pyruvate carboxylase restricts pancreatic β-cell senescence by blocking p53 activation Yang, Yumei, et al.

Drug - Databases and information sources

DrugCentral:

- **Description:** A resource collecting drug information from regulatory agencies such as the FDA, EMA, and PMDA.
- **URL:** <https://drugcentral.org/>
- **Published:** Avram et al., 2021

The screenshot shows the homepage of DrugCentral. At the top right, there are statistics: 4,995 Drugs and 152,476 pharmaceuticals. Below that is a search bar with the placeholder "Enter: Drug, Target, Disease, Uniprot ID, Veterinary Drug". Underneath the search bar are filter options: "All" (radio button selected), "FDA-approved", "EMA-approved", and "PMDA-approved". To the right of the search bar is a magnifying glass icon. Further down, it says "Target Card Uniprot Example: P23975". On the left side of the main content area, there is a chemical structure of a drug molecule: 1-(2-hydroxyethyl)-4-methylimidazolidine-2,5-dione. The background features a geometric pattern of green, blue, and orange triangles.

Featured News

[Drugcentral 2023 NAR Article](#)

[The Latest in Chemistry in Coronavirus Research](#)

Drugs in the News

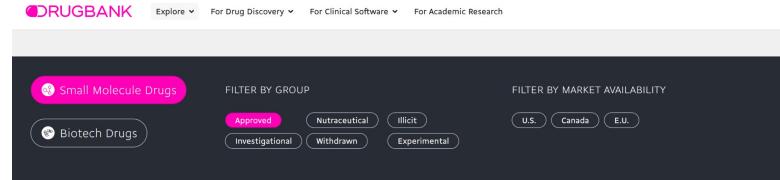
[Makena Venetoclax Dapagliflozin KEYTRUDA Sacubitril LORBRENA Hydroxychloroquine](#)

DrugCentral [Search Overview](#)

Drug - Databases and information sources

DrugBank:

- **Description:** Potentially the most comprehensive database combining detailed drug data with drug target, action, and interaction information.
- **URL:** <https://go.drugbank.com/>
- **Published:** Wishard et al., 2018
- **Noteworthy attributes:**
 - Has by now payed access model for information other than Drugs



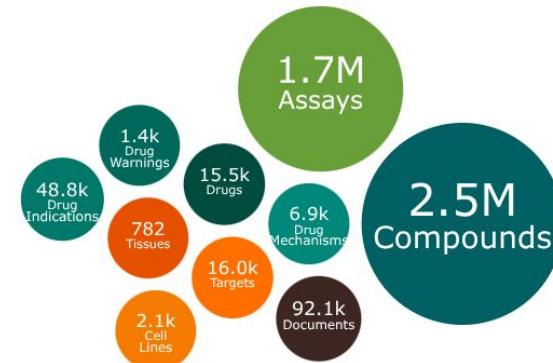
The screenshot shows the DrugBank homepage with various filters for drug categories and market availability. Below the filters, a table displays three drug entries: (-)-menthol, (S)-butane-1,3-diol, and (S)-camphor, along with their chemical structures and brief descriptions.

NAME	WEIGHT	STRUCTURE	DESCRIPTION	CATEGORIES
(-)-menthol	156.269		Not Annotated	Not Annotated
(S)-butane-1,3-diol	90.121		Not Available	Not Available
(S)-camphor	152.2334		Indicated for the temporary symptomatic relief of minor aches and pains of muscles and joints in topical analgesics.	Not Available

Drug - Databases and information sources

ChEMBL:

- **Description:** Stores bioactive drug-like small molecules extracted and curated from literature
- **URL:** <https://www.ebi.ac.uk/chembl/>
- **Published:** Zdrazil et al., 2023
- **Noteworthy attributes:**
 - Structures
 - Calculated properties (e.g. molecular weight)
 - Bioactivities (e.g. binding constants)



Drug - Databases and information sources

PubChem:

- **Description:** Chemical substances and compounds aggregated from several sources
- **URL:** <https://pubchem.ncbi.nlm.nih.gov/>
- **Published:** Kim et al., 2025
- **Noteworthy attributes:**
 - Bioactivity screens
 - Chemical properties
 - Chemical compound groups

PubChem CID: 2244

Structure: 2D, 3D, Crystal

Primary Hazards: Irritant

Molecular Formula: C₉H₈O₄
CH₃COOC₆H₄COOH

Synonyms: aspirin, ACETYL SALICYLIC ACID, 50-78-2, 2-Acetoxbenzoic acid, 2-(Acetoxyl)benzoic acid, View More...

Molecular Weight: 180.16 g/mol
Computed by PubChem 2.2 (PubChem release 2025.04.14)

Dates: Create: 2004-09-16, Modify: 2025-08-23

Description: Aspirin can cause developmental toxicity and female reproductive toxicity according to an independent committee of scientific and health experts.
► California Office of Environmental Health Hazard Assessment (OEHHA)
Aspirin can cause developmental toxicity and female reproductive toxicity according to an independent committee of scientific and health experts.

Disorder - Databases and information sources

ICD-9/10/11:

- **Description:** A WHO-coordinated classification system used globally for health statistics, epidemiology, and clinical reimbursement
- **URL:** <https://icd.who.int/en/>
- **Published:** World Health Organization, 2015

▽ ICD-11 for Mortality and Morbidity Statistics
▷ 01 Certain infectious or parasitic diseases
▷ 02 Neoplasms
▷ 03 Diseases of the blood or blood-forming organs
▷ 04 Diseases of the immune system
▷ Primary immunodeficiencies
▷ 4A20 Acquired immunodeficiencies
▷ Nonorgan specific systemic autoimmune disorders
▷ Autoinflammatory disorders
▷ 4A60 Monogenic autoinflammatory syndromes
▷ 4A61 SAPHO syndrome
▷ 4A62 Behçet disease
▷ 4A8Y Other specified autoinflammatory disorders
▷ 4A6Z Autoinflammatory disorders, unspecified
▷ Allergic or hypersensitivity conditions
▷ Immune system disorders involving white cell lineages
▷ Certain disorders involving the immune system
▷ 4B40 Diseases of thymus
▷ Organ specific autoimmune disorders
▷ Symptoms, signs or clinical findings of blood, blood-forming organs, or the immune system
▷ 4B4Y Other specified diseases of the immune system
▷ 4B4Z Diseases of the immune system, unspecified
▷ 05 Endocrine, nutritional or metabolic diseases
▷ 06 Mental, behavioural or neurodevelopmental disorders
▷ 07 Sleep-wake disorders
▷ 08 Diseases of the nervous system
▷ 09 Diseases of the visual system
▷ 10 Diseases of the ear or mastoid process
▷ 11 Diseases of the circulatory system
▷ 12 Diseases of the respiratory system
▷ 13 Diseases of the digestive system
▷ 14 Diseases of the skin
▷ 15 Diseases of the musculoskeletal system or connective tissue
▷ 16 Diseases of the genitourinary system
▷ 17 Conditions related to sexual health
▷ 18 Pregnancy, childbirth or the puerperium
▷ 19 Certain conditions originating in the perinatal period
▷ 20 Developmental anomalies
▷ 21 Symptoms, signs or clinical findings, not elsewhere classified
▷ 22 Injury, poisoning or certain other consequences of external causes
▷ 23 External causes of morbidity or mortality
▷ 24 Factors influencing health status or contact with health services

4A62 Behçet disease

Code: 4A62

Description
Behçet disease is a disease of incompletely understood aetiopathogen central nervous system inflammatory lesions. Small vessel vasculitis, t Central Asia to China and Japan.

Inclusions
Adamantidis-Behçet disease

Exclusions from above levels Show all [15] ▾

All Index Terms Show all [15] ▾

Coded Elsewhere
Transient neonatal Behçet disease (K407.Y)

Postcoordination ⓘ

Other postcoordination ⓘ (use additional code, if desired.)
search in axis: Other postcoordination

Disorder - Databases and information sources

OMIM:

- **Description:** Online Mendelian Inheritance in Man: A comprehensive, curated catalog of human genes and genetic disorders, focusing on Mendelian inheritance
- **URL:** <https://www.omim.org/>
- **Published:** Amberger et al., 2019
- **Noteworthy attributes:**
 - Claim to cover ALL human mendelian disorders

ICD+

#125853	# 125853				
Table of Contents					
Title	TYPE 2 DIABETES MELLITUS; T2D				
Phenotype-Gene Relationships					
Clinical Synopsis					
Text					
Description	DIABETES MELLITUS, NONINSULIN-DEPENDENT; NIDDM				
Inheritance	NONINSULIN-DEPENDENT DIABETES MELLITUS				
Biochemical Features	DIABETES MELLITUS, TYPE II				
Genotype/Phenotype Correlations	MATURITY-ONSET DIABETES				
Clinical Management					
Pathogenesis					
Mapping					
Molecular Genetics					
Other Features					
Animal Model					
References					
Contributors					
Creation Date					
Edit History					
Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus MIM number
2q24.1	(Type 2 diabetes mellitus, susceptibility to)	125853	AD	3	GPD2 138430
2q31.3	(Type 2 diabetes mellitus, susceptibility to)	125853	AD	3	NEUBOD1 601724
2q36.3	(Type 2 diabetes mellitus, susceptibility to)	125853	AD	3	IRS1 147545
3p25.2	Insulin resistance, severe, digenic	125853	AD	3	PPARG 601487
3p25.2	(Diabetes, type 2)	125853	AD	3	PPARG 601487
3q26.2	(Diabetes mellitus, noninsulin-dependent)	125853	AD	3	SLC2A2 138160
3q27.2	(Diabetes mellitus, noninsulin-dependent, susceptibility to)	125853	AD	3	IGFBP2 608289
4p16.1	(Diabetes mellitus, noninsulin-dependent, association with)	125853	AD	3	WFS1 606201
5q34-q35.2	(Diabetes mellitus, noninsulin-dependent)	125853	AD	2	NIDDM4 608036
6p21.31	(Type 2 diabetes mellitus, susceptibility to)	125853	AD	3	HMG-A1 600701
6q23.2	(Diabetes mellitus, non-insulin-dependent, susceptibility to)	125853	AD	3	ENPP1 173335
7p15.3	(Type 2 diabetes mellitus)	125853	AD	3	IL6 147620
7p13	Diabetes mellitus, noninsulin-dependent, late onset	125853	AD	3	GCK 138079
7q31.1	Insulin resistance, severe, digenic	125853	AD	3	PPARI3A 600917
7q32.1	Diabetes mellitus, type 2	125853	AD	3	PAX4 167413
8q24.11	(Diabetes mellitus, noninsulin-dependent, susceptibility to)	125853	AD	3	SLC30A8 611145
10q25.2	(Diabetes mellitus, type 2, susceptibility to)	125853	AD	3	TGF7L2 602228
q25.3					
11p15.1	Diabetes mellitus, noninsulin-dependent	125853	AD	3	ABCC8 600509
11p11.2	(Diabetes mellitus, noninsulin-dependent)	125853	AD	3	MAPK8IP1 604641
11q14.3	(Diabetes mellitus, type 2, susceptibility to)	125853	AD	3	MTNR1B 600804
12q4.31	(Diabetes mellitus, noninsulin-dependent, 2)	125853	AD	3	HNF1A 142410

Disorder - Databases and information sources

MONDO:

- **Description:** Monarch Disease Ontology: An ontology that harmonizes multiple disease definitions by creating a coherent, logical structure of disease relationships integrating multiple sources
- **URL:** <http://mondo.monarchinitiative.org/>
- **Published:** Mungall et al., 2017

Representation of disease types

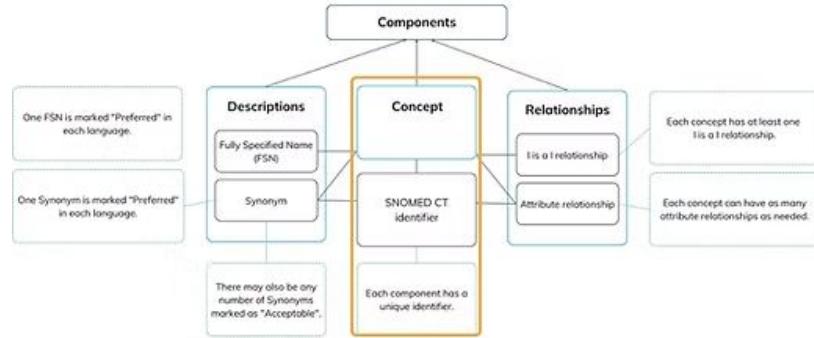
Category	Count (classes)
Total number of diseases	25814
Human diseases	22853
Cancer	4739
Infectious	1074
Mendelian	11600
Rare	15810
Non-human diseases	2960
Cancer	217
Infectious	87
Mendelian	1029

Note: susceptibilities are not included in these counts.

Disorder - Databases and information sources

SNOMED-CT:

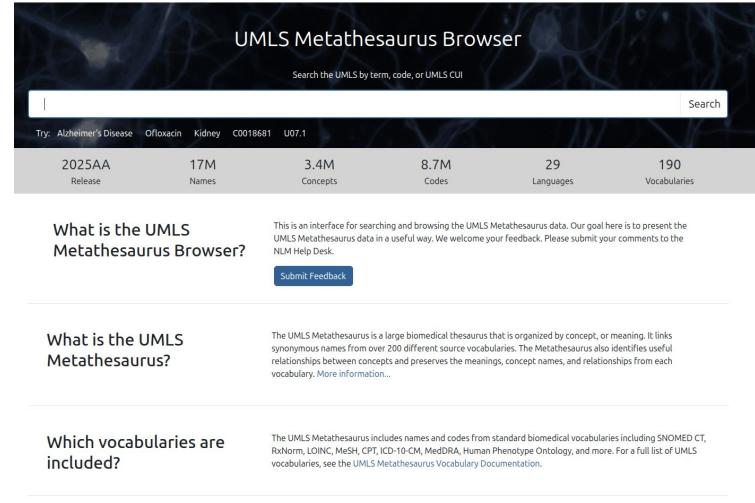
- **Description:** A comprehensive, multilingual clinical terminology designed for standardized use in electronic health records (EHRs)
- **URL:** <https://www.snomed.org/resources>
- **Published:** -
- **Noteworthy attributes:**
 - Requires license by SNOMED international



Disorder - Databases and information sources

UMLS:

- **Description:** Unified Medical Language System: A metathesaurus of clinical and biomedical terminologies designed to enable interoperability between systems
- **URL:**
<https://www.nlm.nih.gov/research/umls/index.html>
- **Published:** Bodenreider, 2004
- **Noteworthy attributes:**
 - Requires free account and license agreement



The screenshot shows the UMLS Metathesaurus Browser interface. At the top, there is a search bar with the placeholder "Search the UMLS by term, code, or UMLS CUI". Below the search bar, there is a navigation bar with links to "Alzheimer's Disease", "Ofloxacin", "Kidney", "C0018681", and "U07.1". To the right of the search bar, there are statistics: "2025AA Release", "17M Names", "3.4M Concepts", "8.7M Codes", "29 Languages", and "190 Vocabularies". Below the search bar, there is a section titled "What is the UMLS Metathesaurus Browser?" with a "Submit Feedback" button. To the right of this section, there is a detailed description of the UMLS Metathesaurus. At the bottom, there is a section titled "Which vocabularies are included?" with a detailed description of the included vocabularies.

Disorder - Databases and information sources

MESH:

- **Description:** Medical Subject Headings is a general medical terminology hierarchy that is driven by literature and used for indexing, cataloging, and searching of information
- **URL:**
<https://www.nlm.nih.gov/mesh/meshhome.html>
- **Published:** National Library of Medicine, 1960-now

Colorectal Neoplasms MeSH Descriptor Data 2025

	Details	Qualifiers	MeSH Tree Structures	Concepts
MeSH Heading	Colorectal Neoplasms			
Tree Number(s)	C04.598.274.476.411.307 C06.301.371.411.307 C06.405.249.411.307 C06.405.469.158.356 C06.405.469.491.307 C06.405.469.860.180			
Unique ID	D015179			
RDF Unique Identifier	http://nlm.nih.gov/mesh/D015179			
Annotation	COLONIC NEOPLASMS and RECTAL NEOPLASMS are also available; coordinate with histological type of neoplasm if given; GENES, DCC (for "Deleted in Colorectal Cancer") & GENES, MCC (for "Mutated in Colorectal Cancer") are available			
Scope Note	Tumors or cancer of the COLON or the RECTUM or both. Risk factors for colorectal cancer include chronic ULCERATIVE COLITIS; FAMILIAL POLYPOSIS COLI; exposure to ASBESTOS; and irradiation of the CERVIX UTERI.			
Entry Version	COLONRECTAL NEOPL			
Entry Term(s)	Colorectal Cancer Colorectal Carcinoma Colorectal Tumors Neoplasms, Colorectal			
NLM Classification #	W1 529			
Previous Indexing	Colonic Neoplasms (1966-1988) Rectal Neoplasms (1966-1988)			
See Also	Genes, DCC Genes, MCC			
Public MeSH Note	89			
History				
Date Established	1988/01/01			
Date of Entry	1988/04/14			
Revision Date	2020/03/03			

page delivered in 0.139s

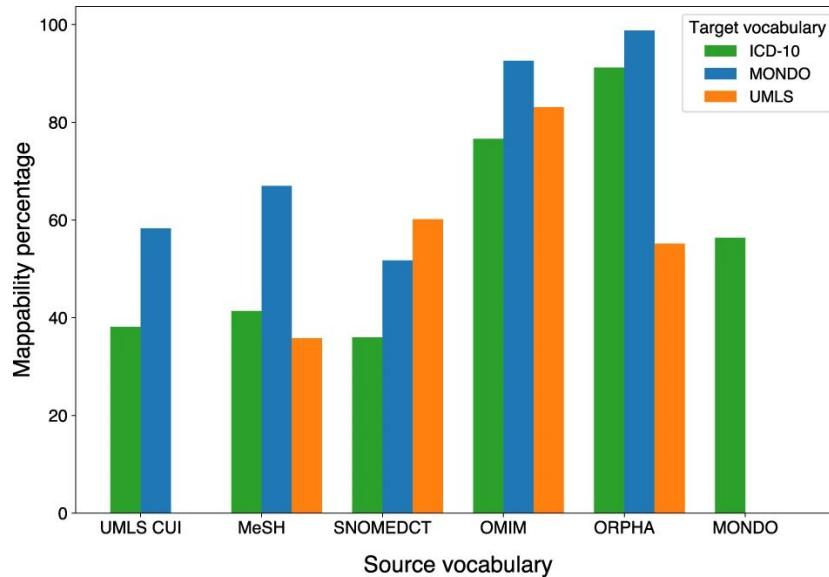
Disorder - Databases and information sources

Name	ICD	MONDO	OMIM	SNOMED CT	UMLS	MESH
Sources	Expert-driven	Integration of various databases	Expert-curated literature-based	Expert-driven	Combination of clinical and biomedical terminologies	Driven by use in literature
Goal	Health statistics, epidemiology, and clinical reimbursement	Creation of a harmonized disorder identification system	Curated catalog of human genetic disorders	Multilingual clinical terminology for standardized use in EHRs	Metathesaurus designed to enable interoperability between systems	Indexing, cataloging, and searching of information

Potential issue: Big differences in definitions of diseases!

Disorder - Databases and information sources

Differences in definitions → differences in interoperability



Sadegh, S., Skelton, J., Anastasi, E. et al. Lacking mechanistic disease definitions and corresponding association data hamper progress in network medicine and beyond. *Nat Commun* 14, 1662 (2023). <https://doi.org/10.1038/s41467-023-37349-4>

Databases for Edges

List of potential evidence/source types of links

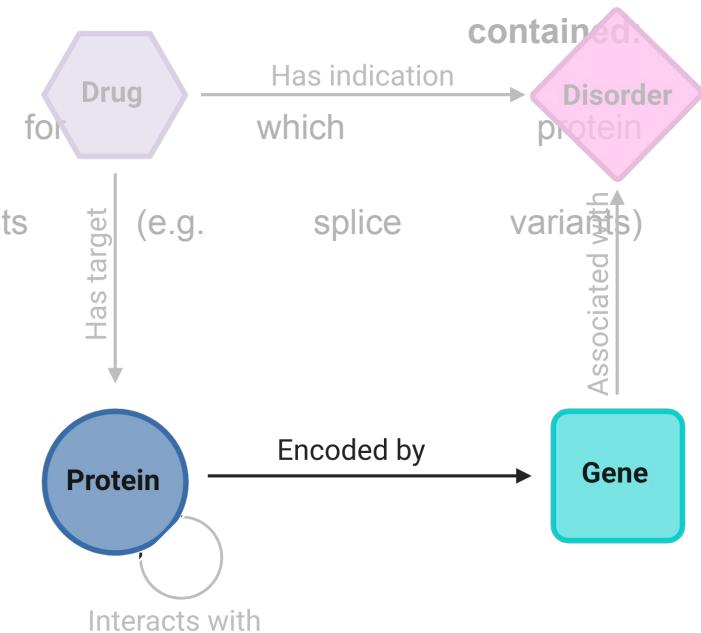
- **Literature-driven:**
 - **Literature Curation:** Information manually extracted from scientific literature by curators
 - **Text-mining:** Automated information extraction from the text of scientific publications using natural language processing techniques
- **Data-driven:**
 - **Experimental:** Information derived from laboratory experiments (e.g., yeast two-hybrid, GWAS, ...)
 - **Sequence Data:** Information derived from genomic or proteomic sequencing efforts
- **Computational:**
 - **Prediction:** Information/Properties generated by computational algorithms, predicting relationships *de novo* (e.g., based on protein structure, domain interactions, or sequence homology)
 - **Inference:** Information derived through logical deduction from other data types, such as inferring interactions via orthology mapping between species
- **Aggregation:** Data collected and compiled from multiple/other secondary databases
 - **Regulatory:** Information from official regulatory bodies (e.g., from the FDA, EMA)

→Origin of edge information can be even very diverse

Gene-Protein - Databases and information sources

- **Potential**

- Which gene encodes information
- Which protein variants do exists (e.g. splice variants)
- Variant prevalence



Gene-Protein - Databases and information sources

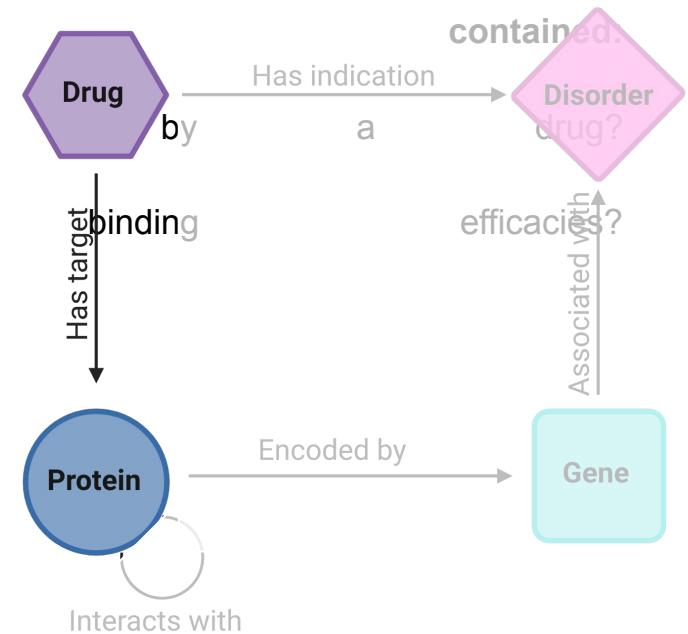
Name	UniProt-KB
URL	www.uniprot.org
Description	Comprehensive resource for protein sequence and annotation data
Primary Evidence Type	Sequence Data
Published	UniProt Consortium, 2025

Drug-Target - Databases and information sources

- **Potential**

- Which proteins are targeted
- What are the
- What is the mode of action?

information

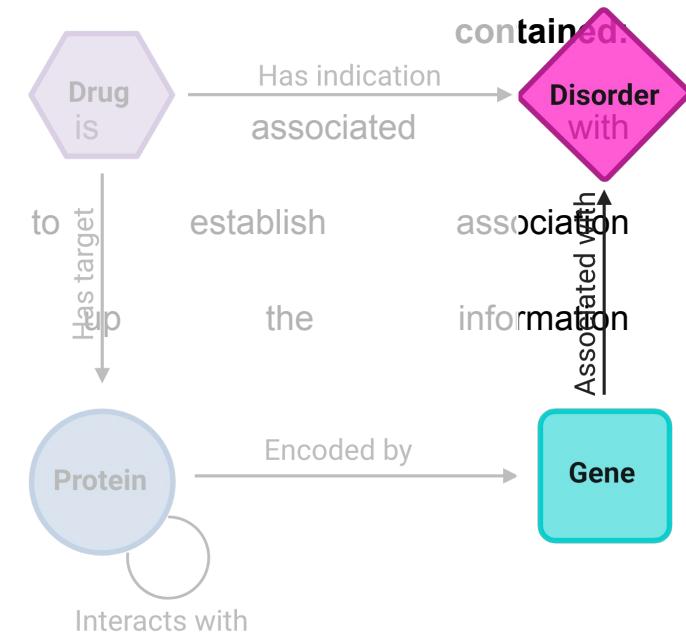


Drug-Target - Databases and information sources

Name	ChEMBL	DGIdb	DrugBank	DrugCentral	BindingDB	STITCH
URL	www.ebi.ac.uk/chembl	dgidb.org	go.drugbank.com	drugcentral.org	www.bindingdb.org/rwd/bind/index.jsp	stitch.embl.de (offline)
Description	Stores bioactive drug-like small molecules extracted and curated from literature containing target information	The Drug-Gene Interaction Database aggregates drug-gene interactions from multiple curated sources	A comprehensive database combining detailed drug data with drug target, action, and interaction information	A resource for drug information that also includes data on veterinary applications	Knowledgebase of measured binding affinities, focusing proteins considered as drug-targets with ligands that are small, drug-like molecules	Stores interactions of chemicals and proteins of several sources.
Primary Evidence Type	Literature Curation, Aggregation, Prediction	Literature Curation, Aggregation	Experimental, Literature Curation, Regulatory	Aggregation, Regulatory, Inference	Experimental, Aggregation, Text-mining	Aggregation, Experimental, Text-mining
Published	Zdrazil et al., 2023	Cannon et al., 2024	Wishart et al., 2018	Avram et al., 2021	Liu et al., 2024	Szklarczyk et al., 2016

Disease-Gene - Databases and information sources

- **Potential information**
 - Which genes a disorder
 - Which methods were used
 - Publications that can back
 - Association scores

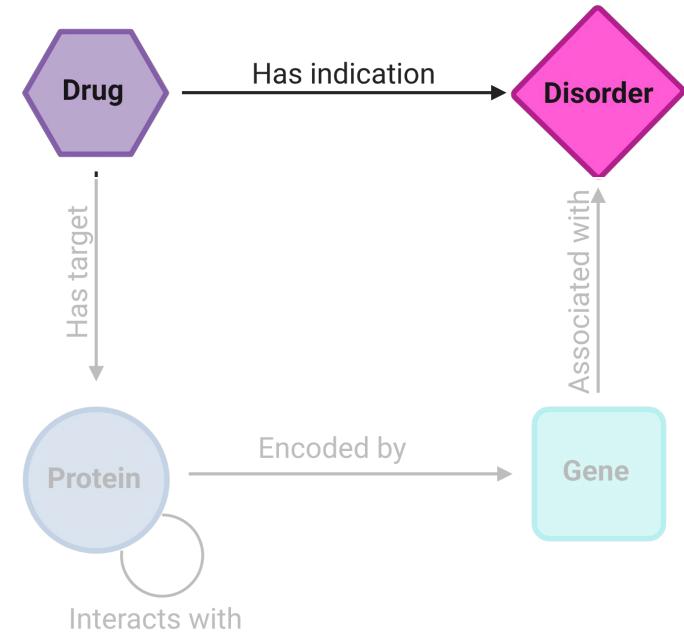


Disease-Gene - Databases and information sources

Name	OMIM	DisGeNet	CTD	DISEASES
URL	www.omim.org	disgenet.com	ctdbase.org	diseases.jensenlab.org
Description	Comprehensive, curated catalog of human genes and genetic disorders, focusing on Mendelian inheritance	Comprehensive discovery platform for exploring the genetic basis of human diseases and their associated variants	Stores data on chemical-gene, chemical-disease, and gene-disease relationships, using MeSH and OMIM identifiers	A daily-updating resource for disease-gene associations, with confidence scores assigned to each relationship
Primary Evidence Type	Literature Curation	Aggregation, Literature Curation, Experimental, Inference	Literature Curation, Aggregation, Inference	Text-mining, Literature Curation, Experimental
Published	Amberger et al., 2019	Pinero et al., 2015	Davis et al., 2021	Grissa et al., 2022

Drug-Disease - Databases and information sources

- Potential information contained:
 - Information about which drugs are approved for which disorder
 - Off-label use information
 - Information about investigated drug disease indications

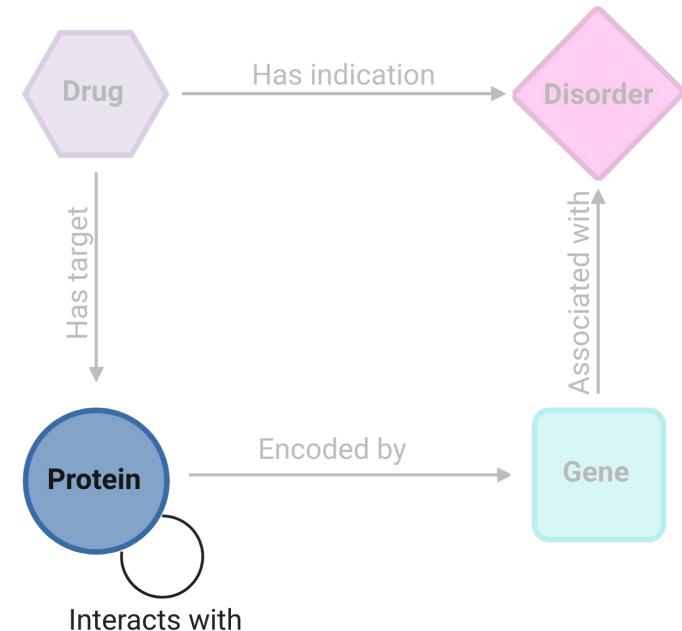


Drug-Disease - Databases and information sources

Name	CTD	DrugBank	DrugCentral	TreatKB	ClinicalTrials
URL	ctdbase.org	go.drugbank.com	drugcentral.org	nlp.case.edu/public/ada/treatKB	clinicaltrials.gov
Description	Stores data on chemical-gene, chemical-disease, and gene-disease relationships, using MeSH and OMIM identifiers	A comprehensive database combining detailed drug data with drug target, action, and interaction information	A resource for drug information that also includes data on veterinary applications	Comprehensive treatment knowledge based using complementary data sources	Clinical Trial registry containing Disease-Drug combinations under investigation
Primary Evidence Type	Literature Curation, Aggregation, Inference	Experimental, Literature Curation, Regulatory	Aggregation, Regulatory, Inference	Text-Mining, Aggregation	Regulatory
Published	Davis et al., 2021	Wishart et al., 2018	Avram et al., 2021	Wang et al., 2018	-

Protein-Protein Interaction - Databases and information sources

- **Potential information contained:**
 - Temporary physical interactions between two proteins
 - Proteins forming complexes together
 - The source of the interaction information (experimental, predicted, text-mined)
 - Methods used to determine the interaction (e.g. yeast two-hybrid)





Protein-Protein Interaction - Databases and information sources

Name	STRING	BioGRID	IntAct	IID	APID
URL	string-db.org	thebiogrid.org	www.ebi.ac.uk/intact	iid.ophid.utoronto.ca	apid.dep.usal.es
Description	Database of known and predicted protein-protein interactions for thousands of organisms	Repository of protein and genetic interactions for all major model organisms, offering datasets with varying levels of validation	Open-source database and analysis toolkit for molecular interaction data partially driven by user-based data submission	Integrated Interactions Database contains human and model organism interactions	Provides PPI networks for over 400 species, focusing on experimentally validated physical interactions
Primary Evidence Type	Aggregation, Experimental, Text-mining, Prediction	Aggregation, Literature Curation	Aggregation, Experimental, Literature Curation, Inference	Aggregation, Experimental, Prediction	Aggregation, Experimental
Published	Szklarczyk et al., 2025	Stark et al., 2006	Orchard et al., 2014	Kotlyar et al., 2022	Alonso-López et al., 2019

Knowledge graphs for DR

A network-based drug repurposing platform

Integrated Network Databases for Drug Repurposing

Requirements for heterogeneous knowledge graphs specifically for mechanism-driven drug repurposing:

- Connecting **drug and disease** information through genetics
- **Integration** of tens to hundreds of resources
- Provide **source/evidence information** about integrated data
- Multiple **options to interact** with stored information (e.g. Web-UI, API, ...)

Well known/relevant examples:

- OpenTargets
- OmniPath
- Het.io
- NeDRex

Integrated Network Databases for Drug Repurposing

OpenTargets:

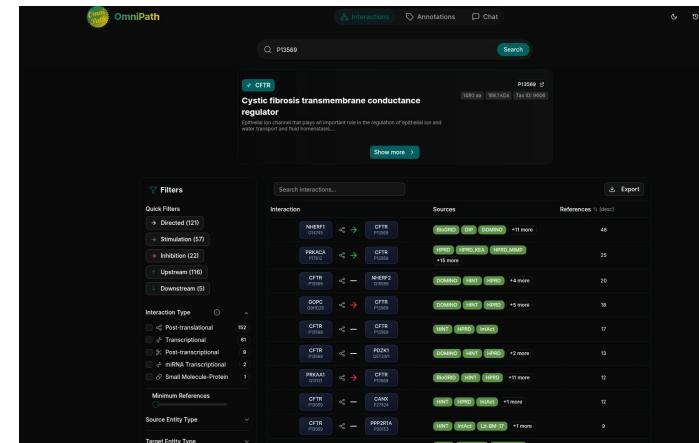
- **Description:** Platform integrating over 40 data sources, backed by EMBL-EBI, Pfizer and other big names
- **Purpose:** Identification and validation of novel therapeutic targets
- **Interaction options:** Web-UI, GraphQL API
- **URL:** platform.opentargets.org
- **Published:** Buniello et al., 2025



Integrated Network Databases for Drug Repurposing

OmniPath:

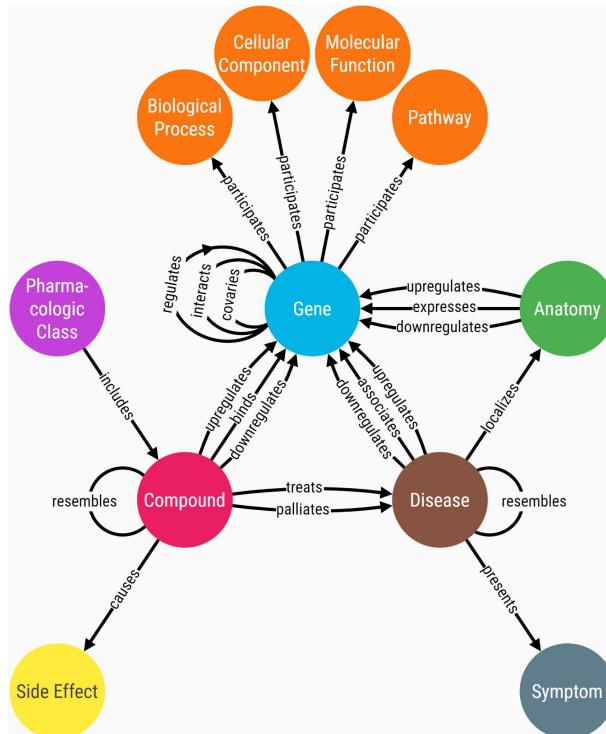
- **Description:** Extensive resource integrating over 100 sources, monthly updated
- **Purpose:** Comprehensive collection of protein interactions, gene regulatory networks, and drug-target interactions for deep mechanistic investigation
- **Interaction options:** Web-UI, Python & R-lang libraries, Cytoscape App
- **URL:** omnipathdb.org
- **Published:** Türei et al, 2021



Integrated Network Databases for Drug Repurposing

Het.io:

- **Description:** Integrates high-quality information of 29 databases, covering 11 node and 24 edge types
 - **Purpose:** Purpose-built network for large-scale systematic evaluation of drug-disease pairs
 - **Interaction Options:** Web-UI, Neo4j graph browser, multiple Python libraries
 - **URL:** <https://het.io/>
 - **Published:** Himmelstein et al., 2017



Integrated Network Databases for Drug Repurposing

NeDRexKG:

- **Description:** Heterogeneous KG, integrating 22 (27) databases, 11 node and 20 edge types
- **Purpose:** Serves as knowledge resource for the NeDRex platform and was built for drug repurposing research
- **Interaction Options:** REST API, Neo4j graph query, Extensive Web-UI, Cytoscape App, Python package
- **URL:** <https://nedrex.net>
- **Published:** Sadegh et al., 2021



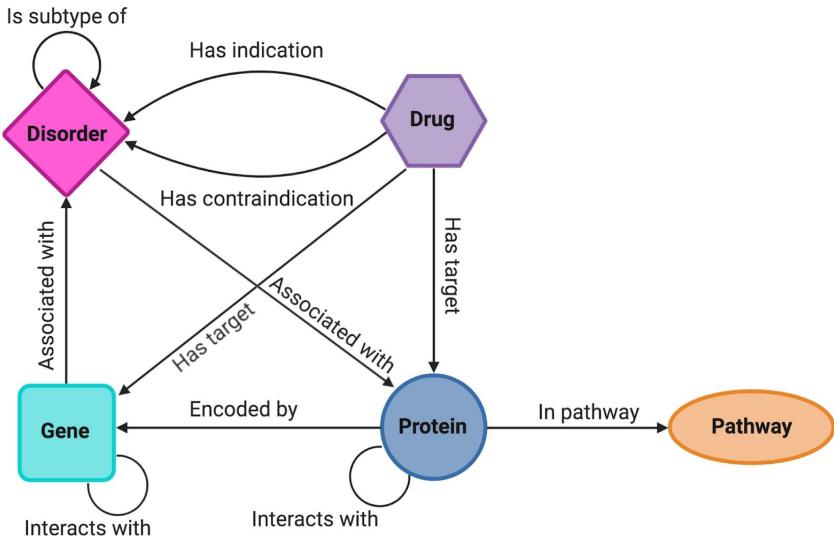
NeDReX

A network-based drug repurposing platform

NeDReXDB metagraph



<https://api.nedrex.net>



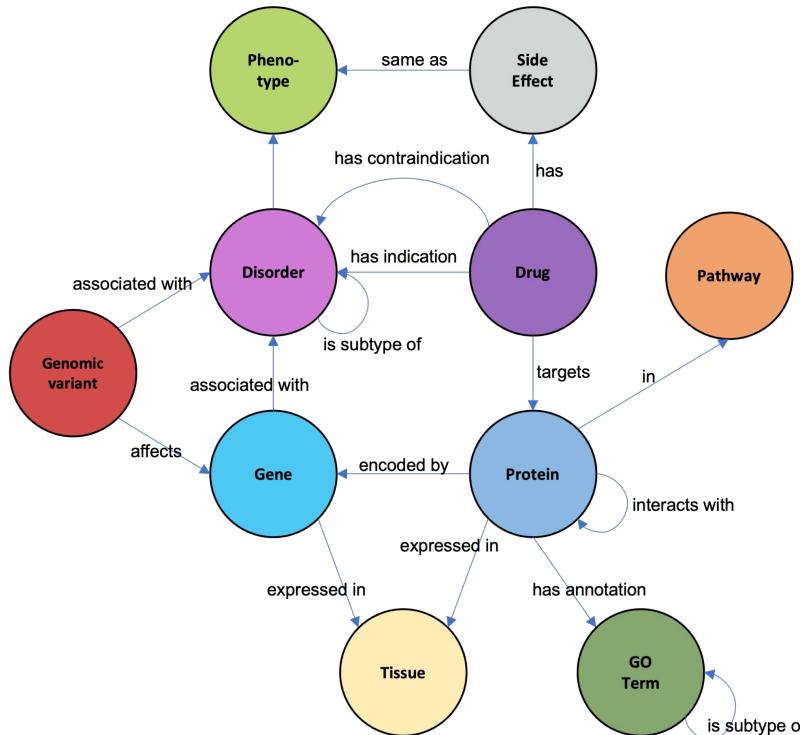
Integrated network:

- 10 databases
- 350K nodes
- 4.4M edges

Edge	Entities	Sources
Disorder - [is subtype of] - Disorder	37,413	MONDO
Drug - [has contraindication] - Disorder	13,760	DrugCentral
Drug - [has indication] - Disorder	16,662	DrugCentral
Drug - [has target] - Gene	26,336	<i>inferred</i>
Drug - [has target] - Protein	26,537	DrugCentral, DrugBank
Gene - [associated with] - Disorder	30,170	DisGeNET, OMIM
Gene - [interacts with] - Gene	1,382,888	<i>inferred</i>
Protein - [associated with] - Disorder	62,305	<i>inferred</i>
Protein - [encoded by] - Gene	32,847	UniProt
Protein - [in] - Pathway	119,715	Reactome
Protein - [interacts with] - Protein	2,662,724	BioGRID, IID, IntAct
Total	4,411,257	-

NeDReX_V2 metagraph

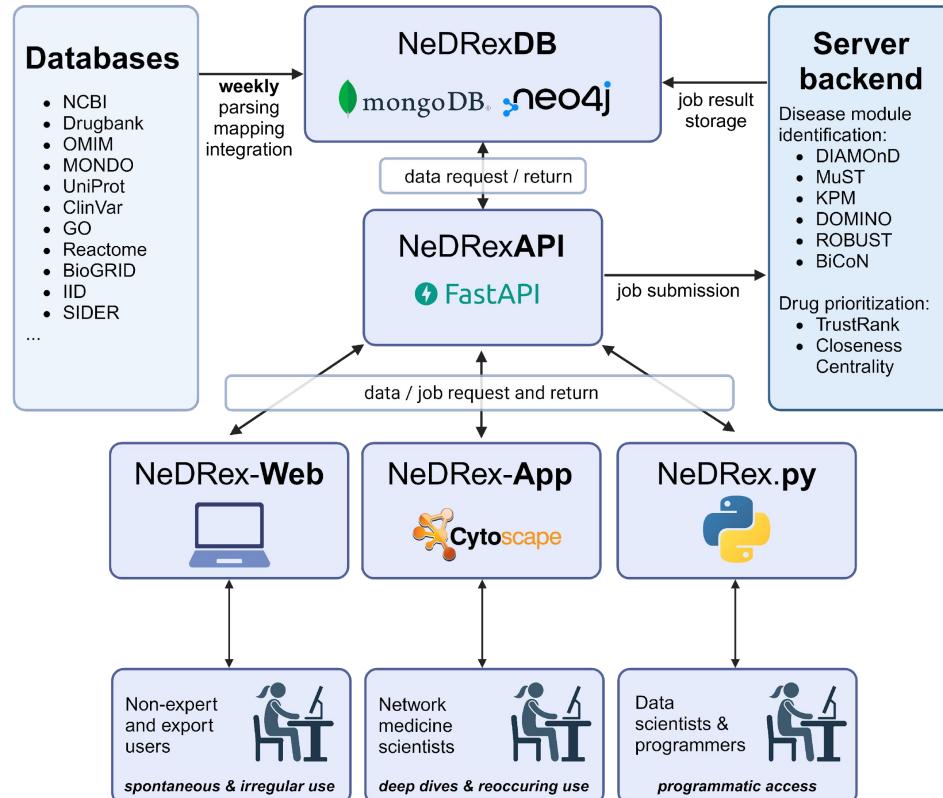
 <https://api.nedrex.net>



Integrated network:

- 19 databases

Edge	Sources
Disorder - [has] - Phenotype	Human Phenotype Ontology
Drug - [has] - Side effect	SIDER
Gene - [expressed in] - Tissue	Human Protein Atlas
Protein - [expressed in] - Tissue	Human Protein Atlas
GO - [is subtype of] - GO	Gene Ontology
Protein - [has] - GO (annotation)	Gene Ontology
Side effect - [matches] - Phenotype	Bioontology.org
Variant - [affects] - Gene	ClinVar
Variant - [associated with] - Disorder	ClinVar



- **NeDrex history:**
 - Developed as part of REPO-TRIAL
 - Continued as part of REPO4EU

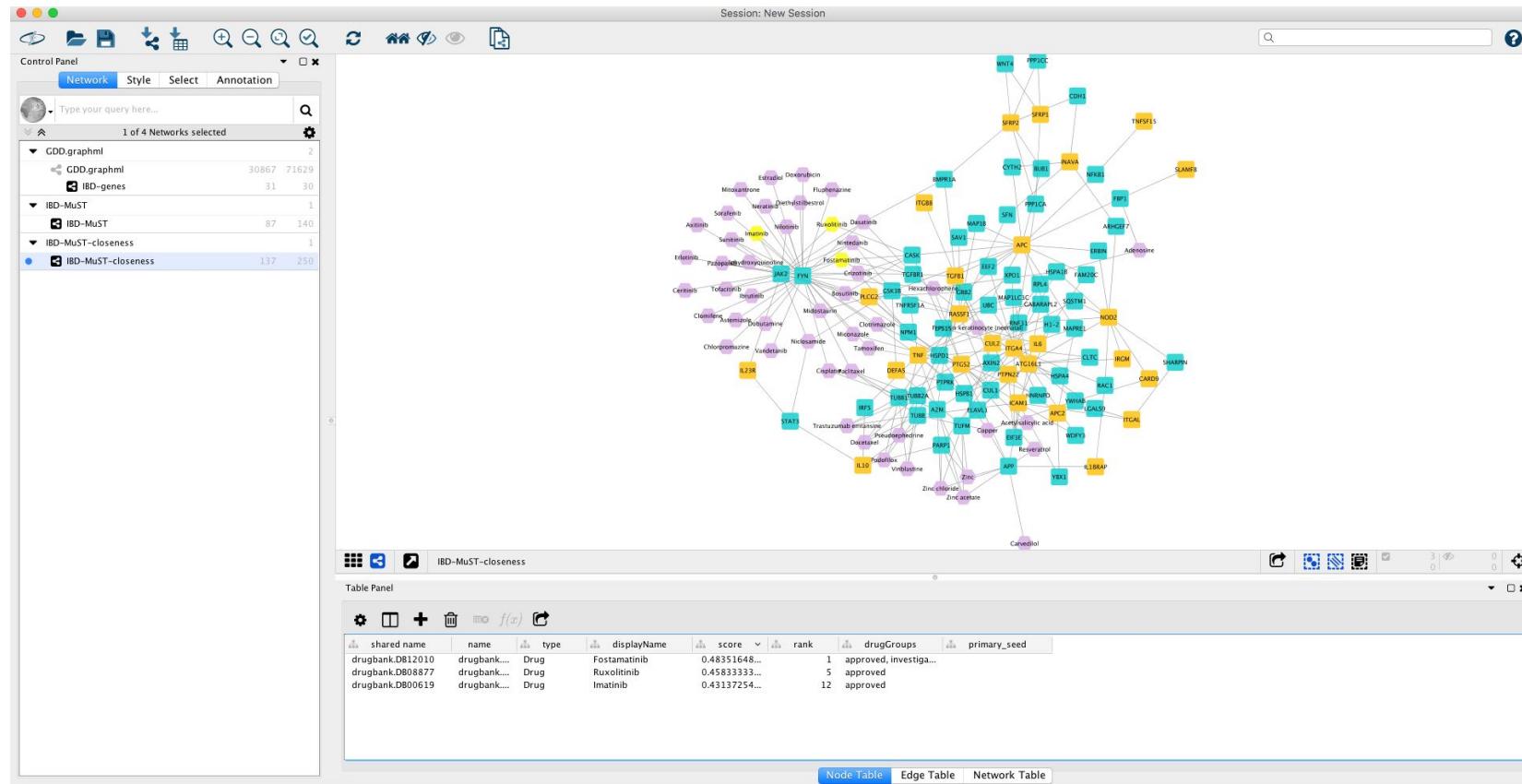


- **Available interfaces:**
 - NeDrex-Python for programmers and data scientists
 - NeDrex-App for network medicine experts and frequent use
 - NeDrex-Web for easy but extensive online exploration options

<https://nedrex.net>

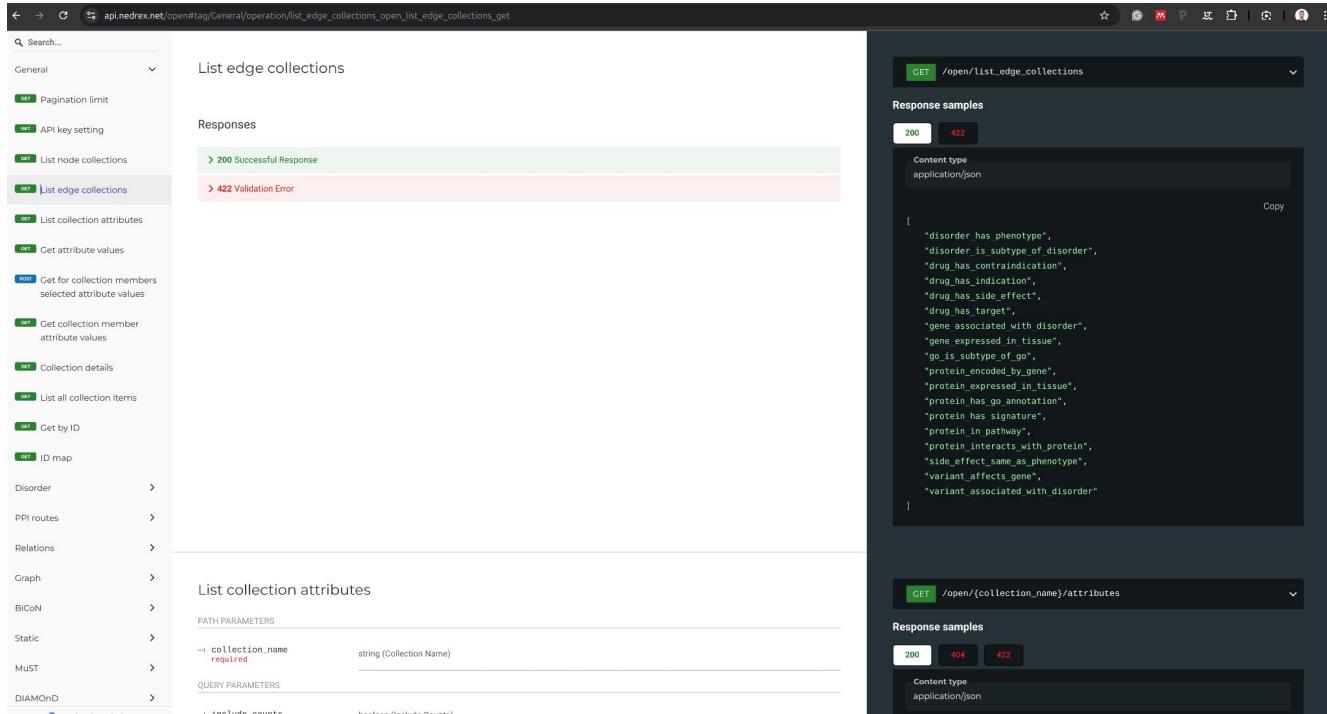
<https://web.nedrex.net>

NeDReXApp (Cytoscape)



Available:

- api.nedrex.net
- dev.api.nedrex.net



The screenshot displays the NeDReXAPI documentation interface, specifically the OpenAPI specification. It shows two main sections: `/open/list_edge_collections` and `/open/{collection_name}/attributes`.

Section 1: /open/list_edge_collections

- Responses:**
 - 200 Successful Response (Content type: application/json)
 - 422 Validation Error
- Response samples:**

```
[ "disorder_has_phenotype", "disorder_is_subtype_of_disorder", "drug_has_contraindication", "drug_has_indication", "drug_has_side_effect", "drug_has_target", "gene_associated_with_disorder", "gene_expressed_in_tissue", "go_is_subtype_of_go", "protein_encoded_by_gene", "protein_expressed_in_tissue", "protein_has_annotation", "protein_has_signature", "protein_in_pathway", "protein_interacts_with_protein", "side_effect_same_as_phenotype", "variant_affects_gene", "variant_associated_with_disorder" ]
```

Section 2: /open/{collection_name}/attributes

- PATH PARAMETERS:**
 - `collection_name` (required, string, Collection Name)
- QUERY PARAMETERS:**
 - `include_attributes`
 - `include_descriptions`
- Responses:**
 - 200
 - 404
 - 422
- Content type:** application/json

NeDRex Python package



The screenshot shows the landing page for the NeDRex Python package. At the top left is a white 3D cube icon composed of smaller cubes, with a yellow square on its front face. To its right is a search bar with the placeholder "Search projects" and a magnifying glass icon. Below this is a large title "nedrex 0.1.4". Underneath the title is a button with the text "pip install nedrex" and a small icon of a clipboard with a document.

A Python library for interfacing with the NeDRex API

Acknowledgments

Researchers

Lisa Spindler	Prof. Dr. David B. Blumenthal
Adrian Hözlwimmer	Prof. Dr. Jan Baumbach
Johannes Kersting	Prof. Dr. Anil Wipat
Dr. Sepideh Sadegh	Prof. Dr. Tim Kacprowski
Dr. James Skelton	Prof. Dr. Markus List
Dr. Elisa Anastasi	
Dr. Gihanna Galindez	

...and many more...

Contact: andi@cosy.bio

Funding



REPO-TRIAL: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777111. This publication reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains.

RePo4EU: This project is funded by the European Union under grant agreement No. 101057619. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held

This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1908A and grant 01ZX1910D) (J.B.). J.B. was partially funded by his VILLUM Young Investigator Grant No. 13154.

Hands-On

Hands-on Session

Focuses on:

- Requesting and handling [network data](#) from [knowledge graphs](#), on the example of [NeDRexDB](#)
- You can **choose** what to start with:
 - [NeDRexAPI](#): Learn how to navigate the NeDRexAPI documentation and generate networks with it
→ [./Session1_databases/NeDRex_api](#)
→ Contact person: [Lisa](#)
 - [NeDRex-Python package](#): Learn how to use the Python API-wrapper and how to construct a heterogeneous network layer-by-layer
→ [./Session1_databases/NeDRex_package](#)
→ Contact person: [Andi \(andi@cosy.bio\)](mailto:andi@cosy.bio)

Instructions are in the Session1_databases folder:
https://github.com/REPO4EU/network_medicine_drug_repurposing_tutorial/tree/main/Session1_databases



If/When you are fully done with one, of course feel free to also check out the other option!