TargetDB: A target information aggregation tool and tractability predictor

Alzheimer's Research UK
The Power to Defeat Dementia

UNIVERSITY OF OXFORD

DRUG DISCOVERY INSTITUTE

Stephane De Cesco
ARUK Oxford Drug Discovery Institute
Post-doctoral researcher, Computational chemist

Features: Overall picture

Targets identification

- Genome wide association studies
- RNAi Screens
- Crispr/Cas9 Screens
- Proteomics
- Knock-out phenotypes

→ Increased number of proposed targets

Targets evaluation

DRUGGABILITY + VALIDITY + TOXICITY = TRACTABILITY

- Binding domains ?
- Crystal structure ?
- Assayability ?
- Small molecules ?

- Confidence in the data
- In vivo models
- k.o. phenotypes
- Target validation experiments ?

- Phenotype data
- Tissue expression

- Can be used to make strategic decisions
- Prioritise/rank lists of targets

1. Gashaw, I., Ellinghaus, P., Sommer, A. & Asadullah, K. What makes a good drug target? *Drug Discov. Today* **17,** (2011).
2. Brown, K. K. et al. Approaches to target tractability assessment – a practical perspective. Medchemcomm 597–748 (2018). doi:10.1039/C7MD00633K

Tools available

Alzheimer's Research UK | University of Oxford
DRUG DISCOVERY INSTITUTE

Aggregated data

UniProt

Open Targets

PHAROS

Biology

THE HUMAN PROTEIN ATLAS

OMIM

InterMine

KEGG

reactome

Chemistry / Structural Biology

BindingDB

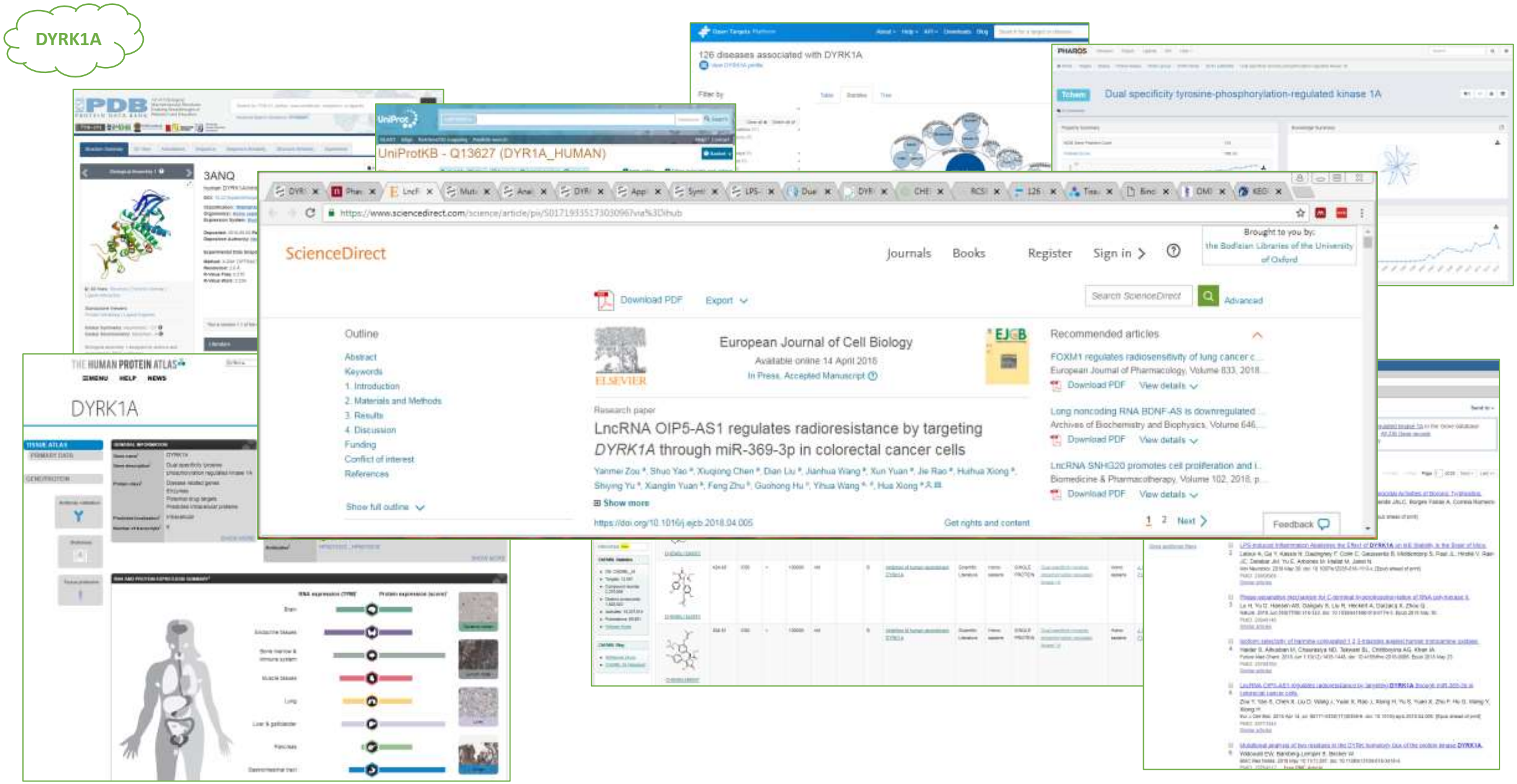
ChEMBL

RCSB PDB PROTEIN DATA BANK

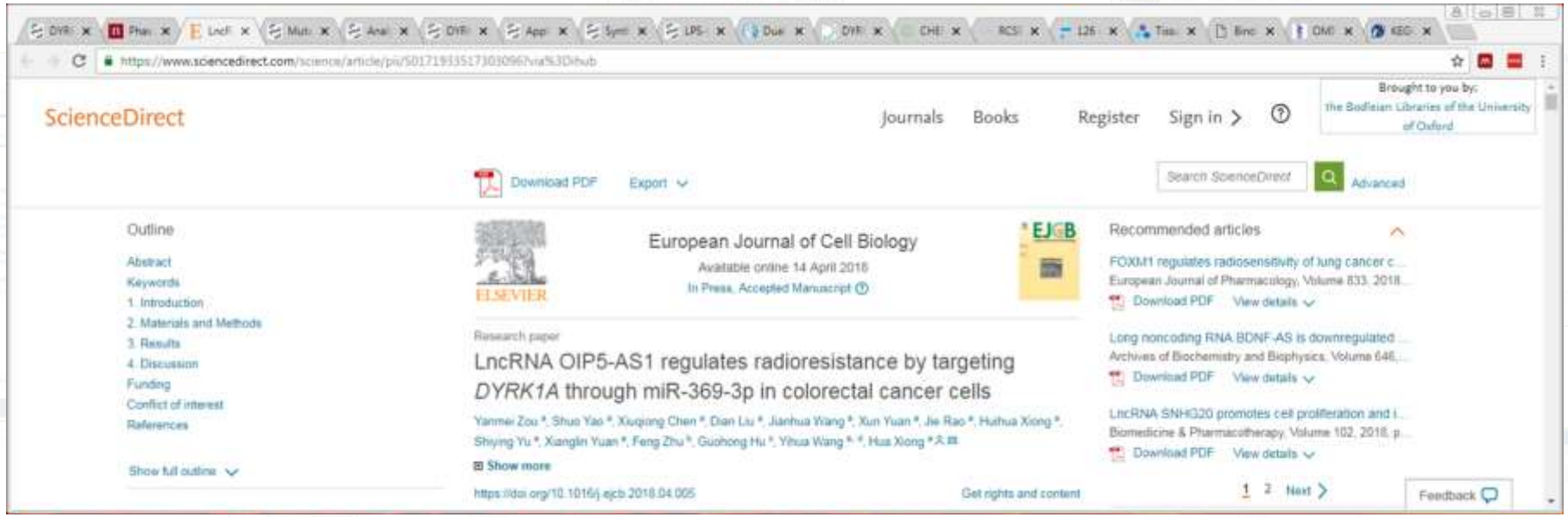
Literature

PubMed

Typical workflow



Typical workflow



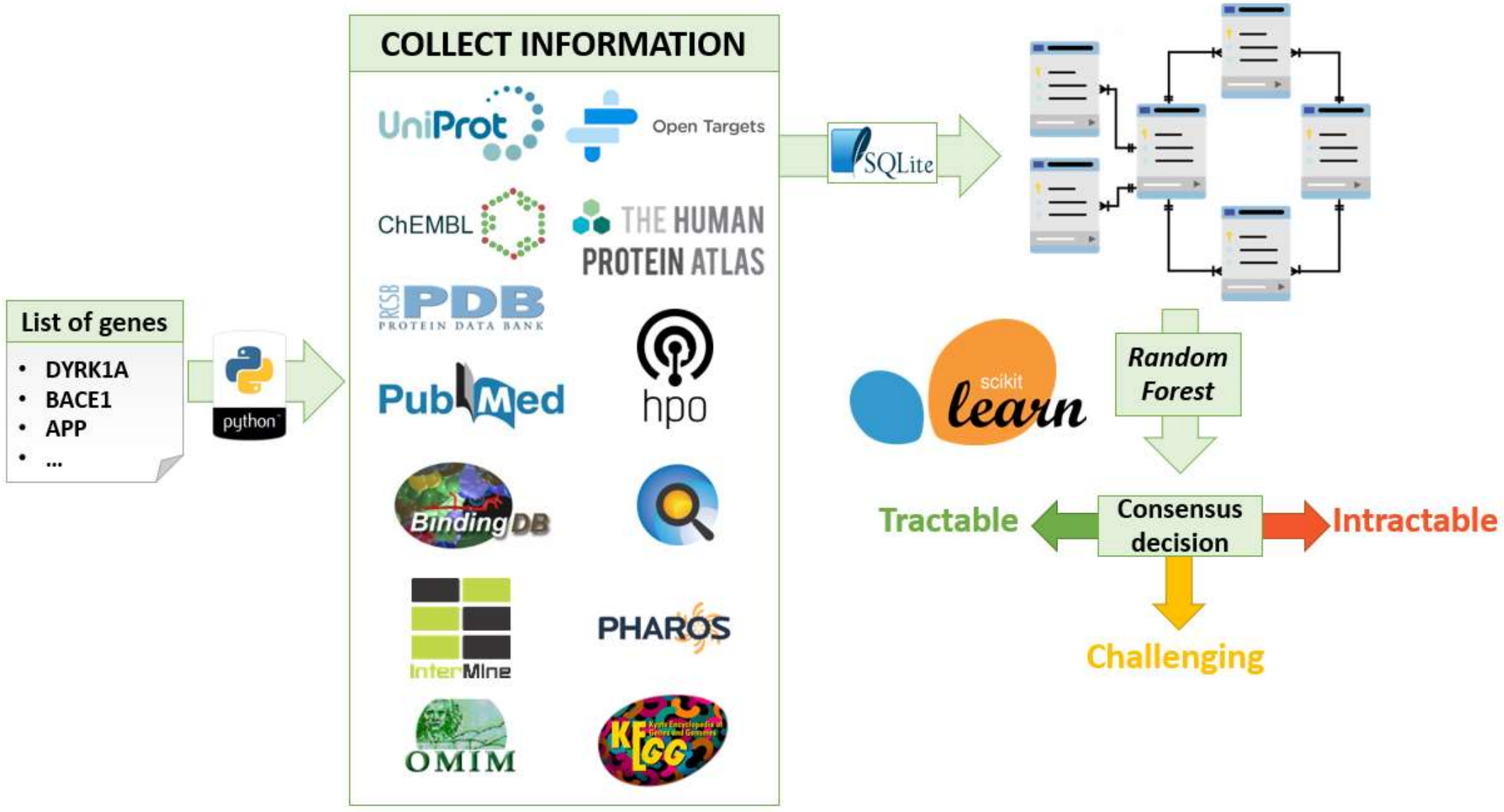
ADVANTAGES

+ Exhaustive search
+ Deep understanding of a target
+ Highly curated data

DRAWBACKS

- Person dependent
- Scattered data
- Difficult to scale up for many targets
- Comparison of targets difficult

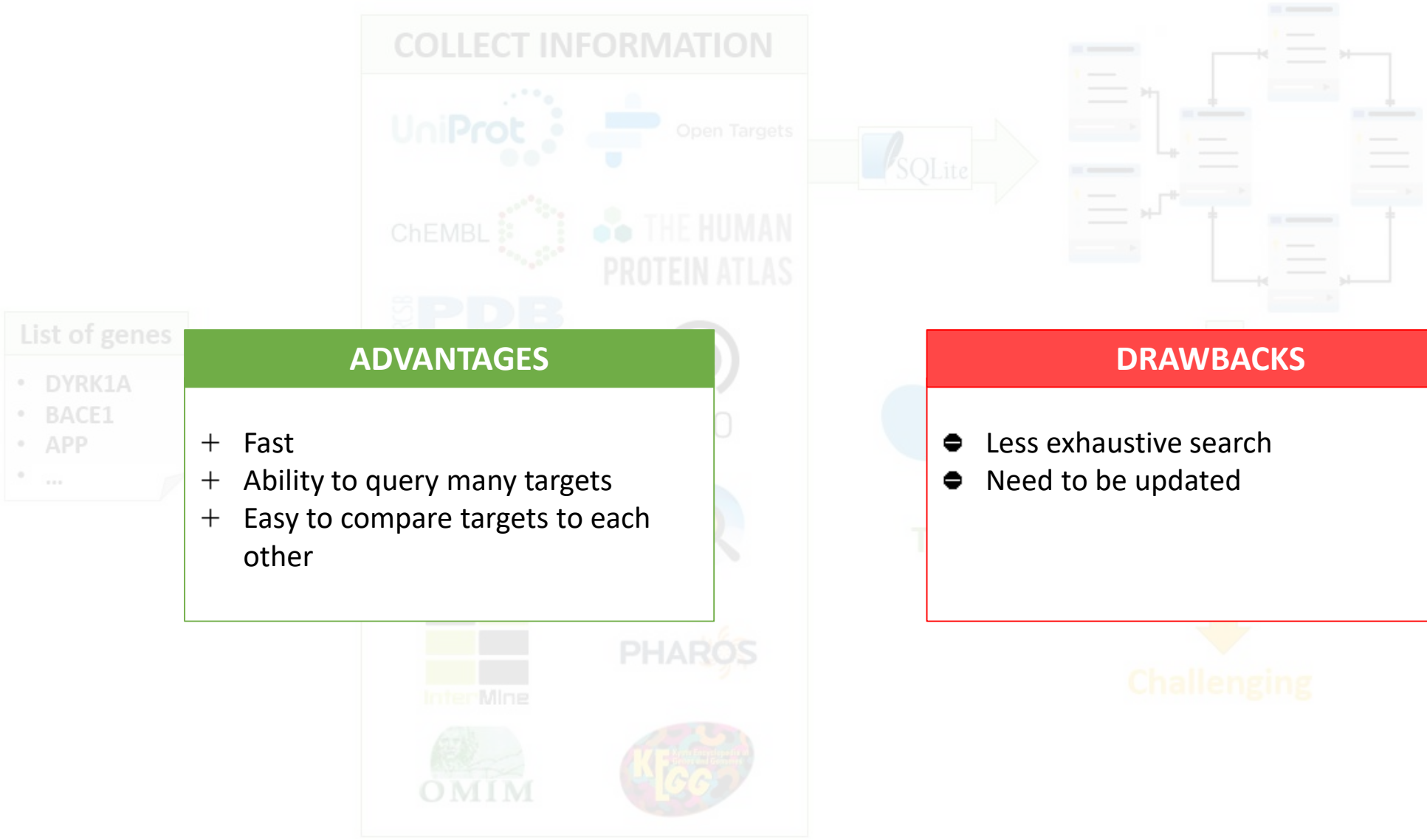
TargetDB: Main workflow



All code is open source and available on GitHub

https://github.com/sdecesco/targetDB

TargetDB: Main workflow



ADVANTAGES

+ Fast
+ Ability to query many targets
+ Easy to compare targets to each other

DRAWBACKS

- Less exhaustive search
- Need to be updated

All code is open source and available on GitHub

https://github.com/sdecesco/targetDB

Step 1: Installing a python distribution

- TargetDB is a python package and needs a python distribution to work.
- It also relies on several other packages to function properly.
- If you already have a python distribution installed you can skip these steps.

requirements

Python version >= 3.4

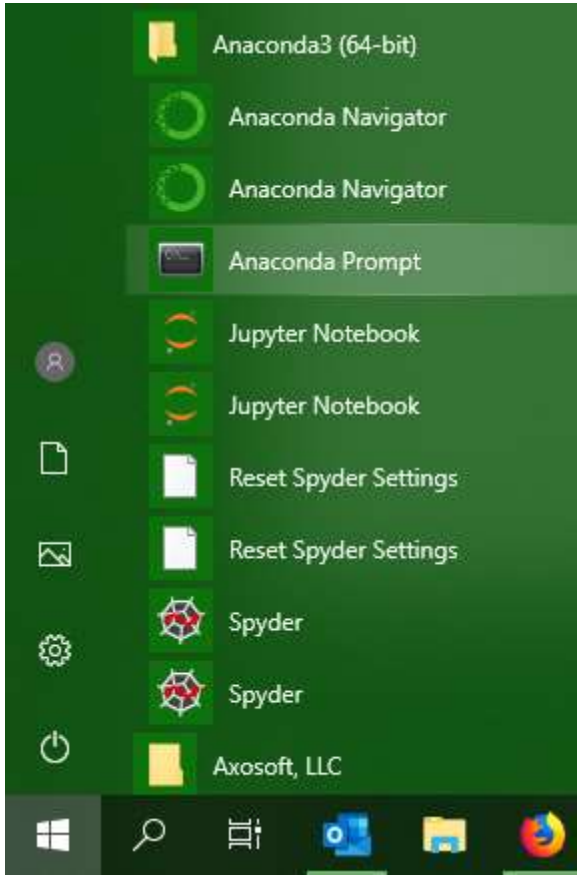
Preferred python distribution

Anaconda 3
(https://www.anaconda.com/distribution/)

After the installation

You should have a folder in your start menu with Anaconda3 (64-bit) in that folder you can **open** the **Anaconda Prompt** software

Note to MacOS and Linux user: Use the terminal app instead



Step 2: Installing TargetDB package

In the Anaconda prompt window simply type: `pip install targetDB`

After some time, you should see a line saying: `Successfully installed targetdb-1.3.0`

Step 3: Downloading TargetDB database

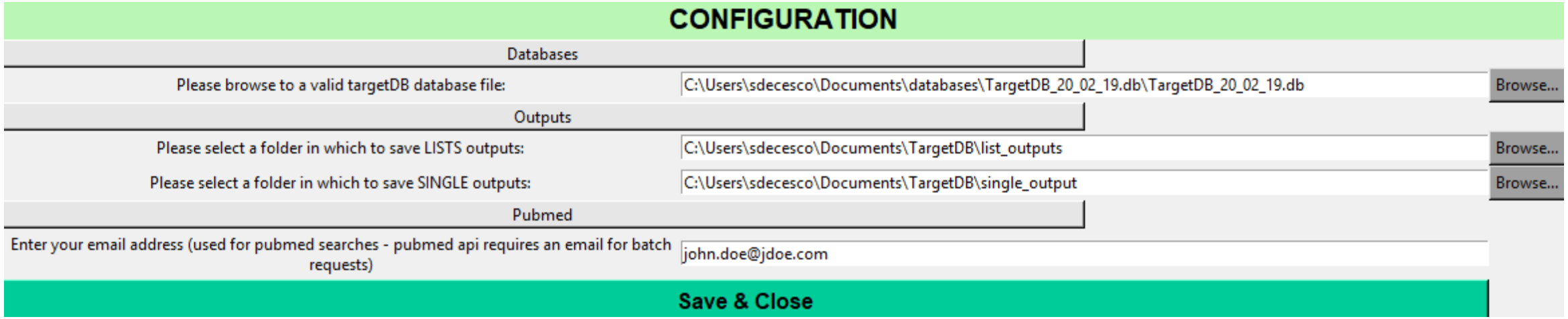
- You can download the database here:
https://github.com/sdecesco/targetDB/releases/download/v1.3.1/TargetDB_20_12_19.db.zip

- After downloading just unzip it in your desired location (unzipped file is > 7Gb)

Note: Please check GitHub page for future update of the database (plan is twice/year)

Step 4: Starting TargetDB for the first time

- Go back to the Anaconda prompt terminal and type: `targetDB`
- After a few seconds a window should appear (with empty boxes in the first launch)

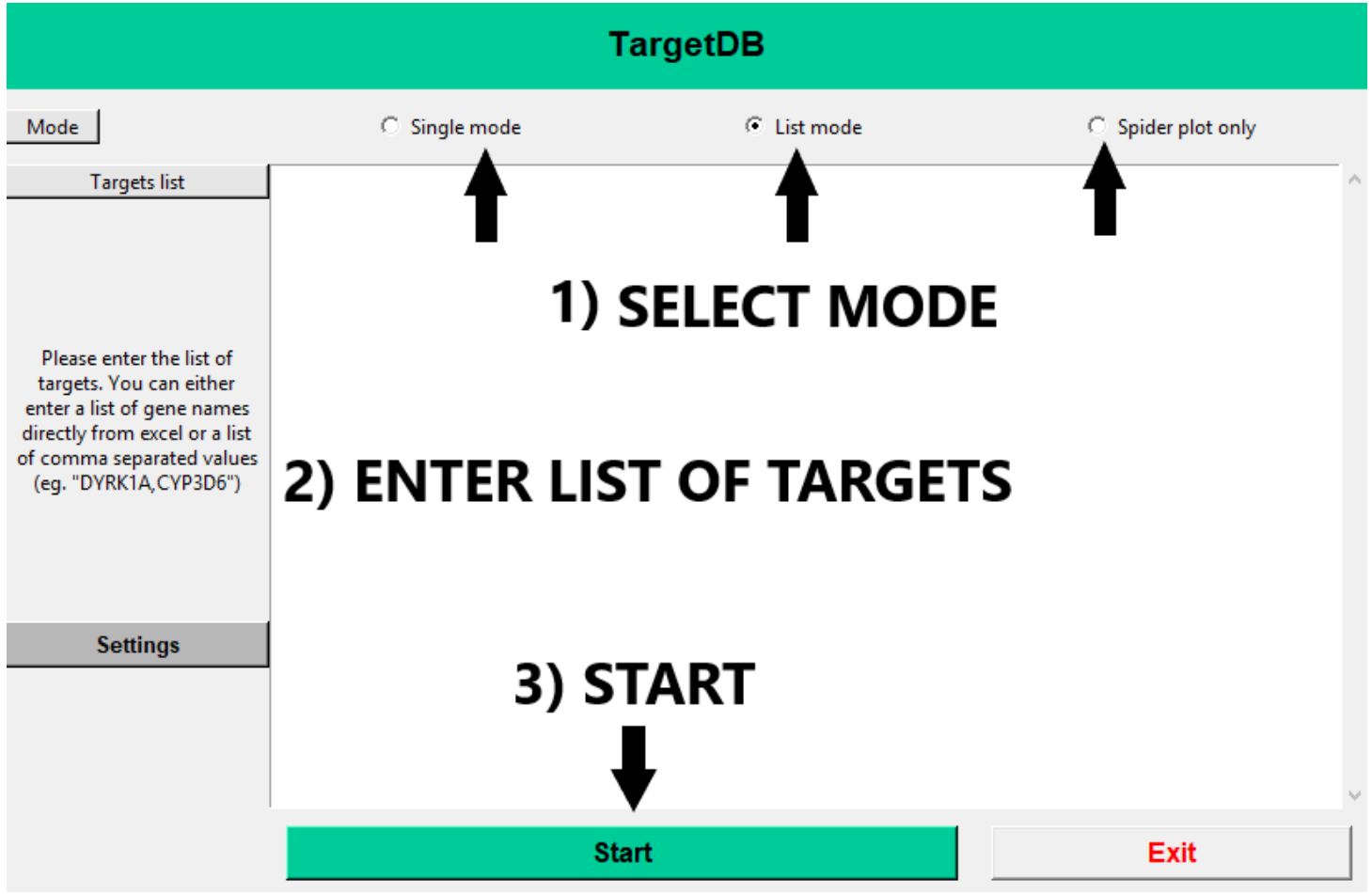


- Fill in the information for each line (4 in total)
  - **Database** line: tell the software where the database file is saved (see step 3)
  - **Outputs** (The program generates two types of outputs depending the mode used)
    - LISTS: Please indicate a folder in which it will save the lists outputs
    - SINGLE: Please indicate a folder in which it will save single outputs
      - *Note: These two folder can be the same*
    - **Pubmed:** TargetDB use a pubmed search to pull out relevant papers on targets or paper numbers if in list mode, it requires the user email address to use this functionality.

- **Click on Save and Close button**
- The main targetDB window should appear

Step 5: Future TargetDB start

The setting window described in step 4 can be called back if you require to make changes to the output folders / database location and/or email address



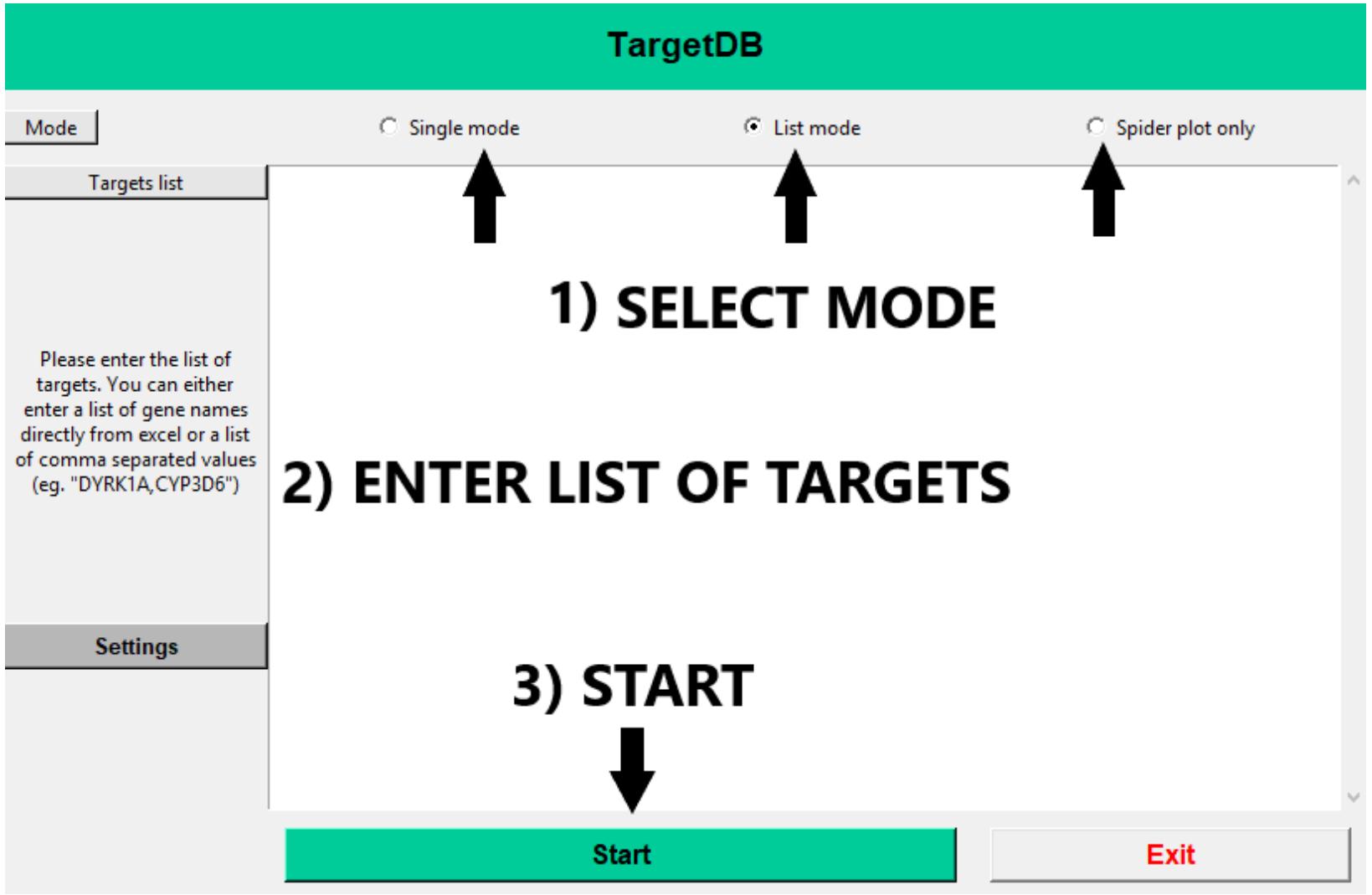
Database creation mode - Not Preferred

- The user can also choose to not use the pre-filled database and generate the database locally
- In order to do so a few extra pieces are required:
  - SQLite database of ChEMBL – Download here
  - Blast software – Instructions
  - Fpocket software – Instructions
- Limitations: will only work on Linux

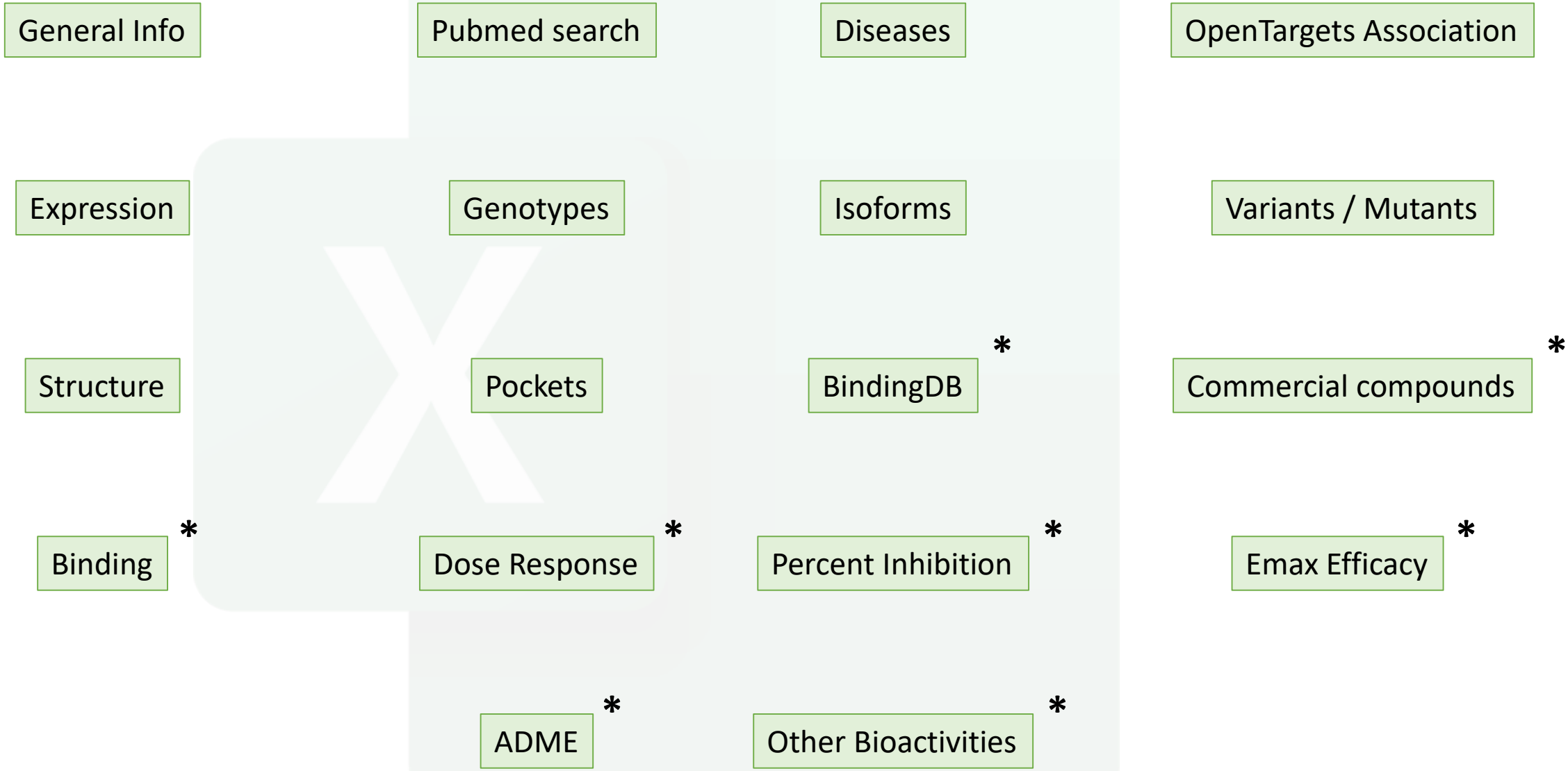
```
usage: target_DB [-h] [-g] [-i] [-l] [-a] [-s] [-v] [-update] [-blastcore]
                 [-update_config] [-create_db]

optional arguments:
  -h, --help            show this help message and exit
  -g , --gene           enter a single gene name
  -i , --in_file        Name of the input file with a list of genes (.txt - 1
                        gene per line)
  -l , --list_genes     Enter a list of gene name separated by a ","
  -a, --do_all          Use this option to create a database with all human
                        genes (list coming from HGNC)
  -s , --sphere_size    enter a value for the probe size of the pocket finder
                        tool (default = 3.0)
  -v, --verbose         Print information
  -update, --update     Update record if already in database (default: No)
  -blastcore , --num_core
                        Enter the value of processor core to use for the blast
                        search (default=8)
  -update_config, --update_config
                        use this if you want to update the config file
  -create_db, --create_db
                        Use this option to create an empty targetDB database
```

TargetDB in practice: 2. Using



TargetDB explained: Single output



* Tab only present if data available

TargetDB explained: Single output

General Info

Gene_name	BACE1
Synonyms	3.4.23.46,ASP2,Asp 2,Aspartyl protease 2,BACE,BACE1,Beta-secretase 1,Beta-site APP cleaving enzyme 1,Beta-site amyloid precursor protein cleaving enzyme 1,KIAA1149,Memapsin-2,Membrane-associated aspartic protease 2
Target_id	P56817
Protein_class	Aspartic protease A1A subfamily
Protein_class_desc	enzyme -> protease -> aspartic -> aa -> a1a
Species	Homo sapiens (Human).
Number_isoforms	6

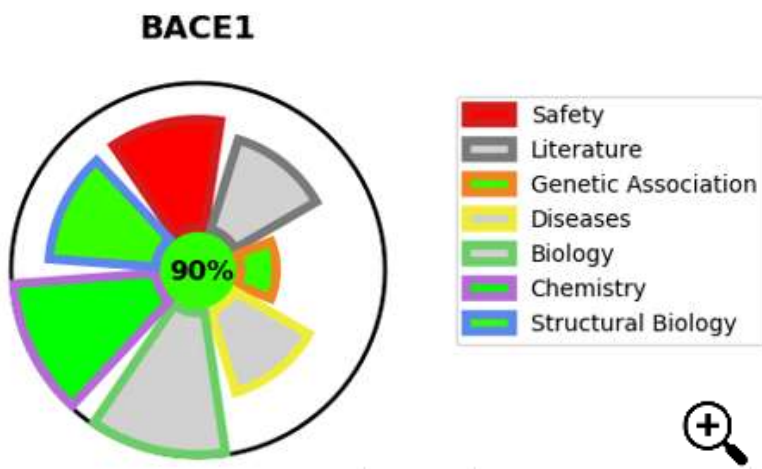
Tractable	Tractability_probability
True	90.92

In Training data set ?
No

OMIM

DISEASE	
disease_id	disease_name

PATHWAYS	
Reactome	KEGG
Amyloid fiber formation	Alzheimer's disease
Metabolism of proteins	



Machine learning model (discussed later)

TargetDB explained: Spider plot

Safety

Height
Amount of information available
Color
Safety issues (red) → No issues (green)

Literature

Height
Amount of information available

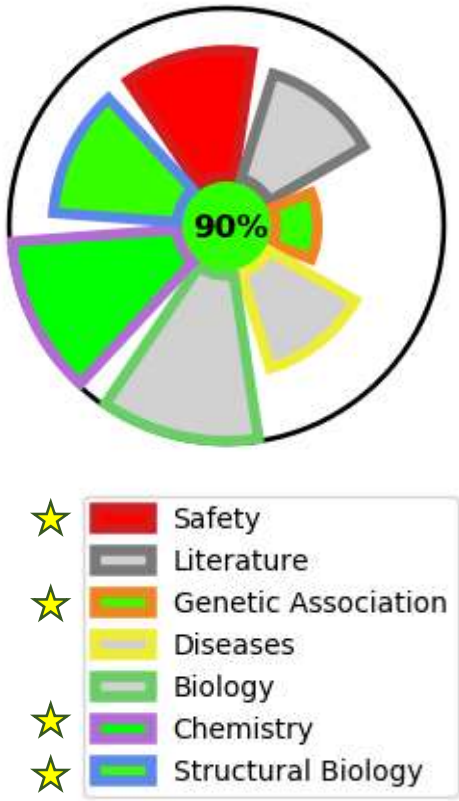
Genetic Association

Height
Amount of information available
Color
Quality of associations (p-value) (green) → (red)

Diseases

Height
Amount of information available

Beta-secretase 1



Central Number

Probability of tractability
(Machine Learning model)

Biology

Height
Amount of information available

Chemistry

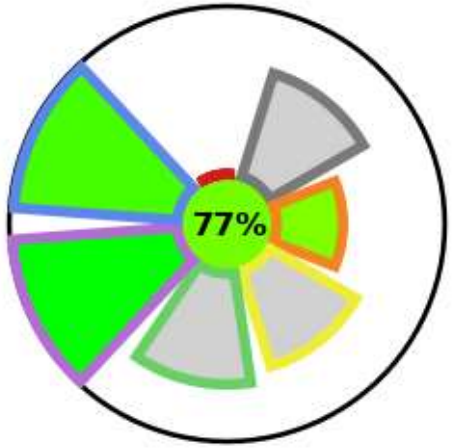
Height
Amount of information available
Color
Quality of compounds
(selectivity/potency)
(green) → (red)

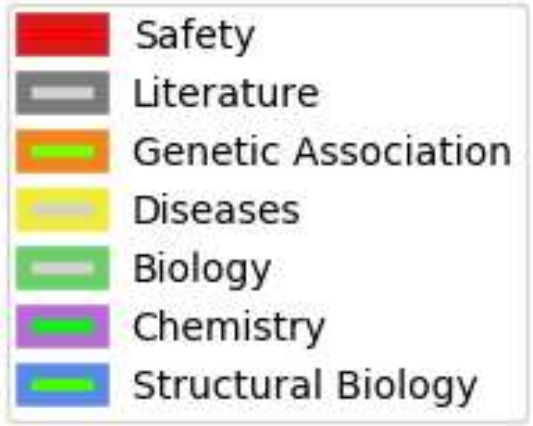
Structural Biology

Height
Amount of information available
Color
Druggability of binding pockets
(green) → (red)

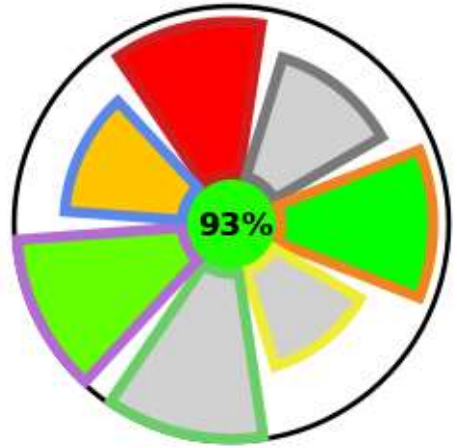
TargetDB explained: Spider plot

Acetylcholinesterase

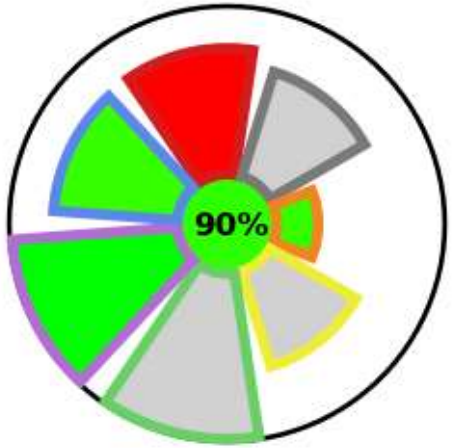




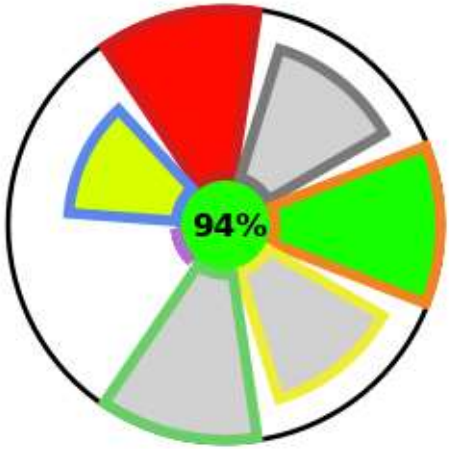
Tau



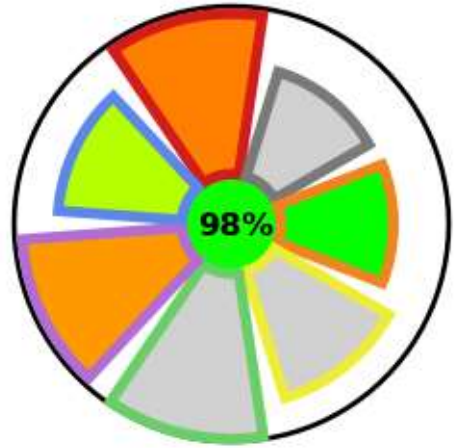
Beta-secretase 1



Apolipoprotein E



Presenilin1



TargetDB explained: Single output

Pubmed search

Simple search using the gene name as query (Caution: could lead to false hits for certain genes (e.g. REST))

Title	Journal Title	Year of Publication	Journal Article	Case Reports	Clinical Trial	Comparative Study	Letter	Meta-Analysis	Review	Neurodegeneration	Chemistry	Major Keywords	Abstract	Author	Affiliation	PMID	MeSH Terms	Other Term
miR-340 reduces accumulation of amyloid-beta through targeting BACE1 (beta-site	Current neurovascul	2020	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE		BACKGRO	['Tan X', 'Luo Y', 'Pi D', 'Xia L	["Department of Neurolog	https://www.ncbi.nlm.nih.gov/pubmed/31957613/		["Alzheime
Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against	SAR and QSAR in env	2020	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE		We have c	['Kumar V', 'Ojha PK', 'Saha	['Drug Theoretics and Cher	https://www.ncbi.nlm.nih.gov/pubmed/31865778/		["Alzheime
Curcumin inhibits BACE1 expression through the interaction between ERbeta and	Molecular and cellul	2020	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE		Alzheime	['Huang P', 'Zheng N', 'Zhou	["Hubei Key Laboratory of	https://www.ncbi.nlm.nih.gov/pubmed/31595422/		['AD', 'BACI
Advanced analytical methodologies in	Journal of pharmace	2020	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE		Despite th	['De Simone A', 'Naldi M', '	['Department for Life Qual	https://www.ncbi.nlm.nih.gov/pubmed/31606562/		["Alzheime
11beta-HSD1 Inhibition Rescues SAMP8	Molecular neurobiol	2020	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE		Ageing an	['Puigoriol-Illamola D', 'Lei	['Pharmacology Section, D	https://www.ncbi.nlm.nih.gov/pubmed/31399953/		['Ageing', "
MicroRNA-298 reduces levels of human amyloid-beta precursor protein (APP), beta-	Molecular psychiatry	2020	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	[]	Alzheime	['Chopra N', 'Wang R', 'Mal	['Laboratory of Molecular I	https://www.ncbi.nlm.nih.gov/pubmed/31942037/		
Activation of PKA/SIRT1 signaling pathway by photobiomodulation therapy reduces Abeta	Aging cell	2020	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE		A hallmar	['Zhang Z', 'Shen Q', 'Wu X'	['MOE Key Laboratory of La	https://www.ncbi.nlm.nih.gov/pubmed/31663252/		['APP proce

Alzheimer's Research UK | University of Oxford
DRUG DISCOVERY INSTITUTE

Diseases

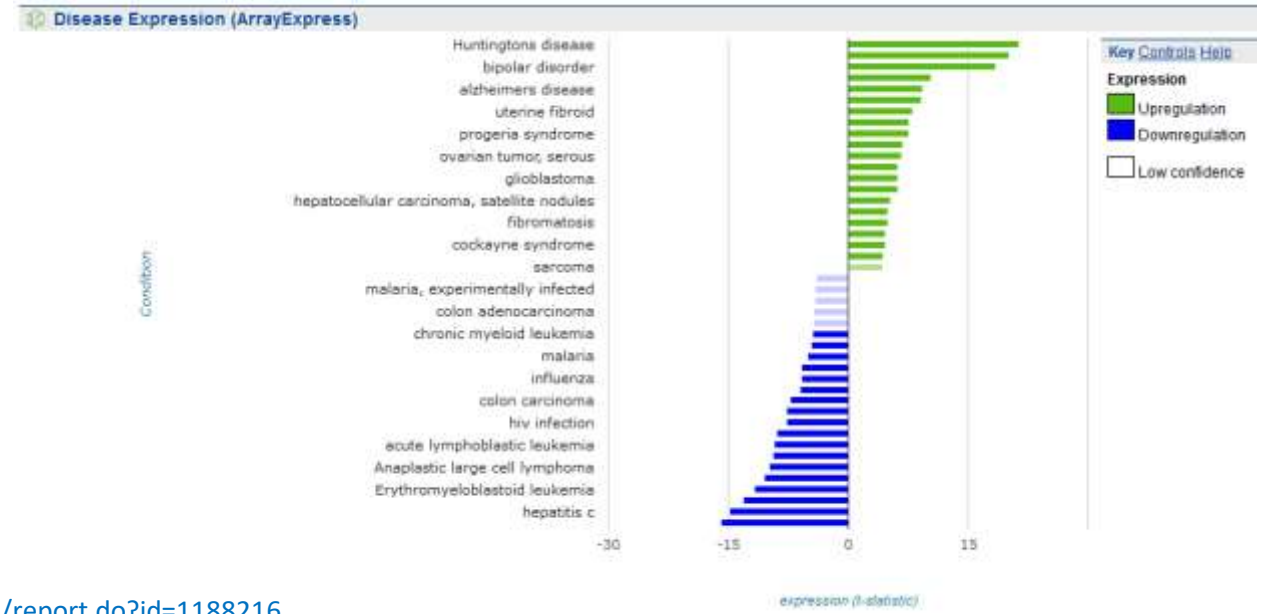
	A	B	C	D	E
1	DISEASE REGULATION				
2	disease	t_stat	std_dev_t	n	direction
3	Huntingtons disease	21.3	0	1	UP
4	brain tumor	20.1	0	1	UP
5	bipolar disorder	18.4	0	1	UP
6	obesity	10.3	0	1	UP
7	alzheimers disease	9.3	0	1	UP
8	neuroblastoma-poorly differentiated	9.2	0	1	UP
9	uterine fibroid	8.1	0	1	UP
10	periodontitis	7.6	0	1	UP
11	progeria syndrome	7.6	0	1	UP
12	hepatocellular carcinoma, no satellite nodules	6.8	0	1	UP
13	ovarian tumor, serous	6.6	0	1	UP
14	breast tumor	6.2	0	1	UP
15	glioblastoma	6.2	0	1	UP

	G	H	I	J	K	L
1	GWAS					
2	phenotype	organism	author	year	p_value	pubmed_id
3	Red cell distribution width	H. sapiens	Kichaev G	2019	0.000000002	30595370

GWAS

InterMine

HumanMine

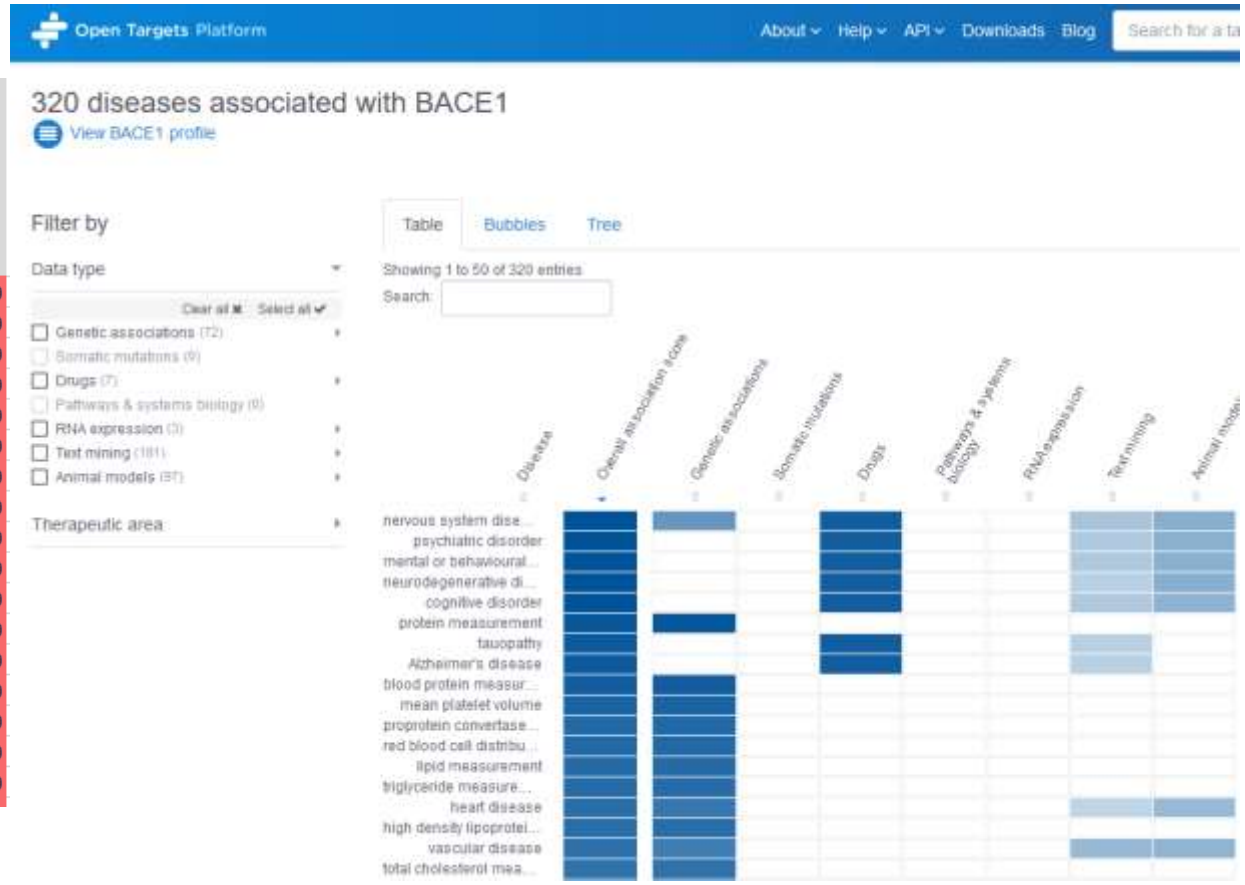


https://www.humanmine.org/humanmine/report.do?id=1188216

TargetDB explained: Single output

OpenTargets Association

disease_name	overall_score	genetic_association	known_drug	litterature_mining	animal_model	affected_pathway	rna_expression	somatic_mutation
nervous system disease	1	0	0.91	0.27	0.33	0	0	0
psychiatric disorder	1	0	0.91	0.17	0.32	0	0	0
mental or behavioural disorder	1	0	0.91	0.17	0.32	0	0	0
neurodegenerative disease	1	0	0.91	0.15	0.32	0	0	0
cognitive disorder	1	0	0.91	0.17	0.29	0	0	0
tauopathy	0.95	0	0.91	0.15	0	0	0	0
Alzheimer's disease	0.95	0	0.91	0.15	0	0	0	0
red blood cell distribution width	0.38	0.38	0	0	0	0	0	0
epilepsy	0.37	0	0	0.23	0.32	0	0	0
neuropathy	0.33	0	0	0.03	0.33	0	0	0
peripheral neuropathy	0.33	0	0	0.02	0.33	0	0	0
vascular disease	0.33	0	0	0.16	0.29	0	0	0
myopathy	0.31	0	0	0.08	0.29	0	0	0
Autosomal recessive non-syndromic intellectual disability	0.31	0	0	0	0.31	0	0	0
amyloidosis	0.31	0	0	0.31	0	0	0	0
Early infantile epileptic encephalopathy	0.3	0	0	0	0.3	0	0	0
diabetes mellitus	0.29	0	0	0.06	0.28	0	0	0



https://www.targetvalidation.org

TargetDB explained: Single output

Expression

THE HUMAN PROTEIN ATLAS

Selectivity			2.45
ORGANS			
organ_name	Total_value	n_tissues	avg_value
Gastrointestinal tract	14	8	1.8
Liver & gallbladder	5	3	1.7
Bone marrow & lymphoid tissues	16	10	1.6
Male tissues	8	5	1.6
Brain	17	11	1.5
Pancreas	3	2	1.5
Kidney & urinary bladder	4	3	1.3
Lung	5	4	1.3
Skin	6	5	1.2
Endocrine tissues	3	3	1.0
Muscle tissues	3	3	1.0
Female tissues	11	13	0.8
Proximal digestive tract	2	3	0.7
Adipose & soft tissue	1	6	0.2

ADIPOSE & SOFT TISSUE		
tissue name	Cell type	Value
Adipose tissue	Adipocytes	0
Soft tissue 1	Chondrocytes	1
Soft tissue 1	Fibroblasts	0
Soft tissue 1	Peripheral nerve	0
Soft tissue 2	Fibroblasts	0
Soft tissue 2	Peripheral nerve	0

BONE MARROW & LYMPHOID TISSUES		
tissue name	Cell type	Value
Appendix	Glandular cells	2
Appendix	Lymphoid tissue	2
Bone marrow	Hematopoietic cells	1
Lymph node	Germinal center cell	2
Lymph node	Non-germinal cente	2
Spleen	Cells in red pulp	1
Spleen	Cells in white pulp	2
Tonsil	Germinal center cell	2
Tonsil	Non-germinal cente	1
Tonsil	Squamous epithelial	1

BRAIN		
tissue name	Cell type	Value
Caudate	Glial cells	2
Caudate	Neuronal cells	1
Cerebellum	Cells in granular laye	2
Cerebellum	Cells in molecular la	2
Cerebellum	Purkinje cells	2
Cerebral cortex	Endothelial cells	0
Cerebral cortex	Glial cells	1
Cerebral cortex	Neuronal cells	2
Cerebral cortex	Neuropil	1
Hippocampus	Glial cells	2
Hippocampus	Neuronal cells	2

KIDNEY & URINARY BLADDER		
tissue name	Cell type	Value
Kidney	Cells in glomeruli	1
Kidney	Cells in tubules	2
Urinary bladder	Urothelial cells	1

LIVER & GALLBLADDER		
tissue name	Cell type	Value
Gallbladder	Glandular cells	2
Liver	Bile duct cells	1
Liver	Hepatocytes	2

LUNG		
tissue name	Cell type	Value
Bronchus	Respiratory epitheli	2
Lung	Macrophages	2
Lung	Pneumocytes	0
Nasopharynx	Respiratory epitheli	1

MALE TISSUES		
tissue name	Cell type	Value
Epididymis	Glandular cells	2
Prostate	Glandular cells	1
Seminal vesicle	Glandular cells	1
Testis	Cells in seminiferou	2
Testis	Leydig cells	2

MUSCLE TISSUES		
tissue name	Cell type	Value
Heart muscle	Myocytes	2
Skeletal muscle	Myocytes	1
Smooth muscle	Smooth muscle cells	0

PANCREAS		
tissue name	Cell type	Value
Pancreas	Exocrine glandular c	2

SELECTIVITY

$$S_{sel} = -\sum_{i}^{T} \rho_i \log \rho_i$$

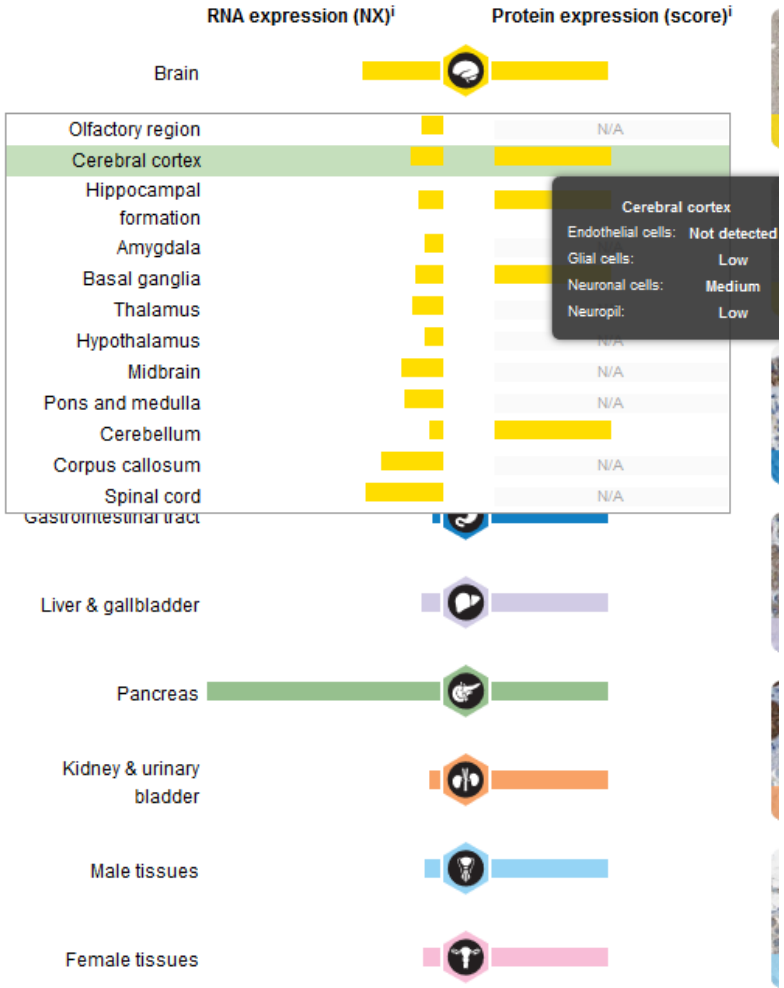
$S_{sel} = Selectivity\ Entropy$

$T = Expression\ in\ different\ tissue$

$\rho_i = Probability\ of\ an\ expression\ value$

$$\rho_{(T)} = \frac{E_T}{\sum_i E_{T\,i}}$$

$E_T = Expression\ value\ in\ tissue\ T$



https://www.proteinatlas.org/ENSG00000186318-BACE1/tissue

TargetDB explained: Single output

Genotypes

	A	B	C
1	Bace1<tm1.1Slr>		Bace1<tm1Nn>
2	Targeted, Null/knockout		Targeted, Null/knockout
3	HOMOZYGOTE		HOMOZYGOTE
4	Bace1<tm1.1Slr>/Bace1<tm1.1Slr>		Bace1<tm1Nn>/Bace1<tm1Nn>
5	no abnormal phenotype detected		abnormal cell physiology
6			
7	Bace1<tm1Bux>		Bace1<tm1Vas>
8	Targeted, Null/knockout|Reporter		Targeted, Null/knockout
9	HOMOZYGOTE		HOMOZYGOTE
10	Bace1<tm1Bux>/Bace1<tm1Bux>		Bace1<tm1Vas>/Bace1<tm1Vas>
11	no abnormal phenotype detected		no abnormal phenotype detected
12			
13	Bace1<tm1Cdi>		Bace1<tm2Psa>
14	Targeted, Reporter|Null/knockout		Targeted, Null/knockout
15	HOMOZYGOTE		HOMOZYGOTE
16	Bace1<tm1Cdi>/Bace1<tm1Cdi>		Bace1<tm2Psa>/Bace1<tm2Psa>
17	abnormal anxiety-related response		decreased anxiety
18	abnormal muscle physiology		hyperactivity
19	decreased exploration in new environment		postnatal lethality
20	improved righting response		
21			

InterMine

HumanMine

Gene (Hum OR Rat) --> Mouse Allele (Phenotype) (60 rows)

Manage Columns | Manage Filters | Manage Relationships | Save as List | Generate Python code | Export

Showing 1 to 25 of 60 rows

Rows per page: 25 | page 1

Gene Symbol	Gene Name	Homologues Homologue . Organism . Short Name	Homologue Symbol	Homologue Name	Alleles Symbol	Alleles Name	Alleles Primary Identifier	Phenotype Terms Name	Phenotype Terms Description	Genotypes Name	Genotypes Zygosity	Genotypes Genetic Background	Phenotype Terms Identifier	Phenotype Terms Ontology . Name
BACE1	beta-secretase 1	M. musculus	Bace1	beta-site APP cleaving enzyme 1	Bace1<tm1.1Slr>	targeted mutation 1.1, Steven L Richards	MGI:2181644	no abnormal phenotype detected	normal viability, fertility, appearance and behavior; reported phenotype is indistinguishable from controls	Bace1<tm1.1Slr>/Bace1<tm1.1Slr>	homozygote	involves: 129P2/OlaHsd * C57BL/6 * CBA	MP:0002169	Mammalian Phenotype Ontology
BACE1	beta-secretase 1	M. musculus	Bace1	beta-site APP cleaving enzyme 1	Bace1<tm1Bux>	targeted mutation 1, Joseph D Buxbaum	MGI:3042562	no abnormal phenotype detected	normal viability, fertility, appearance and behavior; reported phenotype is indistinguishable from controls	Bace1<tm1Bux>/Bace1<tm1Bux>	homozygote	involves: 129S6/SvEvBrd * C57BL/6	MP:0002169	Mammalian Phenotype Ontology

https://www.humanmine.org/humanmine/report.do?id=1188216

TargetDB explained: Single output

Isoforms

UniProt

			BACE1-A			
Is Canonical	Yes		Similarity	100	number of residues	501
			BACE1-B			
Is Canonical	No		Similarity	95.01	number of residues	476
SEQUENCE	MAQALPWLLLWMGAGVLPAHGTQHGIRLPLRSGLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQGYYVEMTVGSPPQTLNILVDTGSSNFAVGAAPHPFL					
start	stop	previous_seq	modification_type	new_seq	in_domains	comments
190	214		remove		Peptidase A1,Asp	(in isoform B, isoform D and isoform 6)
			BACE1-C			
Is Canonical	No		Similarity	91.22	number of residues	457
			BACE1-D			
Is Canonical	No		Similarity	86.23	number of residues	432
			BACE1-5			
Is Canonical	No		Similarity	78.44	number of residues	401
			BACE1-6			
Is Canonical	No		Similarity	73.45	number of residues	376

Variants / Mutants

UniProt

VARIANTS						
start	stop	modification_type	previous_seq	new_seq	in_domains	comments
265	265	replace	V	A	Peptidase A1,Asp	(in dbSNP:rs28989503)
481	481	replace	R	C		(in dbSNP:rs539765)
MUTANTS						
start	stop	modification_type	previous_seq	new_seq	in_domains	comments
93	93	replace	D	N	Peptidase A1,Asp	Decreases beta-cleaved soluble APPproduction. (ECO:0000269|PubMed:10656250)
284	284	replace	D	N	Peptidase A1,Asp	Almost abolishes beta-cleaved soluble APPproduction. (ECO:0000269|PubMed:10656250)
498	498	replace	S	D		No effect on endocytosis form the cellsurface. Decreases recycling from endosomes to the cellsurface. (ECO:0000269|PubMed:15886016)
499	500	replace	LL	AA		Impairs endocytosis and produces a delayedretrograde transport to the trans-Golgi network anddelivery to the lysosmes, decreasinf its degradation.Disrupts location to late endosomes and lysosomes. Locatesmainly at the cell surface. No effect on degradationregulated by GGA3. Effects on protein stability anddefective internalization increases; when associated withR-501. (ECO:0000269|PubMed:15615712, ECO:0000269|PubMed:16033761, ECO:0000269|PubMed:20484053, ECO:0000269|PubMed:23109336)
501	501	replace	K	R		Inhibits ubiquitination. No effect onendocytosis rate. Induced protein stability andacculmulation in early and late endosomes, lysosomes andcell membrane. Effects on protein stability and defectiveinternalization increases; when associated with A-499-500-A. (ECO:0000269|PubMed:20484053, ECO:0000269|PubMed:23109336, ECO:0000269|PubMed:27302062)

TargetDB explained: Single output

Structure

DOMAINS				
Domain_name	**start**	**stop**	**length**	**source**
Peptidase A1	75	416	341	Uniprot
Asp	74	418	344	Pfam-A

DOMAINS - DrugEbillity				
pdb_list	**domain_fold**	**main_superfam**	**tractable**	**druggable**
1FKN,1M4H,1S(	Acid proteases	Acid proteases	1	1
1PY1,1UJJ,1UJK	UNMATCHED	UNMATCHED	1	1
1PY1,1UJJ,1UJK	alpha-alpha su	ENTH/VHS dom	1	0

PDB BLAST						
PDB_code	**Chain**	**similarity**	**gene**	**species**	**SITES_tractable**	**SITES_druggable**
6PZ4	A	100	BACE1	HUMAN		
3ZKM	A	71.5	BACE2	HUMAN		
2EWY	A	71.3	BACE2	HUMAN		
3ZLQ	A	71.2	BACE2	HUMAN		

DrugEbillity: File from ftp://ftp.ebi.ac.uk/pub/databases/chembl/DrugEBIlity/

TargetDB explained: Single output

DRUG DISCOVERY INSTITUTE

Structure

DOMAINS				
Domain_name	start	stop	length	source
Peptidase A1	75	416	341	Uniprot
Asp	74	418	344	Pfam-A

DOMAINS - DrugEbillity				
pdb_list	domain_fold	main_superfam	tractable	druggable
1FKN,1M4H,1S(	Acid proteases	Acid proteases	1	1
1PY1,1UJJ,1UJK	UNMATCHED	UNMATCHED	1	1
1PY1,1UJJ,1UJK	alpha-alpha su	ENTH/VHS dom	1	0

PDB BLAST						
PDB_code	Chain	similarity	gene	species	SITES_tractable	SITES_druggable
6PZ4	A	100	BACE1	HUMAN		
3ZKM	A	71.5	BACE2	HUMAN		
2EWY	A	71.3	BACE2	HUMAN		
3ZLQ	A	71.2	BACE2	HUMAN		

PDB								PDB: Ligand								L - Drugg	
PDB_code	Technique	Resolution	Chain	Domain_name	n_residues	% of full protein	start_stop	type_of_binder	binding_type	operator	value	units	Ligand_name	publication_year	PDBbind_link	SITES_tractable	SITES_druggable
1FKN	X-ray	1.90 A	A,B	Asp	391	inf	46-436	Protein - Ligand	Ki	=	1.6	nM	(7-mer)	2000	1FKN	1	1
1M4H	X-ray	2.10 A	A,B	Asp	391	inf	56-446	Protein - Ligand	Ki	=	0.3	nM	(7-mer)	2002	1M4H	1	1
1PY1	X-ray	2.60 A	F,E,H,G		8	inf	494-501	Protein - Ligand	Kd	=	40	uM	(7-mer) incompl	2003	1PY1	1	0
1SGZ	X-ray	2.00 A	A,B,D,C	Asp	389	inf	58-446									1	1
1TQF	X-ray	1.80 A	A	Asp	405	inf	43-446	Protein - Ligand	IC50	=	1.4	uM	(32P)	2004	1TQF	1	1
1UJJ	X-ray	2.60 A	C		12	inf	490-501	Protein - Ligand	Kd	=	0.8	uM	(12-mer) Kd=0.8	2004	1UJJ	1	0
1UJK	X-ray	1.90 A	D,C		12	inf	490-501	Protein - Ligand	Kd	=	0.27	uM	(12-mer) Kd=0.2	2004	1UJK	1	1
1W50	X-ray	1.75 A	A	Asp	411	inf	43-453									1	1
1W51	X-ray	2.55 A	A	Asp	411	inf	43-453	Protein - Ligand	IC50	=	500	nM	(L01) ligand is co	2004	1W51	1	1
1XN2	X-ray	1.90 A	D,B,C,A	Asp	389	inf	58-446	Protein - Ligand	Ki	=	0.03	nM	(11-mer) incomp	2005	1XN2	1	1
1XN3	X-ray	2.00 A	A,B,D,C	Asp	389	inf	58-446	Protein - Ligand	Ki	=	40	nM	(14-mer)	2005	1XN3	1	1
1XS7	X-ray	2.80 A	D	Asp	389	inf	58-446	Protein - Ligand	Ki	=	25.1	nM	(MMI)	2004	1XS7	1	1
1YM2	X-ray	2.05 A	B,C,A	Asp	402	inf	48-447	Protein - Ligand	IC50	=	0.01	uM	(6-mer)	2006	1YM2	1	1
1YM4	X-ray	2.25 A	B,C,A	Asp	408	inf	48-453	Protein - Ligand	IC50	=	0.039	uM	(5-mer)	2006	1YM4	1	1
2B8L	X-ray	1.70 A	A	Asp	405	inf	43-446	Protein - Ligand	IC50	=	15	nM	(5HA)	2005	2B8L	1	1
2B8V	X-ray	1.80 A	A	Asp	405	inf	43-446	Protein - Ligand	IC50	=	98	nM	(3BN)	2005	2B8V	1	1
2F3E	X-ray	2.11 A	A,B,C	Asp	402	inf	48-447	Protein - Ligand	IC50	=	0.156	uM	(AXQ)	2006	2F3E	1	1

RCSB PDB PROTEIN DATA BANK

PDB bind

DrugEbillity

DrugEbillity: File from ftp://ftp.ebi.ac.uk/pub/databases/chembl/DrugEBIlity/

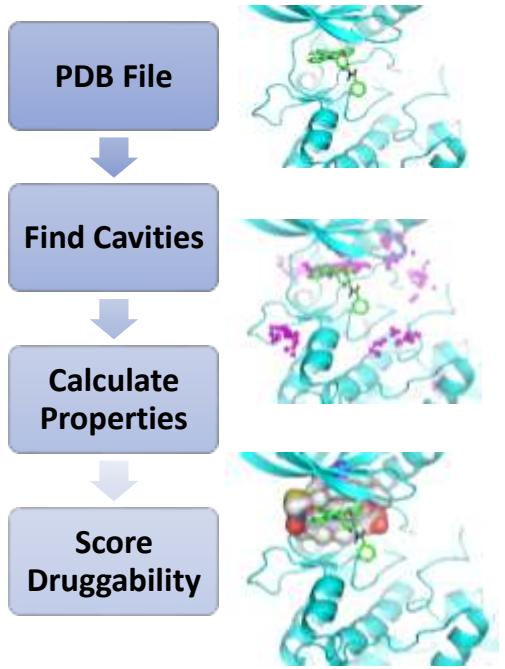
TargetDB explained: Single output

Pockets

DRUGGABLE POCKETS							
PDB_code	druggability_score	pocket_score	pocket_number	volume	area	fraction_apolar	domains
1PY1	0.542	-0.306	p15	2513.7	704.1	50.9	
1PY1	0.85	-0.134	p10	828.2	259.2	62	
1SGZ	0.543	0.128	p5	378.6	100.5	63.7	Peptidase A1 (1.0%),Asp (1.0%)
1SGZ	0.606	0.23	p2	525.9	142.5	69.5	Peptidase A1 (3.0%),Asp (3.0%)
1TQF	0.906	0.536	p1	1424	252.9	34.9	Peptidase A1 (7.0%),Asp (7.0%)
1UJJ	0.67	0.223	p3	357.7	142.5	66.1	
1UJJ	0.681	0.285	p2	465.8	158.4	56.4	
1W51	0.907	0.777	p1	2130.8	354.8	40.2	Peptidase A1 (10.0%),Asp (10.0%)
1XN2	0.736	0.488	p2	320.6	36.7	29.6	Peptidase A1 (2.0%),Asp (2.0%)
1XS7	0.921	0.804	p1	738.1	90.9	38.5	Peptidase A1 (4.0%),Asp (4.0%)
1YM2	0.938	0.615	p1	596.7	100.6	50.4	Peptidase A1 (2.0%),Asp (2.0%)
1YM4	0.927	0.384	p1	663.1	180.4	75	Peptidase A1 (6.0%),Asp (6.0%)
2B8L	0.922	0.878	p1	1417.8	179.2	39.8	Peptidase A1 (7.0%),Asp (7.0%)
2B8V	0.536	-0.133	p12	264.4	110.7	78.5	Peptidase A1 (2.0%),Asp (2.0%)
2B8V	0.572	0.671	p1	1796.7	299.3	42	Peptidase A1 (9.0%),Asp (8.0%)

ALTERNATE DRUGGABLE POCKETS (PDB from blast)									
PDB_code	druggability_score	pocket_score	pocket_number	volume	area	fraction_apolar	gene	species	similarity
2GCD	0.617	8.653	p7	443.8	253.9	67.7	TAOK2	RAT	99.4
3CKW	0.823	37.173	p0	725.1	393.3	72.9	STK24	HUMAN	61
5J5T	0.711	42.643	p0	815.8	405	63.6	M4K3	HUMAN	60.7
4O27	0.696	18.144	p3	518.2	239.4	71.6	STK24	HUMAN	60.6
5AX9	0.851	39.37	p0	699	455.2	54.5	TNIK	HUMAN	60.6
3CKX	0.57	39.585	p0	551.2	343.9	63	STK24	HUMAN	60.6
4U8Z	0.715	40.926	p0	551.8	294.9	63.9	STK24	HUMAN	60.4

fpocket
scalable high performance pocket detection



1. Schmidtke, P. & Barril, X. Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.* **53,** 5858–5867 (2010).

TargetDB explained: Single output

Alzheimer's Research UK | University of Oxford
DRUG DISCOVERY INSTITUTE

Binding / Dose Response / Percent inhibition / Emax Efficacy / ADME / Other bioactivities

ChEMBL

Binding	Ki, Kd
Dose Response	IC50, EC50, Potency
Percent Inhibition	%Activity, Residual activity, %Inhibition
Emax Efficacy	Emax, Efficacy
ADME	ADME
Other Bioactivities	All other

Bioactivity info								Assay info		Structure	Ligand properties													Ligand info					
lig_id	standard_type	operator	value_num	units	pX	Selectivity	number of other targets	best_target_name	assay_description	confidence_score	SMILES	HBA	HBD	LogD	LogP	MW	TPSA	apKa	bpKa	nAr	pass_ro3	ro5_violations	rotB	CNS_MPO	molecular_species	indication_class	class_def	max_phase	oral	assay_ref
CHEMBL3809870	Kd	=	0.4	nM	9.4	0	1	Beta-secretase 1	Binding affinity to BACE1 (un	9	CC(C)(C)CCNC(=O)CCc1cc2cc(	5	3	7.51	7.55	617.77	110	12.41	6.35	5	N	2	11	1.5	NEUTRAL			0	0	10.1021/acs.jmedchem.5b01917
CHEMBL3809662	Kd	=	0.6	nM	9.22	0	1	Beta-secretase 1	Binding affinity to BACE1 (un	9	NC1=NC2(CO1)c3cc(NC(=O)c4	7	2	2.59	2.6	488.93	108.06	10.53	5.88	3	N	0	3	3.7	NEUTRAL			0	0	10.1021/acs.jmedchem.5b01917
CHEMBL3808441	Kd	=	8	nM	8.1	0	1	Beta-secretase 1	Binding affinity to BACE1 (un	9	CC(C)(C)CCNC(=O)CCc1cc2cc(	4	3	7.11	7.13	590.74	97.11		6.21	4	N	2	9	1.9	NEUTRAL			0	0	10.1021/acs.jmedchem.5b01917
CHEMBL3809897	Kd	=	11	nM	7.96	0	1	Beta-secretase 1	Binding affinity to BACE1 (un	9	CC(C)(C)CCNC(=O)CCc1cc2cc(	5	3	8.45	8.47	613.74	110	11.86	6.17	5	N	2	8	1.5	NEUTRAL			0	0	10.1021/acs.jmedchem.5b01917
CHEMBL3808988	Kd	=	16	nM	7.8	0	1	Beta-secretase 1	Binding affinity to BACE1 (un	9	C[C@@H](Cc1cc2cc(ccc2nc1N	4	2	4.93	4.95	404.56	80.9		6.08	3	N	0	6	3.2	NEUTRAL			0	0	10.1021/acs.jmedchem.5b01917
CHEMBL3808967	Kd	=	25	nM	7.6	0	1	Beta-secretase 1	Binding affinity to BACE1 (un	9	CC(C)(C)CCNC(=O)CCc1cc2cc(	4	3	7.31	7.33	590.74	97.11	13.87	6.21	4	N	2	8	1.9	NEUTRAL			0	0	10.1021/acs.jmedchem.5b01917
CHEMBL3808672	Kd	=	140	nM	6.85	0	1	Beta-secretase 1	Binding affinity to BACE1 (un	9	Cc1ccccc1c2ccc3nc(N)c(CCC(=	3	2	5.81	5.83	389.54	68.01		6.38	3	N	1	6	3.3	NEUTRAL			0	0	10.1021/acs.jmedchem.5b01917
CHEMBL2179131	Ki	=	0.017	nM	10.77	0	3	Beta-secretase 1	Inhibition of recombinant BA	8	CC(C)CNC(=O)[C@@H](NC[C	7	5	2.75	2.8	665.86	156.94	13.72	6.51	3	N	1	16	2.6	NEUTRAL			0	0	10.1021/jm3008823

TargetDB explained: Single output

BindingDB

ZincID	IC50(nM)	EC50(nM)	Kd(nM)	Ki(nM)	kon(M-1s-1)	koff(s-1)	pH	Temp	Source	DOI	Patent_number	Institution
ZINC03965949	20						4.8	37.00 C	Curated frc	10.1021/jm061242y		Elan Pharmaceuticals
ZINC10339547	100000						5	22.00 C	Curated frc	10.1021/jm061197u		Astex
ZINC10339550	310000						5	22.00 C	Curated frc	10.1021/jm061197u		Astex
ZINC10339550	310000						5	22.00 C	Curated frc	10.1021/jm0611962		Astex
ZINC11525586	94000						5	22.00 C	Curated frc	10.1021/jm061197u		Astex
ZINC26492127	82						4.6	22.00 C	Curated frc	10.1016/j.bmcl.2005.09.003		Lilly S.A.
				949			5		US Patent	10.1021/jm021079g	US9687494	Merck Sharp & Dohme Corp.
				1261			5		US Patent	10.1021/jm021079g	US9687494	Merck Sharp & Dohme Corp.
	32.9						5	30.00 C	US Patent		US9540359	Shionogi & Co., Ltd.
	20.7						5	30.00 C	US Patent		US9540359	Shionogi & Co., Ltd.
	113						5	30.00 C	US Patent		US9540359	Shionogi & Co., Ltd.

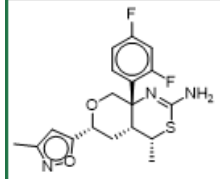
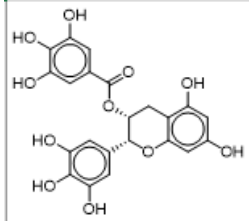
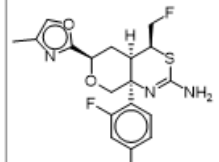
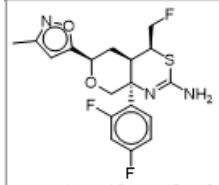
TargetDB explained: Single output

Commercial compounds

smiles	affinity_type	op	affinity_value	affinity_unit	price	Source_0	Source_1	Source_2	Source_3	Source_4	Source_5
C[C@H]1SC(N)=N[C@]2(CO[C@H](C[C@@H]12)c1cc(C)no1)c1ccc(F)cc1F	IC50	=	6	nM		http://www.sigmaaldrich.com/catalog/product/SIGMA/PZ0261?lang=en&region=US					
Oc1cc(O)c2C[C@@H](OC(=O)c3cc(O)c(O)c(O)c3)[C@H](Oc2c1)c1cc(O)c(O)c(O)c1	IC50	=	7.6	nM	($50)/(50 mg) OR ($65)/(100 r	http://www.	https://or	https://w	http://www.medchemexpress.com/ - -Epigallocatechin-Gallate.html		
[H][C@@]12C[C@@H](OC[C@@]1(N=C(N)S[C@@H]2CF)c1ccc(F)cc1F)c1nc(C)co1	IC50	=	10	nM		http://www.sigmaaldrich.com/catalog/product/SIGMA/PZ0260?lang=en&region=US					
Cc1cc(on1)[C@H]1C[C@H]2[C@@H](CF)SC(N)=N[C@]2(CO1)c1ccc(F)cc1F	IC50	=	15	nM		http://www.sigmaaldrich.com/catalog/product/SIGMA/PZ0262?lang=en&region=US					
CCCCOc1ccc(CSC(N)=N)cc1Cl	Kd	=	15	nM	Vitas-M Laboratory, Ltd.: (26.	https://orde	https://www.molport.com/shop/molecule-link/MolPort-002-555-036				
NC1=N[C@@](c2cccc(F)c12)(c1cccc(c1)-c1cncnc1)c1ccnc(c1)C(F)F	IC50	=	17	nM	Selleck Chemicals LLC: (198.9	https://orde	https://www.molport.com/shop/molecule-link/MolPort-039-193-829				
[H][C@@]12COC[C@@]1(N=C(N)SC2)c1cc(NC(=O)c2ccc(F)cn2)ccc1F	IC50	=	20	nM	($140)/(5 mg) OR ($230)/(10 r	https://orde	https://w	http://www.medchemexpress.com/LY2886721.html			
NC1=N[C@](c2cccc(F)c12)(c1ccnc(c1)C(F)(F)F)c1ccc(F)c(c1)-c1cncnc1	IC50	=	22	nM		https://orderbb.emolecules.com/cgi-bin/more?vid=76739946					
CC(C)C[C@H](NC(=O)[C@H](CC(N)=O)NC(=O)[C@@H](NC(=O)[C@@H](N)CCC(O)=O)C(C)C)[C@@H](O)C[C@@	Kd	=	40	nM		http://www.sigmaaldrich.com/catalog/product/SIGMA/G8291?lang=en&region=US					
COc1ccc2c(CCC2(Cc2ccc(nc2)C(F)(F)F)c2cn(Cc3ccccc3)c(N)n2)c1	IC50	=	63	nM		https://orderbb.emolecules.com/cgi-bin/more?vid=49436464					
O=C(CSc1nc2ccc(Nc3nc(nc(n3)N3CCCC3)N3CCCC3)cc2s1)Nc1ccccc1	IC50	=	120	nM	InterBioScreen Ltd.: (32.00 US	http://www.	https://or	https://www.molport.com/shop/molecule-link/MolPort-002-579-749			
Oc1cc(O)c2C[C@H](OC(=O)c3cc(O)c(O)c(O)c3)[C@H](Oc2c1)c1cc(O)c(O)c(O)c1	Ki	=	170	nM		http://www.	https://orders.emolecules.com/cgi-bin/more?vid=71007646				
Oc1cc(O)c2C[C@H](OC(=O)c3cc(O)c(O)c(O)c3)[C@@H](Oc2c1)c1cc(O)c(O)c(O)c1	Ki	=	210	nM		https://orders.emolecules.com/cgi-bin/more?vid=1984371					
C[C@@H](NC(=O)c1cc(cc(c1)C(=O)N[C@@H](Cc1ccccc1)[C@H](O)CNC1CC1)N(C)S(C)(=O)=O)c1ccccc1	Ki	=	233	nM		https://orderbb.emolecules.com/cgi-bin/more?vid=29914527					
C[C@]1(CCSC(N)=N1)c1cc(c(F)cc1F)-c1cncnc1	IC50	=	240	nM	Angene: (253.00 USD)/(5 mg)	http://www.	https://or	http://ww	https://www.molport.com/shop/molecule-link/MolPort-009-679-511		
O[C@@H]1O[C@@H]2COC(=O)c3cc(O)c(O)c(O)c3-c3c(O)c(O)c4oc(=O)c5c(c(O)c(O)c6oc(=O)c3c4c56)-c3c(O)c(O	IC50	=	410	nM	AK Scientific, Inc.: (98.00 USD	http://www.	https://or	https://www.molport.com/shop/molecule-link/MolPort-023-220-646			
O=C(CSc1nc2ccc(Nc3nc(nc(n3)N3CCCC3)N3CCCC3)cc2s1)NCc1ccco1	IC50	=	690	nM	InterBioScreen Ltd.: (44.00 US	http://www.	https://or	https://www.molport.com/shop/molecule-link/MolPort-002-577-515			

TargetDB explained: Single output

Commercial compounds

smiles	affinity_type	op	affinity_value	affinity_unit	price	Source_0	Source_1	Source_2	Source_3	Source_4	Source_5
 C[C@H]1SC(N)=N[C@]2(CO[C@H](C[C@@H]12	IC50	=	6	nM		http://www.sigmaaldrich.com/catalog/product/SIGMA/PZ0261?lang=en&region=US					
 Oc1cc(O)c2C[C@@H](OC(=O)c3cc(O)c(O)c(O)c3	IC50	=	7.6	nM	($50)/(50 mg) OR ($65)/(100 r	http://www.	https://or	https://w	http://www.medchemexpress.com/ - -Epigallocatechin-Gallate.html		
 [H][C@@]12C[C@@H](OC[C@@]1(N=C(N)S[C@	IC50	=	10	nM		http://www.sigmaaldrich.com/catalog/product/SIGMA/PZ0260?lang=en&region=US					
 Cc1cc(on1)[C@H]1C[C@H]2[C@@H](CF)SC(N)=	IC50	=	15	nM		http://www.sigmaaldrich.com/catalog/product/SIGMA/PZ0262?lang=en&region=US					

TargetDB explained: Single output - summary

General Info

OMIM KEGG reactome

Pubmed search

PubMed

Diseases

InterMine

OpenTargets Association

Open Targets

Expression

THE HUMAN PROTEIN ATLAS

Genotypes

InterMine

Isoforms

UniProt

Variants / Mutants

UniProt

Structure

RCSB PDB PROTEIN DATA BANK PDBbind

DrugEbillity

Pockets

fpocket

BindingDB

BindingDB

Commercial compounds

BindingDB

Binding / Dose Response / Percent inhibition / Emax Efficacy / ADME / Other bioactivities

ChEMBL

TargetDB explained: List view – information categories

Druggability

- Pockets druggability
- Domain druggability
- Druggability of similar targets' pockets

Structure

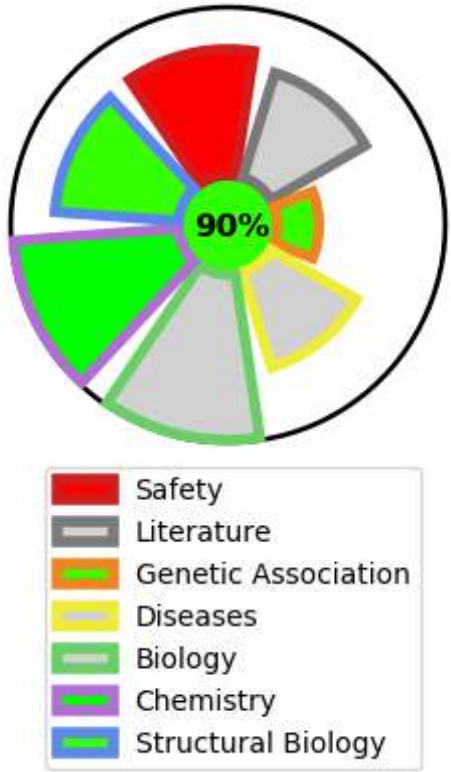
- % of sequence covered
- % of domains covered
- Number of PDB
- Number of PDB of similar targets

Chemistry

- BindingDB potent ligands
- BindingDB ligands in phase 2 clinic
- ChEMBL potent ligands
- ChEMBL selective ligands

Biology

- Protein expression levels
- Number of antibodies
- Variants
- Mutants
- Mice genotypes
- KEGG/Reactome



Disease Link

- Number of disease areas
- Max association score
- Diseases count

Genetic links

- GWAS association count
- Genetic association score

Information

- JensenLab Pubmed score

Safety

- Heart protein expression
- Liver protein expression
- Kidney protein expression
- Lethal phenotypes observed in mice

TargetDB explained: List view – information categories

Structure

- **% of sequence covered**
- **% of domains covered**
- **Number of PDB**
- **Number of PDB of similar targets**

- % of sequence covered
  - Percent of the sequence covered by PDB structures
- % of domains covered
  - Percentage of the domains that are covered by PDB structures
- Number of PDB
  - if count >=1 ➔ Score = 0.25
  - if count >=2 ➔ Score = 0.5
  - if count >=3 ➔ Score = 1
- Number of alternate (BLAST) PDB
  - Same as number of PDB * max similarity

Structure Score

Average of all these components

Druggability

- **Pockets druggability**
- **Domain druggability**
- **Druggability of similar targets' pockets**

- Pockets druggability scores
  - Average of pockets with a druggability score > 0.5
- Domain druggability
  - Average of domains druggability and tractability coming from DrugEBIlity
- Alternate (BLAST) pockets druggability scores
  - Average of pockets with a druggability score > 0.5 * max similarity

Druggability Score

Average of all these components

TargetDB explained: List view – information categories

Chemistry

- **BindingDB potent ligands**
- **BindingDB ligands in phase 2 clinic**
- **ChEMBL potent ligands**
- **ChEMBL selective ligands**

- BindingDB potent cpds (Score = log(count), normalized (0->1))
  - Compounds with activity < 100 nM
- ChEMBL potent cpds (Score = log(count), normalized (0->1))
  - Compounds with activity < 100 nM

- Quality
  - Score = 1
    - Count of BindingDB ligands that are labelled phase 2 > 0
  - Score = 0.8
    - Count of ChEMBL ligands with great selectivity > 0
  - Score = 0.7
    - Count of ChEMBL ligands with good selectivity > 0
  - Score = 0.6
    - Count of ChEMBL ligands with moderate selectivity > 0
  - Score = 0.3
    - Count of potent ChEMBL ligands > 0

      OR
    - Count of potent BindingDB ligands > 0
  - Score = 0
    - All of the above not met

Chemistry Score

Average of all these components

TargetDB explained: List view – information categories

Biology

- **Protein expression levels**
- **Number of antibodies**
- **Variants**
- **Mutants**
- **Mice genotypes**
- **KEGG/Reactome**

- Protein expression levels (bio_EScore = 1 if available)
- Number of antibodies (bio_AScore = 1 if count >50)
- Variants (bio_VScore = 1 if count >0)
- Mutants (bio_MScore = 1 if count >0)
- Mice genotypes (bio_GScore = 1 if count >0)
- KEGG/Reactome (bio_PScore = 1 if KEGG+Reactome data =0.5 if KEGG or Reactome = 0 if none available)

Biology Score

Average of all these components

Information

- **JensenLab Pubmed score**

- Log(JensenLab Pubmed score) capped at 12 then normalized (0->1)

Druggability Score

Only component

TargetDB explained: List view – information categories

Disease Link

- **Number of disease areas**
- **Max association score**
- **Diseases count**

- Number of disease area (OpenTargets)
  - Score = 0.5 if count = 1 / Score = 1 if count > 1
- Max association score (OT) – **weight of 2 in average**
  - OpenTargets max association score
- Diseases count uniprot
  - Score = 1 if any
- Diseases count tcrd
  - Score = 1 if any

Disease Score

Weighted average of all these components

Genetic links

- **GWAS association count**
- **Genetic association score**

- Log_{10}(gwas count*10) capped at 2, normalized (0->1)
- OpenTargets MAX genetic association score * normalisation depending on the number of associations with this maximum score
  - Normalisation factor
    - Log_2(count of association with max score) capped at 5 normalized (0->1)
- Count of significant gwas associations (p-value <= 5e-9) / total count of gwas associations
- Avg score for the top10 open-targets associations

Druggability Score

Average of all these components

TargetDB explained: List view – information categories

Safety

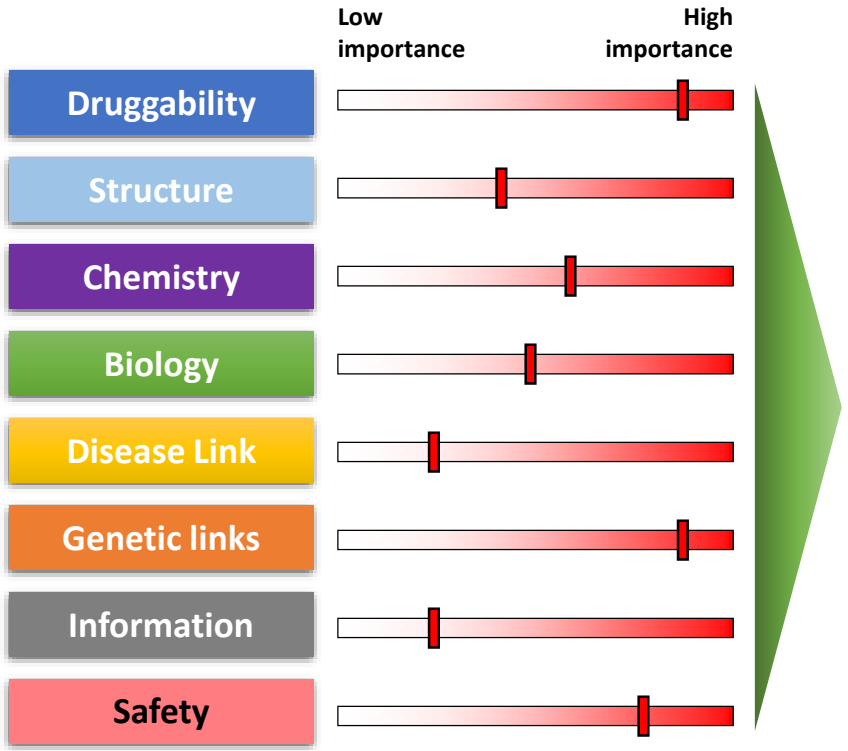
- **Heart protein expression**
- **Liver protein expression**
- **Kidney protein expression**
- **Lethal phenotypes observed in mice**

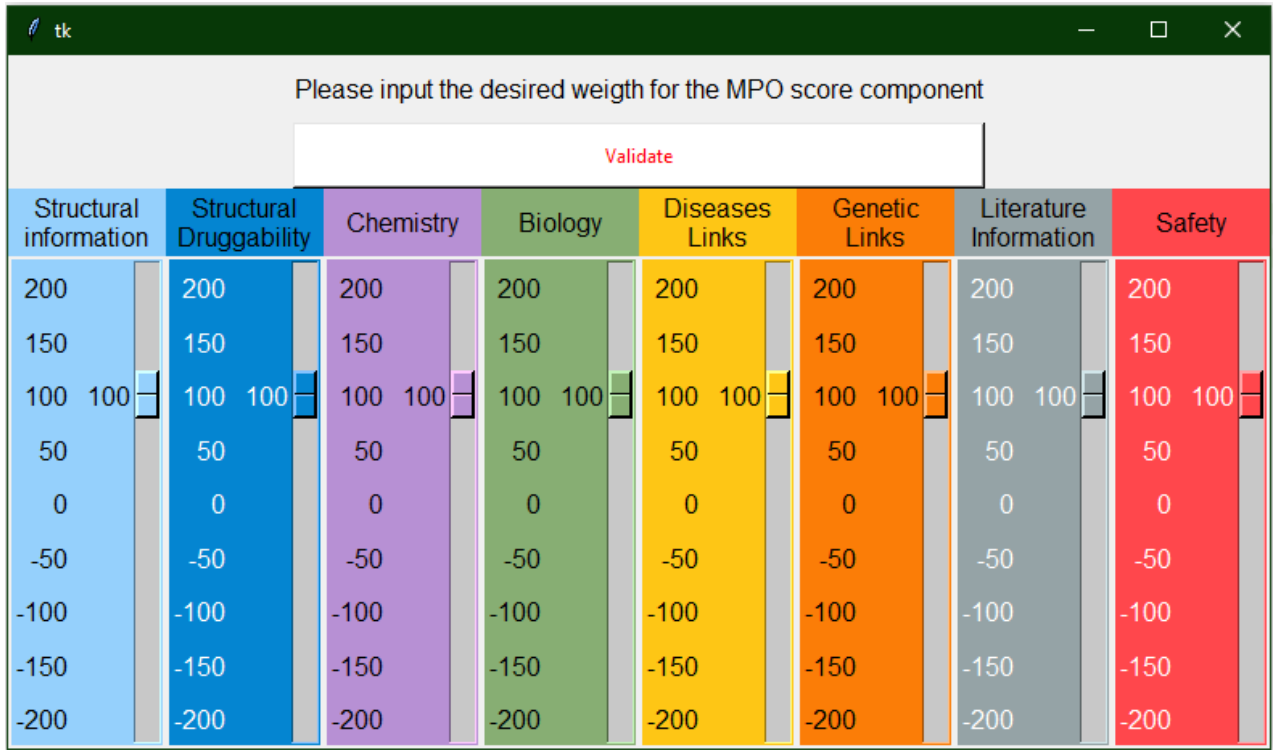
- Log(Number of genotypes) capped at 6, normalized (0->1) + 0.3 if expression data available, capped to 1
- Genotypes (safe_GScore), capped to 1
  - 2 * Count of homozygote genotype with lethal phenotype + count of heterozygote genotype with lethal phenotype – count of heterozygote genotype with normal phenotype – 2 * count of homozygote genotype with normal phenotype
- Expression profile (safe_EScore)
  - Heart protein expression
  - Liver protein expression
  - Kidney protein expression
    - If any of the above is high or 1stddev higher than all tissue average ➔ score = 1
    - Else ➔ score = 0

Safety Score

Average of all these components

TargetDB explained: List view – MPO Score





$$CScoreS = (CScore - 0.5) * 2 \qquad \{-1 \rightarrow 1\}$$

$$W = \frac{Weights}{100} \qquad \{-2 \rightarrow 2\}$$

$$MPO\ Score = \frac{\sum_i CScoreS_i * W_i}{\sum_i |W_i|} \qquad \{-1 \rightarrow 1\}$$

$$MPO\ Score\ (scaled) = \frac{MPO\ Score}{2} + 0.5 \qquad \{0 \rightarrow 1\}$$

TargetDB explained: List view – MPO Score

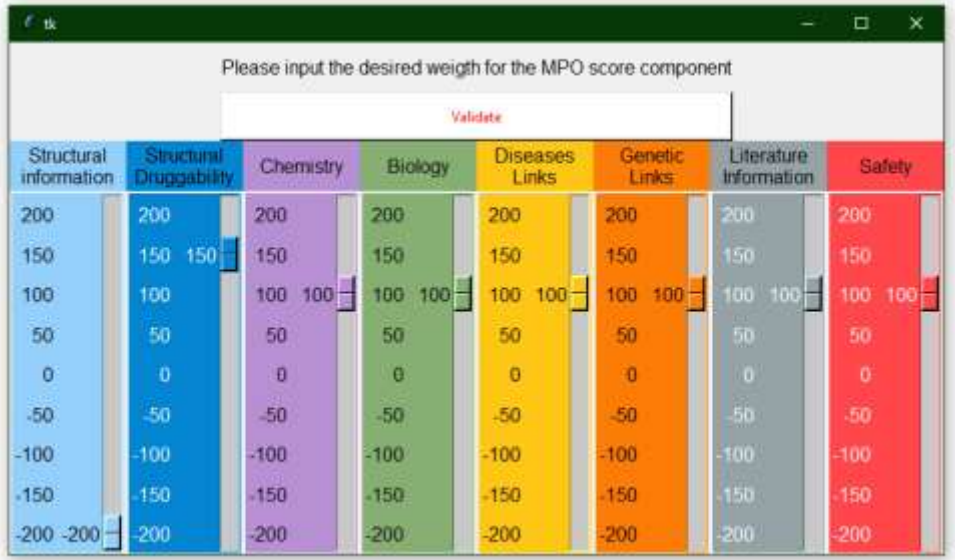
Medicinal chemists

AMP-AD

Agora

95 Targets

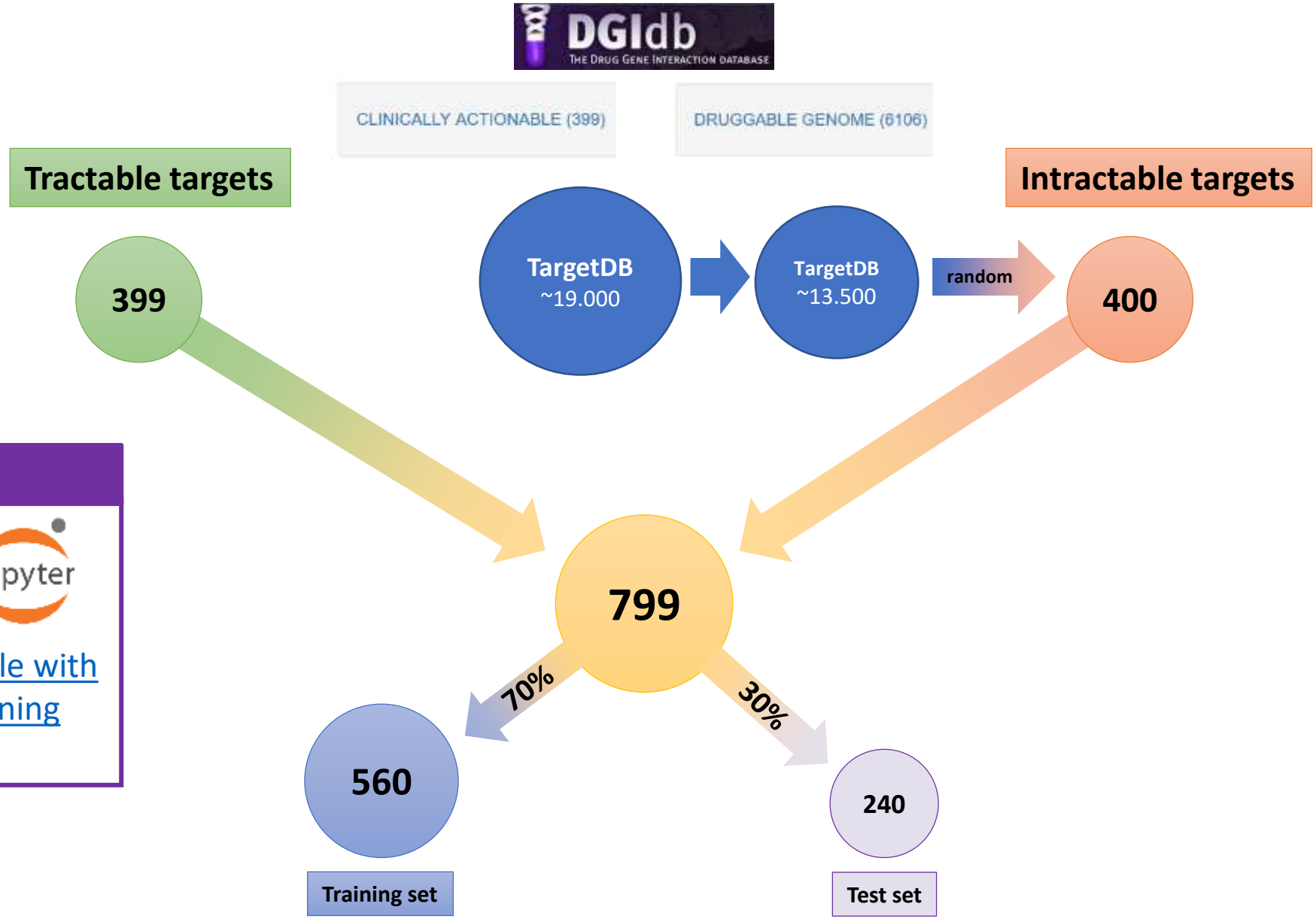
Crystallographers



TOP 10	MPO Score
GRIN2A	0.8
PLEC	0.79
TGFBR2	0.78
PLCG2	0.78
CFH	0.78
TGFB1	0.76
AP2B1	0.76
MSN	0.74
ERBB3	0.74
TREM2	0.73

TOP 10	MPO Score
SGPL1	0.73
ALK	0.69
SYNGAP1	0.68
S1PR1	0.67
NEFL	0.67
CSF1R	0.65
PLCG2	0.63
NR1H4	0.62
GFAP	0.62
PPARA	0.62

TargetDB explained: List view – building a model



DATA

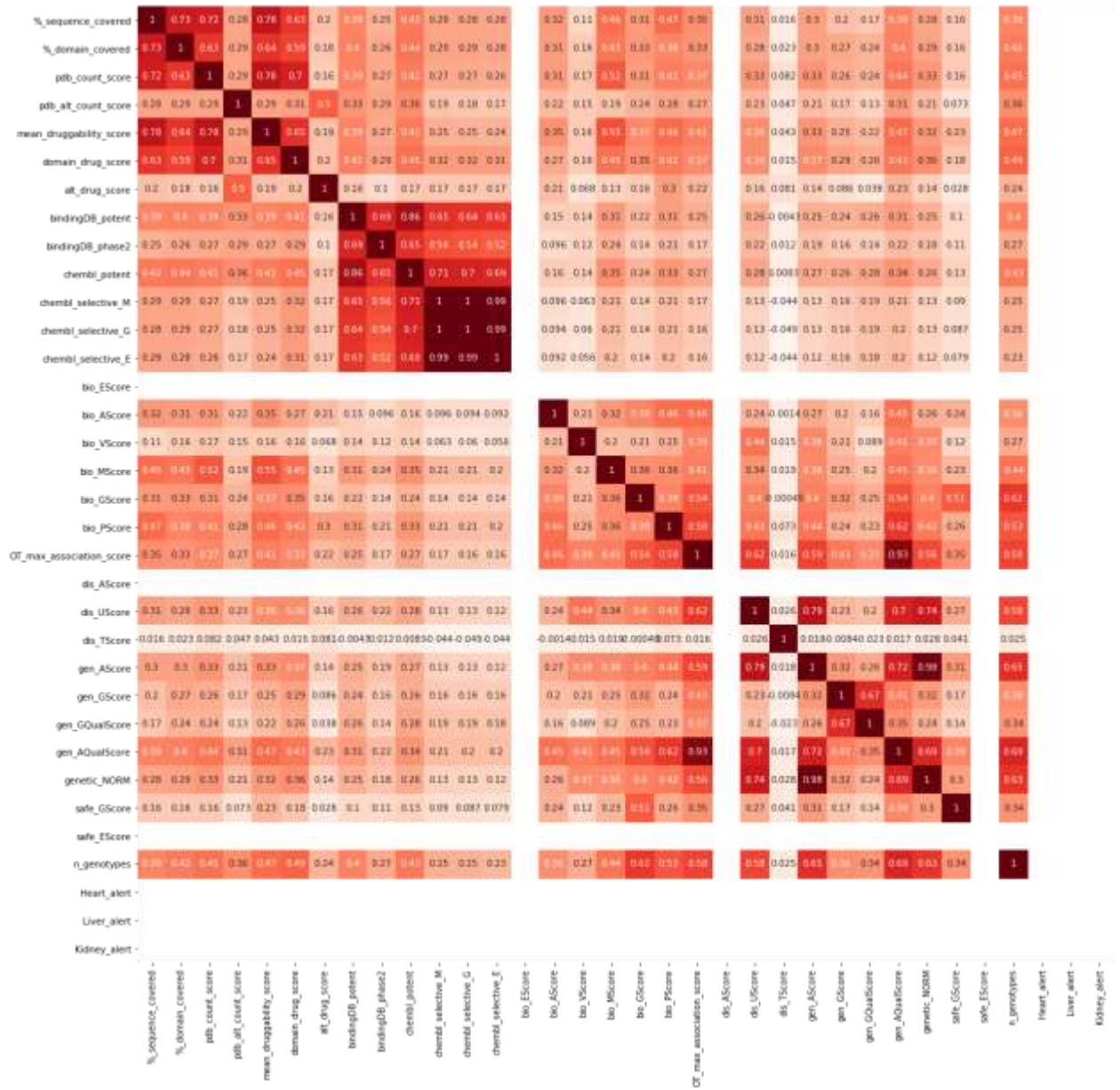
GitHub jupyter

[Jupyter Notebook file with all the model training process]

http://www.dgidb.org/

TargetDB explained: List view – building a model

Step 1: Removed all colinear features



Alzheimer's Research UK | OXFORD DRUG DISCOVERY INSTITUTE

Step 2: Testing different machine learning algorithms

	AU_ROC	accuracy	f1_score	precision	recall
sgd	0.909	0.82	0.821	0.816	0.825
forest	0.921	0.854	0.851	0.864	0.839
svm_prob	0.918	0.841	0.841	0.84	0.843
svm	0.919	0.841	0.841	0.84	0.843
Gaussian_process	0.909	0.827	0.83	0.814	0.846
kNeighbor	0.868	0.802	0.801	0.805	0.796
Naive_bayes	0.914	0.664	0.508	0.951	0.346
QDA	0.863	0.638	0.444	0.953	0.289
AdaBoost	0.92	0.85	0.847	0.863	0.832

Predicted Class

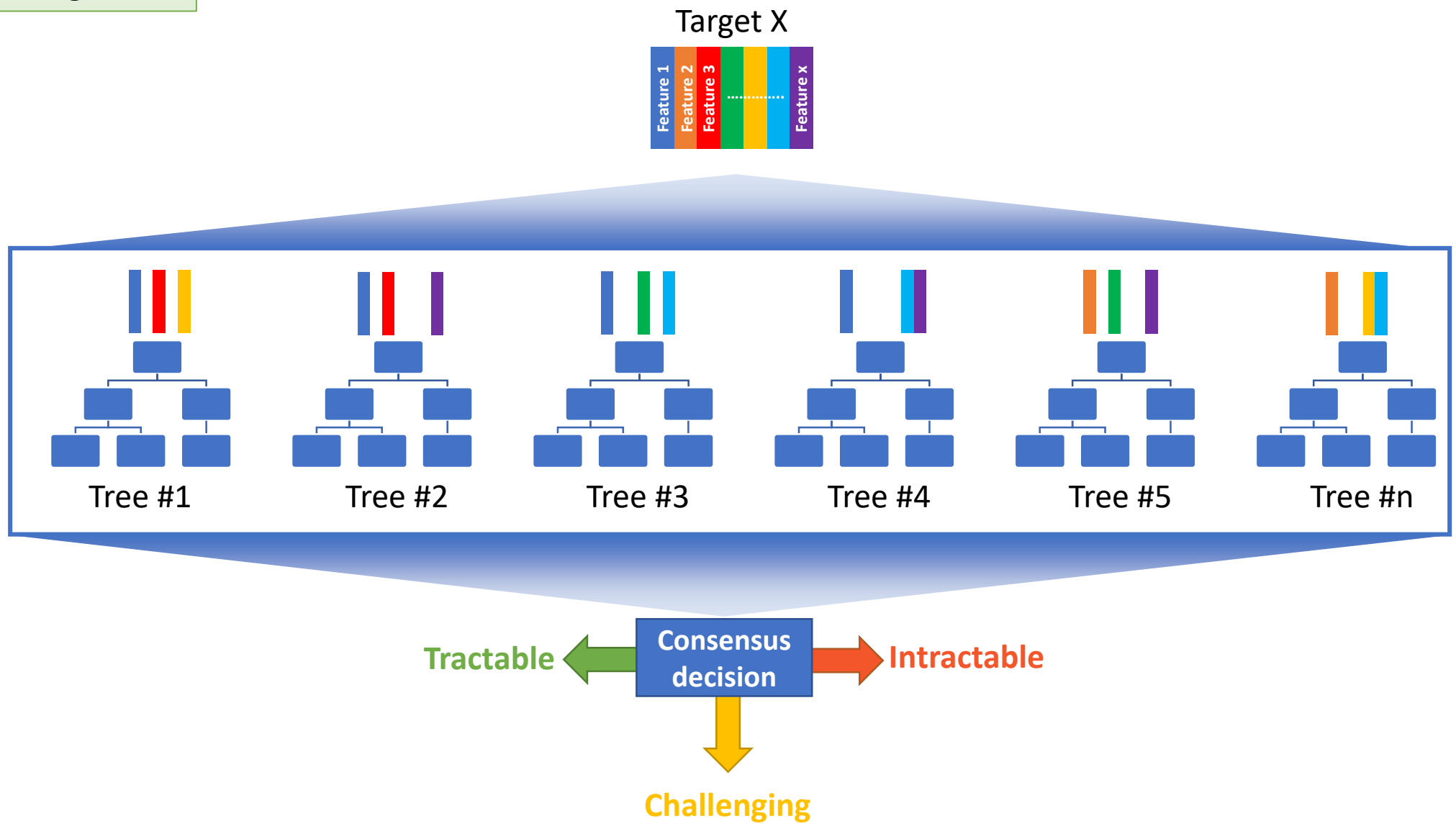
Actual Class	Positive	Negative	Sensitivity = Recall
Positive	True Positive (TP)	False Negative (FN) **Type II Error**	**Sensitivity** $\frac{TP}{(TP+FN)}$
Negative	False Positive (FP) **Type I Error**	True Negative (TN)	**Specificity** $\frac{TN}{(TN+FP)}$
	Precision $\frac{TP}{(TP+FP)}$	**Negative Predictive Value** $\frac{TN}{(TN+FN)}$	**Accuracy** $\frac{TP+TN}{(TP+TN+FP+FN)}$

$\boldsymbol{Fall\text{-}out} - FP\ rate \qquad \frac{FP}{FP+TN}$

$\boldsymbol{F1-score} \qquad 2 * \frac{1}{\frac{1}{precision}+\frac{1}{recall}}$

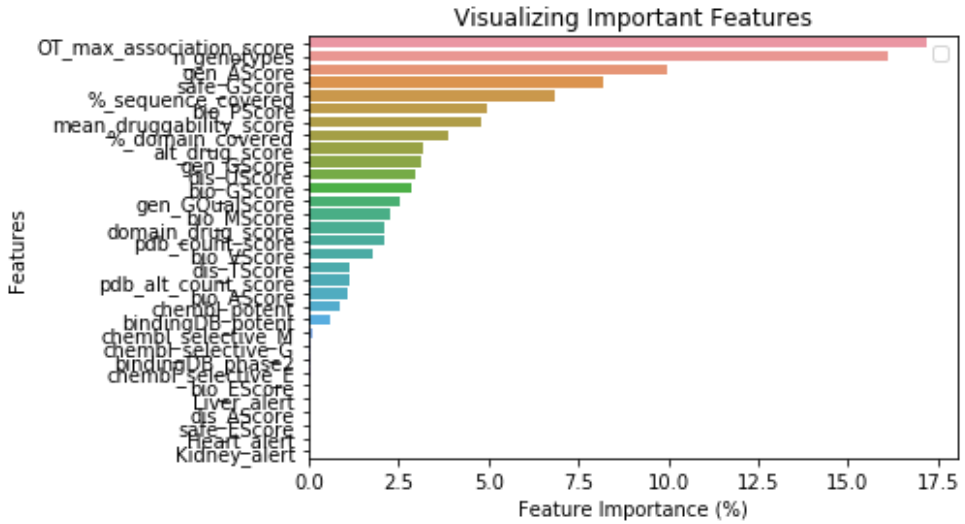
$\boldsymbol{AU-ROC}$ *Area under curved when the TP rate and FP rate are plotted at different threshold values*

Random Forest algorithm

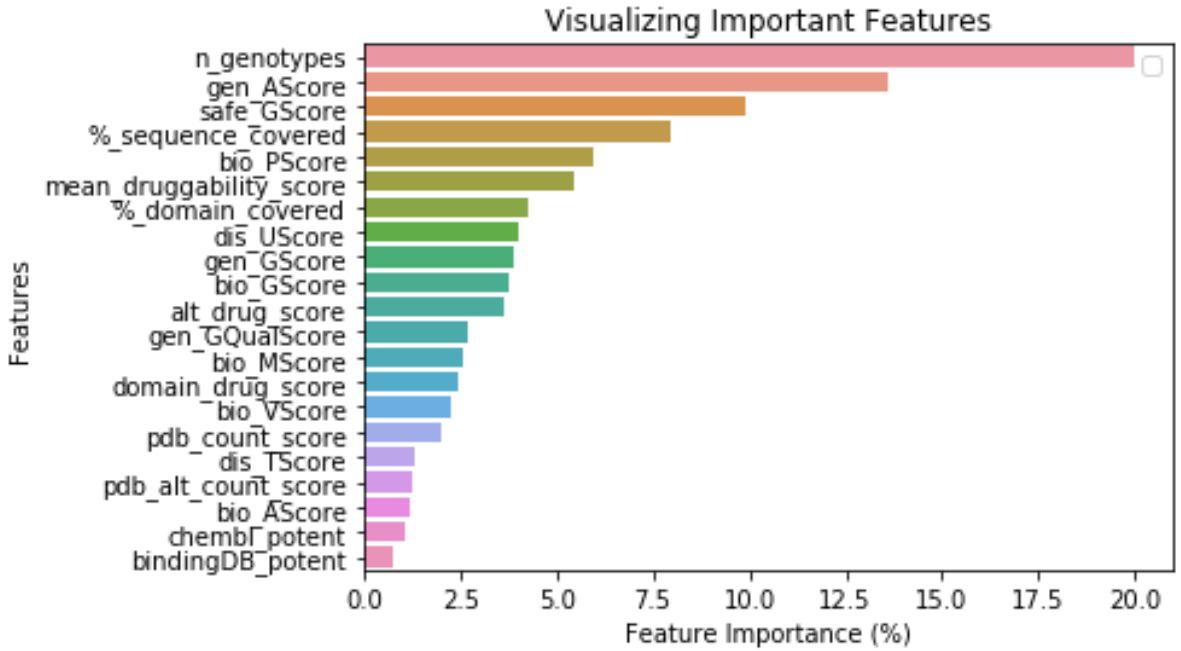


Detailed information on random forests: [Towards data science - Random Forest explained]

Step 3: Looking at feature importance

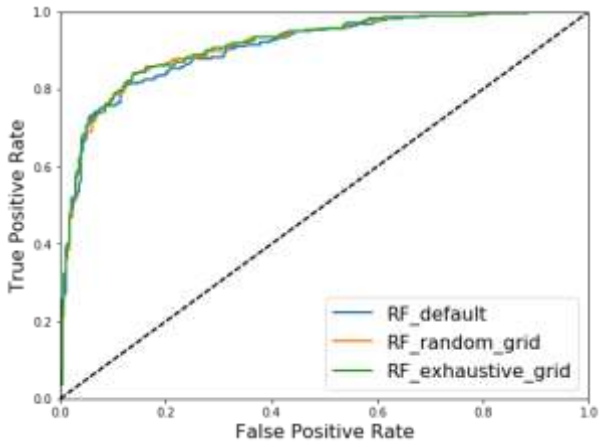


Remove



Step 4: Final tweaking

	AU_ROC	accuracy	f1_score	precision	recall
RF_default	0.927	0.838	0.838	0.835	0.842
RF_random_grid	0.925	0.825	0.825	0.825	0.825
RF_exhaustive_grid	0.924	0.829	0.828	0.832	0.825



TargetDB explained: List view

GENERAL INFO							SCORES												
Target_id	Gene_name	Synonyms	Pharos_class	protein_family	protein_family_detail	Number_isoforms	mpo_score	Tractable	Tractability_probability	In_training_set	structure_info_score [mpo coeff= 1.0]	structural_drug_score [mpo coeff= 1.5]	chemistry_score [mpo coeff= -1.0]	biology_score [mpo coeff= 1.0]	disease_score [mpo coeff= 1.0]	genetic_score [mpo coeff= 1.5]	information_score [mpo coeff= -1.0]	safety_score [mpo coeff= 0.0]
Q12879	**GRIN2A**	GRIN2A,G	Tclin	IC	IC	2	0.8	Tractable	96.52	Yes	0.84	0.68	0	1	0.8	0.88	0.59	0.68
Q15149	**PLEC**	HD1,Hemi	Tbio			9	0.79	Tractable	96.87	No	0.83	0.82	0	0.75	0.6	0.92	0.49	0.74
P37173	**TGFBR2**	2.7.11.30,T	Tchem	Kinase	Kinase	2	0.78	Tractable	99.5	No	0.95	0.66	0	0.83	0.6	0.97	0.55	0.6
P16885	**PLCG2**	1-phospha	Tchem	Enzyme		0	0.78	Tractable	91.7	No	0.52	1	0	0.83	0.6	0.83	0.46	0.79
P08603	**CFH**	CFH,Comp	Tbio			2	0.78	Tractable	97.85	No	0.75	0.81	0	0.83	0.6	0.99	0.62	0.65
P01137	**TGFB1**	LAP,Laten	Tchem			0	0.76	Tractable	93.77	No	0.98	0.85	0	0.83	0.6	0.72	0.72	0.91
P63010	**AP2B1**	ADTB2, CL	Tbio			3	0.76	Tractable	73.54	No	0.98	0.83	0	0.83	0.6	0.5	0.3	0.57
P26038	**MSN**	Not found	Tbio			0	0.74	Tractable	79.19	No	0.94	0.79	0	1	0.6	0.48	0.5	0.42
P21860	**ERBB3**	2.7.10.1,E	Tchem	Kinase	Kinase	5	0.74	Tractable	95.32	Yes	0.94	0.9	0.42	0.92	0.6	0.73	0.57	0.82
Q9NZC2	**TREM2**	Not found	Tbio			3	0.73	Tractable	87.92	No	0.67	0.63	0	1	0.6	0.76	0.49	0.6

TargetDB explained: List view

		LITTERATURE/PATENT INFORMATION									
Target_id	Gene_name	EBI Total Patent Count	JensenLab PubMed Score	NCBI Gene PubMed Count	PubTator Score	total_patent_count	year_max_patents	count_patents_max_year	novelty_score	total # publications	number of Dementia publications
Q12879	**GRIN2A**	14768	1197.83	172	508.67	14768	2013	1341	7.53E-04	267	5
Q15149	**PLEC**		359.01	117	262.31				2.92E-03	364	0
P37173	**TGFBR2**	16731	776.68	464	653.46	16731	2012	2478	1.29E-03	5454	15
P16885	**PLCG2**		250.87	113	124.53				3.71E-03	127	5
P08603	**CFH**		1613.64	744	1744.96				6.08E-04	1583	17
P01137	**TGFB1**		5796.97	4042	16789.31				1.40E-04	7290	24
P63010	**AP2B1**		36.27	52	24.78				1.12E-02	28	0
P26038	**MSN**		381.49	182	963.22				2.31E-03	11257	123
P21860	**ERBB3**	263248	961.8	440	864.45	263248	2014	40081	9.76E-04	2015	3
Q9NZC2	**TREM2**		352.94	151	243.95				2.96E-03	776	302

TargetDB explained: List view

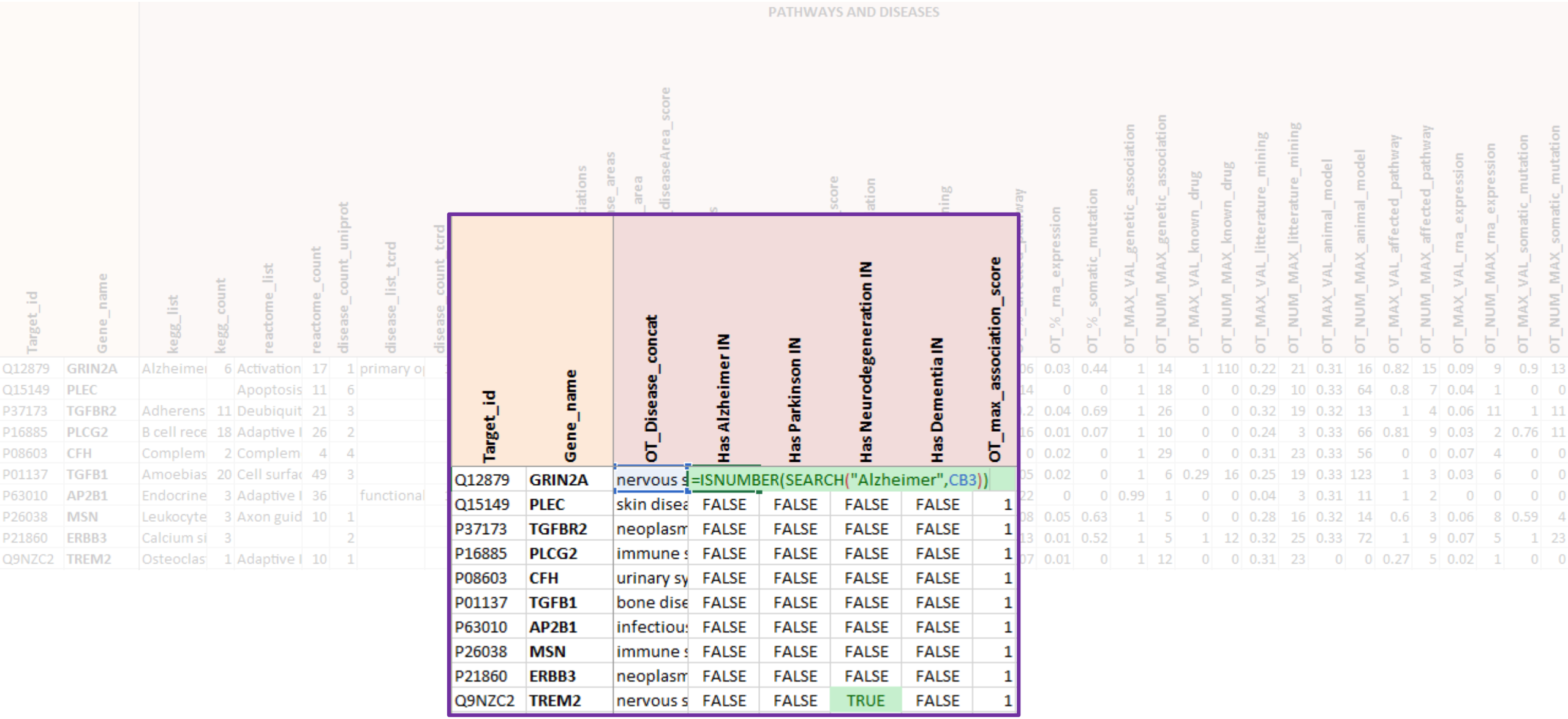
Alzheimer's Research UK | University of Oxford
DRUG DISCOVERY INSTITUTE

		BIOLOGY																																						
Target_id	Gene_name	Brain	Adipose & soft tissue	Bone marrow & lymphoid tissues	Endocrine tissues	Female tissues	Gastrointestinal tract	Kidney & urinary bladder	Liver & gallbladder	Lung	Male tissues	Muscle tissues	Pancreas	Proximal digestive tract	Skin	Expression_Selectivity	tissue_max_expression	expression_max_tissue	EXP_LVL_AVG	EXP_LVL_STDDEV	Heart_alert	Heart_value	Liver_alert	Liver_value	Kidney_alert	Kidney_value	variants_count	mutants_count	gwas_count	number_of_genotypes	phenotypes_heterozygotes_lethal_count	phenotypes_homozygotes_lethal_count	phenotypes_heterozygotes_normal_count	phenotypes_homozygotes_normal_count	Ab Count	MAb Count
Q12879	GRIN2A	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0.35	Brain	2	0.21	0.58	FALSE		FALSE		FALSE		87	1	23	10					715	208
Q15149	PLEC	1.5	1.3	1.8	1.7	2.1	2.2	2.5	2.5	2.3	2	2	2	1.7	2	2.5	Kidney & urinary bladder	2.5	1.97	0.35	FALSE		FALSE		FALSE		19		43	14	1	3		1	180	66
P37173	TGFBR2																										38	1	10	37	9	3	1	2	734	62
P16885	PLCG2	1	0	3	1	1	1.3	1	2	1	1	1	0	0	3	2.15	Bone marrow & lymphoid tissues	3	1.16	0.95	FALSE		FALSE		FALSE		7		16	19	3	2			932	180
P08603	CFH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	Adipose & soft tissue	0	0.00	0.00	FALSE		FALSE		FALSE		59	7	136	8	1	1	1		855	368
P01137	TGFB1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	Adipose & soft tissue	0	0.00	0.00	FALSE		FALSE		FALSE		13	11	7	38	11	18			1690	795
P63010	AP2B1	3	3	1.7	2	1.6	1.6	2	3	1	2	1	3	1.3	2	2.44	Adipose & soft tissue	3	2.01	0.73	FALSE		FALSE		FALSE			9	20	5		1			242	18
P26038	MSN	1.8	0	2.6	1.3	1.3	2	1.5	2	2	1.5	1	1	2	2	2.31	Bone marrow & lymphoid tissues	2.6	1.57	0.64	FALSE		FALSE		FALSE		1	4	1	2				2	925	461
P21860	ERBB3	2.8	2.7	2.2	2.7	2.2	3	2.5	2.5	2.7	3	2	3	2.3	2.5	2.52	Gastrointestinal tract	3	2.58	0.32	FALSE		FALSE		FALSE		15	2	11	22	1	10	2	1	2081	750
Q9NZC2	TREM2	2	1.7	2	2	1.8	2	2	1.5	2	2	1.5	1	1	0	2.4	Bone marrow & lymphoid tissues	2	1.61	0.59	FALSE		FALSE		FALSE		29	7	4	6					521	177

TargetDB explained: List view

		PATHWAYS AND DISEASES																																					
Target_id	Gene_name	kegg_list	kegg_count	reactome_list	reactome_count	disease_count_uniprot	disease_list_tcrd	disease_count_tcrd	max_disease_score	name_max_disease	OT_number_of_associations	OT_number_of_disease_areas	OT_list_max_disease_area	OT_max_association_diseaseArea_score	OT_list_max_diseases	OT_TOP10_diseases	OT_max_association_score	OT_%_genetic_association	OT_%_known_drug	OT_%_litterature_mining	OT_%_animal_model	OT_%_affected_pathway	OT_%_rna_expression	OT_%_somatic_mutation	OT_MAX_VAL_genetic_association	OT_NUM_MAX_genetic_association	OT_MAX_VAL_known_drug	OT_NUM_MAX_known_drug	OT_MAX_VAL_litterature_mining	OT_NUM_MAX_litterature_mining	OT_MAX_VAL_animal_model	OT_NUM_MAX_animal_model	OT_MAX_VAL_affected_pathway	OT_NUM_MAX_affected_pathway	OT_MAX_VAL_rna_expression	OT_NUM_MAX_rna_expression	OT_MAX_VAL_somatic_mutation	OT_NUM_MAX_somatic_mutation
Q12879	**GRIN2A**	Alzheimer	6	Activation	17	1	primary o	1	2.31	primary o	331				nervous s	nervous s	1	0.15	0.5	0.35	0.05	0.06	0.03	0.44	1	14	1	110	0.22	21	0.31	16	0.82	15	0.09	9	0.9	13
Q15149	**PLEC**			Apoptosis	11	6					207				skin disea	skin disea	1	0.17	0	0.4	0.63	0.14	0	0	1	18	0	0	0.29	10	0.33	64	0.8	7	0.04	1	0	0
P37173	**TGFBR2**	Adherens	11	Deubiquit	21	3					255				neoplasm	neoplasm	1	0.2	0	0.64	0.13	0.2	0.04	0.69	1	26	0	0	0.32	19	0.32	13	1	4	0.06	11	1	11
P16885	**PLCG2**	B cell rece	18	Adaptive I	26	2					152				immune s	immune s	1	0.22	0	0.33	0.69	0.16	0.01	0.07	1	10	0	0	0.24	3	0.33	66	0.81	9	0.03	2	0.76	11
P08603	**CFH**	Complem	2	Complem	4	4					176				urinary sy	urinary sy	1	0.35	0	0.59	0.59	0	0.02	0	1	29	0	0	0.31	23	0.33	56	0	0	0.07	4	0	0
P01137	**TGFB1**	Amoebias	20	Cell surfac	49	3					363				bone dise	bone dise	1	0.07	0.1	0.33	0.8	0.05	0.02	0	1	6	0.29	16	0.25	19	0.33	123	1	3	0.03	6	0	0
P63010	**AP2B1**	Endocrine	3	Adaptive I	36		functional	1	1.09	functional	59				infectious	infectious	1	0.19	0	0.05	0.59	0.22	0	0	0.99	1	0	0	0.04	3	0.31	11	1	2	0	0	0	0
P26038	**MSN**	Leukocyte	3	Axon guid	10	1					158				immune s	immune s	1	0.04	0	0.59	0.23	0.08	0.05	0.63	1	5	0	0	0.28	16	0.32	14	0.6	3	0.06	8	0.59	4
P21860	**ERBB3**	Calcium si	3			2					334				neoplasm	neoplasm	1	0.07	0.29	0.43	0.38	0.13	0.01	0.52	1	5	1	12	0.32	25	0.33	72	1	9	0.07	5	1	23
Q9NZC2	**TREM2**	Osteoclas	1	Adaptive I	10	1					72				nervous s	nervous s	1	0.22	0	0.94	0	0.07	0.01	0	1	12	0	0	0.31	23	0	0	0.27	5	0.02	1	0	0

TargetDB explained: List view



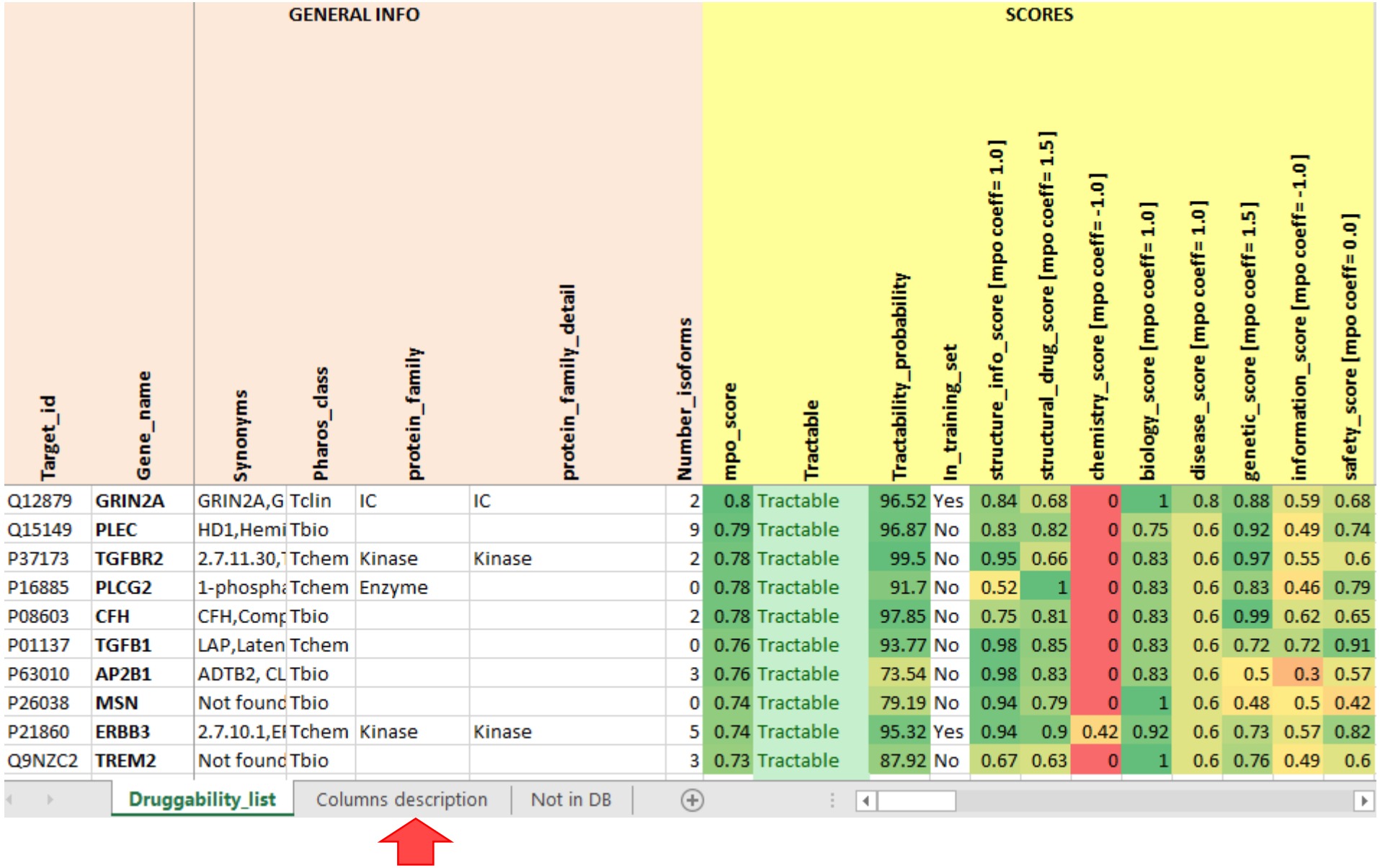
TargetDB explained: List view

		STRUCTURAL INFORMATION																													
Target_id	Gene_name	PDB_total_count	PDB_with_Ligand_count	%_sequence_covered	%_domain_covered	PDB_sites_tractable_count	PDB_sites_druggable_count	PDB_blast_close_count	PDB_blast_max_similarity	domains_count	domain_tractable	domain_druggable	mean_druggability_score	stddev_druggability_score	mean_area	mean_volume	mean_fraction_apolar	mean_pocket_score	pdb_with_druggable_pocket	druggable_pockets_total	mean_alt_druggability_score	alt_stddev_druggability_score	mean_alt_area	mean_alt_volume	mean_alt_fraction_apolar	mean_alt_pocket_score	mean_alt_similarity	max_alt_similarity	alt_pdb_with_druggable_pocket	alt_druggable_pockets_total
Q12879	**GRIN2A**	15	1	59%	76%	1		51	100	4	1	0	0.81	0.16	287.23	1023	77.92	-0.02	14	82	0.79	0.16	420.19	1263.38	78.75	-0.4	92.65	100	10	90
Q15149	**PLEC**	14	1	32%	100%	5	2	40	100	8	1	1	0.63	0.13	147.38	472.73	74.67	0.19	5	6										
P37173	**TGFBR2**	14	4	81%	100%	2		13	79	3	1	0	0.81	0.11	173.83	905.99	56.67	0.47	7	7										
P16885	**PLCG2**	3	1	9%	0%	3	1	16	100	12	1	1	0.99		170.2	1080.1	68.3	0.72	1	1										
P08603	**CFH**	46	8	99%	100%	8	4	4	98.4	21	1	1	0.7	0.13	233.28	618.04	87.44	-0.1	17	39	0.74		212.5	1000	57.4	0.35	98.4	98.4	1	1
P01137	**TGFB1**	7	2	93%	100%	4	1	8	100	2	1	1	0.8	0.17	242.35	823.27	68.77	0.19	4	6	0.75	0.14	286.5	1017.38	69.55	0.08	98.2	100	2	4
P63010	**AP2B1**	13	3	95%		8	5	6	100		1	1	0.76	0.14	209.39	741.49	65.77	0.05	10	19	0.74	0.13	194.99	682.37	78.51	0.05	97.78	98.4	5	19
P26038	**MSN**	3		76%	100%	3	2	20	99.7	1	1	1	0.68	0.07	245.12	875.75	68.95	0.17	2	4	0.71	0.16	180.34	708.84	66.11	0.23	96.76	99.7	6	7
P21860	**ERBB3**	14	3	76%	100%	5	2	74	99.7	5	1	1	0.79	0.13	142.02	692.37	68.17	0.32	11	18	0.9		166.3	1274.2	61.7	0.74	99.7	99.7	1	1
Q9NZC2	**TREM2**	4		68%	100%					1			0.63	0.13	235.37	886.53	56.93	0.05	2	3										

TargetDB explained: List view

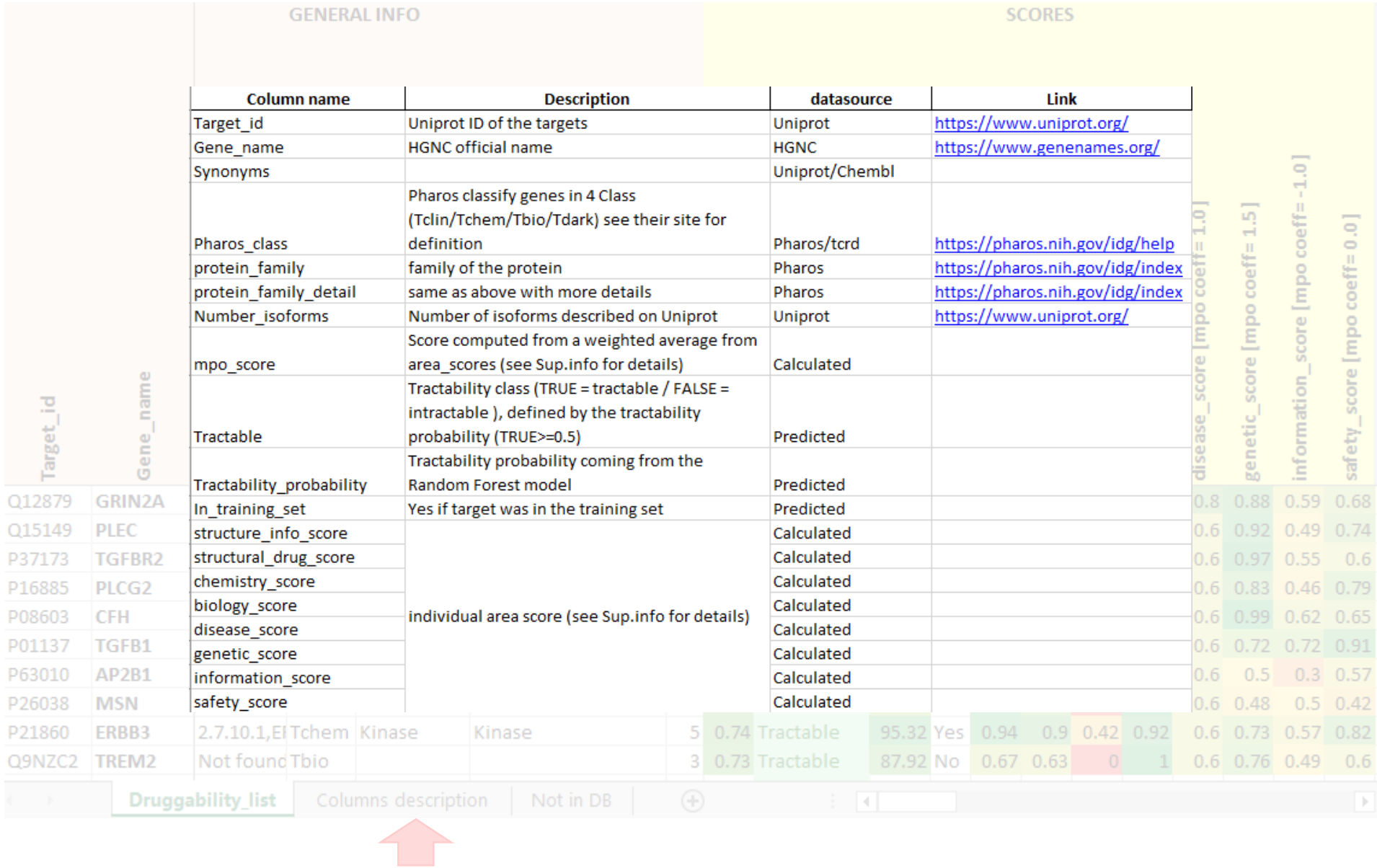
		CHEMISTRY									
Target_id	Gene_name	BindingDB_count	BindingDB_potent_count	BindingDB_potent_phase2_count	ChEMBL_bioactives_count	ChEMBL_bioactives_potent_count	ChEMBL_bioactives_moderate_selectivity_count	ChEMBL_bioactives_good_selectivity_count	ChEMBL_bioactives_great_selectivity_count	commercial_total	commercial_potent_total
Q12879	GRIN2A				11					1	1
Q15149	PLEC	234			78					37	
P37173	TGFBR2	305			73	1				23	
P16885	PLCG2				14					9	
P08603	CFH				1						
P01137	TGFB1				7					1	
P63010	AP2B1										
P26038	MSN										
P21860	ERBB3				24	3				15	3
Q9NZC2	TREM2										

TargetDB explained: List view



		GENERAL INFO					SCORES												
Target_id	Gene_name	Synonyms	Pharos_class	protein_family	protein_family_detail	Number_isoforms	mpo_score	Tractable	Tractability_probability	In_training_set	structure_info_score [mpo coeff= 1.0]	structural_drug_score [mpo coeff= 1.5]	chemistry_score [mpo coeff= -1.0]	biology_score [mpo coeff= 1.0]	disease_score [mpo coeff= 1.0]	genetic_score [mpo coeff= 1.5]	information_score [mpo coeff= -1.0]	safety_score [mpo coeff= 0.0]
Q12879	**GRIN2A**	GRIN2A,G	Tclin	IC	IC	2	0.8	Tractable	96.52	Yes	0.84	0.68	0	1	0.8	0.88	0.59	0.68
Q15149	**PLEC**	HD1,Hemi	Tbio			9	0.79	Tractable	96.87	No	0.83	0.82	0	0.75	0.6	0.92	0.49	0.74
P37173	**TGFBR2**	2.7.11.30,T	Tchem	Kinase	Kinase	2	0.78	Tractable	99.5	No	0.95	0.66	0	0.83	0.6	0.97	0.55	0.6
P16885	**PLCG2**	1-phospha	Tchem	Enzyme		0	0.78	Tractable	91.7	No	0.52	1	0	0.83	0.6	0.83	0.46	0.79
P08603	**CFH**	CFH,Comp	Tbio			2	0.78	Tractable	97.85	No	0.75	0.81	0	0.83	0.6	0.99	0.62	0.65
P01137	**TGFB1**	LAP,Laten	Tchem			0	0.76	Tractable	93.77	No	0.98	0.85	0	0.83	0.6	0.72	0.72	0.91
P63010	**AP2B1**	ADTB2, CL	Tbio			3	0.76	Tractable	73.54	No	0.98	0.83	0	0.83	0.6	0.5	0.3	0.57
P26038	**MSN**	Not found	Tbio			0	0.74	Tractable	79.19	No	0.94	0.79	0	1	0.6	0.48	0.5	0.42
P21860	**ERBB3**	2.7.10.1,E	Tchem	Kinase	Kinase	5	0.74	Tractable	95.32	Yes	0.94	0.9	0.42	0.92	0.6	0.73	0.57	0.82
Q9NZC2	**TREM2**	Not found	Tbio			3	0.73	Tractable	87.92	No	0.67	0.63	0	1	0.6	0.76	0.49	0.6

TargetDB explained: List view



Column name	Description	datasource	Link
Target_id	Uniprot ID of the targets	Uniprot	https://www.uniprot.org/
Gene_name	HGNC official name	HGNC	https://www.genenames.org/
Synonyms		Uniprot/Chembl	
Pharos_class	Pharos classify genes in 4 Class (Tclin/Tchem/Tbio/Tdark) see their site for definition	Pharos/tcrd	https://pharos.nih.gov/idg/help
protein_family	family of the protein	Pharos	https://pharos.nih.gov/idg/index
protein_family_detail	same as above with more details	Pharos	https://pharos.nih.gov/idg/index
Number_isoforms	Number of isoforms described on Uniprot	Uniprot	https://www.uniprot.org/
mpo_score	Score computed from a weighted average from area_scores (see Sup.info for details)	Calculated	
Tractable	Tractability class (TRUE = tractable / FALSE = intractable), defined by the tractability probability (TRUE>=0.5)	Predicted	
Tractability_probability	Tractability probability coming from the Random Forest model	Predicted	
In_training_set	Yes if target was in the training set	Predicted	
structure_info_score	individual area score (see Sup.info for details)	Calculated	
structural_drug_score		Calculated	
chemistry_score		Calculated	
biology_score		Calculated	
disease_score		Calculated	
genetic_score		Calculated	
information_score		Calculated	
safety_score		Calculated	

TargetDB: Test cases

AMP-AD

Agora

95 Targets

						SCORES													
Target_id	Gene_name	Pharos_class	protein_family	protein_family_detail	Number_isoforms	mpo_score	Tractable	Tractability_probability	In_training_set	structure_info_score [mpo coeff= 1.0]	structural_drug_score [mpo coeff= 1.5]	chemistry_score [mpo coeff= -1.0]	biology_score [mpo coeff= 1.0]	disease_score [mpo coeff= 1.0]	genetic_score [mpo coeff= 1.5]	information_score [mpo coeff= -1.0]	safety_score [mpo coeff= 0.0]
Q12879	GRIN2A	Tclin	IC	IC	2	0.8	Tractable	96.85	Yes	0.84	0.68	0	1	0.8	0.88	0.59	0.68
Q15149	PLEC	Tbio			9	0.79	Tractable	97.47	No	0.83	0.82	0	0.75	0.6	0.92	0.49	0.74
P37173	TGFBR2	Tchem	Kinase	Kinase	2	0.78	Tractable	99.48	No	0.95	0.66	0	0.83	0.6	0.97	0.55	0.6
P16885	PLCG2	Tchem	Enzyme		0	0.78	Tractable	90.12	No	0.52	1	0	0.83	0.6	0.83	0.46	0.79
P08603	CFH	Tbio			2	0.78	Tractable	98.76	No	0.75	0.81	0	0.83	0.6	0.99	0.62	0.65
P01137	TGFB1	Tchem			0	0.76	Tractable	94.5	No	0.98	0.85	0	0.83	0.6	0.72	0.72	0.91
P63010	AP2B1	Tbio			3	0.76	Tractable	71.55	No	0.98	0.83	0	0.83	0.6	0.5	0.3	0.57
P26038	MSN	Tbio			0	0.74	Tractable	79.97	No	0.94	0.79	0	1	0.6	0.48	0.5	0.42
P21860	ERBB3	Tchem	Kinase	Kinase	5	0.74	Tractable	95.5	Yes	0.94	0.9	0.42	0.92	0.6	0.73	0.57	0.82
Q9NZC2	TREM2	Tbio			3	0.73	Tractable	88.52	No	0.67	0.63	0	1	0.6	0.76	0.49	0.6
P47870	GABRB2	Tclin	IC	IC	4	0.73	Tractable	74.13	No	0.59	0.72	0	1	0.6	0.56	0.3	0.48
Q13572	ITPK1	Tbio	Kinase	Kinase	2	0.73	Tractable	65.16	No	0.7	0.95	0	0.83	0.4	0.5	0.27	0.3
Q06413	MEF2C	Tbio	TF	TF	6	0.72	Tractable	87.23	No	0.5	0.63	0	1	0.6	0.82	0.48	0.78
P07333	CSF1R	Tclin	Kinase	Kinase	2	0.72	Tractable	94.01	Yes	0.98	0.95	1	1	0.8	0.77	0.59	0.53
P04792	HSPB1	Tchem			0	0.71	Tractable	75.65	No	1	0.34	0	1	0.6	0.74	0.58	0.48
Q9UQB8	BAIAP2	Tbio			6	0.71	Tractable	84.59	No	0.67	0.88	0	1	0.31	0.5	0.38	0.53
P14867	GABRA1	Tclin	IC	IC	0	0.71	Tractable	99.14	No	1	0.7	0.72	1	0.6	0.74	0.35	0.76
O43157	PLXNB1	Tbio			3	0.69	Tractable	76.21	No	0.7	0.83	0	1	0.32	0.39	0.36	0.65
P20020	ATP2B1	Tbio	Transporter	cation_transpo	6	0.69	Tractable	84.91	No	0.33	0.74	0	0.83	0.4	0.8	0.39	0.65
Q13651	IL10RA	Tbio			0	0.68	Tractable	78.45	No	0.49	0.93	0	0.67	0.6	0.46	0.39	0.3

E3 Ligases

372

>25% In Dark Proteome

>60% without structure

Tractable	Challenging	Intractable
47 (13%)	57 (15%)	268 (72%)

Disease link

AD	PD
4	1
2 2	1

Literature

64 targets with dementia linked literature

TargetDB: Conclusions

Scope and limitations

TargetDB is …

… quick

… good at comparing targets

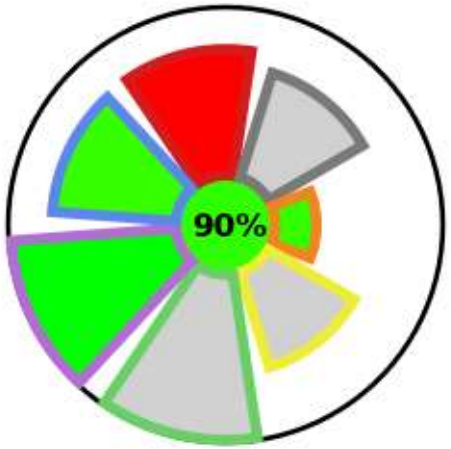
… giving you a good overview on targets

TargetDB is **NOT** …

… perfect

… exhaustive and 100% accurate

… completely replacing reading papers



Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

TargetDB: A target information aggregation tool and tractability predictor

Stephane De Cesco, John B. Davis, Paul E. Brennan

doi: https://doi.org/10.1101/2020.04.21.052878 Link here

This article is a preprint and has not been certified by peer review [what does this mean?].