

Open Targets: enabling systematic target discovery and validation

Takeda California (TCAL)

San Diego

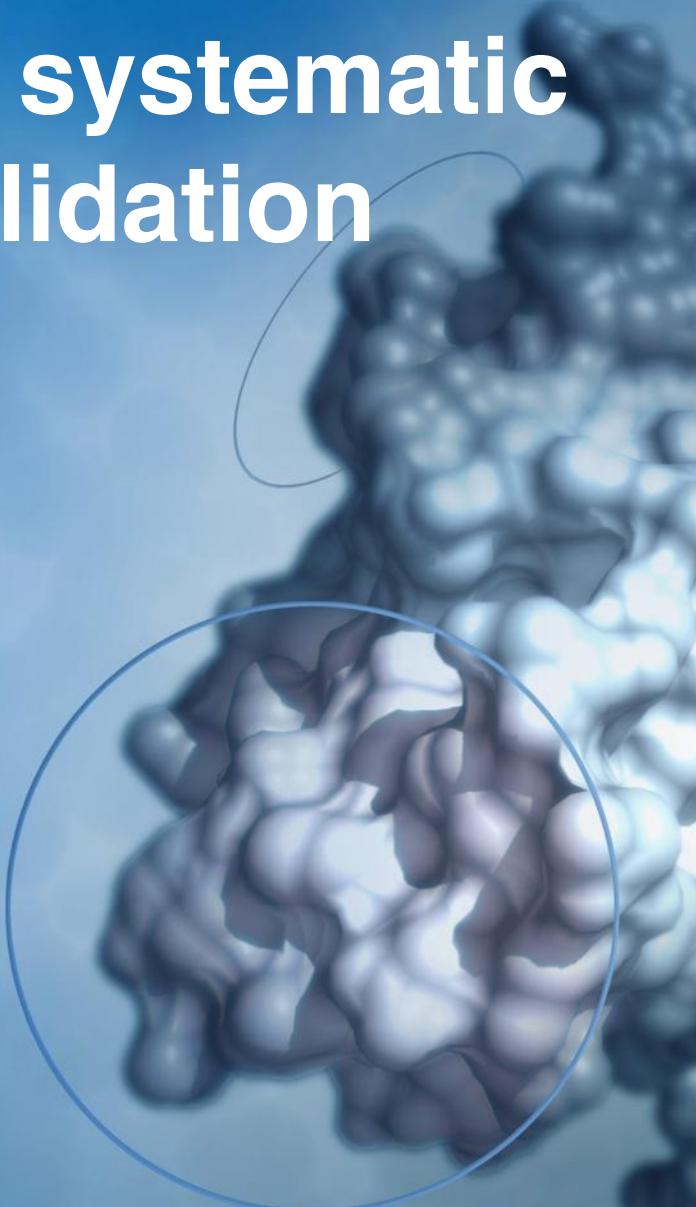
March 16th 2018

Denise Carvalho-Silva, PhD

Scientific Outreach Lead

Open Targets / EMBL-EBI

Wellcome Genome Campus, United Kingdom



Aims



Open Targets Partnership

How to navigate the
Open Targets Platform

How to query the Open
Targets REST server

How to get help

Materials

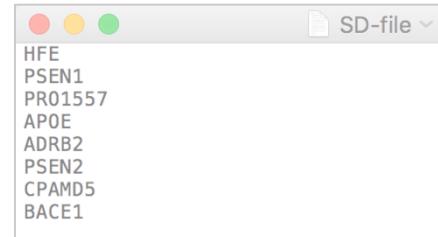
- bit.ly/SD-slides



- bit.ly/SC-exercises



- bit.ly/SD-file-takeda
- bit.ly.com/SD-batch-takeda



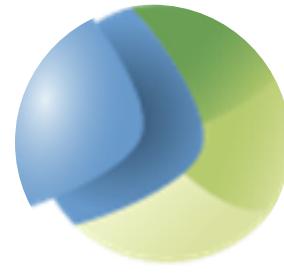
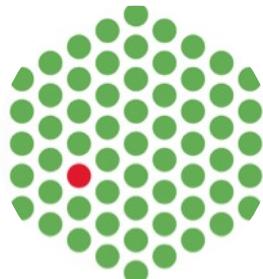
In these ~ 3.5 hours

- Introduction
- Live demos
- Exercises
- Wrap up



Our Vision

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



Open 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



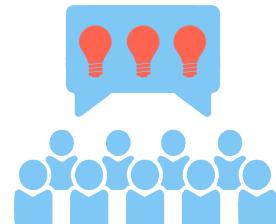
Human cell
models



Advanced
data
analysis



Open source
software, rapid
data release, open
publication



Make target
decisions
together



Open Targets

What do users want?



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

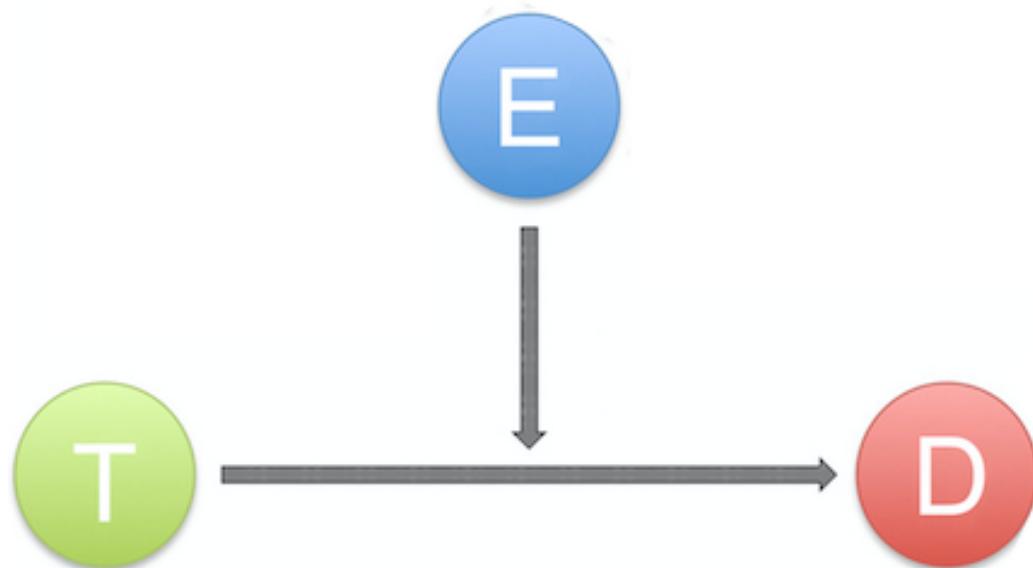


I know Open Targets Platform is all you need! It's a one-stop shop for as much data as possible, plus links to **the original source** for my own assessment.

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

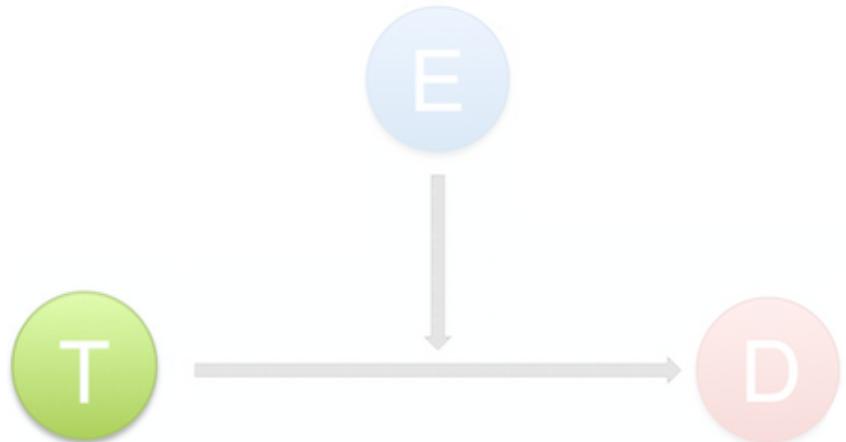
Open Targets Platform

- Resource for associations between targets and diseases
- Drug discovery, disease biology, translational medicine
- Data driven



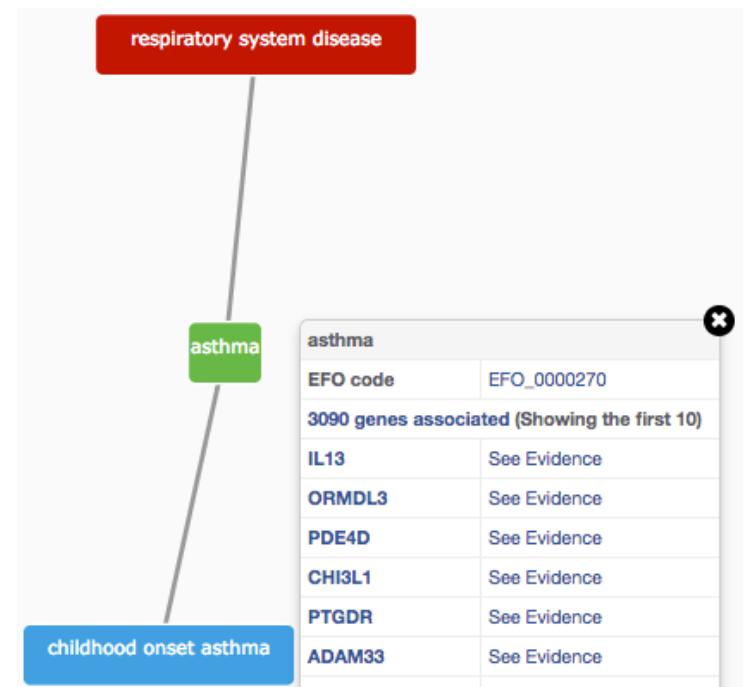
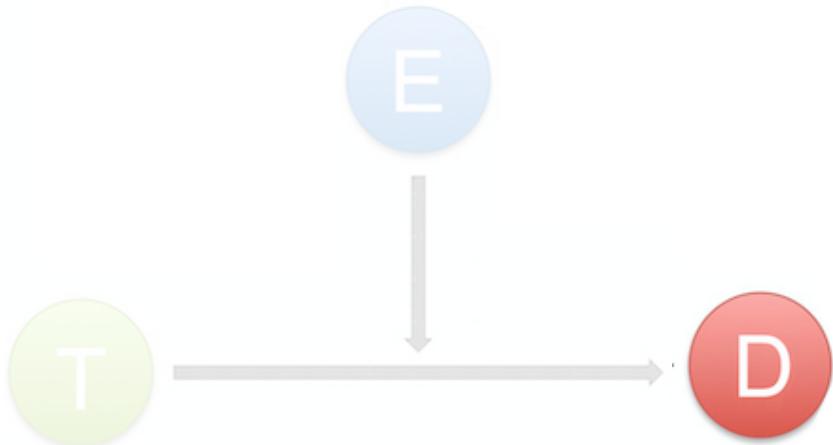
Our targets → genes or proteins

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

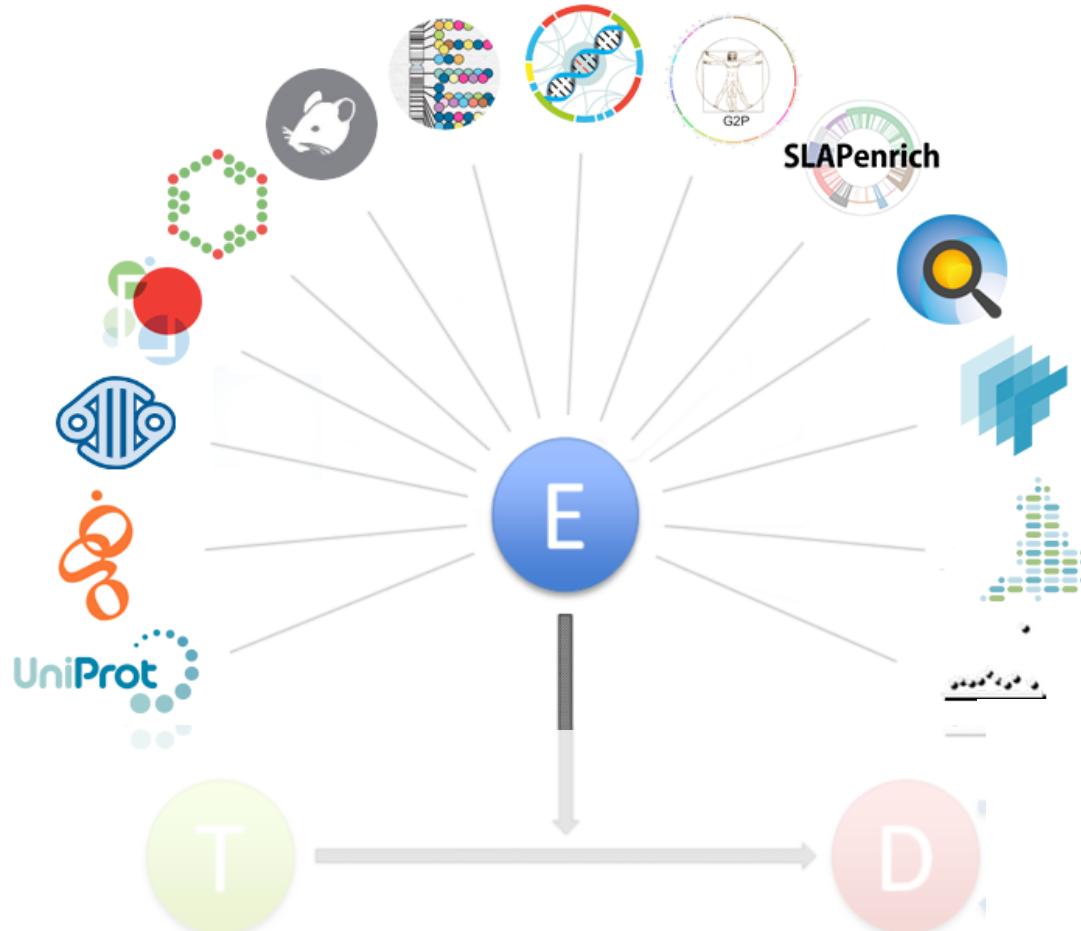


Our diseases

- 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 the associations

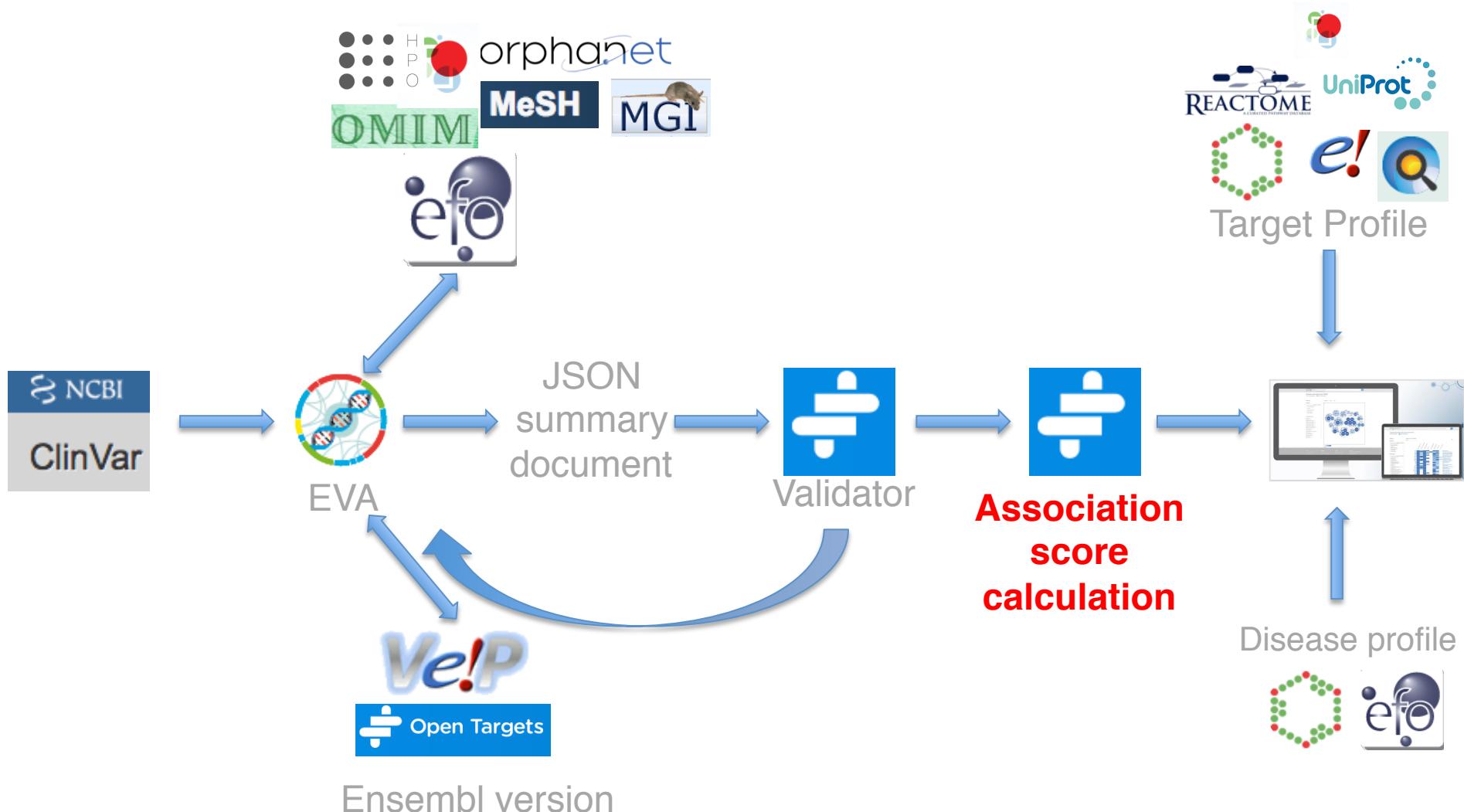


www.targetvalidation.org/data_sources

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						

T-D association data* flow



* e.g. genetic variants from EVA

Association score

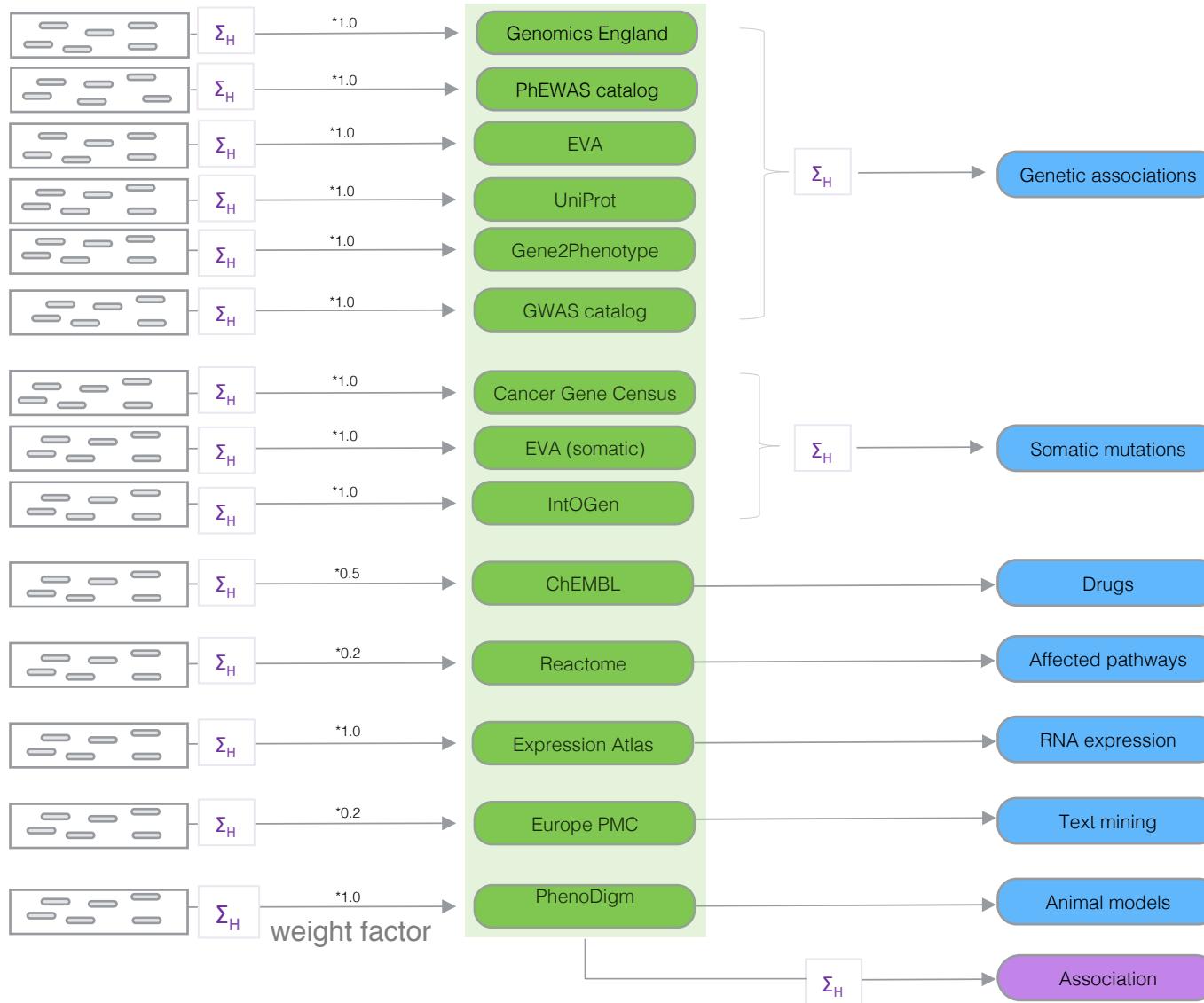


How confident can I be in
the associations that
Open Targets reports?

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)



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

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

Factors affecting the relative strength of an evidence

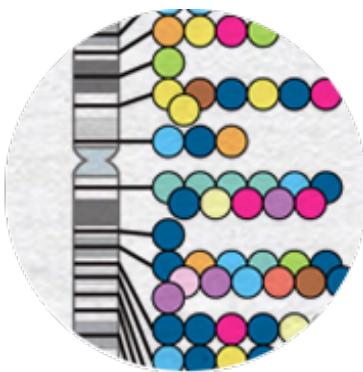
e.g. *GWAS Catalog*

$$S = f * s * c$$

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

Aggregating scores across the data

- Using a mathematical function, the harmonic sum*

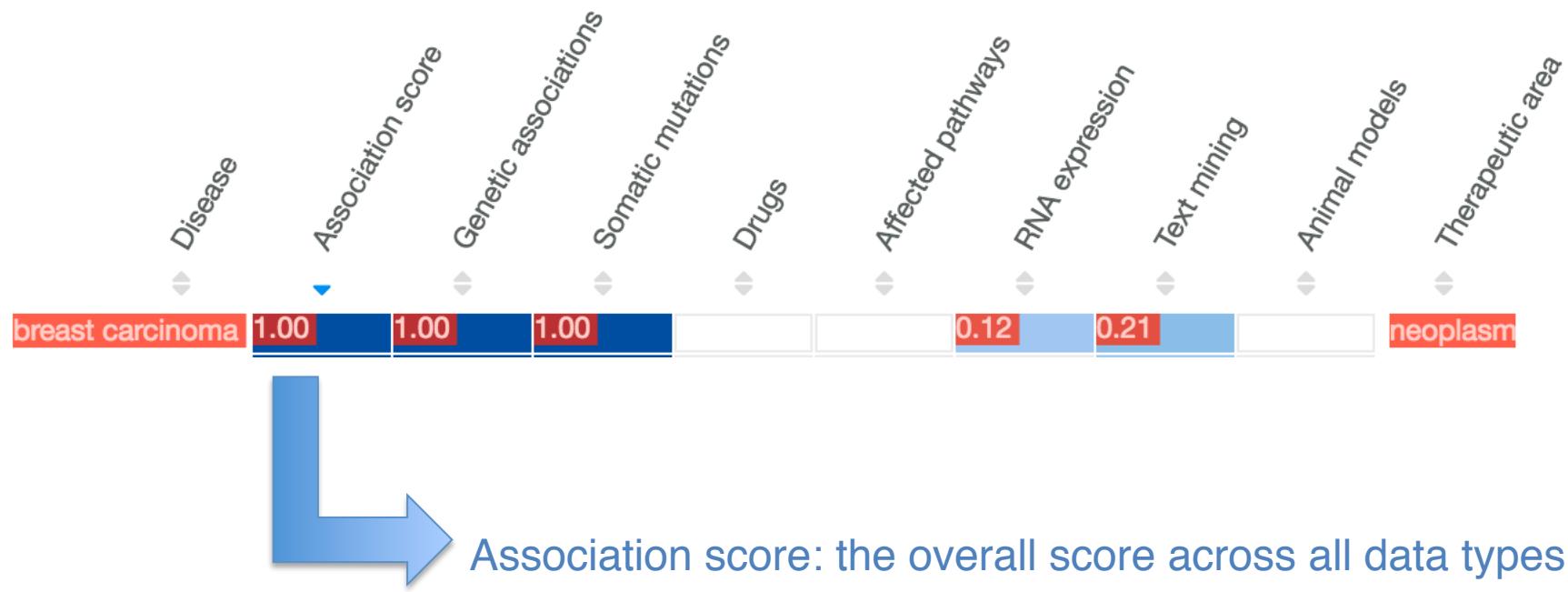
$$S_{1..i} = S_1 + \frac{S_2}{2^2} + \frac{S_3}{3^2} + \frac{S_4}{4^2} \dots + \frac{S_i}{i^2}$$

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

* PMID: 19107201, PMID: 20118918

Ranking the target-disease association



- Based on the data sources
- Different weight applied:

genetic association = drugs = mutations = pathways > RNA expression > animal models = 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...)

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". To the right, there's a "Target profile page" for "CD86" (CD86 molecule, Synonyms: B7.2, B7-2, CD28LG2). Further down, there's a "Disease profile page" for "multiple sclerosis" (Synonyms: MS (Multiple Sclerosis), MS, MULTIPLE SCLEROSIS ACUTE FULMINATING, Disseminated Sclerosis, Sclerosis...). In the center, there's a circular diagram with several overlapping ovals representing different evidence sources: "Text Mining", "RNA", "Pathways", "Drugs", "Signatures", "Mouse", "Somatic", and "Genetics".

Evidence for CD86 in multiple sclerosis

CD86
CD86 molecule
Synonyms: B7.2, B7-2, CD28LG2

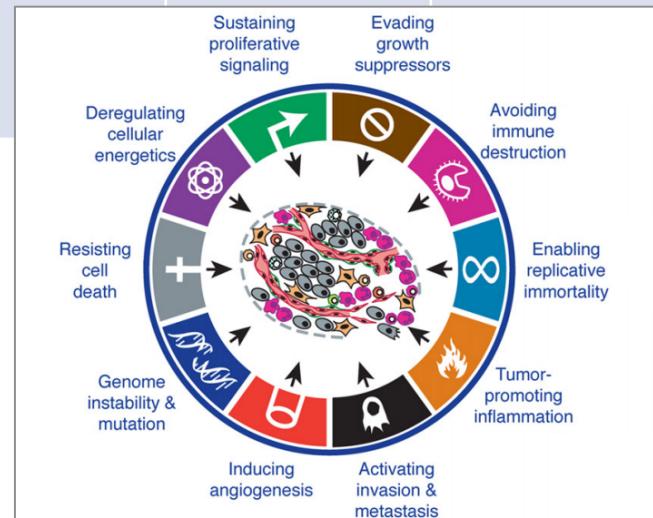
multiple sclerosis
Synonyms: MS (Multiple Sclerosis), MS, MULTIPLE SCLEROSIS ACUTE FULMINATING, Disseminated Sclerosis, Sclerosis...

Target profile page

Disease profile page

Text Mining
RNA
Pathways
Drugs
Signatures
Mouse
Somatic
Genetics

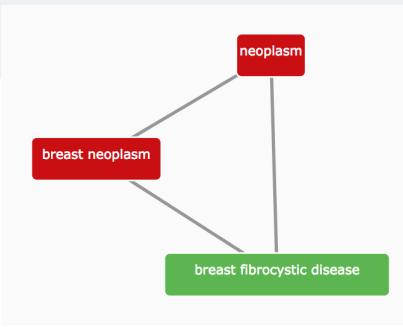
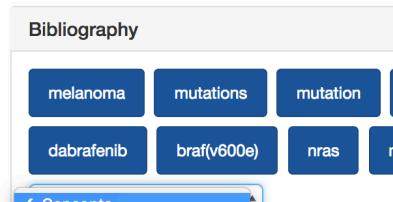
Profile of a drug target

Protein	Drugs	Pathways interactions	RNA and protein baseline expression	Variants, isoforms and genomic context	Mouse phenotypes	Bibliography
		Protein Interactions 	Expression Atlas 	e!Ensembl 		
 Protein Data Bank in Europe		reactome 	GTEX 	Gene tree		Library/LINK
Description Synonyms Gene Ontology Protein Structure		Similar Targets 	THE HUMAN PROTEIN ATLAS 	 <p>The diagram illustrates the ten hallmarks of cancer, represented by a central cell surrounded by ten colored segments, each associated with a specific biological process:</p> <ul style="list-style-type: none"> Sustaining proliferative signaling Evading growth suppressors Avoiding immune destruction Enabling replicative immortality Tumor-promoting inflammation Activating invasion & metastasis Inducing angiogenesis Genome instability & mutation Resisting cell death Deregulating cellular energetics 		

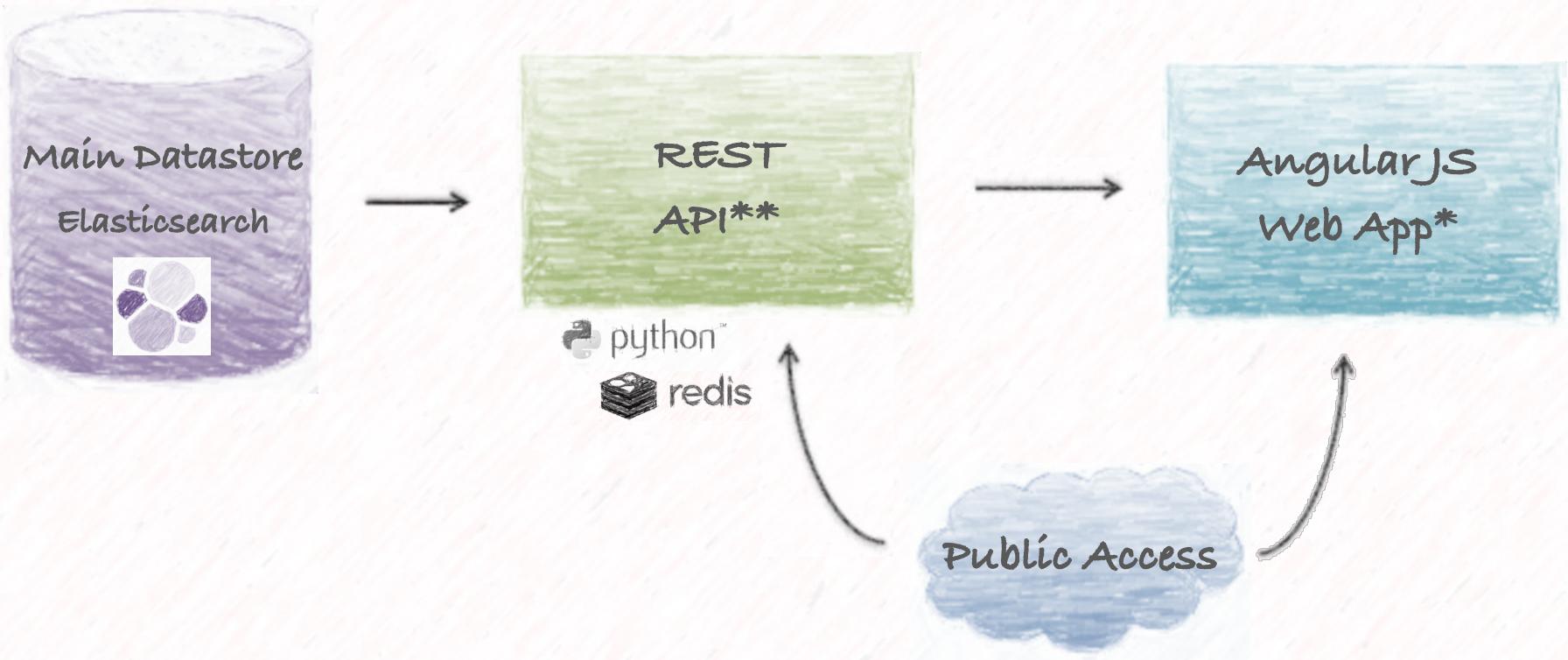
Extra! Extra! Extra!

Cancer hallmarks in our latest release!

Profile of a disease

Classification	Drugs	Similar diseases	Bibliography
  <pre> graph TD A[neoplasm] --> B[breast neoplasm] A --> C[breast fibrotic disease] </pre>		 	<p>Open Targets Library/LINK</p>  <p>of targeting the Journal of cellular physiology 2018 233(3):2162-2169</p> <p>Majid Khazaee, Seyyed</p>

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>

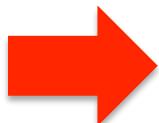
In these ~ 3.5 hours

- Introduction
- Live demos
- Exercises
- 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 ~ 3.5 hours

- Introduction
- Live demos
- Exercises
- Wrap up

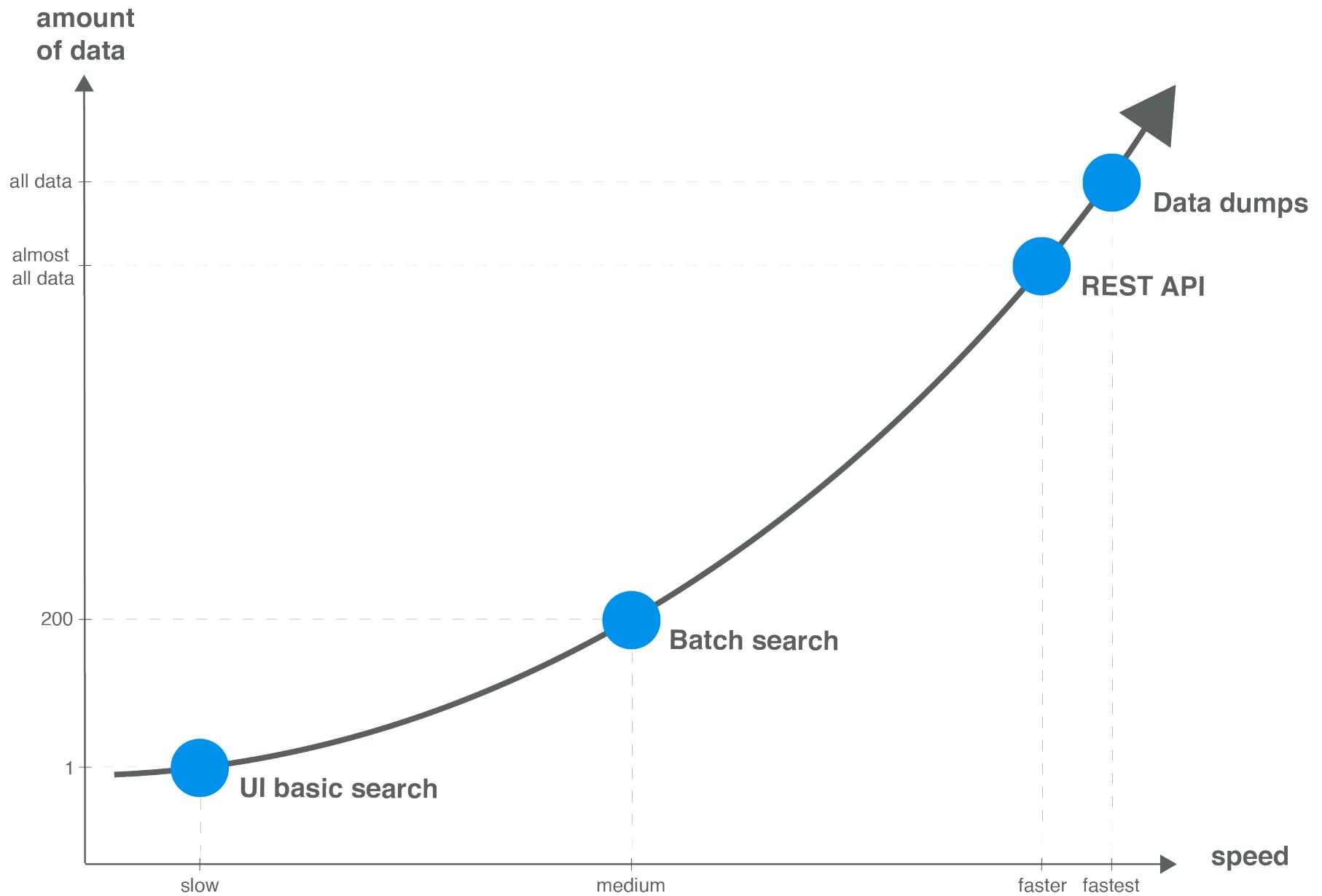


Coursebook

tinyurl.com/SD-exercises

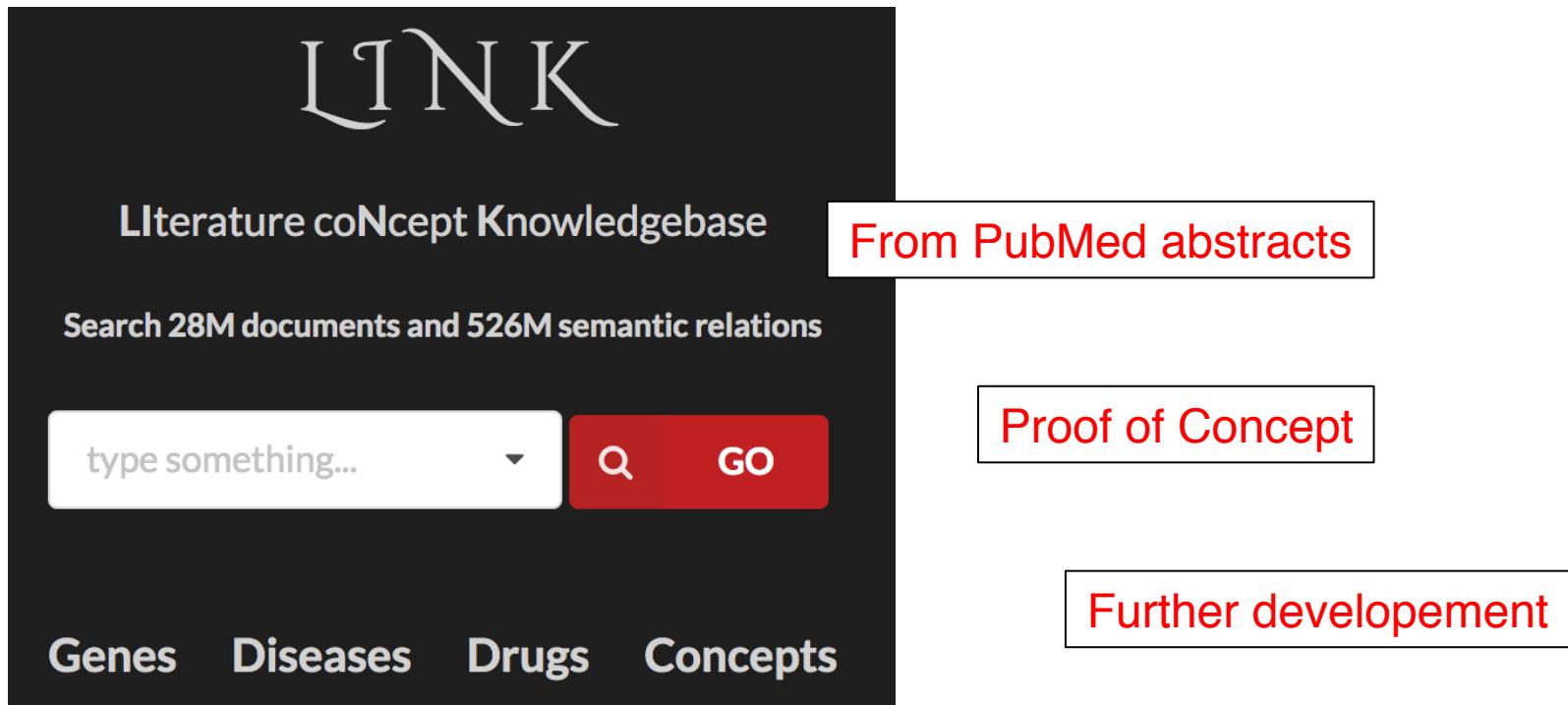
pages 29-33

Other modes of data access

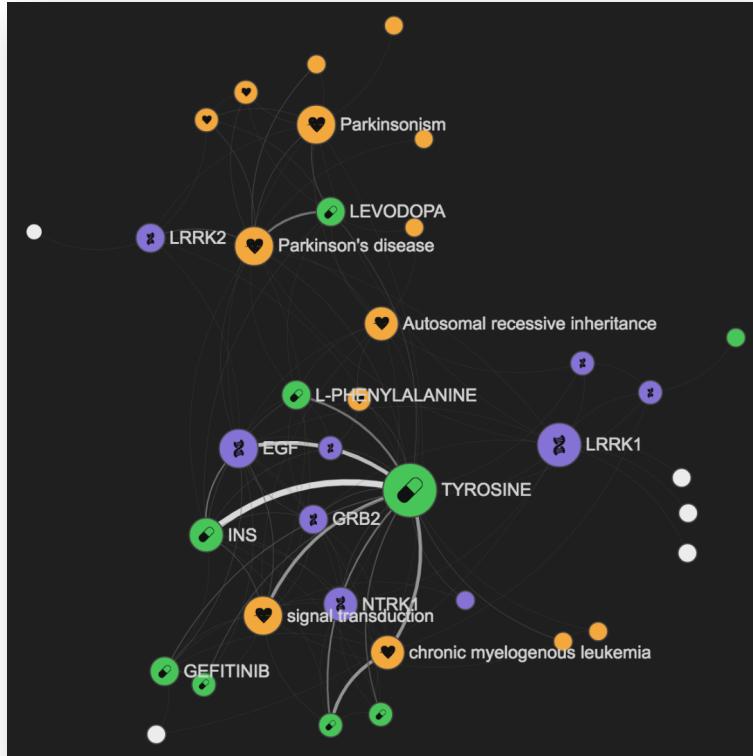


Beyond the Open Targets Platform

- LINK: Literature coNcept Knowledgebase



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



- Entities: genes, diseases, drugs
- Concepts extracted via NLP (Natural Language Processing)
- <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

In these ~ 3.5 hours

- Introduction
- Live demos
- Exercises
- Wrap up



Feedback survey

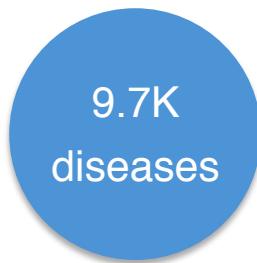
<https://bit.ly/SD-160318>

Help shape the navigation bar

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

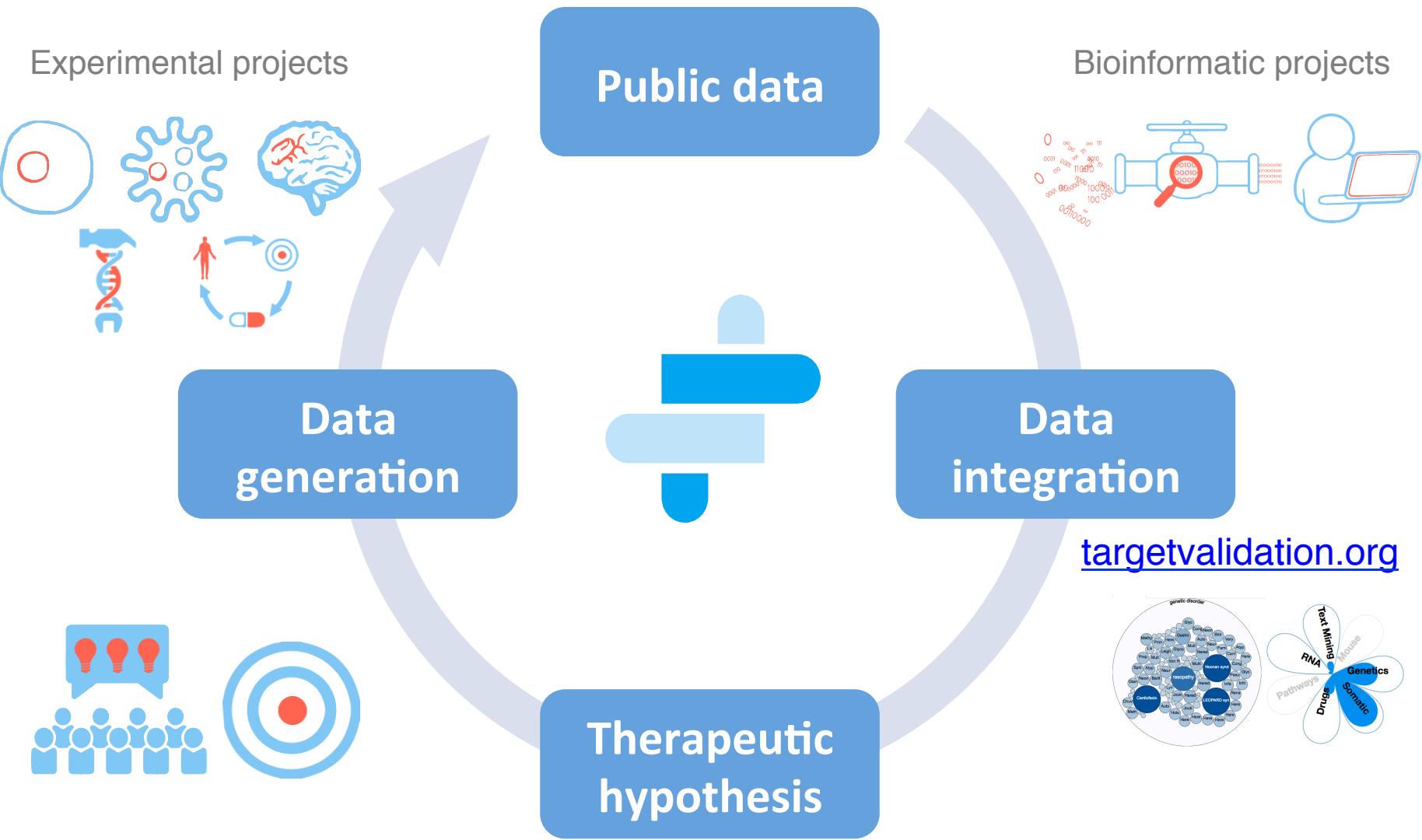
Open Targets Platform

- Resource of integrated multiomics data
- Added value (e.g. score) and links to original sources
- Graphical web interface: easy to use
- Drug discovery, translational medicine, disease biology



February 2018 release

Virtuous cycle in Open Targets



Concurrent
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?



...

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/

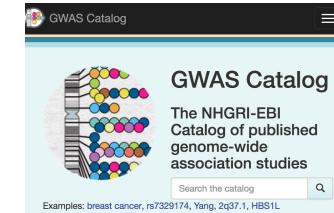
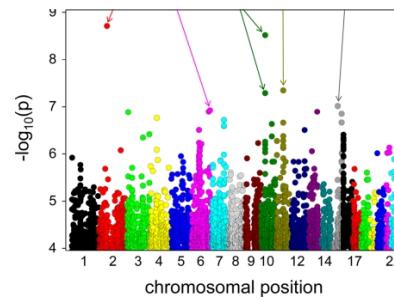
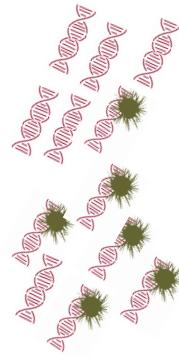


Open Targets

Extra Extra Extra

Data sources: GWAS catalog

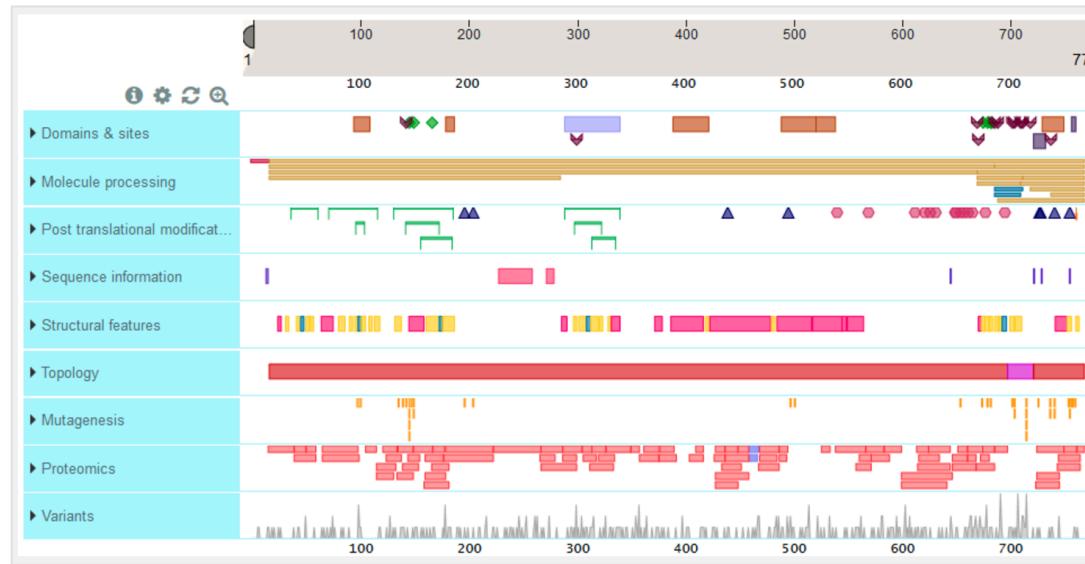
- Genome Wide Association Studies
- Array-based chips → genotyping 100,000 SNPs genomewide



Open Targets

Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients

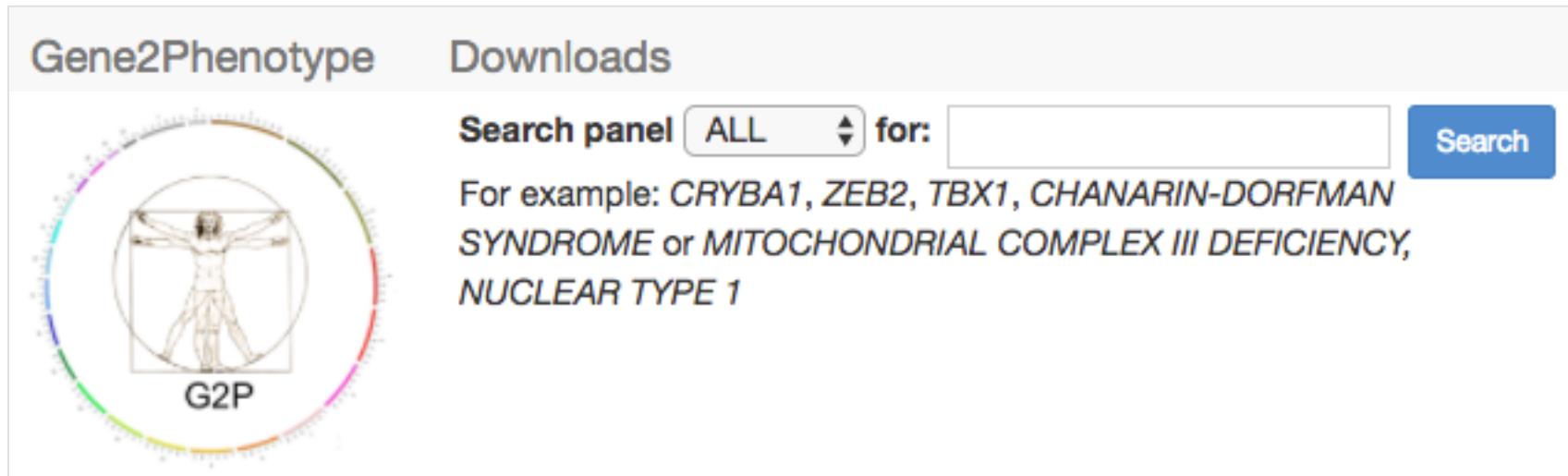


EMBL-EBI train online



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



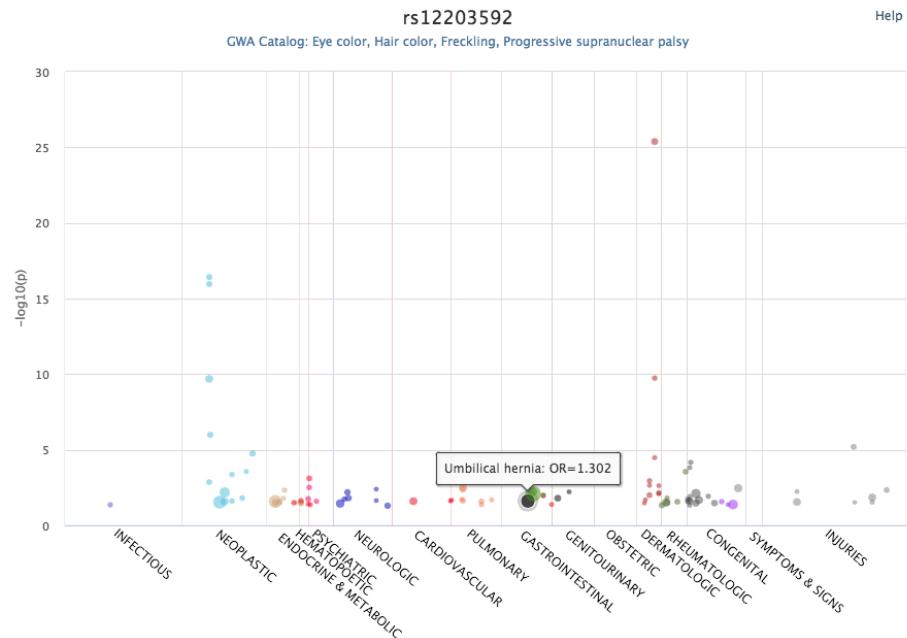
EMBL-EBI train online



Open Targets

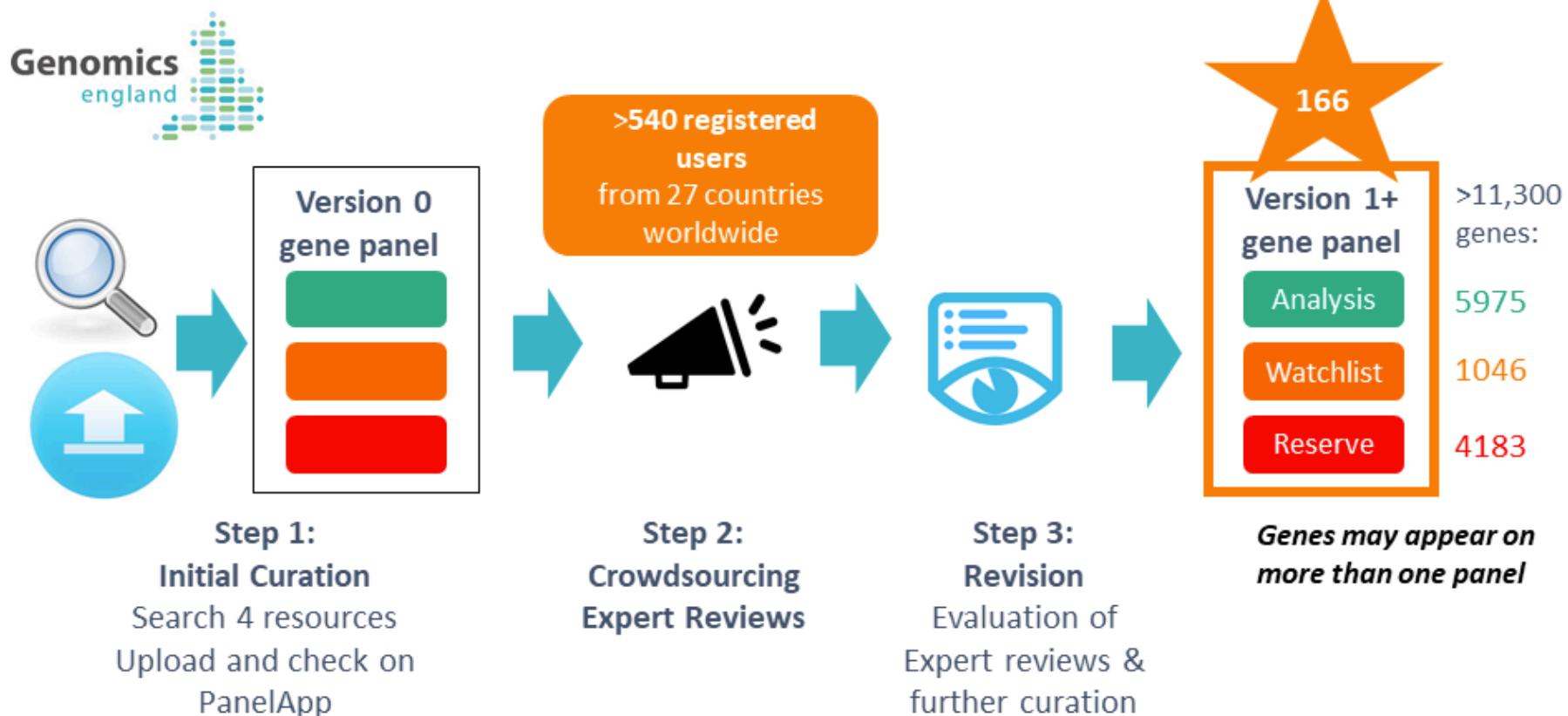
Data sources: PheWAS

- 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

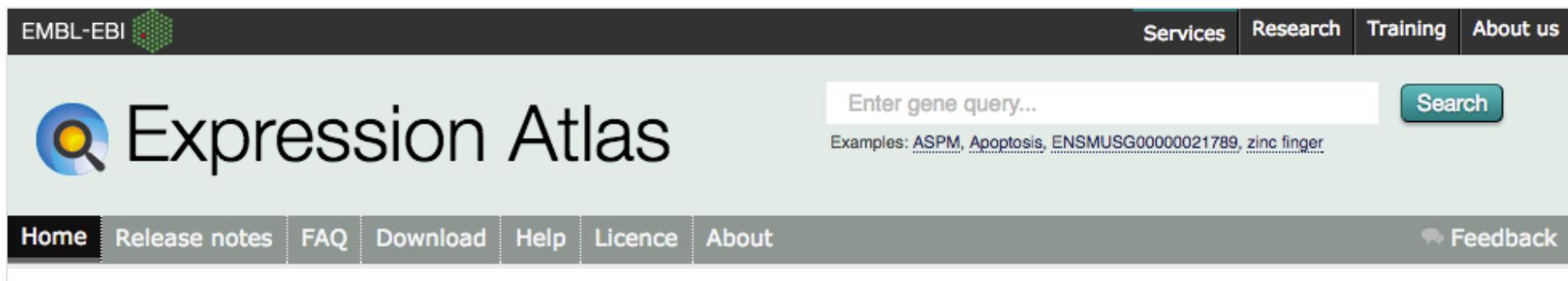


EMBL-EBI train online



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".

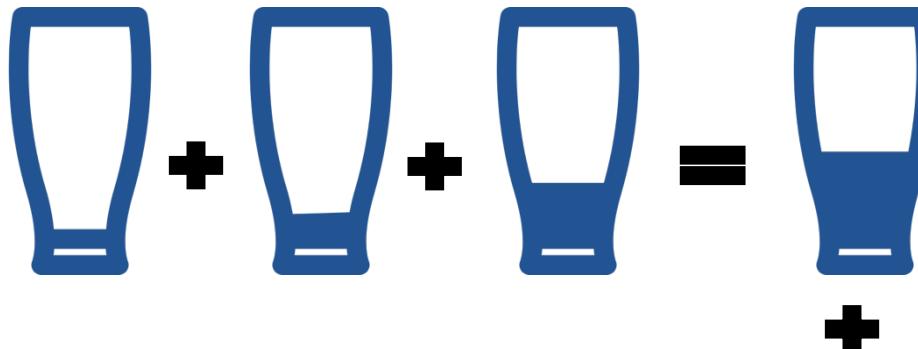
Welcome to PhenoDigm (PHENOtype comparisons for DIsease and Gene Models)

Diseases Tissue phenotype associations Secondary phenotypes

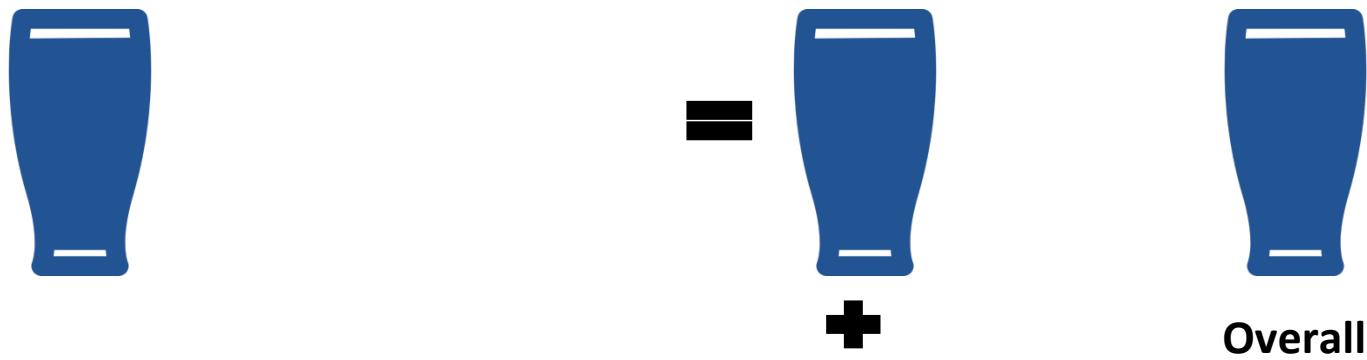
- Semantic approach to associate mouse models with diseases

Target-Disease Association Score

EuropePMC
(Text Mining)

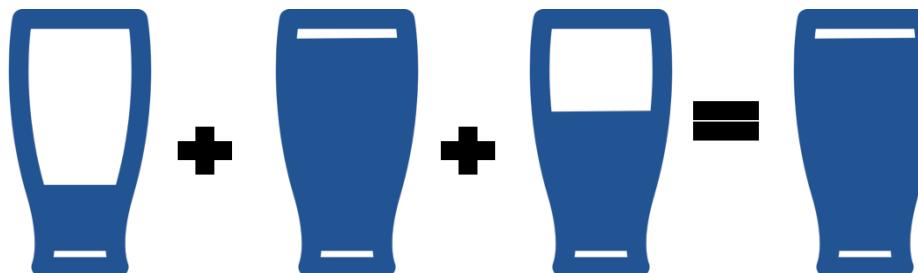


UniProt
(Manual Curation)



Overall

ChEMBL
(Manual Curation)



VERY simplified diagram

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