

# Open Targets: enabling systematic target discovery and validation

Takeda

March 19<sup>th</sup> 2018

**Denise Carvalho-Silva, PhD**

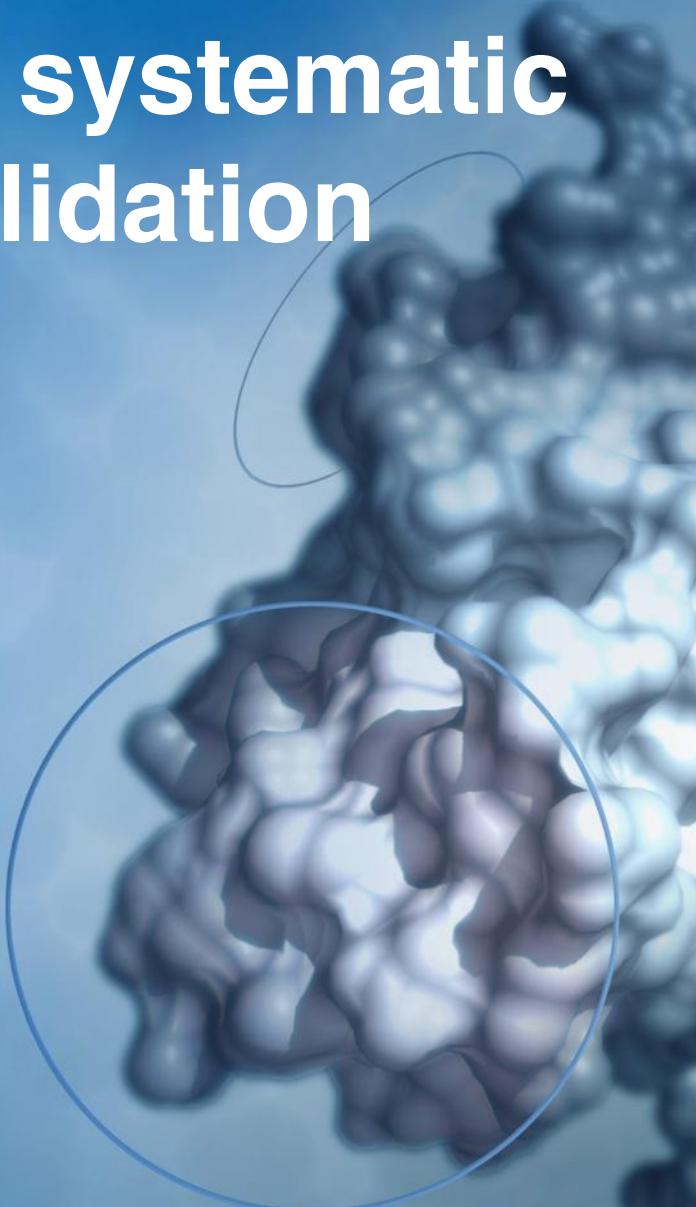
Scientific Outreach Lead

Open Targets / EMBL-EBI

Wellcome Genome Campus, United Kingdom



Open Targets



# Aims



What is the Open Targets Partnership?

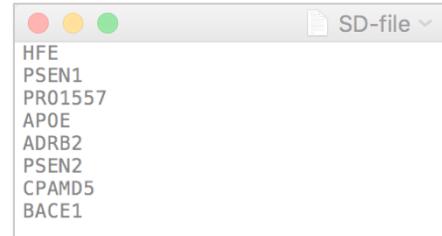
How to navigate the Open Targets Platform?

Are there other Open Targets tools?

Where do I get help?

# Materials

- [bit.ly/TBOS\\_slides](https://bit.ly/TBOS_slides)
- [bit.ly/TBOS\\_exercises](https://bit.ly/TBOS_exercises)
- [bit.ly/SD-file-takeda](https://bit.ly/SD-file-takeda)
- [bit.ly/SD-batch-takeda](https://bit.ly/SD-batch-takeda)



# In these ~3.5 hours

- Introduction
- Open Targets Platform: live demos
- Exercises
- Beyond the Platform
- Wrap up

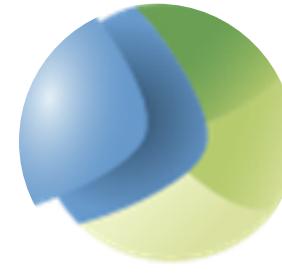
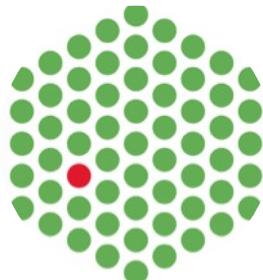
# In these ~3.5 hours

- Introduction
- Open Targets Platform: live demos
- Exercises
- Beyond the Platform
- Wrap up

# Our Vision

<https://www.opentargets.org>

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



# What does Open Targets want?

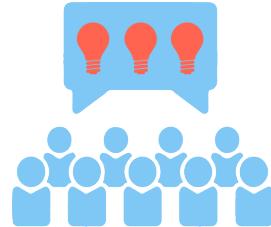
- Systematically find the best targets for safe and effective medicines
- Help others find good targets
- Get these targets adopted into drug discovery pipelines



High throughput  
human genetics



Organoids  
IPS cells



Make target  
decisions  
together



Advanced  
data analysis

# What do users want?



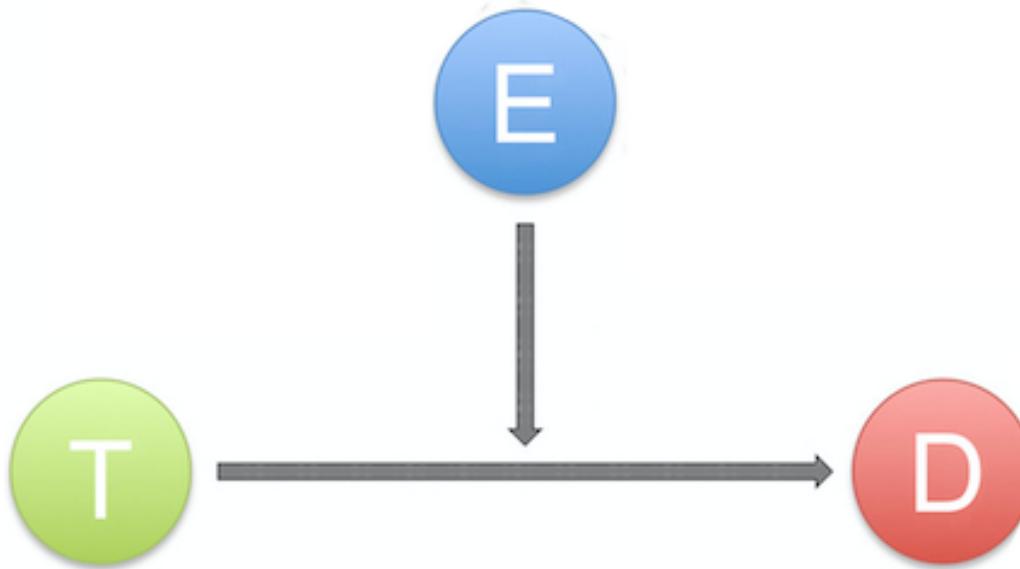
**Open Targets Platform**  
is all you need!

Many datasets in one place  
**one-stop shop**  
**new analysis**  
original data  
**comprehensive**  
**trustworthy**  
**up-to-date**  
**sustainable**  
**easy-to-use**  
**resource**

# Open Targets Platform

<https://www.targetvalidation.org>

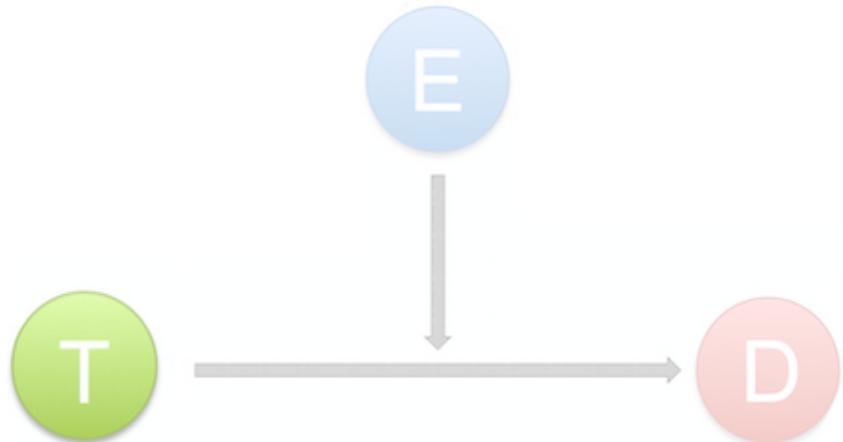
- Associations between targets and diseases



- Target and disease information

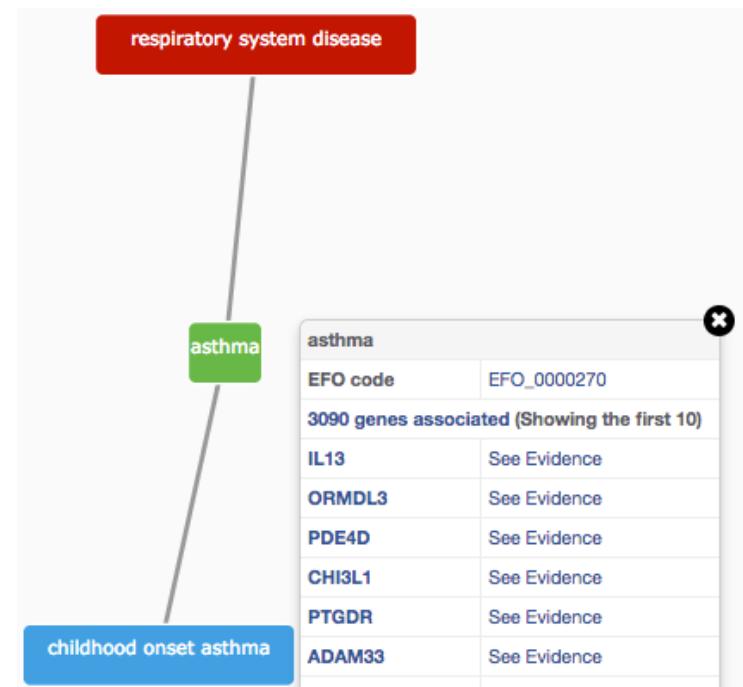
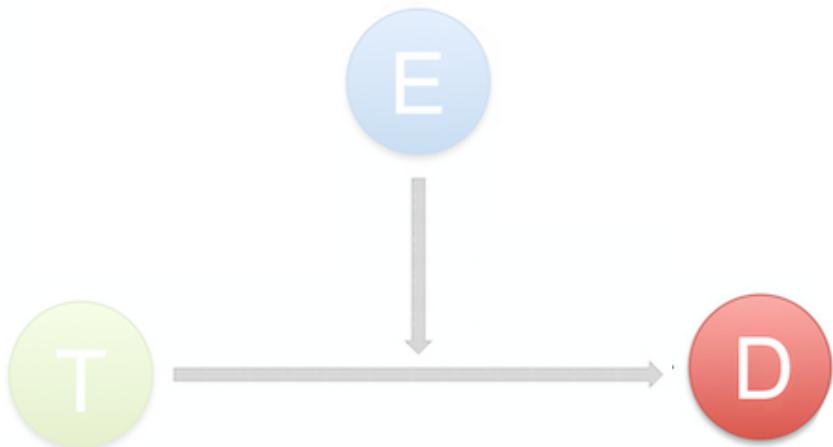
# Our targets → genes or proteins

- Ensembl Gene IDs e.g. ENSGXXXXXXXXXXXX
- UniProt IDs e.g. P15056
- HGNC names e.g. DMD
- Also non-coding RNA genes
- Metabolites: to come

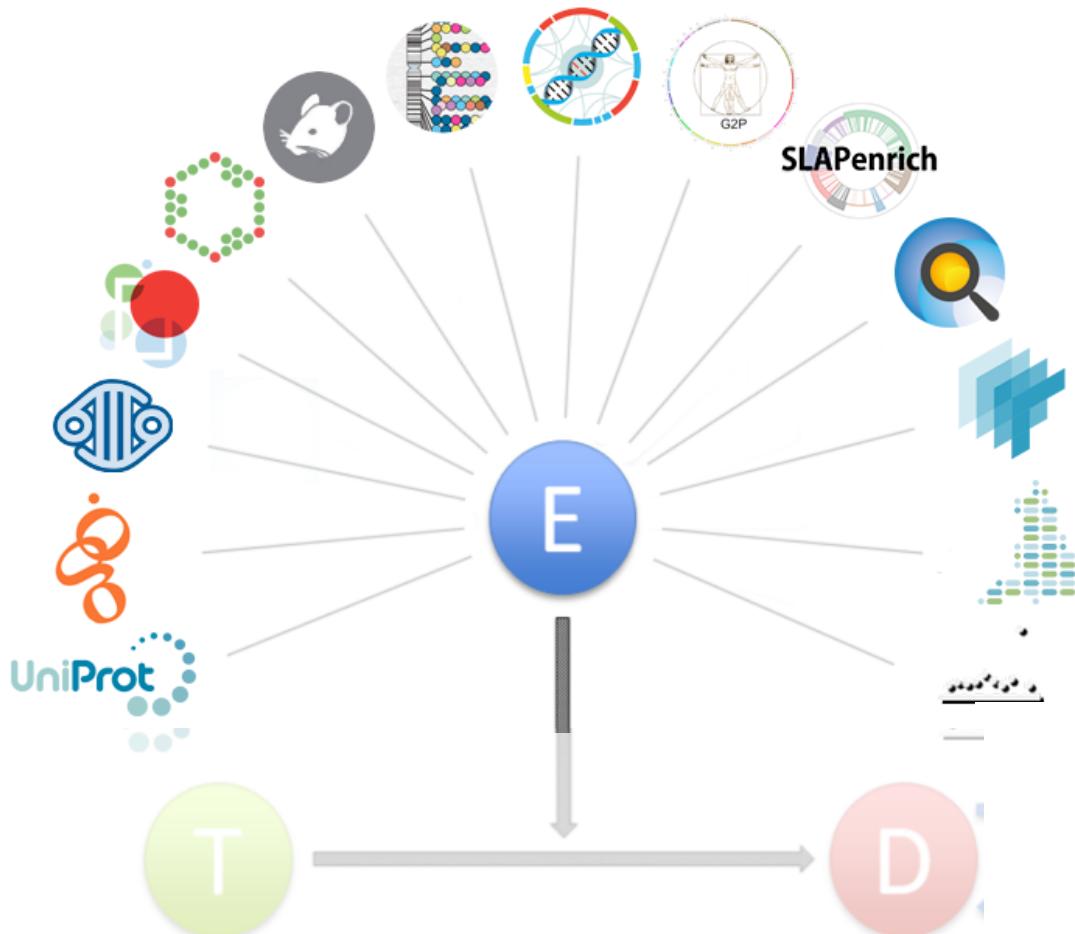


# Our diseases

- Experimental Factor Ontology (EFO)
- Controlled vocabulary (Alzheimers versus alzheimer's)
- Hierarchy (relationships)
  - Promotes consistency
  - Increases the richness of annotation
  - Allow for easier and automatic integration

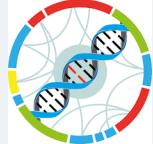


# Our evidence for T-D associations



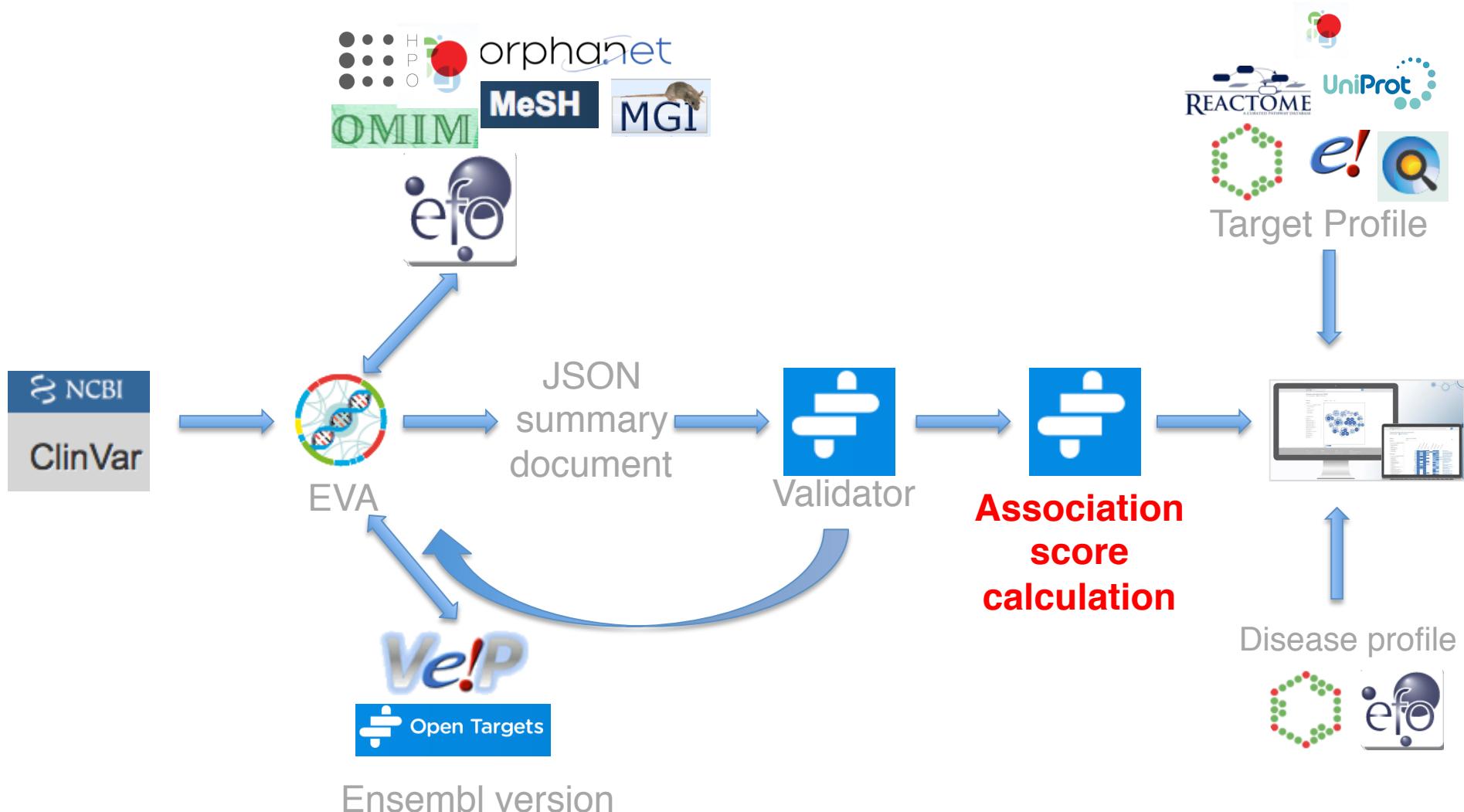
[www.targetvalidation.org/data\\_sources](http://www.targetvalidation.org/data_sources)

# Data sources grouped into seven data types

Genetic Associations	Somatic Mutations	Drugs	Affected Pathways	Differential RNA expression	Animal Models	Text Mining
 UniProt	 UniProt	 chEMBL	 reactome	 Expression Atlas	 PhenoDigm	 Europe PMC
 EVA	 EVA			 SLAPenrich		
 Genomics England	 COSMIC Catalogue Of Somatic Mutations In Cancer					
 GWAS Catalog	 intOgen					
 PheWAS						
 G2P						

More data sources to come

# How the data\* flows



\* 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

# Association score



How confident can I be  
in the Open Targets  
T-D associations?

$$S_1 + S_2/2^2 + S_3/3^2 + S_4/4^2 + S_i/i^2$$

## 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

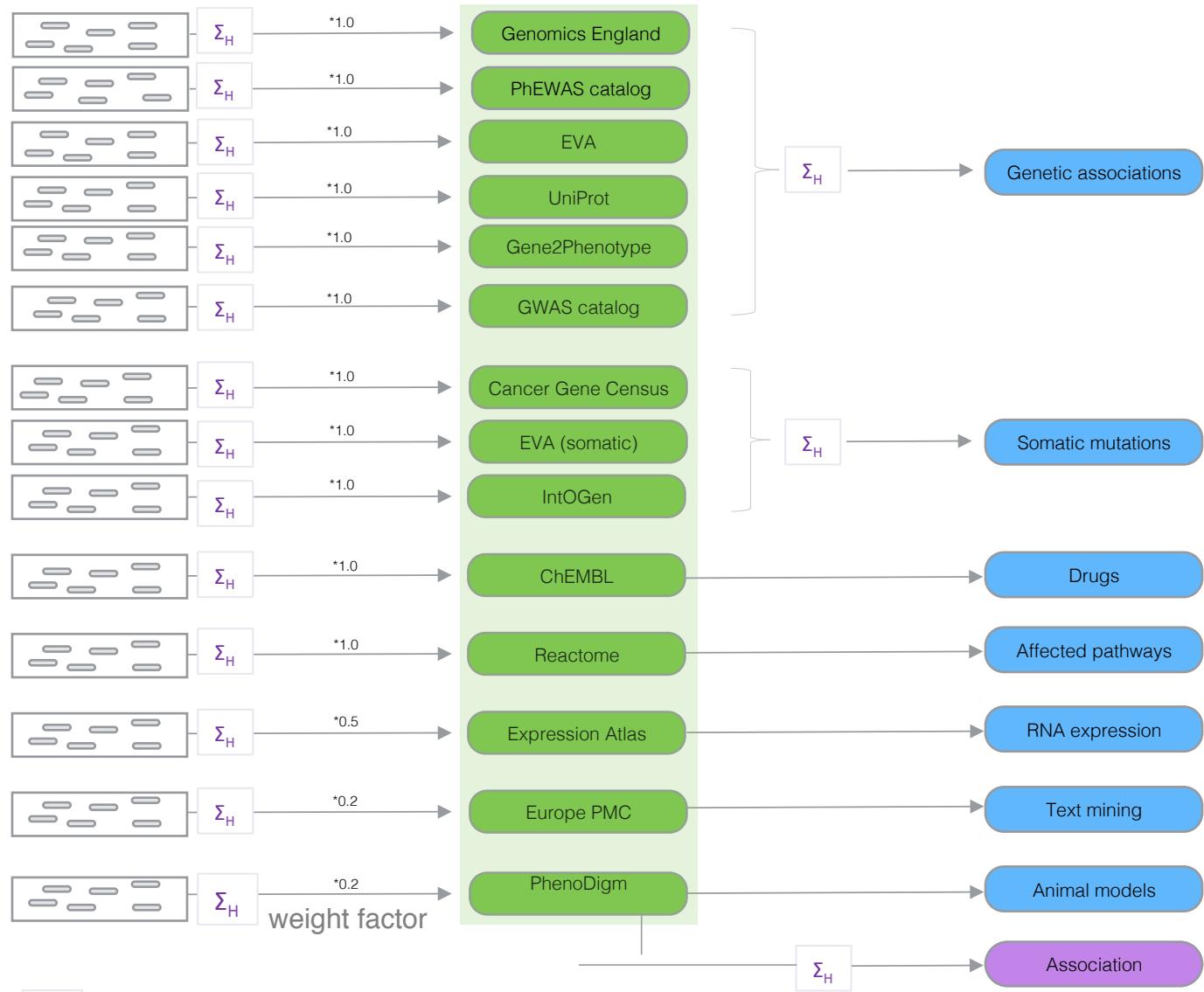
# Four-tier framework

Score: 0 to 1 (max)

Calculated at 4 levels:

- Evidence
- Data source
- Data type
- Overall

Aggregation with  
(harmonic sum)  $\Sigma_H$



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

$$\Sigma_H = S_1 + S_2/2^2 + S_3/3^2 + S_4/4^2 + \dots + S_i/i^2$$

# Factors affecting the relative strength of an evidence

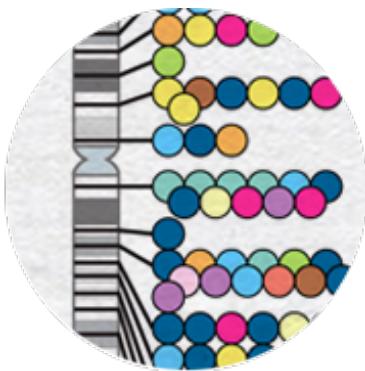
e.g. *GWAS Catalog*

$$S = f * s * c$$

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

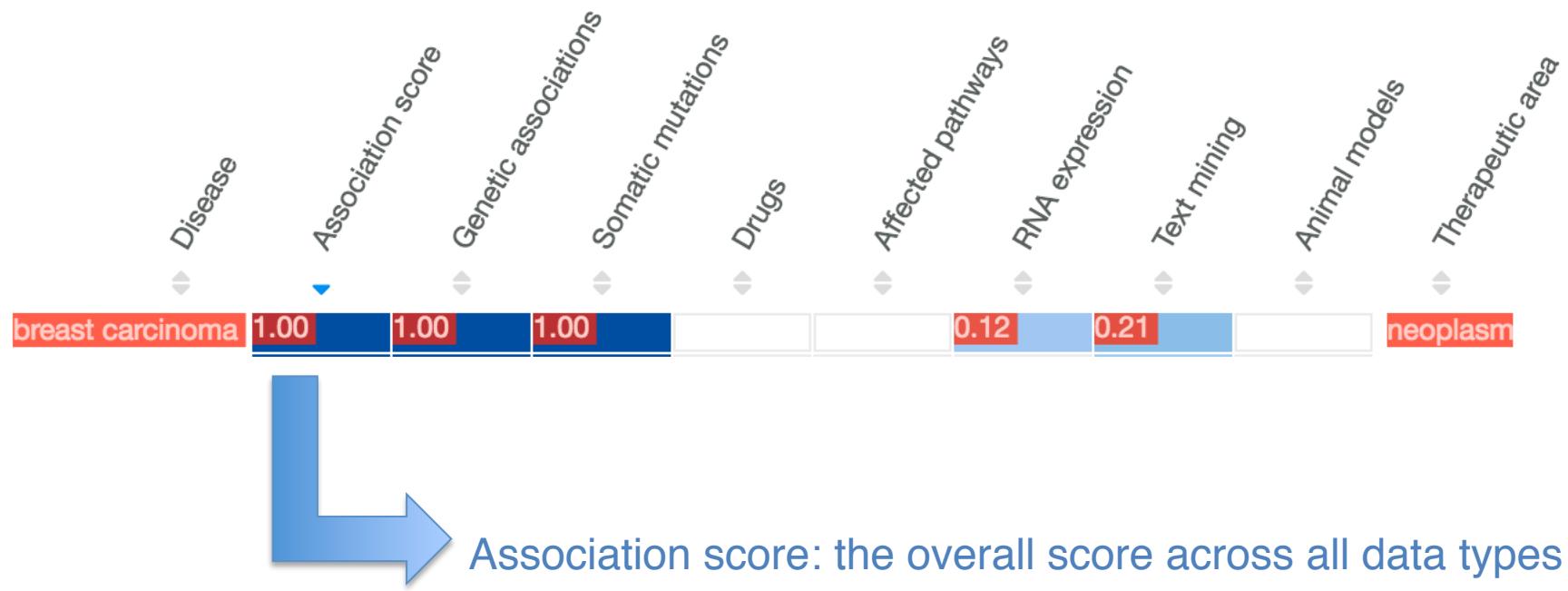


f = sample size (cases versus controls)

s = predicted functional consequence (VEP)

c = *p* value reported in the paper

# Ranking the target-disease association



- Based on the data sources
- Different weight applied:

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

# In addition to T-D associations

- Everything you wanted to know about...



... but were afraid to ask.

The screenshot shows the Open Targets Platform interface. At the top, there's a blue header bar with the "Open Targets Platform" logo and a search icon. Below the header, the main content area has a title "Evidence for CD86 in multiple sclerosis". On the left, there's a circular diagram with several overlapping ovals representing different data sources: "Text Mining", "RNA", "Pathways", "Drugs", "Mouse", "Somatic", and "Genetics". To the right of this diagram, there's a box for the "CD86" molecule, which includes its name, synonym information ("Synonyms: B7.2, B7-2, CD28LG2"), and a link to a "Target profile page". Further down, there's another box for "multiple sclerosis", listing its synonyms ("MS (Multiple Sclerosis), MS, MULTIPLE SCLEROSIS ACUTE FULMINATING, Disseminated Sclerosis, Sclerosis...") and a link to a "Disease profile page".

Evidence for CD86 in multiple sclerosis

CD86

CD86 molecule

Synonyms: B7.2, B7-2, CD28LG2

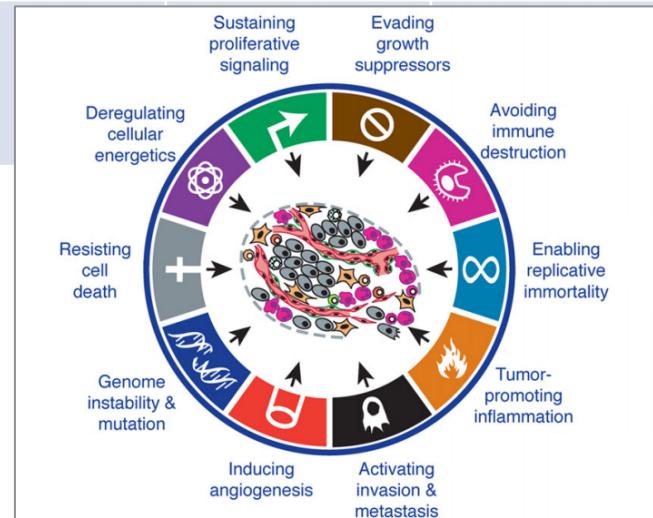
Target profile page

multiple sclerosis

Synonyms: MS (Multiple Sclerosis), MS, MULTIPLE SCLEROSIS ACUTE FULMINATING, Disseminated Sclerosis, Sclerosis...

Disease profile page

# Profile of a drug target

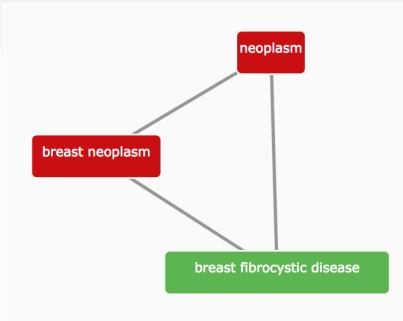
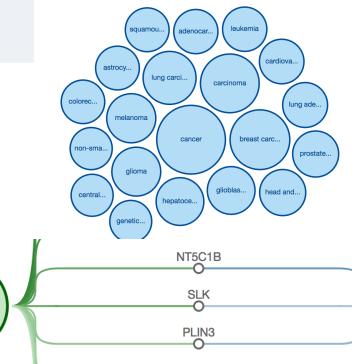
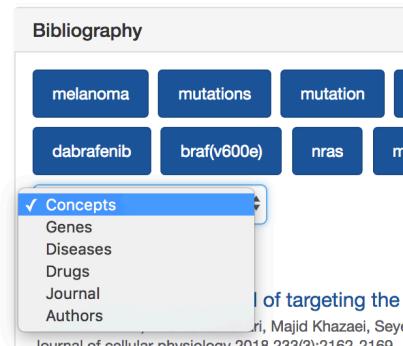
Protein	Drugs	Pathways interactions	RNA and protein baseline expression	Variants, isoforms and genomic context	Mouse phenotypes	Bibliography
		Protein Interactions 	 Expression Atlas			
 Protein Data Bank in Europe				Gene tree		Library/LINK
Description Synonyms Gene Ontology Protein Structure		Similar Targets 				

Extra, extra, extra!

Cancer hallmarks in our latest release!

<http://www.targetvalidation.org/target/ENSG00000141510>

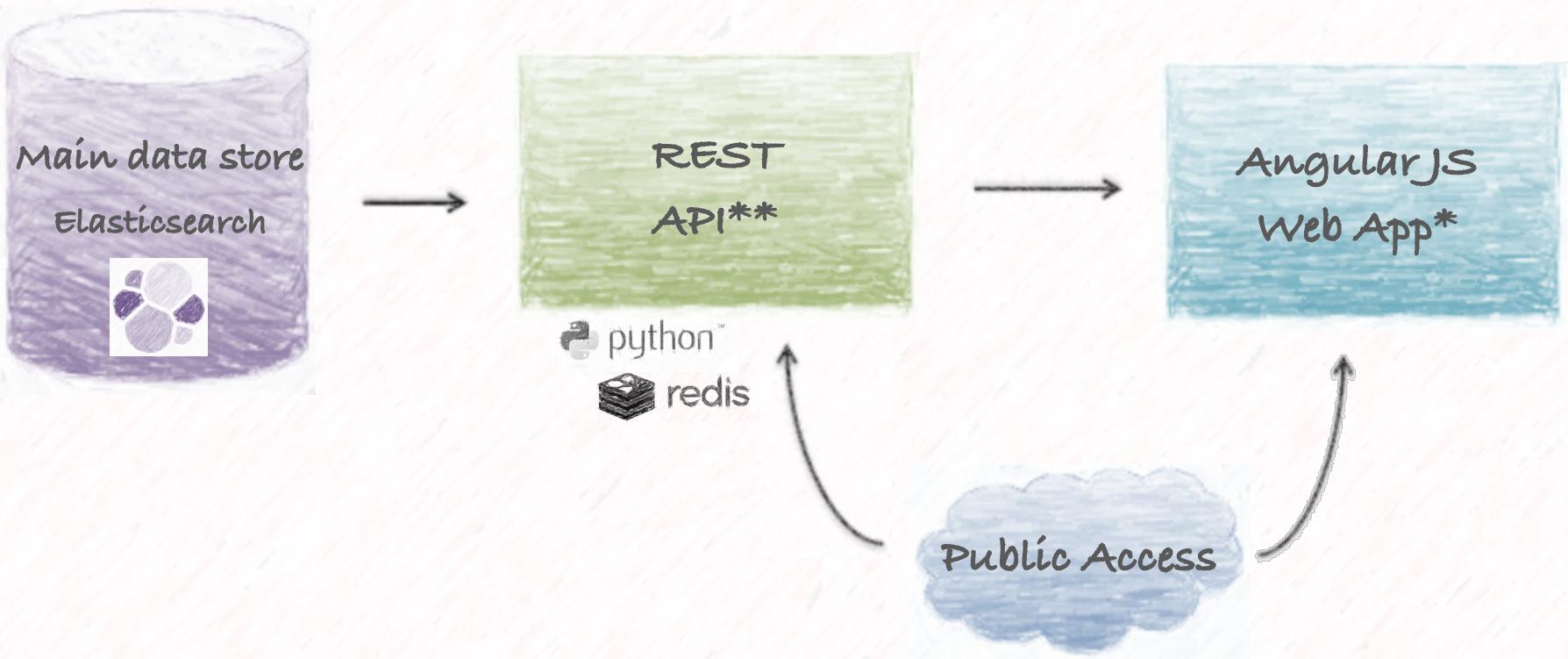
# Profile of a disease

Classification	Drugs	Similar diseases	Bibliography
 		 	 Open Targets Library/LINK 

[http://www.targetvalidation.org/disease/Orphanet\\_262](http://www.targetvalidation.org/disease/Orphanet_262)

I of targeting the  
Journal of cellular physiology 2018 233(3):2162-2169  
Majid Khazaei, Seyyed

# How to access all of this



\* UI: first released in December 2015

<https://www.targetvalidation.org>

\*\* API first release in April 2016

<https://api.opentargets.io>

# In these ~3.5 hours

- Introduction
- Open Targets Platform: live demos
- Exercises
- Beyond the Platform
- 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 20 possible targets for multiple myeloma.

Are these targets represented in other diseases?

Which pathways are represented in this set of targets?



# In these ~3.5 hours

- Introduction
- Open Targets Platform: live demos
- **Exercises**
- Beyond the Platform
- Wrap up

# Coursebook

[bit.ly/TBOS\\_exercises](https://bit.ly/TBOS_exercises)

pages 29-33

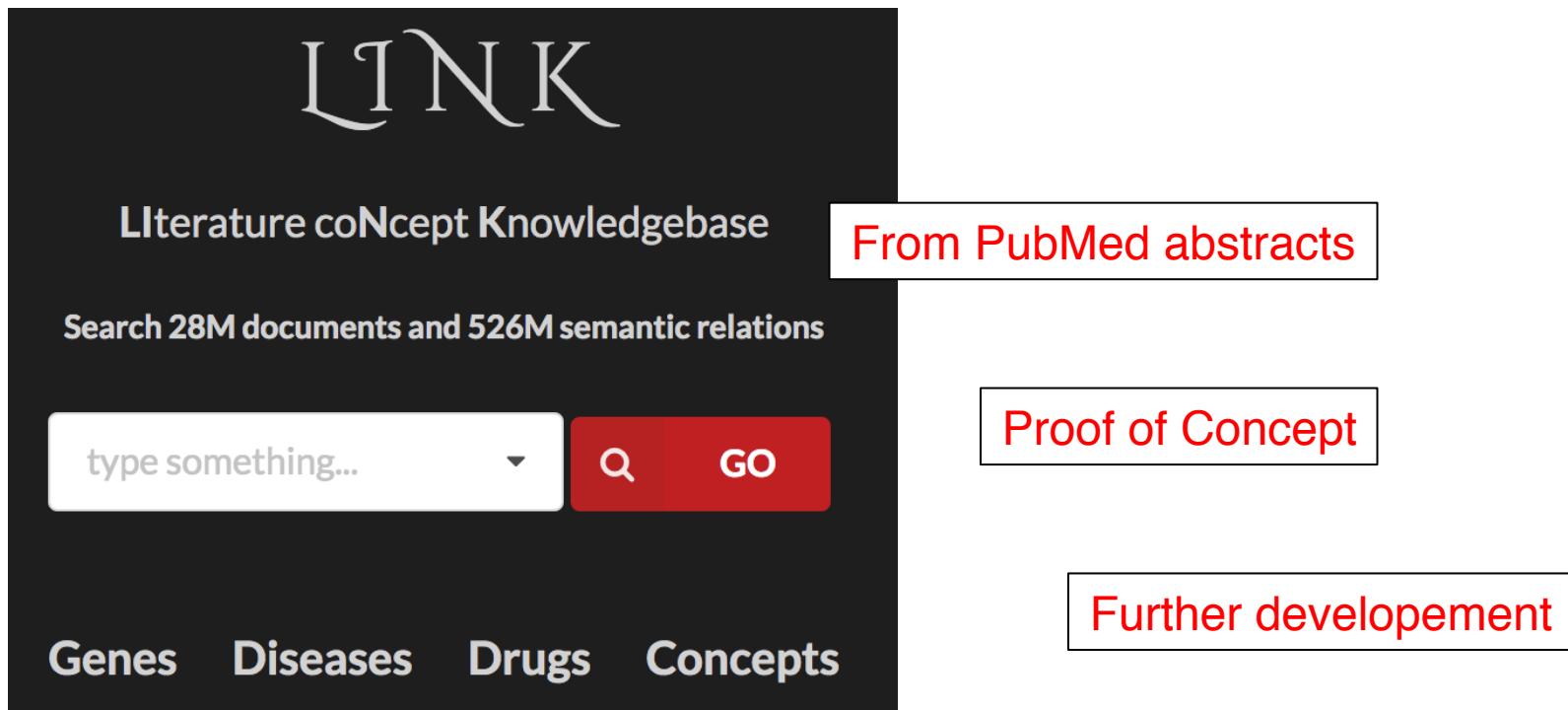
# In these ~3.5 hours

- Introduction
- Open Targets Platform: live demos
- Exercises
- Beyond the Platform
- Wrap up

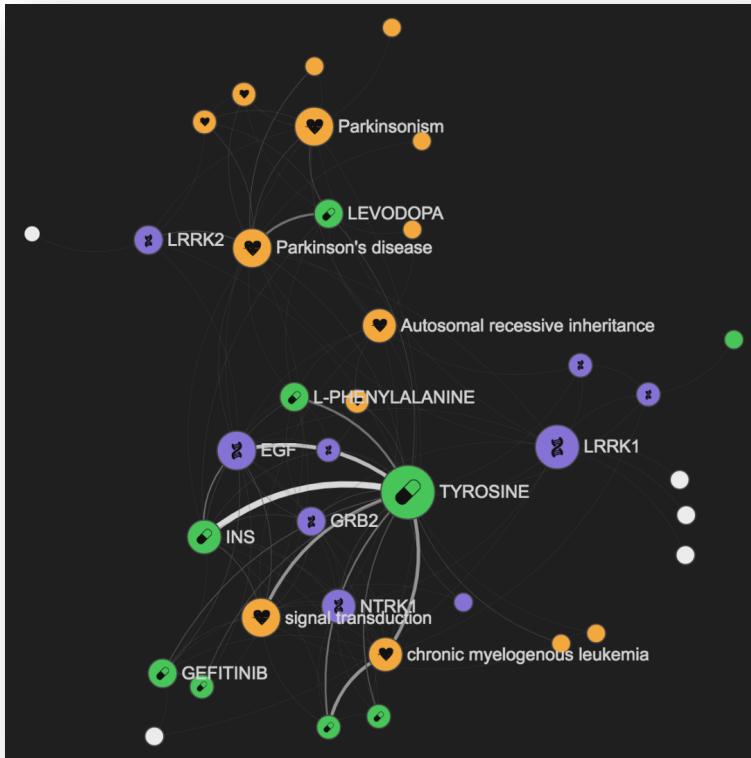
# LINK

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

- LINK: Literature coNcept Knowledgebase
- Subject / predicate / object structured relations



# Addressing text mining shortcomings



- Entities: genes, diseases, drugs
- Concepts extracted via NLP  
(Natural Language Processing)
- 28 M documents, 500 M relations
- <http://blog.opentargets.org/link/>

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

# DoRothEA

dorothea.opentargets.io



## DoRothEA Discriminant Regulon Expression Analysis

Home

TF-Drug Interactions

TF-Pharmacogenomic Marker Interactions

- Candidate TF-drug interactions in cancer
- 1000 cancer cell lines
- 265 anti-cancer compounds
- 127 transcription factors

# Example: Rapamycin

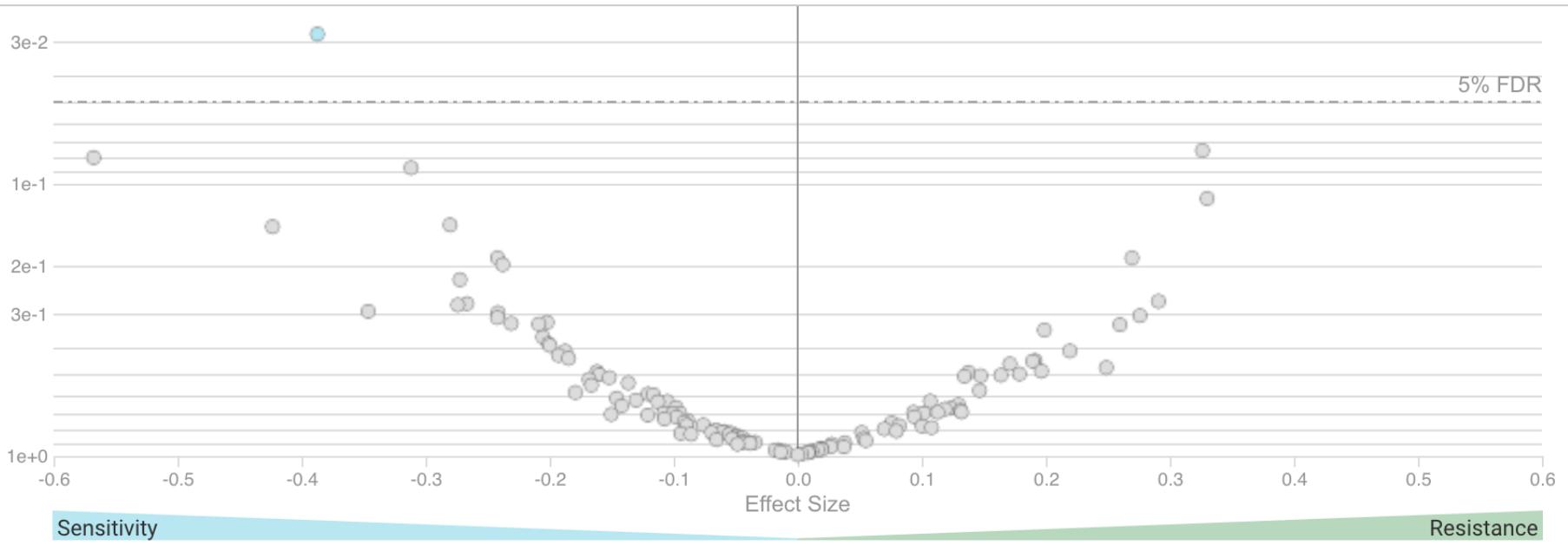
## Filter interactions

- No filtering
- Filter by drug
- Filter by transcription factor

Drug

Rapamycin

- ~ 1000 cancer cell lines
- 265 anti-cancer compounds
- 127 transcription factors



# In these ~3.5 hours

- Introduction
- Live demos
- Exercises
- Wrap up

# Open Targets Platform

- Resource of integrated multiomics
- Added value (e.g. see)
- Graphical
- Not yet a bespoke resource that works for you?

User design research.

targeted

diseases

2.3 M  
associations

5.9 M  
evidence

February 2018 release

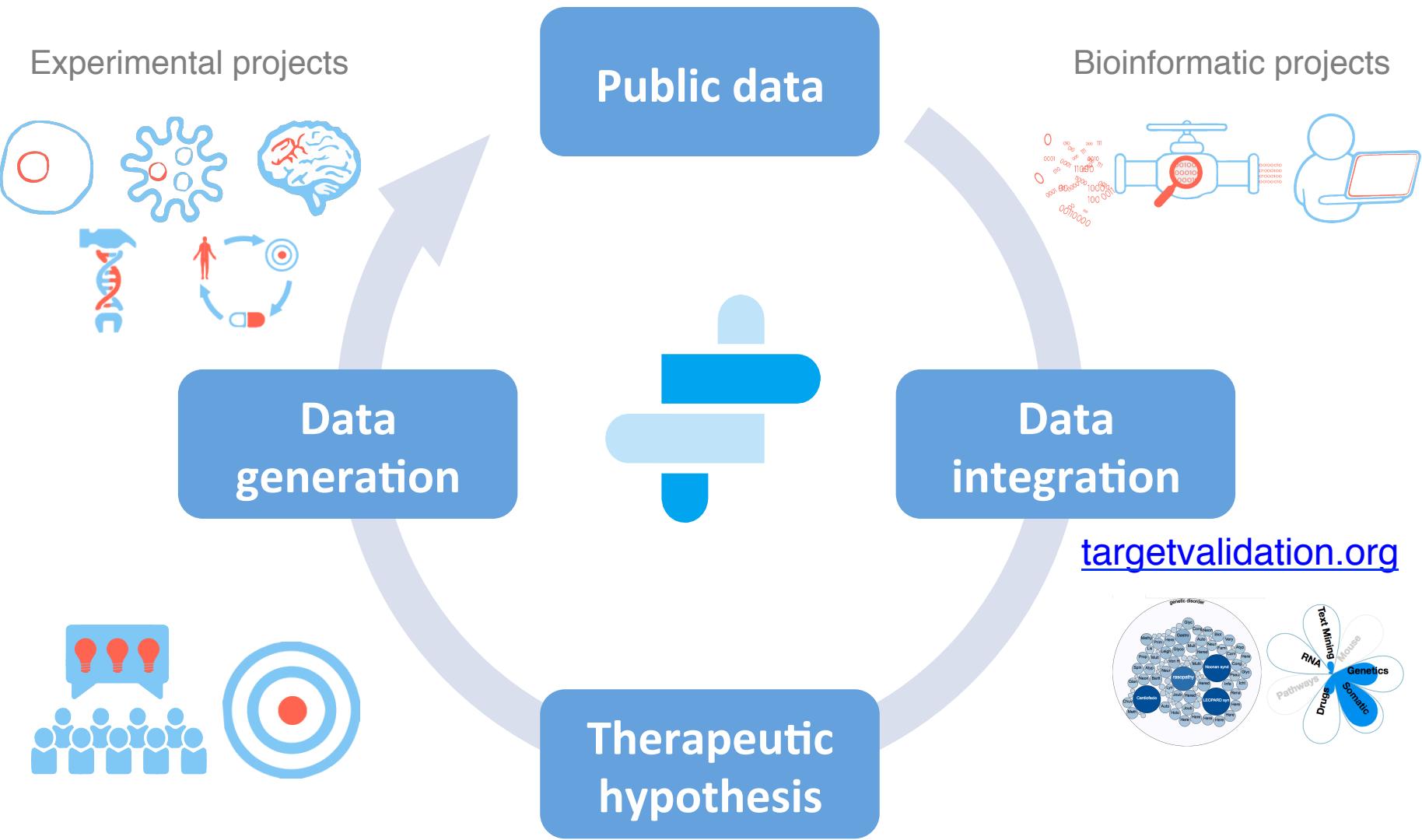
# Your feedback is important

<https://bit.ly/takeda-190318>

# Help shape the navigation bar

<http://bit.ly/ot-nav-cs>

# Virtuous cycle in Open Targets



Concurrent  
[www.opentargets.org/projects](http://www.opentargets.org/projects)

# We support decision-making

Which targets are associated with a disease?

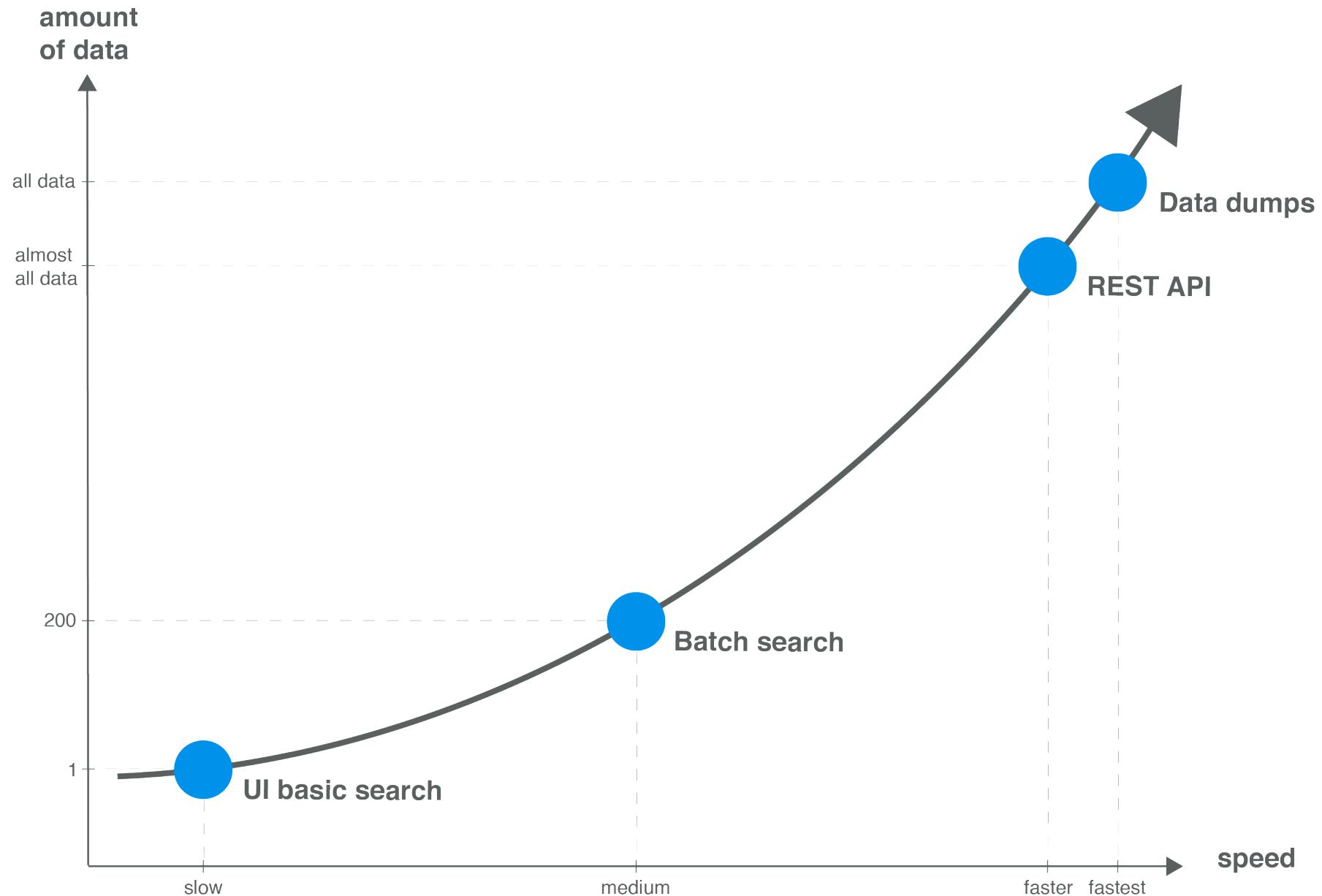
Can I find out about the mechanisms of the disease?

Are there FDA drugs for this association?



...

# Other modes of Platform data access



# How to cite us

Published online 8 December 2016

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

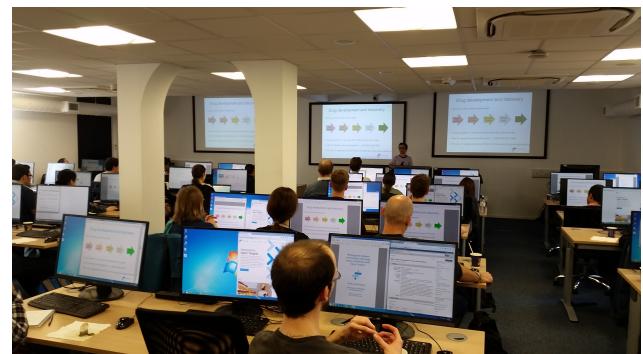
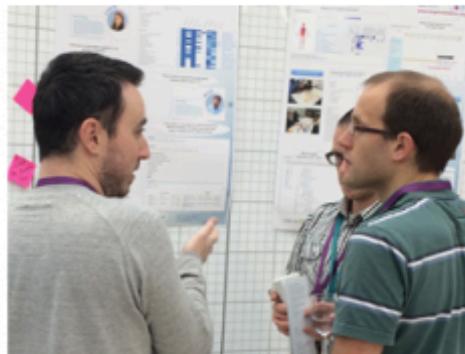
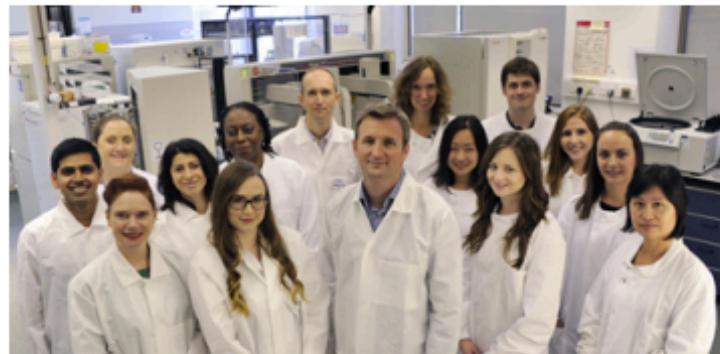
## Open Targets: a platform for therapeutic target identification and validation

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

<sup>1</sup>Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, <sup>2</sup>GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK, <sup>3</sup>Biogen, Cambridge, MA 02142, USA, <sup>4</sup>European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, <sup>5</sup>Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK and <sup>6</sup>National 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](mailto:support@targetvalidation.org)



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



[@targetvalidate](#)



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

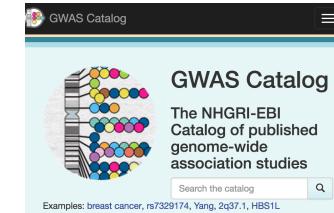
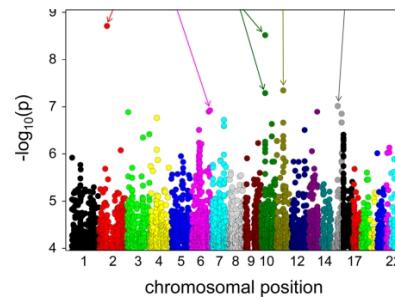
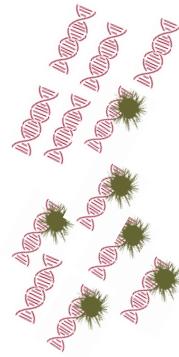
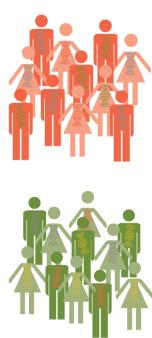


[blog.opentargets.org/](http://blog.opentargets.org/)

# Specs on data sources for the Open Targets Platform

# Data sources: GWAS catalog

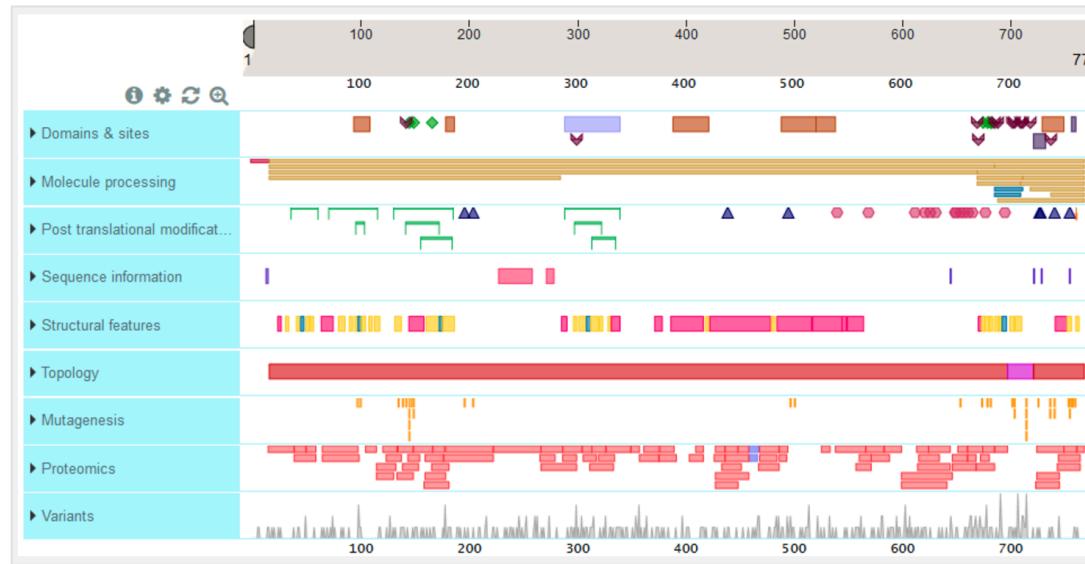
- Genome Wide Association Studies
- Array-based chips → genotyping 100,000 SNPs genomewide



Open Targets

# Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients

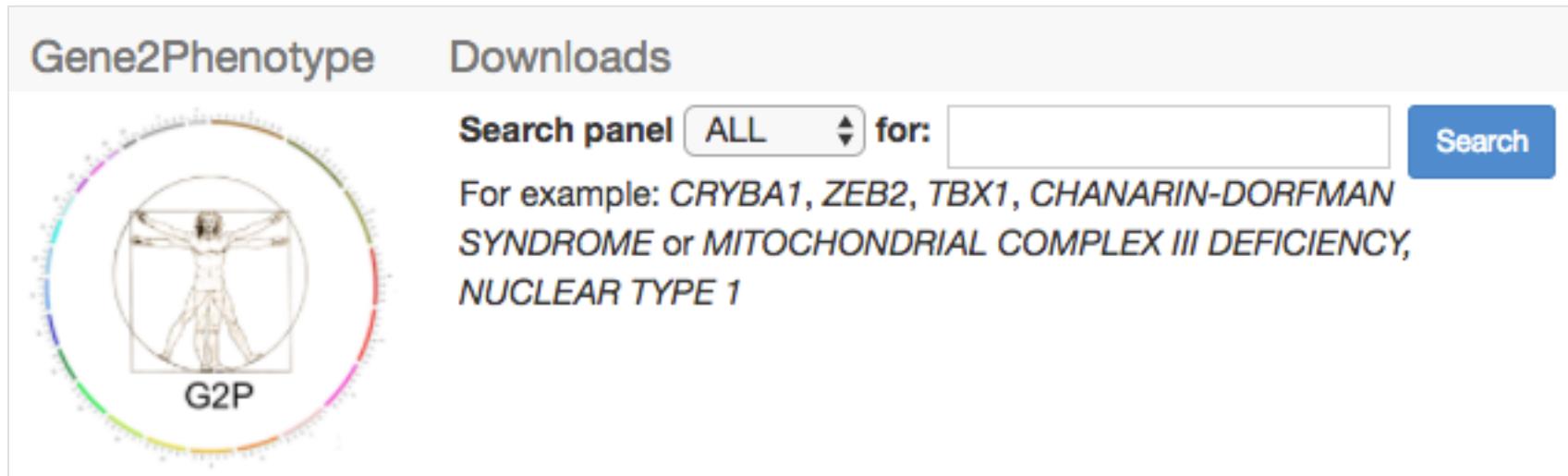


EMBL-EBI train online



Open Targets

# Data sources: Gene2Phenotype



Gene2Phenotype      Downloads

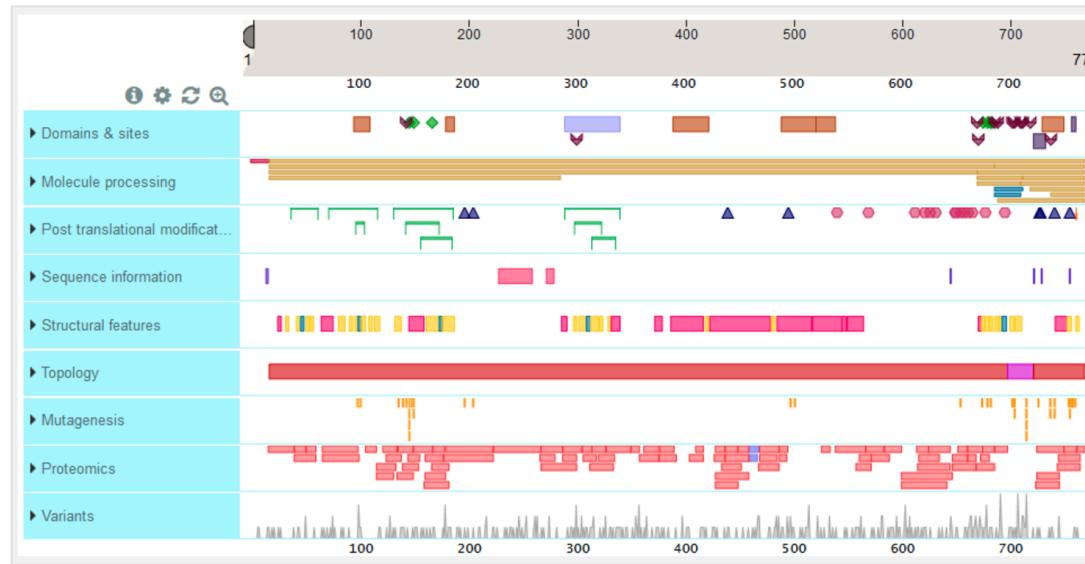
Search panel ALL for:  **Search**

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



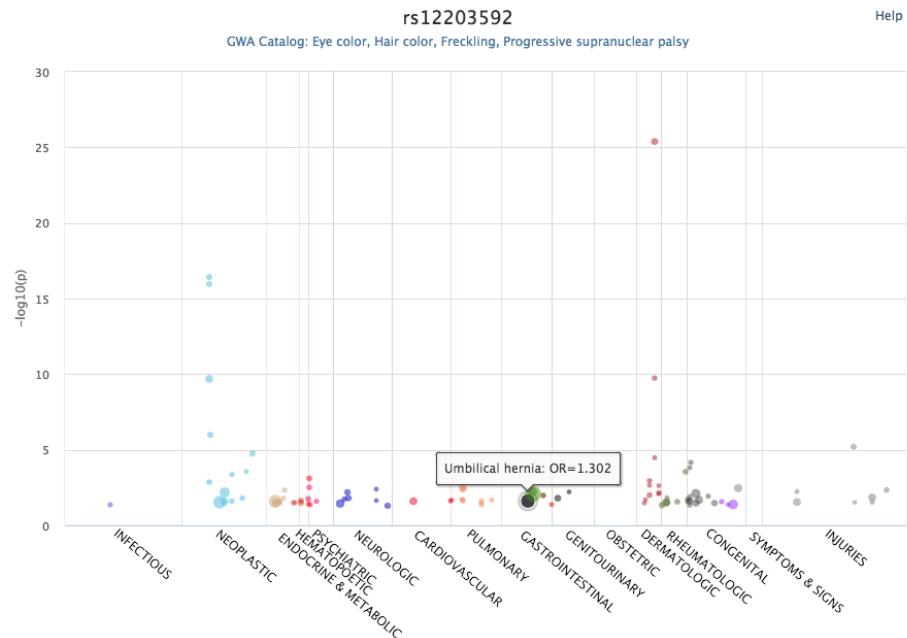
EMBL-EBI train online



Open Targets

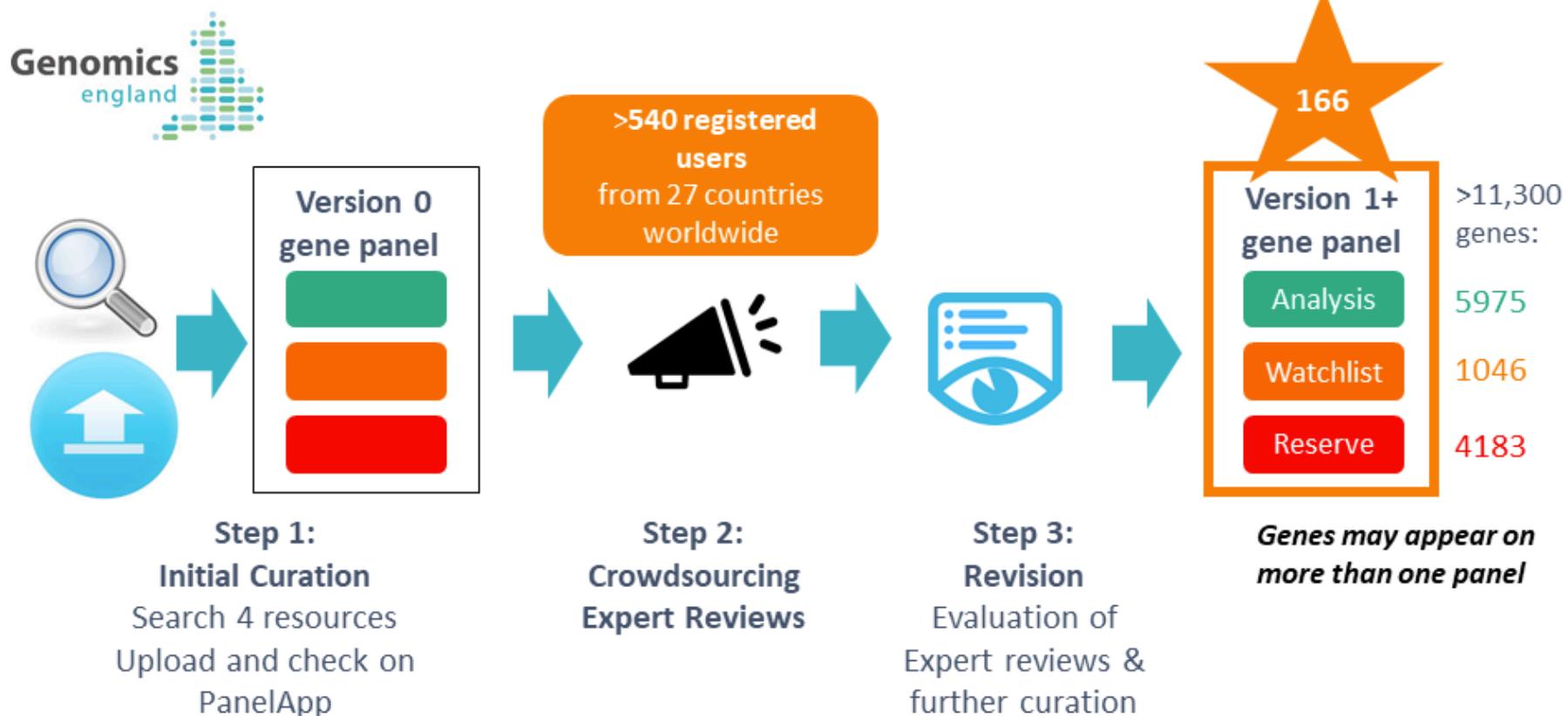
# Data sources: PheWAS

- 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 dark blue), GA4GH, API, FAQ, and Feedback. Below the navigation is a search bar with a magnifying glass icon and a "Filter" button. To the right of the search bar is the title "ClinVar Browser" with an information icon. The main content area displays a table of ClinVar variants. The table has columns for Position, Affecte..., Most Severe Consequence..., Trait, Clinical Significance, and ClinVar ID. The first few rows show variants for the MSH6 gene on chromosome 2, with positions ranging from 480,000,000 to 490,000,000. The clinical significance for most variants is listed as Benign.

...	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

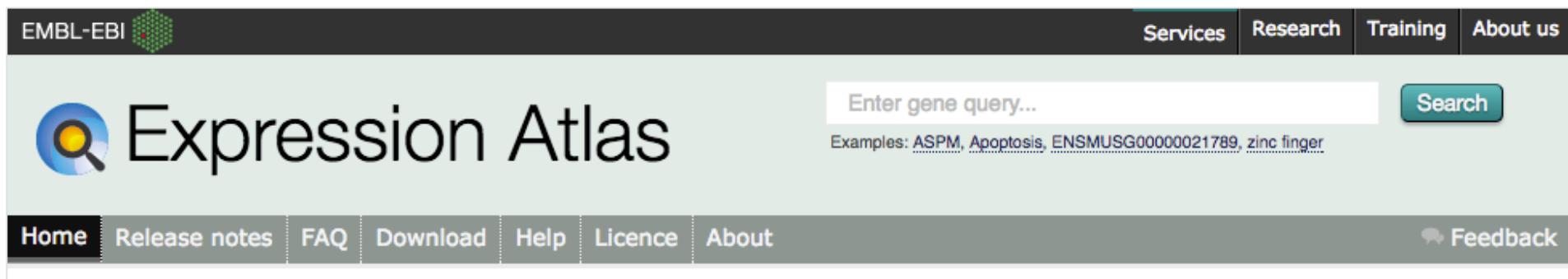


EMBL-EBI train online



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

# Aggregating scores across the data

- Using a mathematical function, the harmonic sum\*

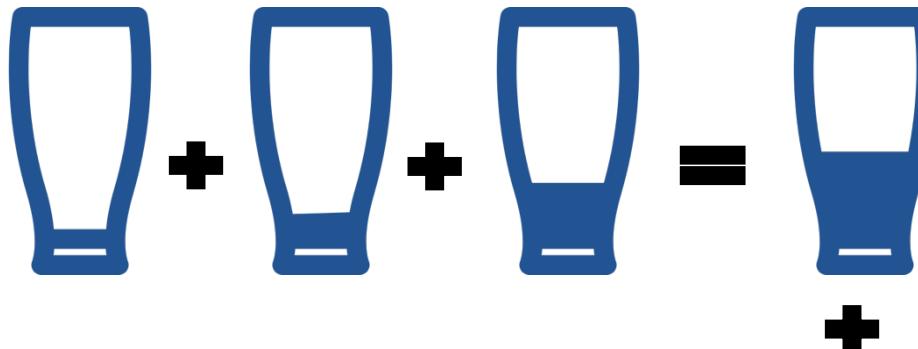
$$S_{1..i} = S_1 + \frac{S_2}{2^2} + \frac{S_3}{3^2} + \frac{S_4}{4^2} \dots + \frac{S_i}{i^2}$$

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

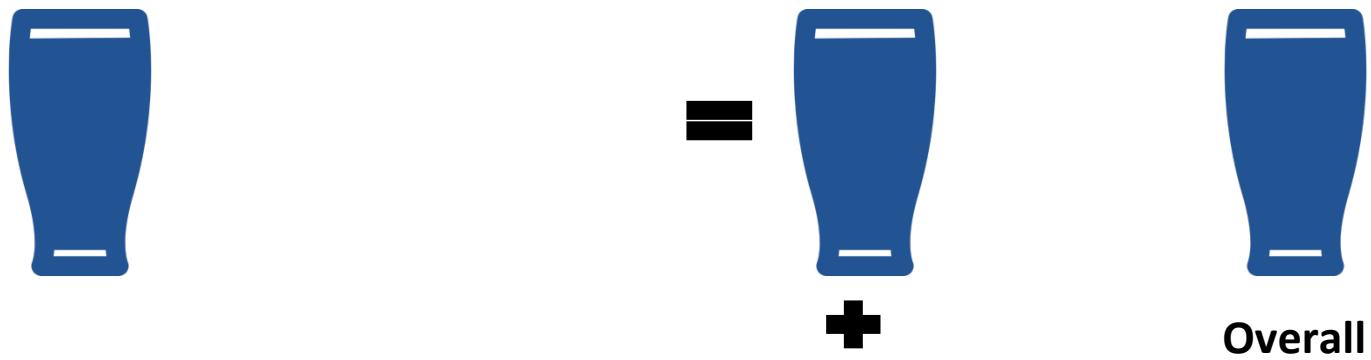
- Advantages:
  - A) account for replication
  - B) deflate the effect of large amounts of data e.g. text mining

# Target-Disease Association Score

EuropePMC  
(Text Mining)

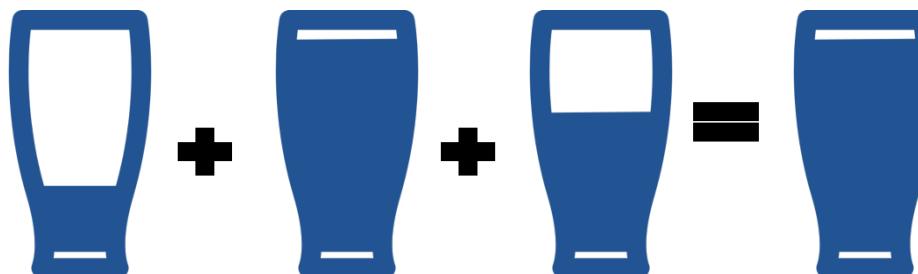


UniProt  
(Manual Curation)



Overall

ChEMBL  
(Manual Curation)



VERY simplified diagram

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

# REST API calls: some examples\*

GET

/public/search

[https://api.opentargets.io/v3/platform/public/search?q=EFO\\_0003767](https://api.opentargets.io/v3/platform/public/search?q=EFO_0003767)

<https://api.opentargets.io/v3/platform/public/search?q=asthma>

GET

/public/association/filter

[https://api.opentargets.io/v3/platform/public/association/filter?  
target=ENSG00000110324&direct=false&fields=is\\_direct&fields=disease.efo\\_info.lab  
el&size=100](https://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000110324&direct=false&fields=is_direct&fields=disease.efo_info.label&size=100)

GET

/public/evidence/filter

[https://api.opentargets.io/v3/platform/public/evidence/filter?  
target=ENSG00000141867&disease=EFO\\_0000565&datatype=expression\\_atl  
as&size=100&format=json](https://api.opentargets.io/v3/platform/public/evidence/filter?target=ENSG00000141867&disease=EFO_0000565&datatype=expression_atlas&size=100&format=json)

# Several ways to run our REST endpoints

- Paste the URL in the location bar in a browser
- Use the terminal window (e.g. with CURL command)
- Use our free clients (i.e. Python\* and R)
- Call them from your own application/workflow

\* <http://opentargets.readthedocs.io/en/stable/index.html>

# Python and R clients for the REST API

opentargets  
latest

Search docs

Tutorial  
High Level API  
Low Level API  
Code Documentation  
Changelog

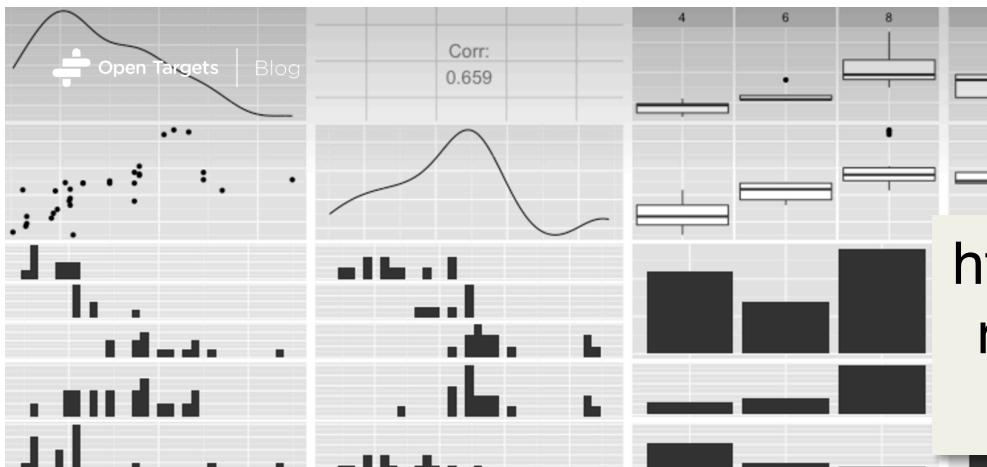
Docs » opentargets - Python client for targetvalidation.org

Edit on GitHub

## opentargets - Python client for targetvalidation.org

opentargets is the official python client for the [Open Targets REST API](#) at [targetvalidation.org](#)

<http://opentargets.readthedocs.io>



[https://blog.opentargets.org/  
rest-api-exploration-using-  
an-r-client/](https://blog.opentargets.org/rest-api-exploration-using-an-r-client/)

How to access Open Targets  
with R