

Open Targets: integrating multiomics data for drug discovery

Part of “Introduction to Multiomics Data Integration”

February 22nd 2018, EMBL-EBI

Denise Carvalho-Silva, PhD

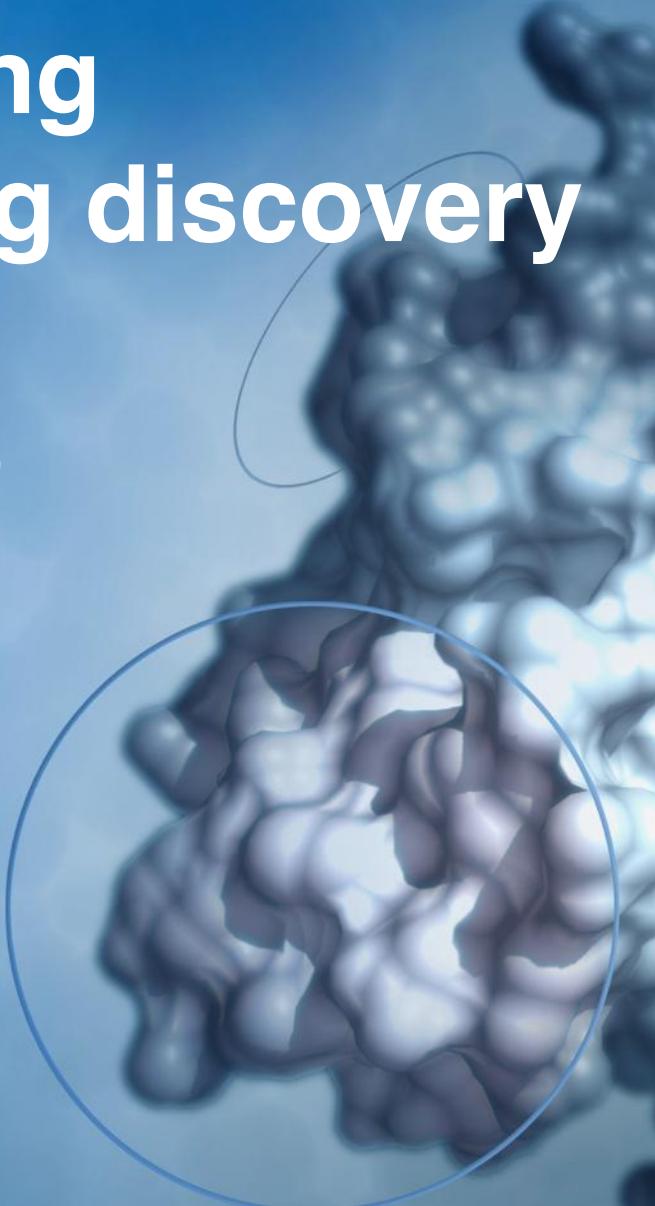
Scientific Outreach Lead

Open Targets / EMBL-EBI

Wellcome Genome Campus, United Kingdom



Open Targets



Objectives

Why should we integrate



Open Targets Partnership

How to navigate the
Open Targets Platform

Official schedule

Overview

Train at EMBL-EBI

Train outside EMBL-EBI

Train online

Webinars

Ab

Day 3 - Thursday 22 February 2018

09:00 - 10:00	Open Targets	Denise Carvalho-Silva
10:00 - 10:30	Tea/Coffee Break	
10:30 - 11:30	Open Targets - hands on example	Denise Carvalho-Silva

In these 2.5 hours

- Introduction
- Live demos
- Wrap up



10:30-11:00

In these 2.5 hours

- Introduction
- Live demos
- Wrap up

An example



Where to start?



Where can we find more about the disease?

Where can we get the drugs currently on clinical trials?

Where can we find information on the current targets (i.e genes, proteins)?

Where can we get the latest scientific literature on this disease from?

About 71,200,000 results (0.59 seconds)

Alzheimer's Association: Alzheimer's Disease and Dementia

<https://www.alz.org/> ▾

Alzheimer's Association national site – information on Alzheimer's disease and dementia symptoms, diagnosis, stages, treatment, care and support resources.

Alzheimer's Disease & Dementia | Alzheimer's Association

https://www.alz.org/alzheimers_disease_what_is_alzheimers.asp ▾

Alzheimer's and dementia basics. Alzheimer's is the most common form of dementia, a general term for memory loss and other cognitive abilities serious enough to interfere with daily life.

Alzheimer's disease accounts for 60 to 80 percent of dementia cases.

[Stages](#) · [Younger/Early Onset](#) · [Risk Factors](#) · [2016 Facts and Figures](#)

Know the 10 Signs of Alzheimer's Disease - Alzheimer's Association

<https://www.alz.org/10-signs-symptoms-alzheimers-dementia.asp> ▾

10 Early Signs and Symptoms of Alzheimer's. Memory loss that disrupts daily life may be a symptom of Alzheimer's or another dementia. Alzheimer's is a brain disease that causes a slow decline in memory, thinking and reasoning skills.

Alzheimer's disease - Wikipedia

https://en.wikipedia.org/wiki/Alzheimer%27s_disease ▾

Alzheimer's disease (AD), also referred to simply as Alzheimer's, is a chronic neurodegenerative disease that usually starts slowly and worsens over time. It is the cause of 60% to 70% of cases of dementia. The most common early symptom is difficulty in remembering recent events (short-term memory loss). As the disease ...

Causes: Poorly understood **Symptoms:** Difficulty in remembering recent ev...

Usual onset: Over 65 years old **Deaths:** 1.9 million (2015)

Alzheimer's disease - Alzheimer's Society

https://www.alzheimers.org.uk/info/20007/types_of_dementia/.../alzheimers_disease ▾

Alzheimer's disease is the most common cause of dementia. The word dementia describes a set of symptoms that can include memory loss and difficulties with thinking, problem-solving or language. These symptoms occur when the brain is damaged by certain diseases, including Alzheimer's disease. This page describes ...

Alzheimer's disease

Also called: senile dementia

ABOUT

SYMPTOMS

TREATMENTS

Memory loss



A progressive disease that destroys memory and other important mental functions.

Very common

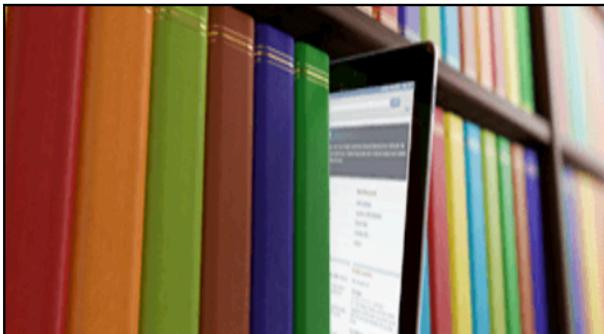
More than 2 million cases per year (Brazil)

Requires a medical diagnosis

Lab tests or imaging not required

Chronic: can last for years or be lifelong

Brain cell connections and the cells themselves degenerate and die,



PubMed

PubMed comprises more than 28 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Search results

Items: 1 to 20 of 135639

<< First < Prev Page 1 of 6782 Next > Last >>

- [CA1 pyramidal neuron gene expression mosaics in the Ts65Dn murine model of Down syndrome and Alzheimer's disease following maternal choline supplementation \(MCS\).](#)

Alldred MJ, Chao HM, Lee SH, Beilin J, Powers BE, Petkova E, Strupp BJ, Ginsberg SD. Hippocampus. 2018 Feb 2. doi: 10.1002/hipo.22832. [Epub ahead of print]
PMID: 29394516

- [Interpreting Alzheimer's disease polygenic scores.](#)

2. Tan CH, Desikan RS.
Ann Neurol. 2018 Feb 2. doi: 10.1002/ana.25164. [Epub ahead of print] No abstract available.
PMID: 29394507

- [Sleep and Cognitive Decline: A Prospective Non-demented Elderly Cohort Study.](#)

3. Suh SW, Han JW, Lee JR, Byun S, Kwon SJ, Oh SH, Lee KH, Han G, Hong JW, Kwak KP, Kim BJ, Kim SG, Kim JL, Kim TH, Ryu SH, Moon SW, Park JH, Seo J, Youn JC, Lee DY, Lee DW, Lee SB, Lee JJ, Jhoo JH, Kim KW.
Ann Neurol. 2018 Feb 2. doi: 10.1002/ana.25166. [Epub ahead of print]
PMID: 29394505

Sort by:

[Best match](#)

[Most recent](#)

Results by year

Related searches

Titles with your search terms

Find related data

Search details

Recent Activity



WIKIPEDIA
The Free Encyclopedia

Main page

Not logged in Talk Contributions Create account Log in

Article

Talk

Read

View source

View history

Search Wikipedia



Alzheimer's disease

From Wikipedia, the free encyclopedia



"Alzheimer" redirects here. For other uses, see [Alzheimer \(disambiguation\)](#).

In 2015, there were approximately 29.8 million people worldwide with AD.^{[8][2]} It most often begins in people over 65 years of age, although 4% to 5% of cases are [early-onset Alzheimer's](#) which begin before this.^[3] It affects about 6% of people 65 years and older.^[1] In 2015, dementia resulted in about 1.9 million deaths.^[9] It was first described by, and later named after, German psychiatrist and pathologist [Alois Alzheimer](#) in 1906.^[18] In developed countries, AD is one of the most financially costly diseases.^{[19][20]}

Contents [hide]

1 Signs and symptoms

- 1.1 Pre-dementia
- 1.2 Early
- 1.3 Moderate
- 1.4 Advanced

2 Cause

- 2.1 Genetic
- 2.2 Cholinergic hypothesis
- 2.3 Amyloid hypothesis
- 2.4 Tau hypothesis
- 2.5 Other hypotheses

conditions

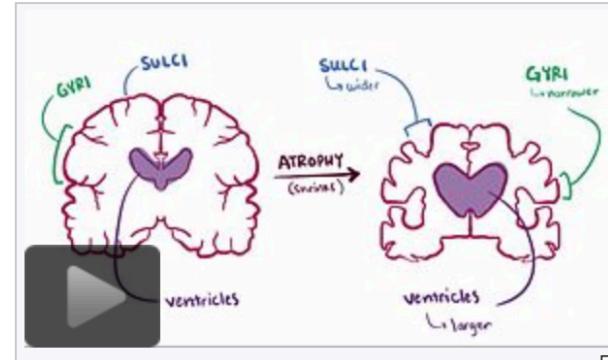
Medication Acetylcholinesterase inhibitors, NMDA receptor antagonists (small benefit)^[6]

Prognosis Life expectancy 3–9 years^[7]

Frequency 29.8 million (2015)^{[2][8]}

Deaths 1.9 million (2015)^[9]

[\[edit on Wikidata\]](#)



Alzheimer disease video



Public databases for drug discovery

- EMBL-EBI (European Bioinformatics Institute)



- Elsewhere



Fit everything together



- Time consuming
- Possible lack of resources or expertise
- ...



I wish I did not have to go to all those **different places** to get the information I'm after.

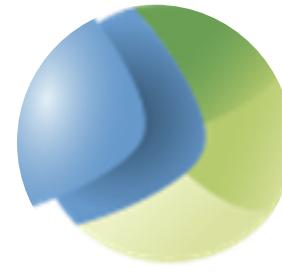
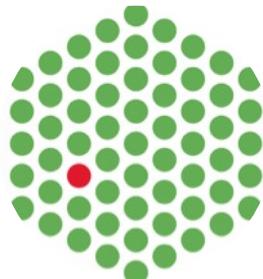


Open Targets is all you need!
I know. If you're like me, you want a one-stop shop where I can get as much data as possible, plus analyses and links to the original source for my own assessment.

A resource that is **comprehensive, trustworthy, up-to-date, sustainable, easy-to-use** and free.

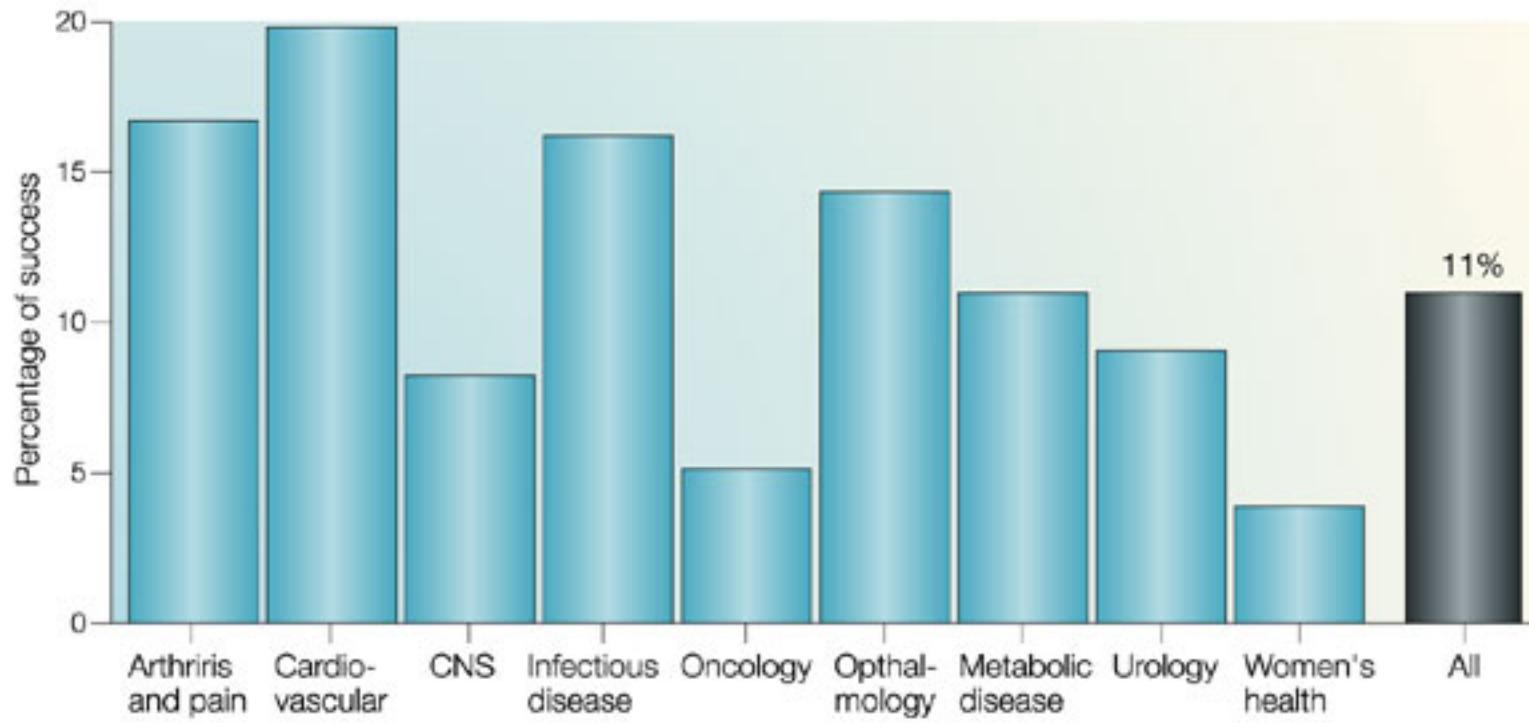
Our Vision

A partnership to transform drug discovery
through the systematic identification and
prioritisation of targets



Open Targets

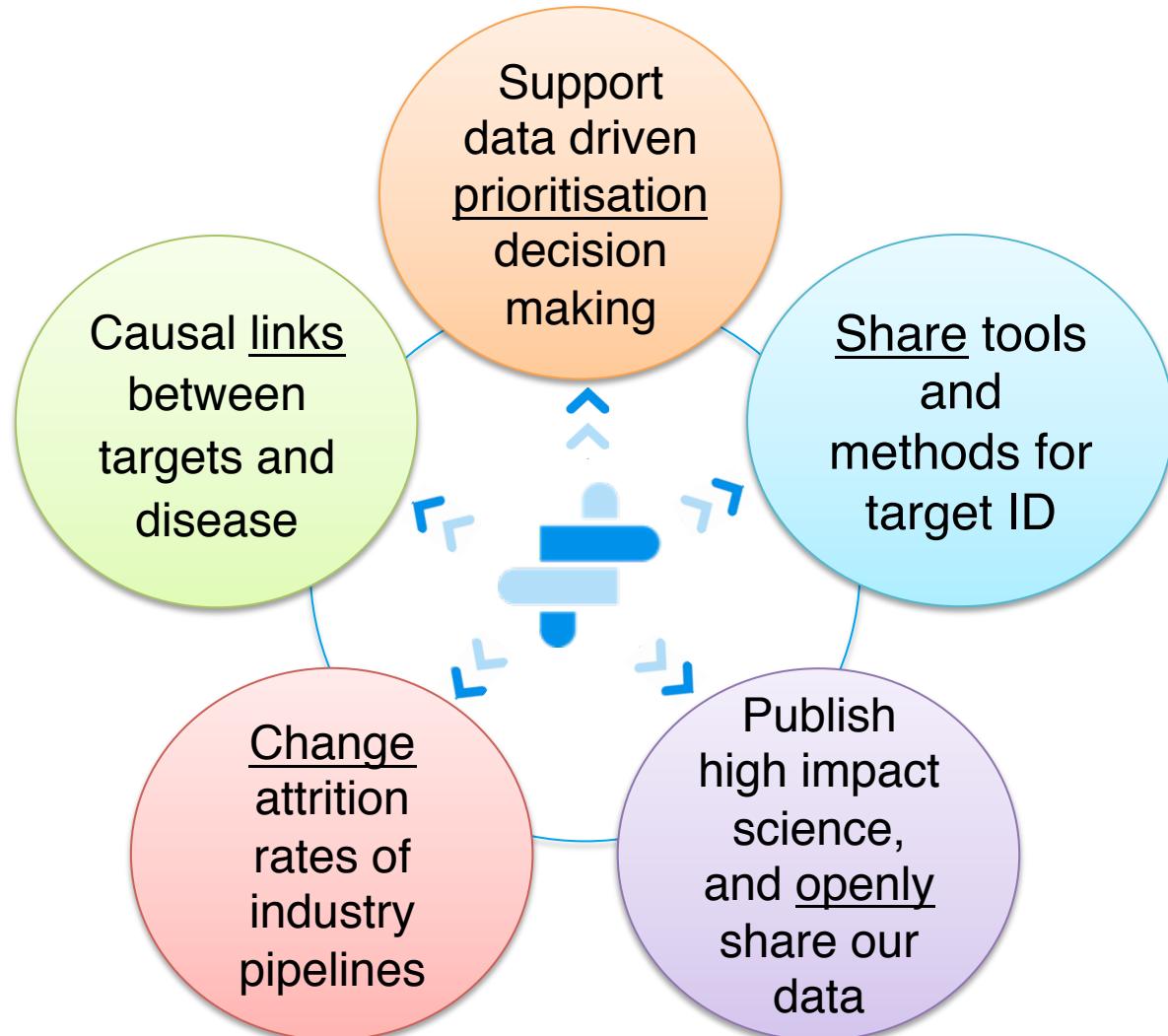
We need to address current challenges



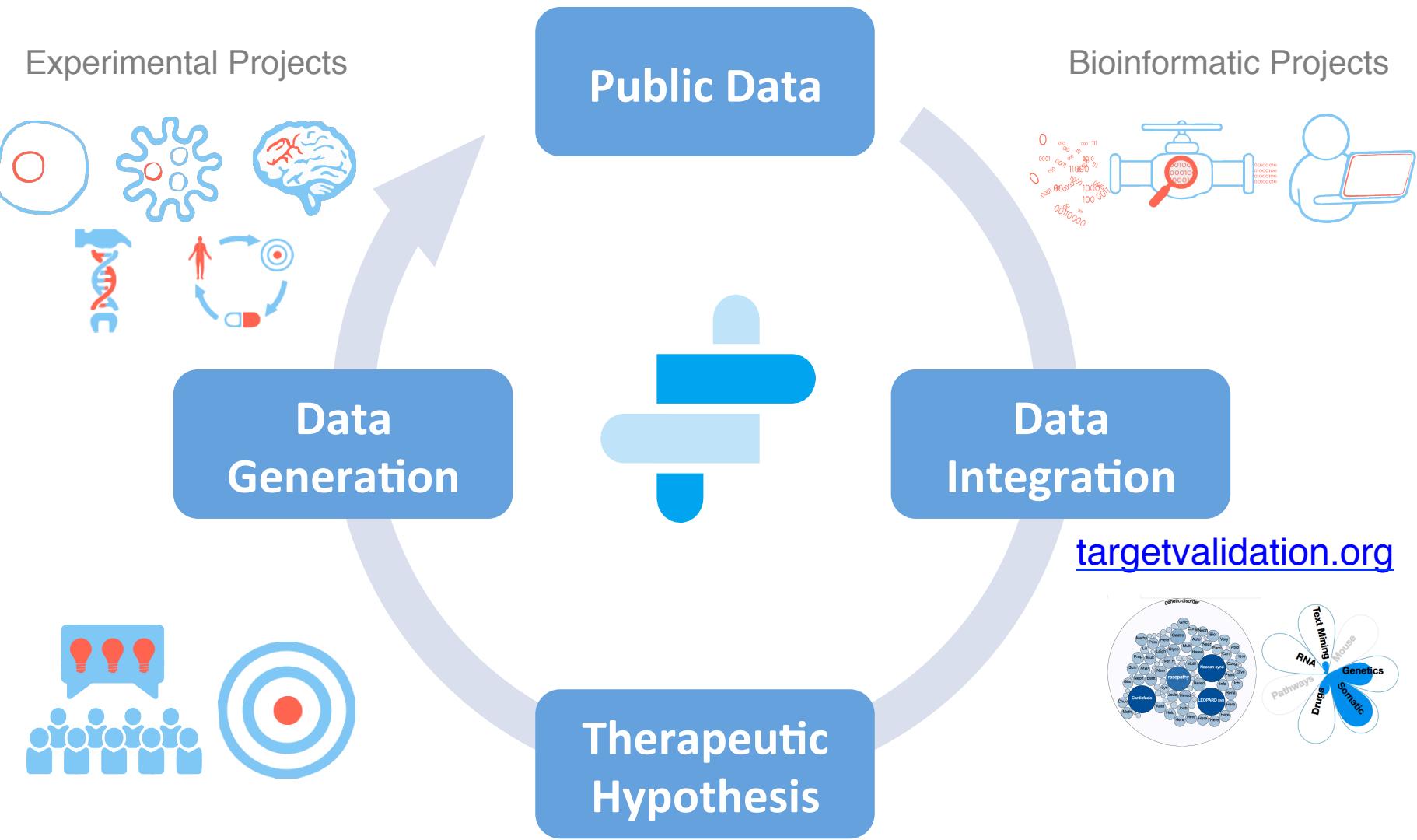
Lengthy, costly, low success rate, **HIGH ATTRITION RATES**

Source: doi:10.1038/nrd1470

World leader for human target discovery



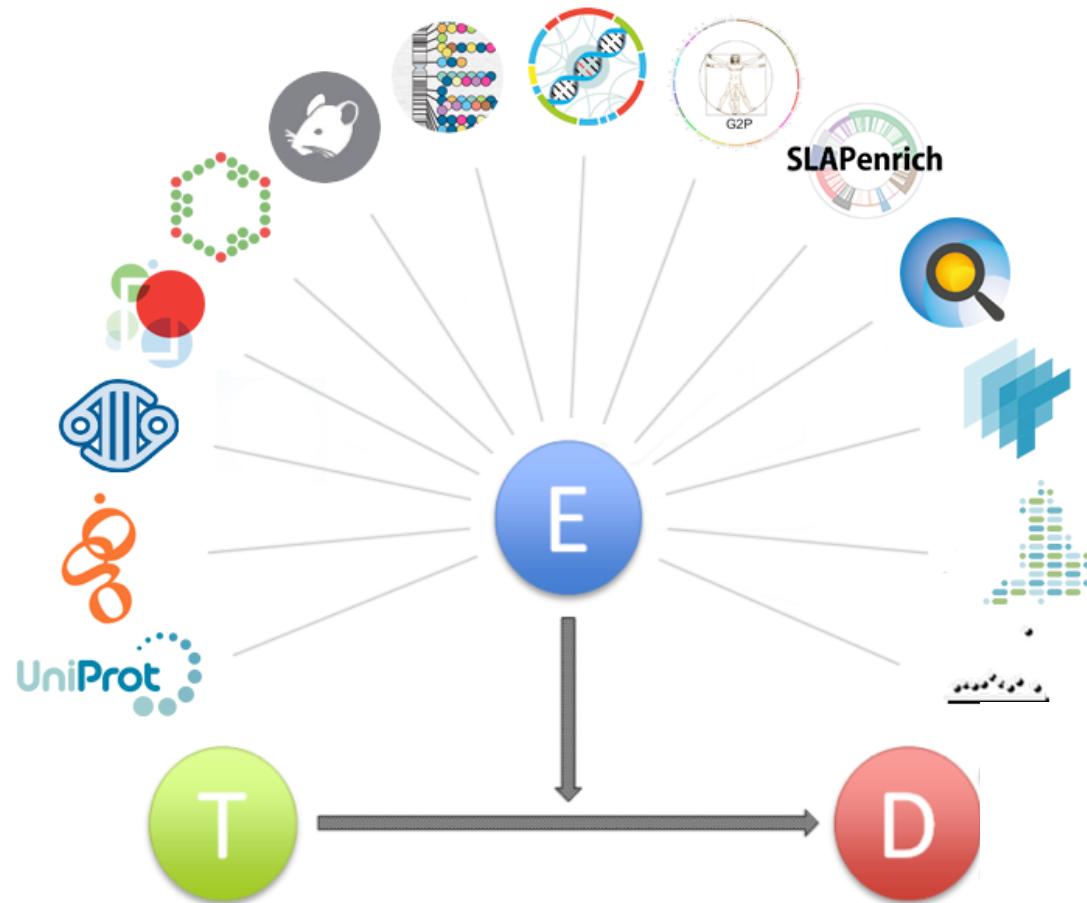
Virtuous cycle in Open Targets



Concurrent
www.opentargets.org/projects

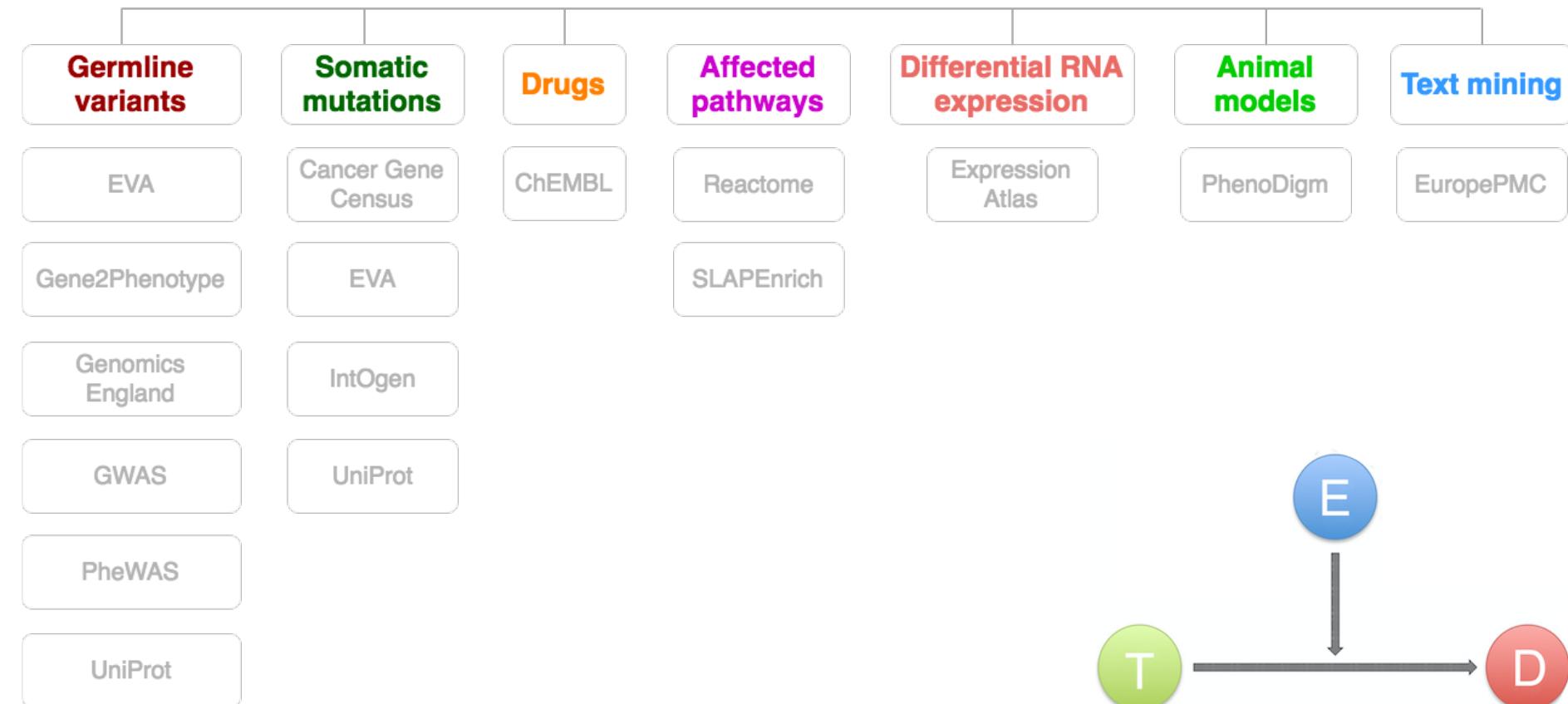
Integration – the data

- 100 GSK biologists interviewed
- 10 data sources then
- More interviews
- 15 data sources now
- Future: crowdsourcing



Data Sources → Data types

Open Targets Platform

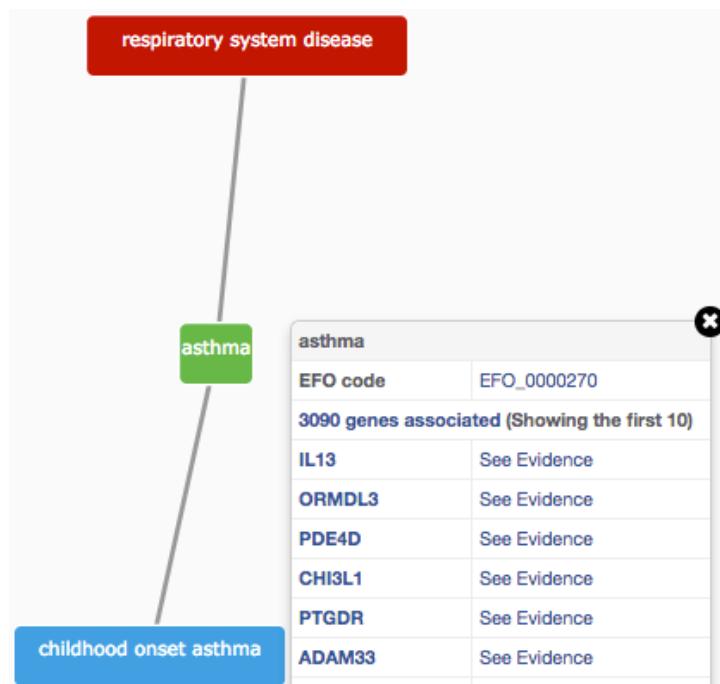


Which multiomics data to integrate?

- Suggest a causal link between the target and disease
- Aid an existing target hypothesis
- Be open, free and of easy access
- Be sustainable and likely maintained in the future
- Can be easily mapped to our disease and target IDs

Disease ontology

- Ontology: dictionary of relationships between entities
- EFO: way to organise experimental variables (e.g. diseases)



controlled vocabulary
(alzheimers x alzheimer's)
+
hierarchy (relationship)

Promotes consistency
Increases the richness of annotation
Allow for easier and automatic integration

Target IDs (genes or proteins)

- Ensembl Gene IDs: ENSGXXXXXXXXXXXX
- UniProt IDs: P15056

Not in Ensembl gene ID? Go back to Pablo's tips.

Day 2 - Wednesday 21 February 2018

09:00 - 10:30	ID Mapping	Pablo Porras
---------------	------------	-----------------

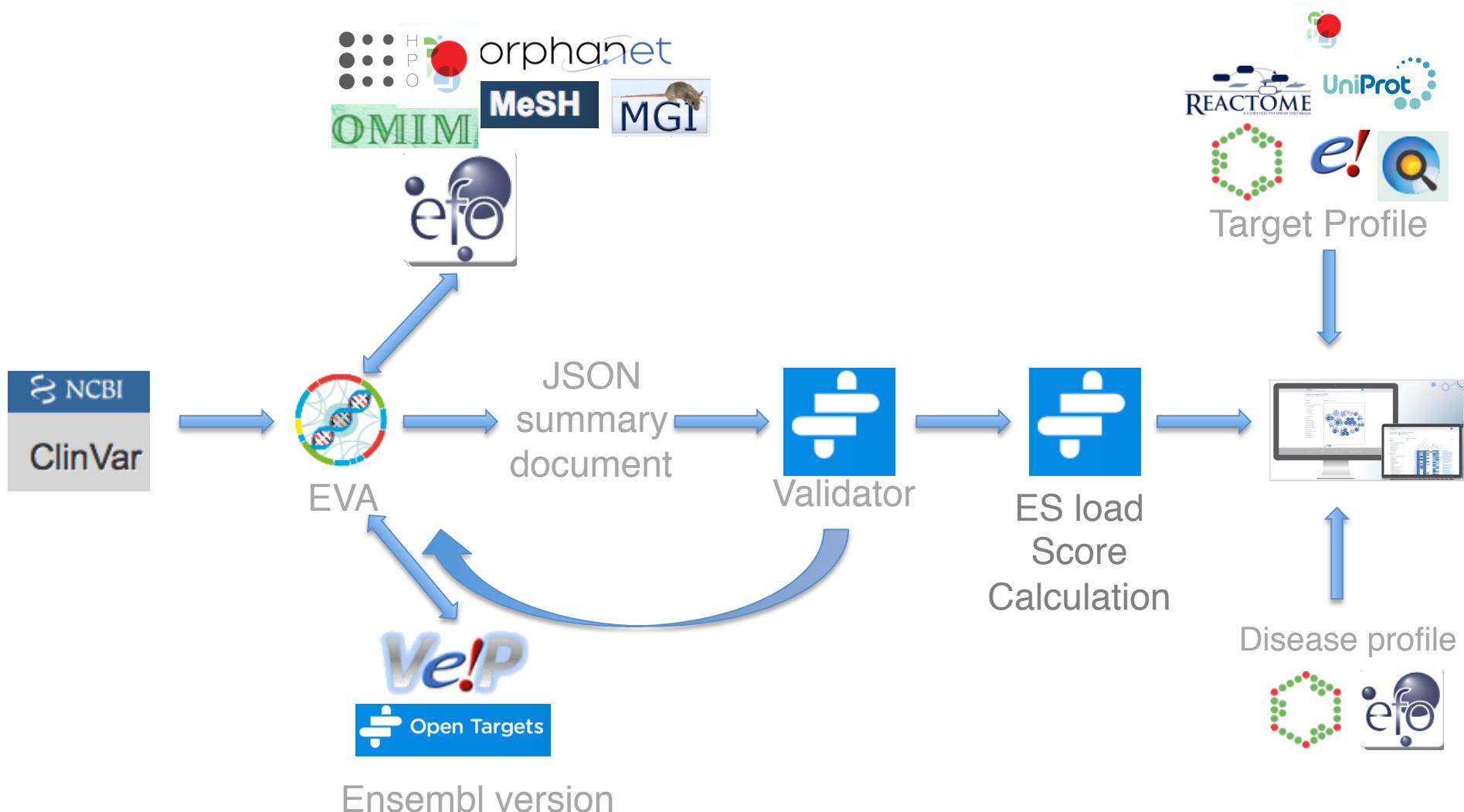


Integration – the infrastructure

- Big data? Not quite, but learning from it
- Allow for quick grow
- Data processing time: days → hours
- Run pipeline in the cloud
- Must be open source



Integration – the data* flow



* e.g. genetic variants from EVA

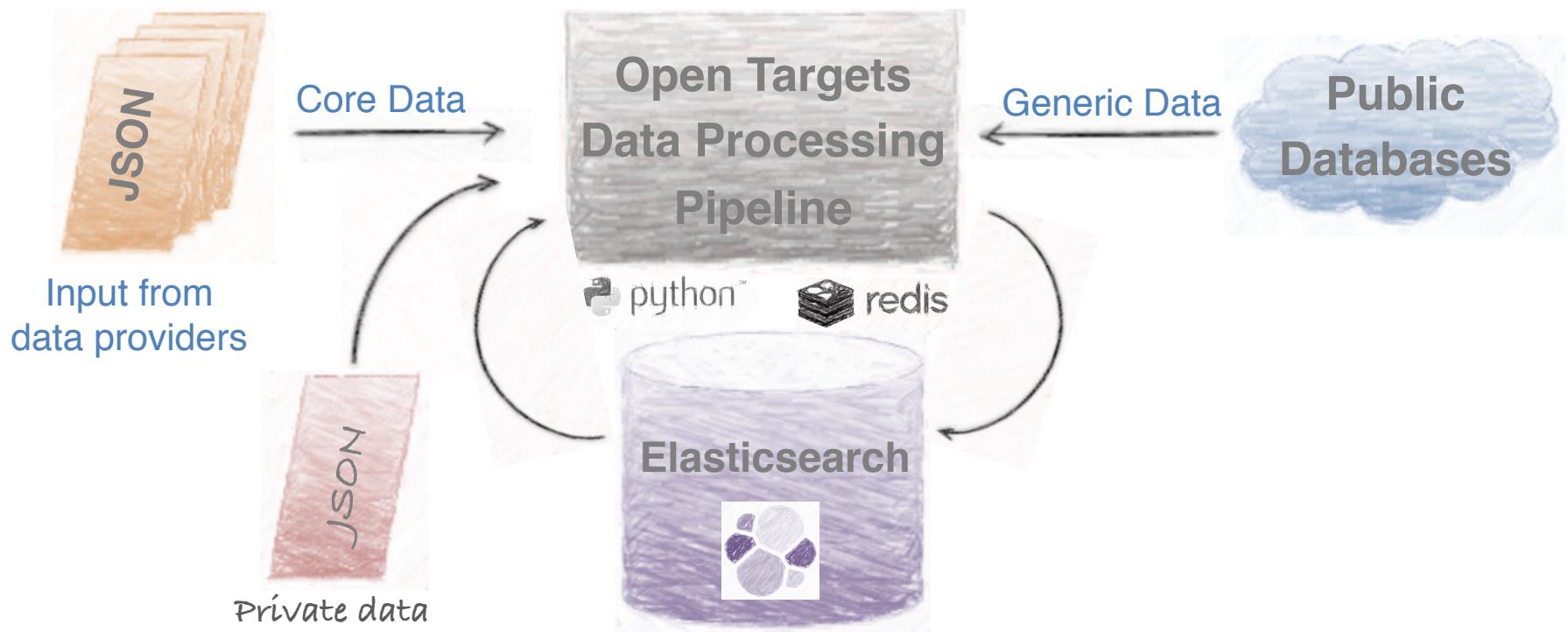


JSON summary document

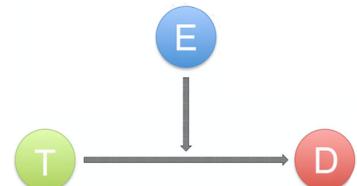
```
hort_name": "Franke A"}, {"last_name": "Alizadeh", "full_name": "Alizadeh Behrooz Z", "short_name": "Alizadeh BZ"}, {"last_ rkes Miles", "short_name": "Parkes M"}, {"last_name": "B K", "full_name": "B K Thelma", "short_name": "B K T"}, {"last_name' k J", "short_name": "Daly MJ"}, {"last_name": "Kubo", "full_name": "Kubo Michiaki", "short_name": "Kubo M"}, {"last_name": 'n Carl A", "short_name": "Anderson CA"}, {"last_name": "Weersma", "full_name": "Weersma Rinse K", "short_name": "Weersma RK" al_data": {"medlineAbbreviation": "Nat. Genet.", "title": "Nature genetics"}, "target": {"activity": "predicted_damaging", "name": "integrin subunit alpha L", "geneid": "ENSG00000005844"}, "id": "ENSG00000005844", "target_type": "gene_evidence"}, ceID": "gwas_catalog", "variant": {"type": "snp single", "id": "http://identifiers.org/dbsnp/rs11150589"}, "disease": {"efo_bel": ["immune system disease", "digestive system disease"], "codes": [{"efo_0000405", "efo_0000540"}], "path": [{"efo_00040", "efo_0005140", "efo_0003767"}]}, "efo_id": "http://www.ebi.ac.uk/efo/EF0_0003767", "label": "inflammatory bowel disease'_association_fields": {"pubmed_refs": "http://europepmc.org/abstract/MED/26192919", "object": "http://www.ebi.ac.uk/efo/EF0_ ntifiers.org/dbsnp/rs11150589", "study_name": "cttv009_gwas_catalog", "sample_size": "96486", "gwas_panel_resolution": "900": "http://identifiers.org/ensembl/ENSG00000005844"}, "evidence": {"variant2disease": {"gwas_sample_size": 96486, "unique_ex_ pepmc.org/abstract/MED/26192919", "gwas_panel_resolution": 9000000, "provenance_type": {"literature": {"references": [{"lit act/MED/26192919"}]}}, "expert": {"status": true, "statement": "Primary submitter of data", "database": {"dbxref": {"version": "id": "http://identifiers.org/gwascatalog"}, "id": "GWAS Catalog", "version": "2017-03-23T03:44:36+00:00"}}, "is_associated": ":" pvalue", "method": {"description": "pvalue for the snp to disease association."}, "value": 9e-07}, "evidence_codes": [{"http://purl.obolibrary.org/obo/ECO_0000205", "http://purl.obolibrary.org/obo/ECO_0000205", "http://identifiers.org/eco/ctt rted": "2017-03-23T03:44:36+00:00"}, "evidence_codes_info": [{"{"eco_id": "GWAS", "label": "Genome-wide association study ev: ing_pipeline", "label": "CTTV-custom annotation pipeline"}], [{"{"eco_id": "ECO_0000205", "label": "curator inference"}], [{"{"eco_id": "upstream_gene_variant"}]}, "gene2variant": {"functional_consequence": "http://purl.obolibrary.org/obo/SO_0001631", "provenan true, "statement": "Primary submitter of data", "database": {"dbxref": {"version": "2017-03-23T03:44:36+00:00", "id": "http , "id": "GWAS Catalog", "version": "2017-03-23T03:44:36+00:00"}}, "is_associated": true, "resource_score": {"type": "probab codes": [{"http://purl.obolibrary.org/obo/ECO_0000205", "http://identifiers.org/eco/cttv_mapping_pipeline"}, "date_asserted": evidence_codes": [{"GWAS", "cttv_mapping_pipeline", "ECO_0000205", "SO_0001631"}], "validated_against_schema_version": "1.2.5 res": {"association_score": 0.24183029962242697}, "type": "genetic_association", "id": "f8aa5612c7f01940f3958914fc6074ba"}} loads denote$
```

* IDs (gene, disease, papers) + curation (e.g. manual) + evidence + source + stats for the score

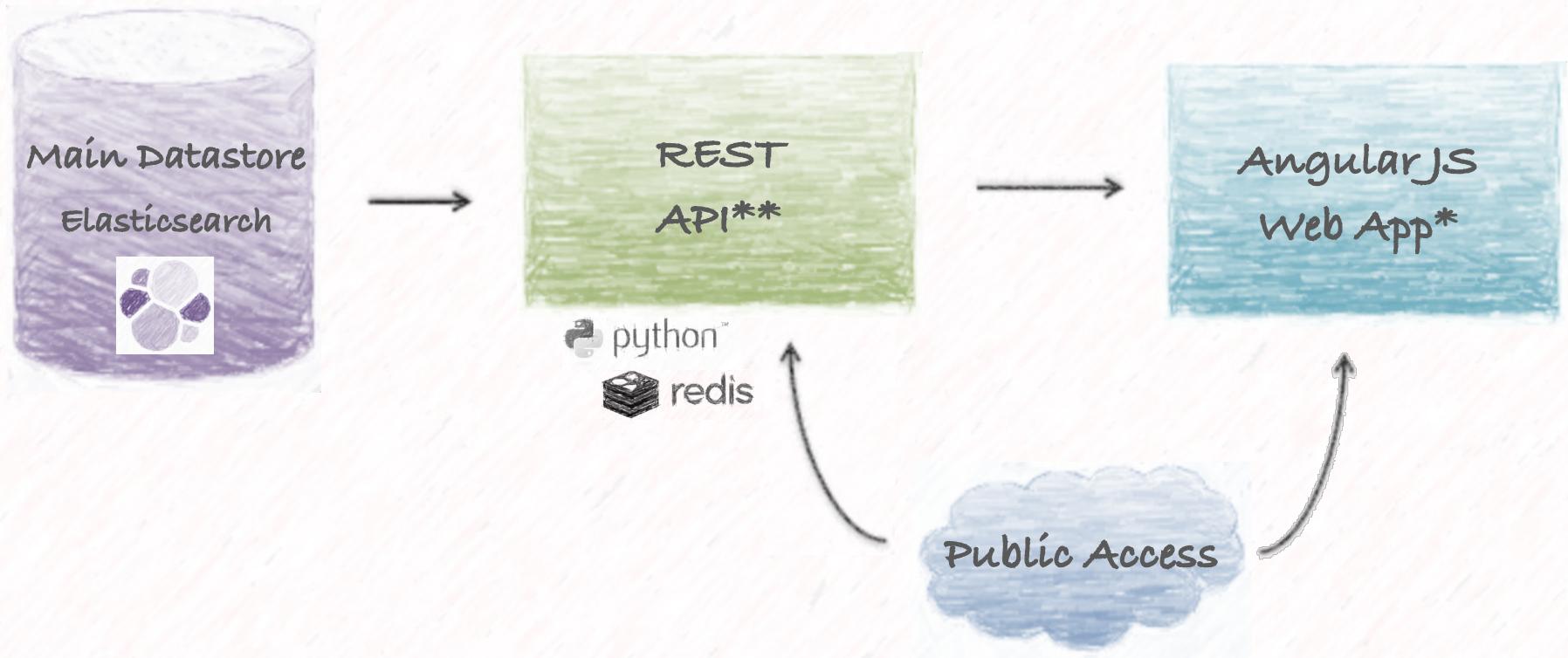
Integration - the components



1. Gather generic data from public resources (target, disease, pathways, etc...)
2. Validate Input
3. Analyse data and fit it to data model
4. Partners add private data



Integration – user access



*UI: first released in December 2015

<https://www.targetvalidation.org>

** API first release in April 2016

<https://api.opentargets.io>

In these 2.5 hours

- Introduction
- Live demos
- Wrap up

Demo 1: Disease centric workflow

Which targets are associated with a disease?



What is the evidence for the association between a target and a disease?

Find new targets for drug discovery

multiple sclero 🔍

multiple sclerosis
2697 targets associated

💡 Disease

An autoimmune disorder mainly affecting young adults and characterized by destruction of myelin in the central nervous system. Pathologic findings include multiple sharply demarcated areas of demyelination throughout the white matter of the central nervous system. Clinical manifestations include vis...

Targets
MBP myelin basic protein

Diseases
relapsing-remitting **multiple sclerosis**
autoimmune disease > multiple sclerosis > relapsing-remitting multiple ...

<https://www.targetvalidation.org/>

Choose your favourite internet browser*

*Supported ones: Internet Explorer 11 (not earlier versions), Chrome, Firefox, Safari

Demo 2: Several targets at once



We have a list of 26 possible targets for IBD (inflammatory bowel disease).

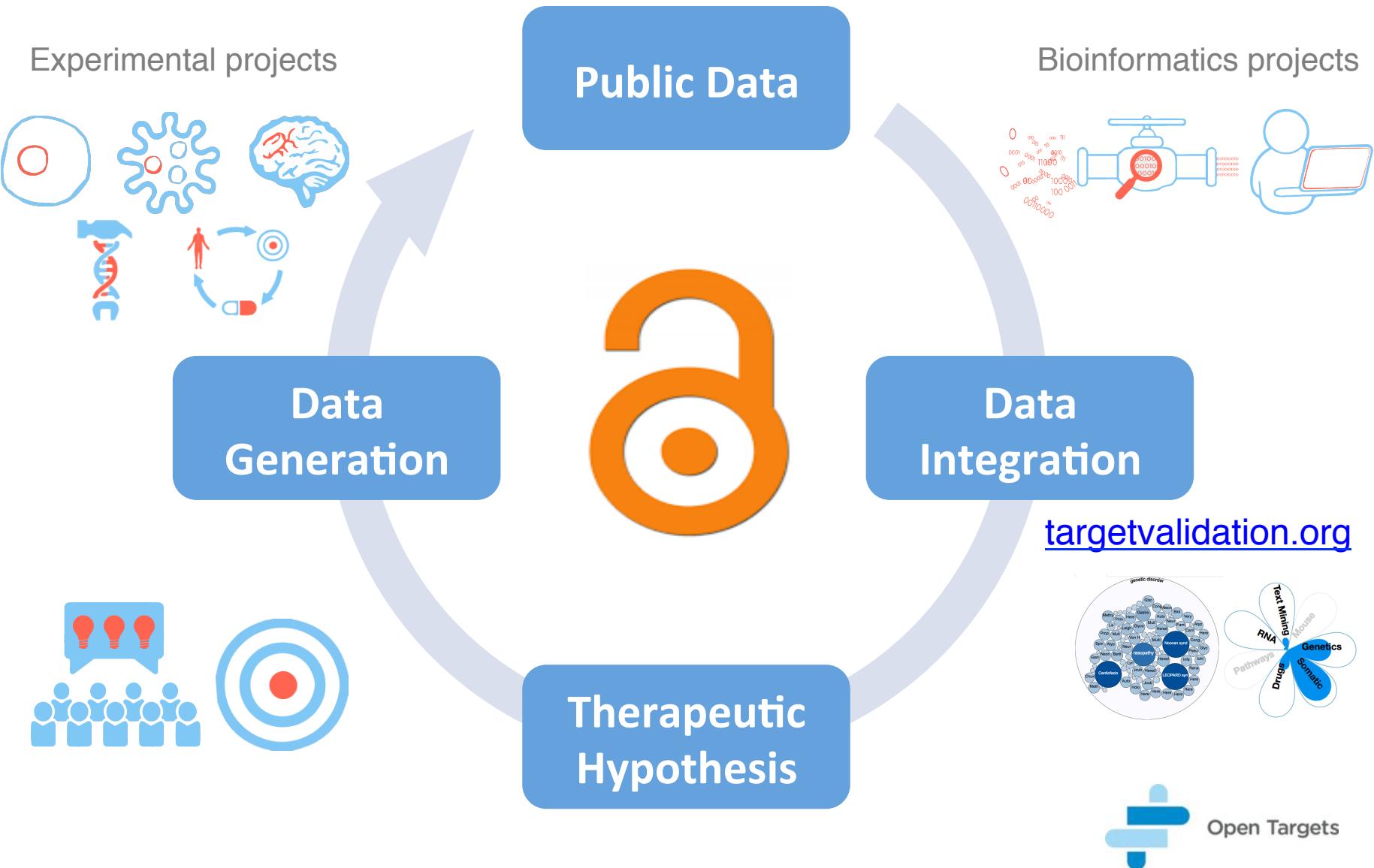
Are these targets represented in other diseases?

Which pathways are represented in this set of targets?

In these 2.5 hours

- Introduction
- Live demos
- Wrap up

Open Targets Partnership



Open Targets Platform

- Resource of integrated multiomics data
- Added value (e.g. score)
- Graphical web interface
- All data from multiple sources

Oh Yes!
And all is 100% free

20K targets

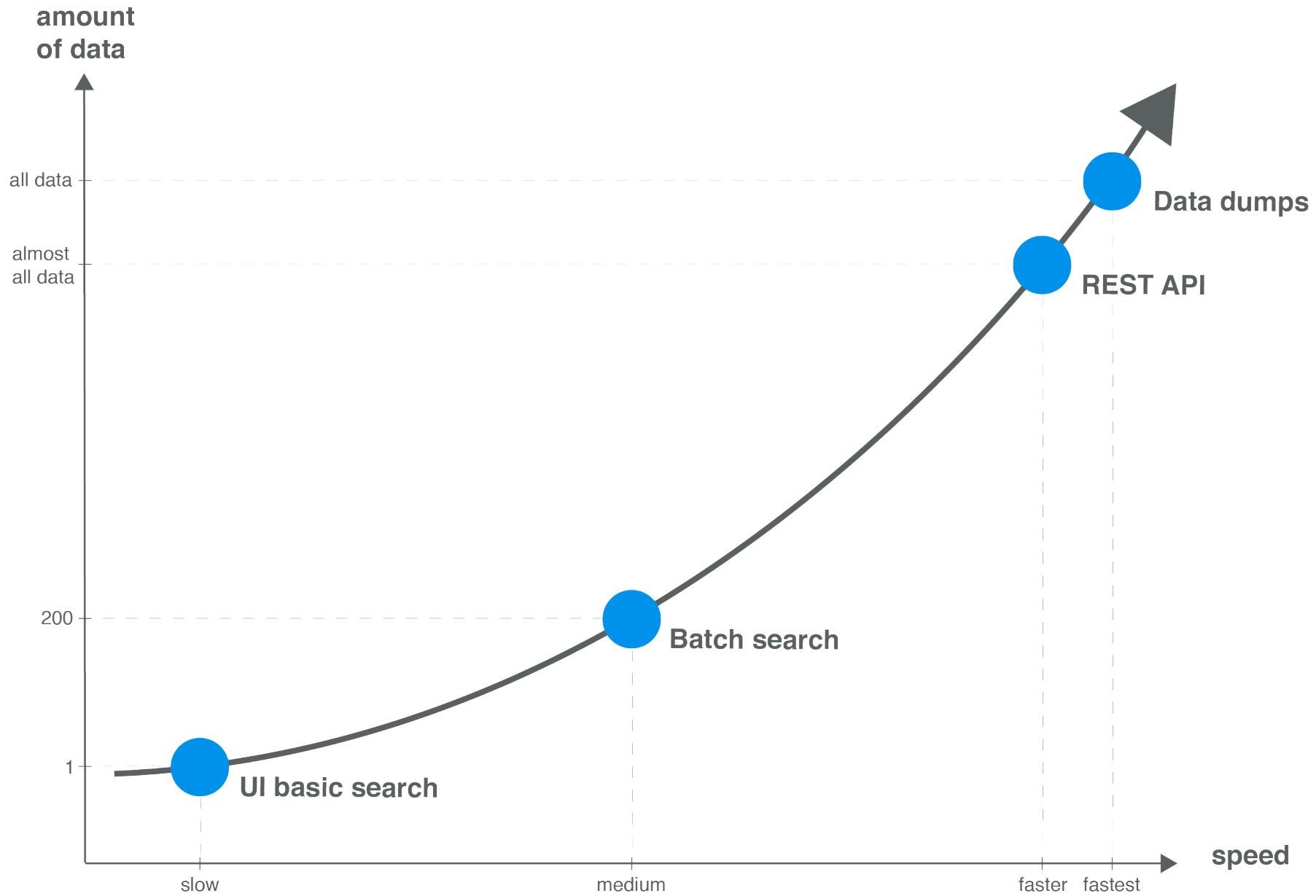
9.2K diseases

2.2 M associations

5.7 M evidence

December 2017 release

Modes of data access



Beyond the Platform

- Why should you go to   
- If you can use LINK (**L**iterature co**N**cept **K**nowledgebase)
- 28M documents and 526M semantic relations from PubMed abstracts
- Entities: genes, diseases, drugs
- Concepts extracted via Natural Language Processing

<http://link.opentargets.io/>

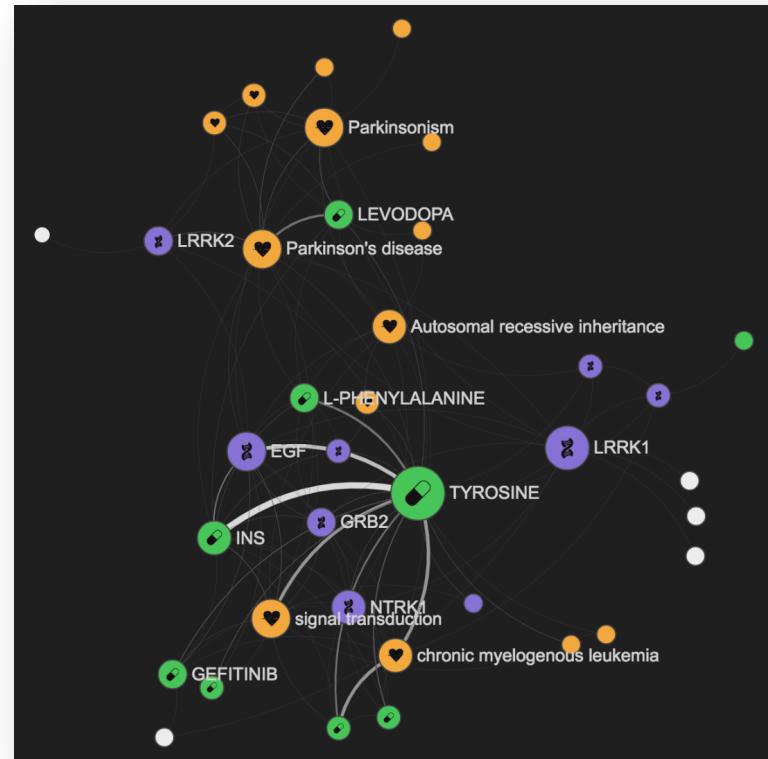
LINK

LIterature coNcept Knowledgebase

Search 28M documents and 526M semantic relations

 **GO**

Genes **Diseases** **Drugs** **Concepts**



LRRK1 Irrk1 play in Parkinson's disease pd

In contrast, LRRK1 **GENE**, the closest homologue to LRRK2, does not play any role in **PD** **DISEASE** **CONCEPT**.

PubMed: 28819229 2017-08-17

We support decision-making

Which targets are associated with a disease?

Can I find out about the mechanisms of the disease?

Are there FDA drugs for this association?



...

How to cite us

Published online 8 December 2016

Nucleic Acids Research, 2017, Vol. 45, Database issue D985–D994
doi: 10.1093/nar/gkw1055

Open Targets: a platform for therapeutic target identification and validation

Gautier Koscielny^{1,2,*}, Peter An^{1,3}, Denise Carvalho-Silva^{1,4}, Jennifer A. Cham^{1,4}, Luca Fumis^{1,4}, Rippa Gasparyan^{1,3}, Samiul Hasan^{1,2}, Nikiforos Karamanis^{1,4}, Michael Maguire^{1,4}, Eliseo Papa^{1,3}, Andrea Pierleoni^{1,4}, Miguel Pignatelli^{1,4}, Theo Platt^{1,3}, Francis Rowland^{1,4}, Priyanka Wankar^{1,3}, A. Patrícia Bento^{1,4}, Tony Burdett^{1,4}, Antonio Fabregat^{1,4}, Simon Forbes^{1,5}, Anna Gaulton^{1,4}, Cristina Yenyxe Gonzalez^{1,4}, Henning Hermjakob^{1,4,6}, Anne Hersey^{1,4}, Steven Jupe^{1,4}, Şenay Kafkas^{1,4}, Maria Keays^{1,4}, Catherine Leroy^{1,4}, Francisco-Javier Lopez^{1,4}, Maria Paula Magarinos^{1,4}, James Malone^{1,4}, Johanna McEntyre^{1,4}, Alfonso Munoz-Pomer Fuentes^{1,4}, Claire O'Donovan^{1,4}, Irene Papatheodorou^{1,4}, Helen Parkinson^{1,4}, Barbara Palka^{1,4}, Justin Paschall^{1,4}, Robert Petryszak^{1,4}, Naruemon Pratanwanich^{1,4}, Sirarat Sarntivijal^{1,4}, Gary Saunders^{1,4}, Konstantinos Sidiropoulos^{1,4}, Thomas Smith^{1,4}, Zbyslaw Sondka^{1,5}, Oliver Stegle^{1,4}, Y. Amy Tang^{1,4}, Edward Turner^{1,4}, Brendan Vaughan^{1,4}, Olga Vrousou^{1,4}, Xavier Watkins^{1,4}, Maria-Jesus Martin^{1,4}, Philippe Sanseau^{1,2}, Jessica Vamathevan⁴, Ewan Birney^{1,4}, Jeffrey Barrett^{1,4,5} and Ian Dunham^{1,4,*}

¹Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, ²GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK, ³Biogen, Cambridge, MA 02142, USA, ⁴European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, ⁵Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK and ⁶National Center for Protein Research, No. 38, Life Science Park Road, Changping District, 102206 Beijing, China

Received August 19, 2016; Revised October 19, 2016; Editorial Decision October 20, 2016; Accepted November 03, 2016

Acknowledgements



Open Targets

Help!



support@targetvalidation.org



<https://tinyurl.com/opentargets-youtube>



[@targetvalidate](#)



<http://tinyurl.com/opentargets-in>

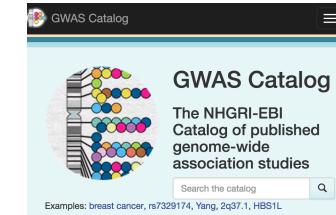
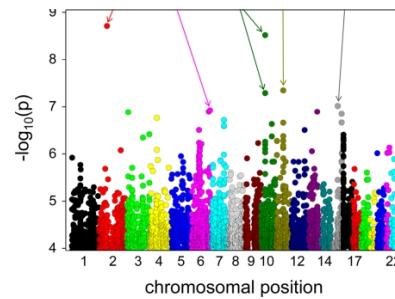
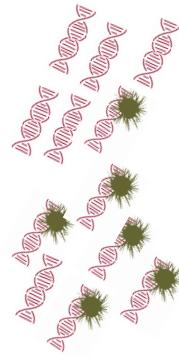
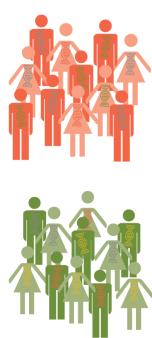


blog.opentargets.org/

Extra Extra Extra

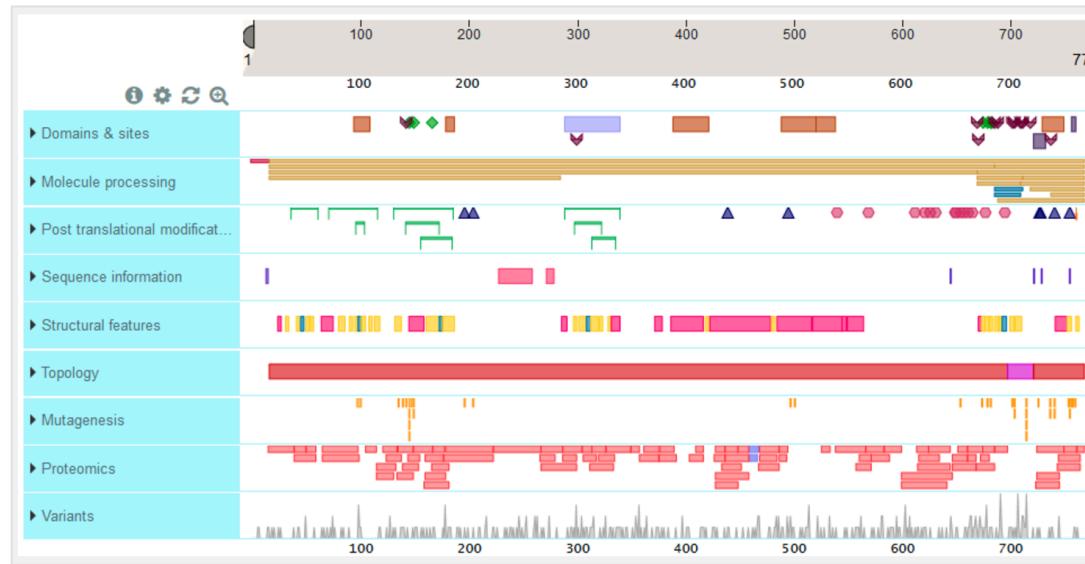
Data sources: GWAS catalog

- Genome Wide Association Studies
- Array-based chips → genotyping 100,000 SNPs genomewide



Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients

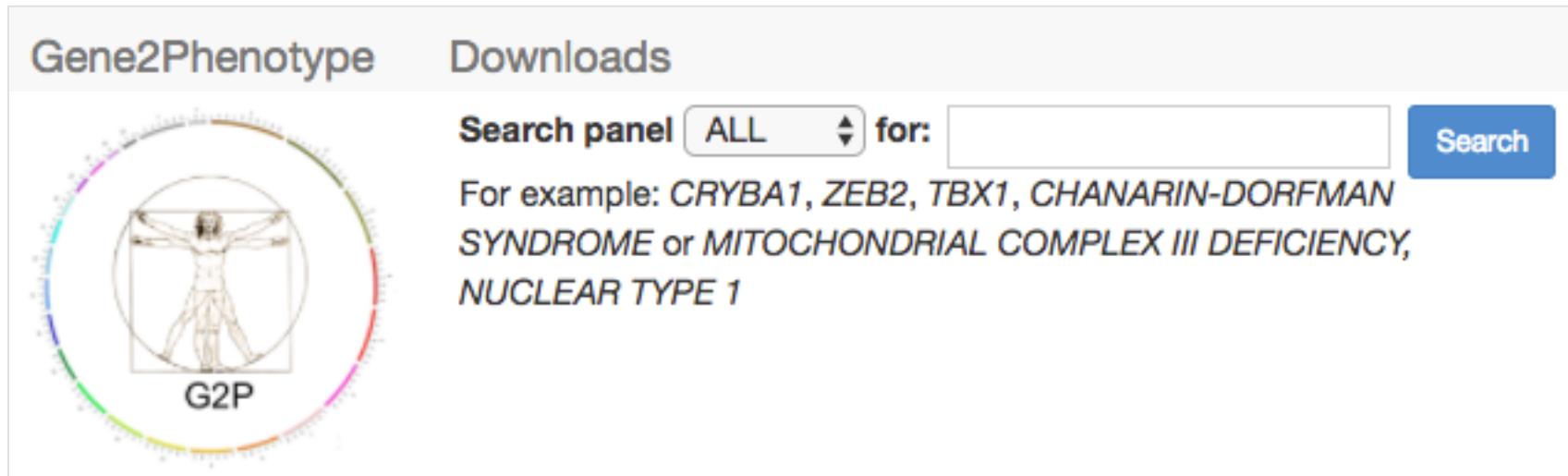


EMBL-EBI train online



Open Targets

Data sources: Gene2Phenotype



Gene2Phenotype Downloads

Search panel ALL for: **Search**

For example: **CRYBA1, ZEB2, TBX1, CHANARIN-DORFMAN SYNDROME or MITOCHONDRIAL COMPLEX III DEFICIENCY, NUCLEAR TYPE 1**

- Variants, genes, phenotypes in rare diseases
- Literature curation → consultant clinical geneticists in the UK

Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients



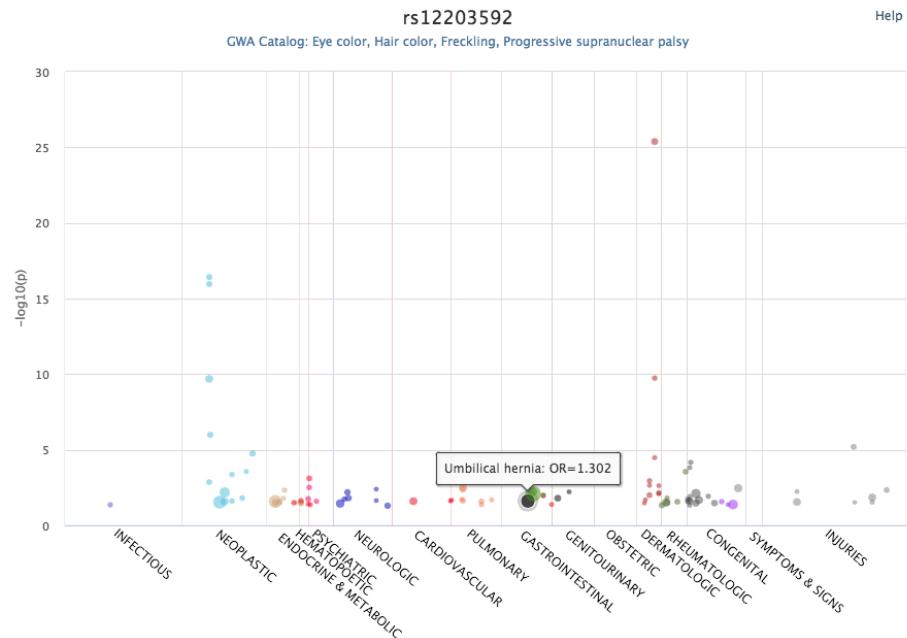
EMBL-EBI train online



Open Targets

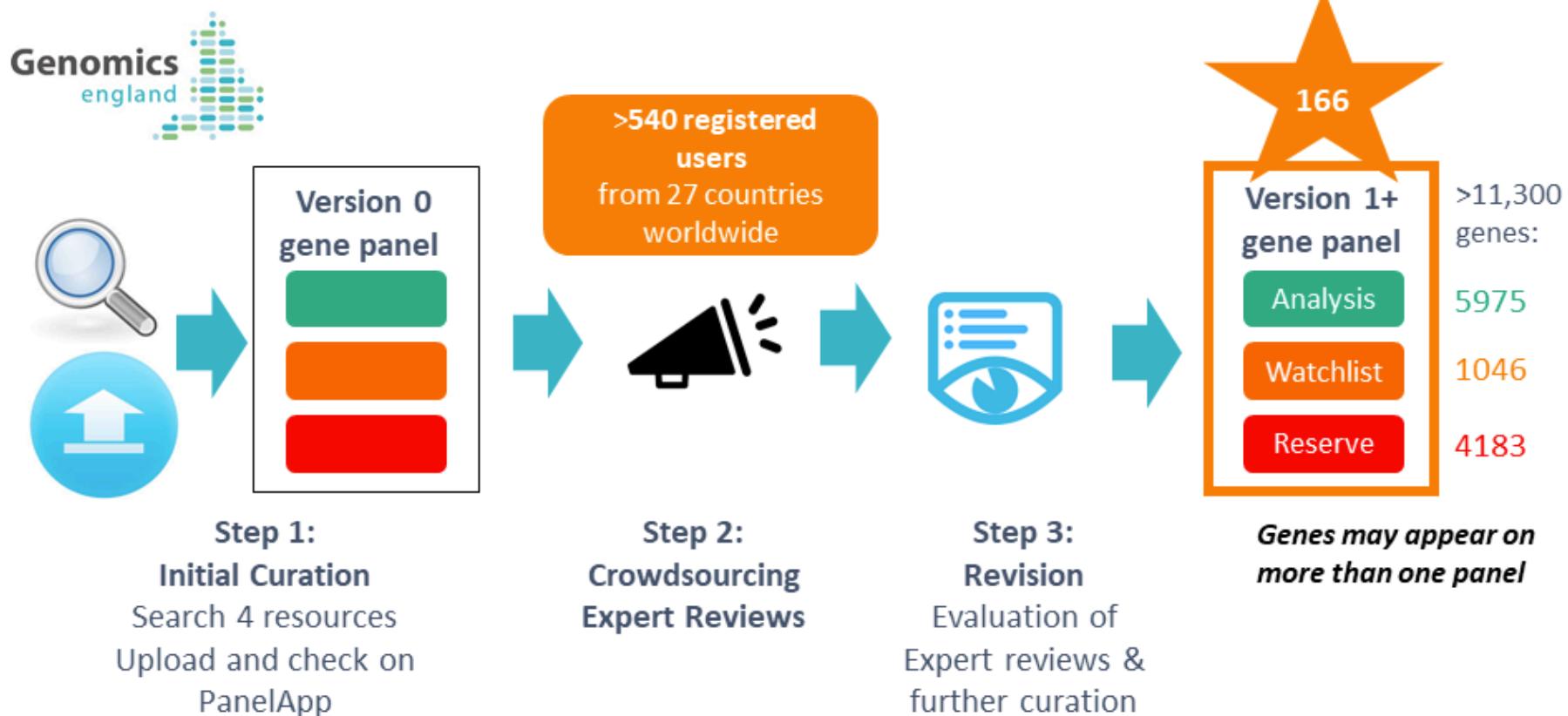
Data sources: PheWAS

- Phenome Wide Association Studies
- A variant associated with multiple phenotypes
- Clinical phenotypes derived from EMR-linked biobank BioVU
- ICD9 codes mapped to EFO



Data sources: GE PanelApp

- Aid clinical interpretation of genomes for the 100K project
- We include ‘green genes’ from version 1+ and phenotypes



Germline
variants

Somatic
mutations

Data sources: EVA

- With ClinVar information for rare diseases
- Clinical significance: pathogenic, protective

The screenshot shows the European Variation Archive (EVA) Clinical Browser interface. The top navigation bar includes links for Home, Submit Data, Study Browser, Variant Browser, Clinical Browser (which is highlighted in black), GA4GH, API, FAQ, and Feedback. Below the navigation is a search bar with a magnifying glass icon and a "Filter" button. The main content area is titled "ClinVar Browser" with an information icon. It features a table with 960 records, showing columns for Position, Affecte..., Most Severe Consequence..., Trait, Clinical Significance, and ClinVar The first few rows of data are as follows:

...	Posi...	Affecte... i	A...	Most Severe Consequence...	Trait	Clinical Significance	ClinVar ...
2	480...	MSH6	T/G	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	G/A	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	C/T	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	C/T	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	G/T	5_prime_UTR...	Lynch synd...	Uncertain s...	RCV000...
2	480...	MSH6	G/T	5_prime_UTR...	Hereditary ...	conflicting ...	RCV000...



Data sources: The Cancer Gene Census

Census

Breakdown

Abbreviations

The cancer Gene Census is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer. The original census and analysis was published in [Nature Reviews Cancer](#) and supplemental analysis information related to the paper is also available.

- Genes with mutations causally implicated in cancer
- Gene associated with a cancer plus other cancers associated with that gene

Data sources: IntOGen

The screenshot shows the homepage of the intOGen website. At the top is a navigation bar with an orange gradient background. From left to right, it contains: the intOGen logo (a stylized orange 'i' icon followed by the word 'intOGen'), a 'Search' button with a magnifying glass icon, a 'Downloads' button with a download icon, an 'Analysis' button with a gear icon, an 'About' button with a speech bubble icon, and a 'Sign In' button with a user profile icon.

The main content area features the intOGen logo again, this time with the full name 'intOGen' in a large serif font next to a smaller orange 'i' icon. To the right of the name is the tagline 'Integrative Onco Genomics' in orange text.

- Genes and somatic (driver) mutations, 28 cancer types
- Involvement in cancer biology
- Rubio-Perez et al. 2015

Data sources: ChEMBL

EMBL-EBI

ChEMBL

EBI > Databases > Small Molecules > ChEMBL Database > Home

Search ChEMBL... Compounds Targets Assays

Ligand Search Target Search Browse Targets Browse Drugs Browse Drug Targets

- Known drugs linked to a disease and a known target
- FDA approved for clinical trials or marketing



EMBL-EBI train online



Open Targets

Affected
pathways

Data sources: Reactome



The image shows the Reactome homepage. At the top left, there is a purple box containing the text "Affected pathways". The main title "REACTOME" is displayed in large white letters, with "A CURATED PATHWAY DATABASE" in smaller text below it. To the right of the title is a decorative graphic featuring a red sphere, blue and green abstract shapes, and arrows. Below the title is a navigation bar with links: "About", "Content", "Documentation", "Tools", "Community", "Download", and "Contact". To the right of the navigation bar is a search bar containing the placeholder text "e.g. O95631, NTN1, signalin" and a "Search" button.

- Biochemical reactions and pathways
- Manual curation of pathways affected by mutations

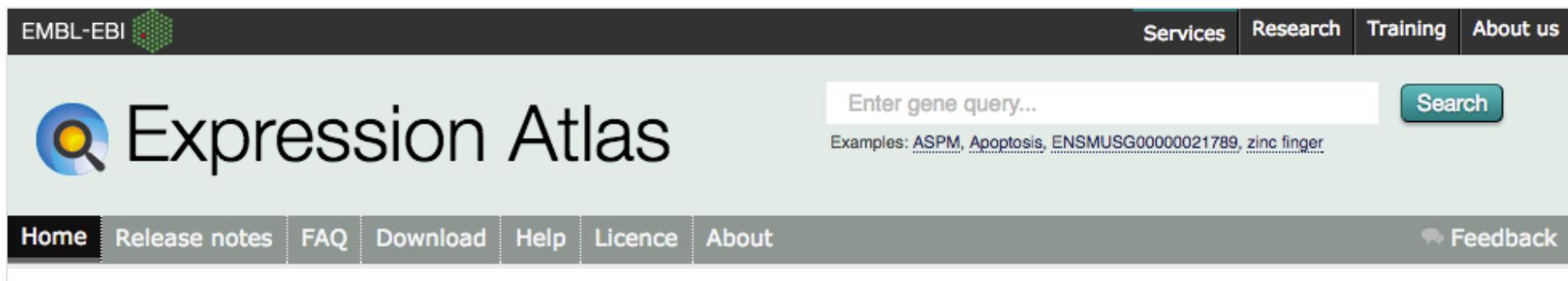


EMBL-EBI train online



Open Targets

Data sources: Expression Atlas



The screenshot shows the Expression Atlas website. At the top, there's a dark header bar with the EMBL-EBI logo and links for Services, Research, Training, and About us. Below the header is a search bar with a placeholder "Enter gene query..." and a "Search" button. To the left of the search bar is a magnifying glass icon. The main title "Expression Atlas" is displayed prominently. Below the title is a navigation bar with links for Home, Release notes, FAQ, Download, Help, Licence, and About. On the far right of this bar is a "Feedback" link with a speech bubble icon.

- Baseline expression for human genes
- Differential mRNA expression (*healthy versus diseased*)



EMBL-EBI train online



Open Targets

Data sources: Europe PMC



Europe PMC

About

Tools

Developers

Help

Europe PMC plus

Search worldwide, life-sciences literature

- Mining titles, abstracts, full text in research articles
- Target and disease co-occurrence in the same sentence



EMBL-EBI train online



Open Targets

Data sources: PhenoDigm

The screenshot shows the homepage of the PhenoDigm website. At the top, there is a dark header bar with the Wellcome Trust Sanger Institute logo on the left. To the right of the logo is a blue navigation bar with the following links: "ABOUT" (with a dropdown arrow), "Who we are", "Careers", "Study", "Sex in Science", "Groups", and "Campus". On the far right of the blue bar is a magnifying glass icon representing a search function. Below the header, the main title "Welcome to PhenoDigm (PHENOtype comparisons for DIsease and Gene Models)" is displayed in large, bold, black font. Underneath the title, there is a horizontal menu bar with three items: "Diseases" (which is highlighted in blue), "Tissue phenotype associations", and "Secondary phenotypes".

Welcome to PhenoDigm (PHENOtype comparisons for DIsease and Gene Models)

Diseases Tissue phenotype associations Secondary phenotypes

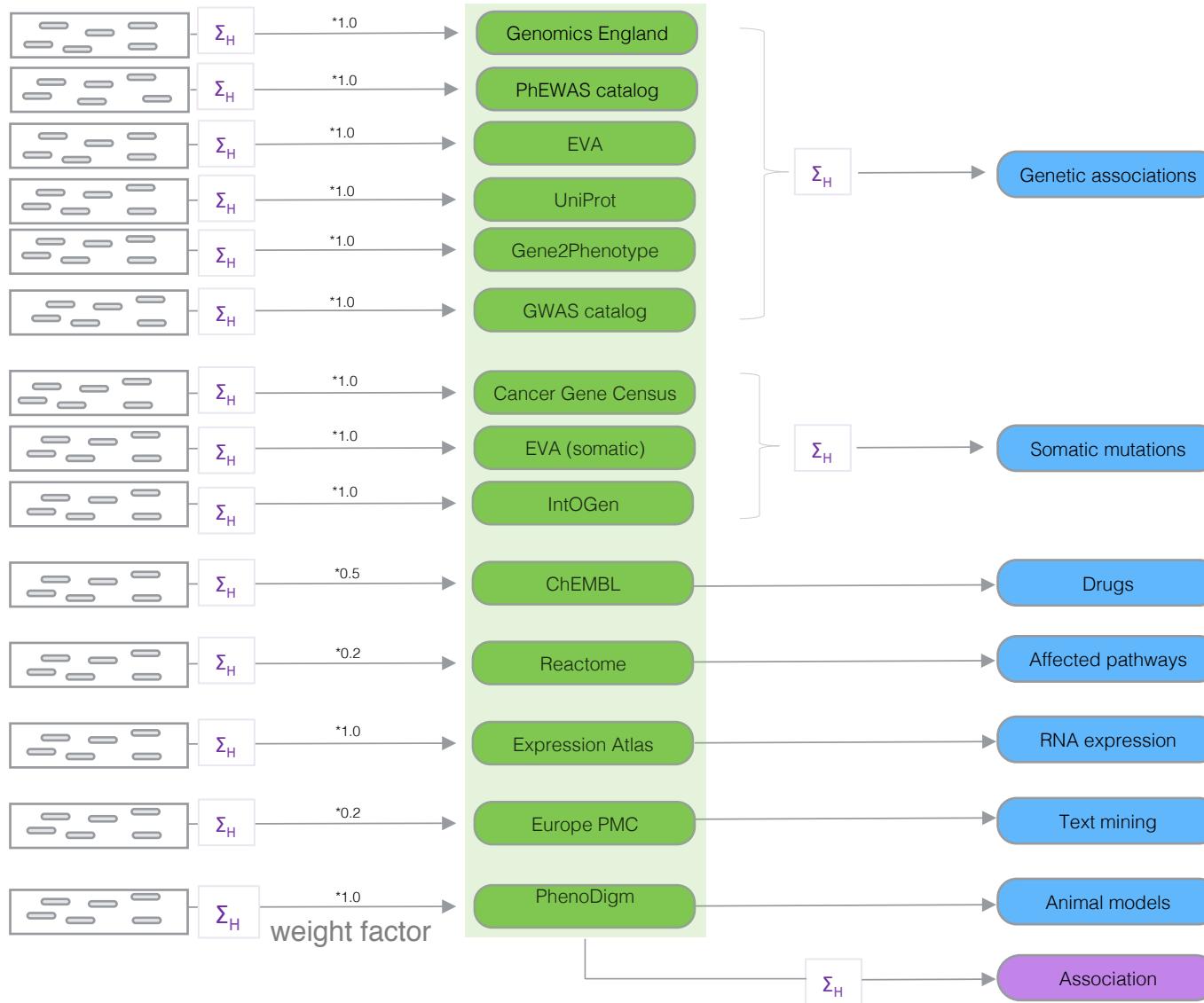
- Semantic approach to associate mouse models with diseases

How confident can you be of the target-disease associations in Open Targets?

Statistical integration, aggregation and scoring*

- A) per evidence (e.g. one SNP from a GWAS paper)
- B) per data source (e.g. SNPs from the GWAS catalog)
- C) per data type (e.g. Genetic associations)
- D) overall

How do we score the associations?



Score: 0 to 1 (max)

Calculated at 4 levels:

- Evidence
- Data source
- Data type
- Overall

Aggregation with (harmonic sum)



Note: Each data set has its own scoring and ranking scheme
Open Targets

Factors affecting the relative strength of an evidence

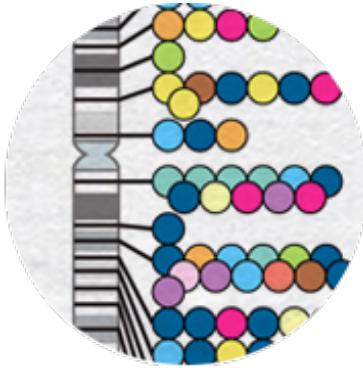
e.g. *GWAS Catalog*

$$S = f * s * c$$

f, relative occurrence of a target-disease evidence

s, strength of the effect described by the evidence

c, confidence of the observation for the target-disease evidence



f = sample size (cases versus controls)

s = predicted functional consequence

c = *p* value reported in the paper



Open Targets

Aggregating scores across the data

- Using a mathematical function, the harmonic sum*

$$S_{1..i} = S_1 + \frac{S_2}{2^2} + \frac{S_3}{3^2} + \frac{S_4}{4^2} \dots + \frac{S_i}{i^2}$$

where S_1, S_2, \dots, S_i are the individual sorted evidence scores in descending order

- Advantages:
 - A) account for replication
 - B) deflate the effect of large amounts of data e.g. text mining

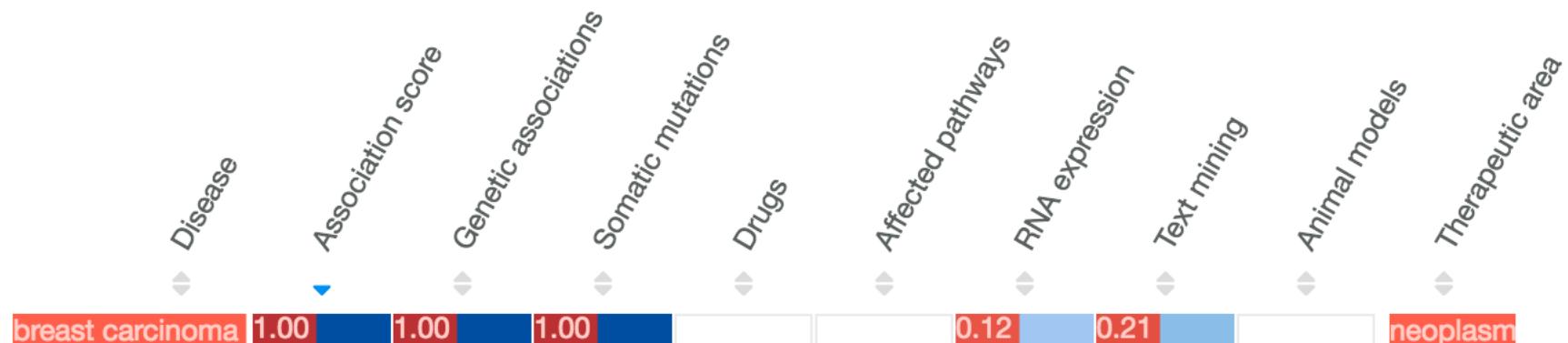
Disclaimer: score, dos and don'ts

- It's a ranking of target-disease associations
- It shows how confident we are in the association
- It's based on data sources, publicly available



- It can help you to design your null hypothesis
- It can help you to decide which target to pursue
- It is NOT sufficient on its own (use it in combination with...)

Ranking the target-disease association



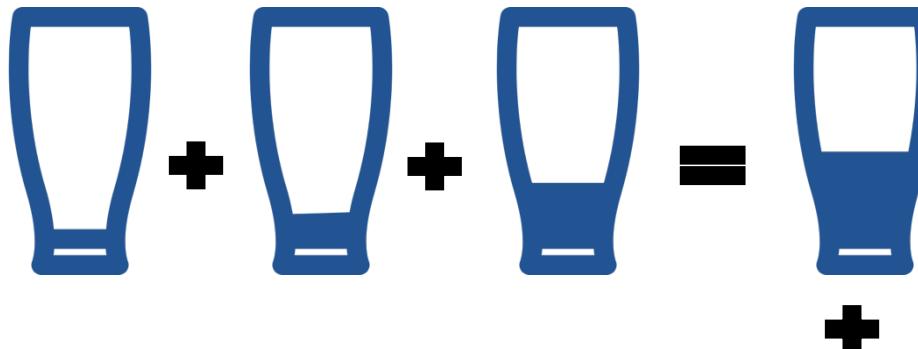
Association score: the overall score across all data types

- Based on the data sources
- Different weight applied:

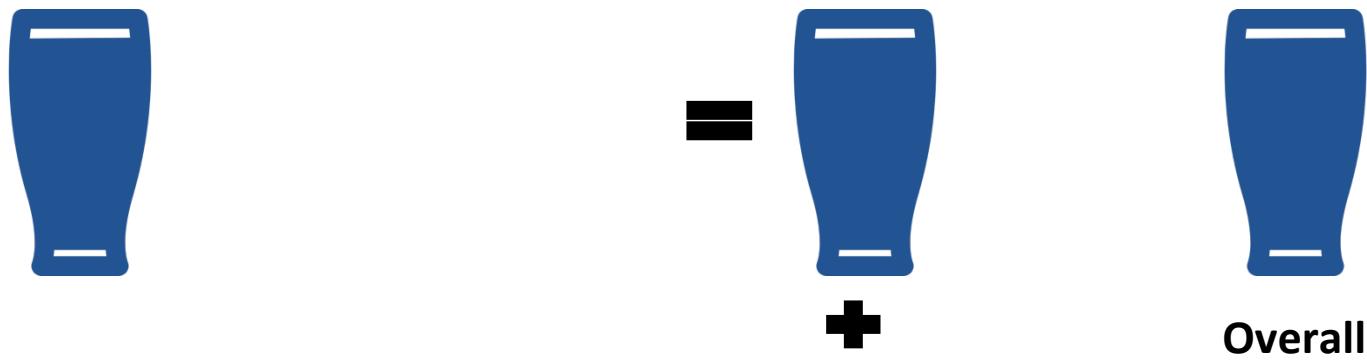
genetic association = drugs = mutations = pathways > RNA expression > animal models = text mining

Target-Disease Association Score

EuropePMC
(Text Mining)

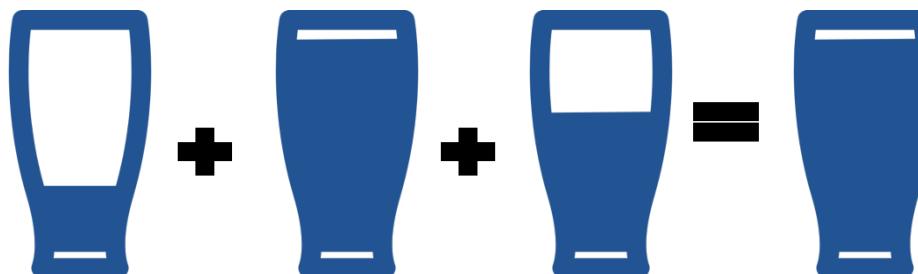


UniProt
(Manual Curation)



Overall

ChEMBL
(Manual Curation)



VERY simplified diagram

<https://www.targetvalidation.org/scoring>