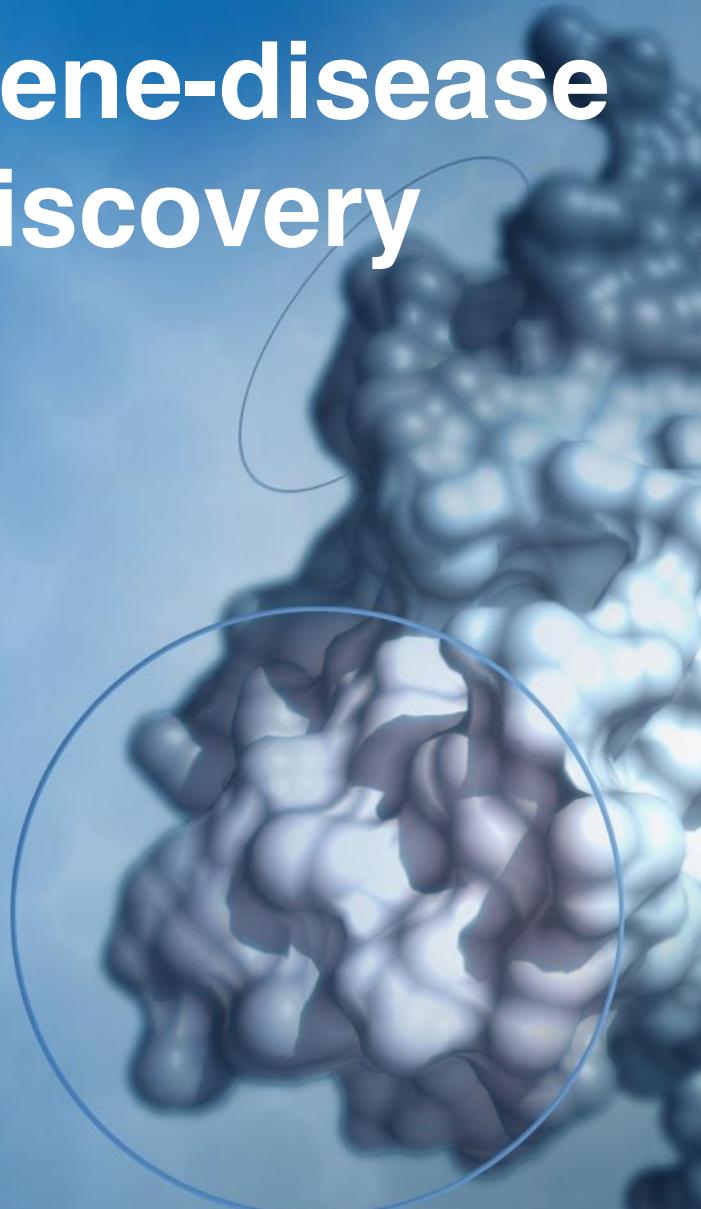


Open Targets: mining gene-disease associations for drug discovery

Alzheimer's Research UK – Oxford DDI
October 23rd 2017

Denise Carvalho-Silva, PhD
Scientific Outreach Lead
Open Targets
Wellcome Genome Campus, United Kingdom



Today 13:00-16:00

- Introduction
- Live demos

14:30-14:45 break?

- Hands-on exercises
- Wrap up and feedback survey

Objectives

What is the Open Targets Consortium?

What is the Open Targets Platform?

How to navigate the Platform?

How to connect with the team



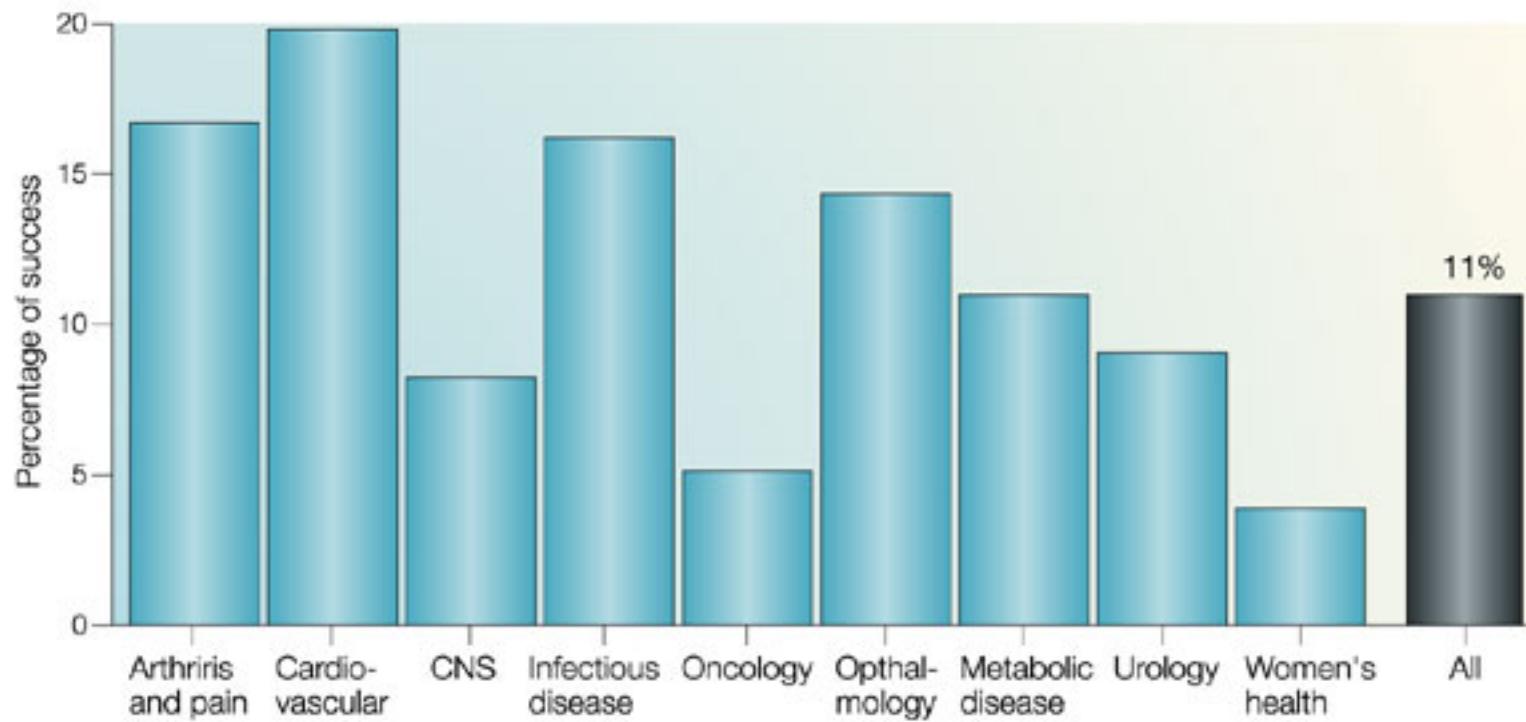
Today 13:00-16:00

- Introduction
- Live demos

14:30-14:45 break?

- Hands-on exercises
- Wrap up and feedback survey

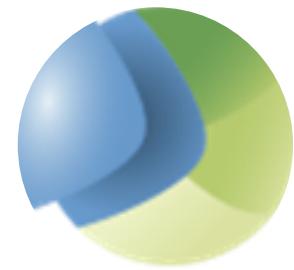
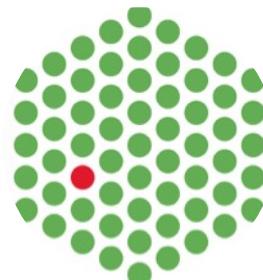
Drug discovery: the challenges



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

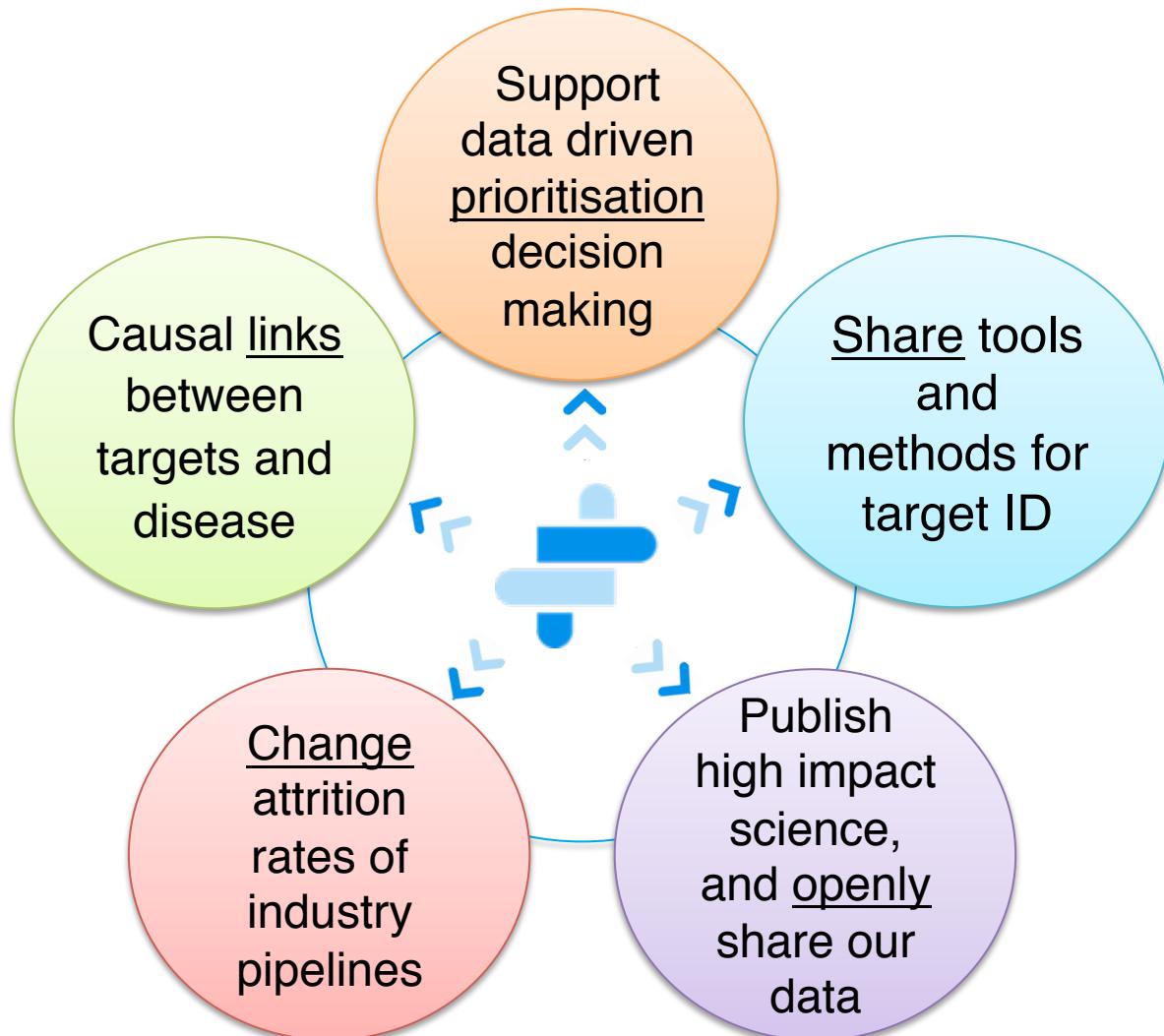
Our Vision

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



Open Targets

World leader for human target discovery



Two major areas of work in Open Targets

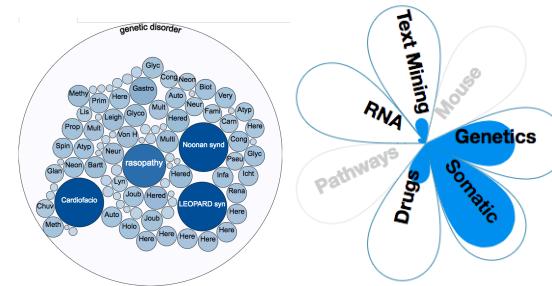
Experimental projects



Generate new evidence
CRISPR/Cas9
Organoids and IPS cells
(cellular models for disease)

Concurrent
www.opentargets.org/projects

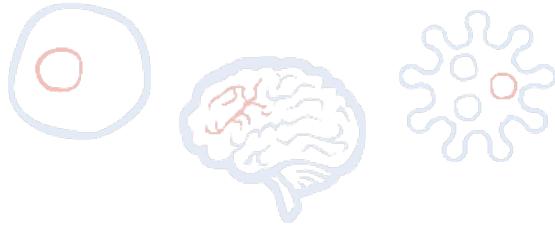
Core bioinformatics pipelines



Integration of available data
Web interface
REST API
Data dumps

Two major areas of work in Open Targets

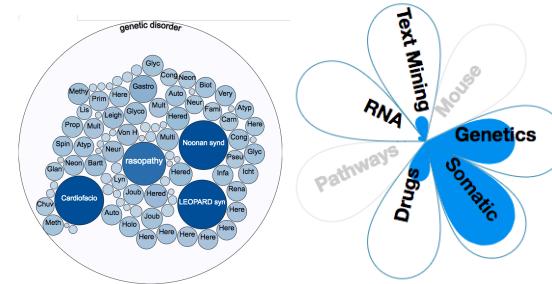
Experimental projects



Generate new evidence
CRISPR/Cas9
Organoids and IPS cells
(cellular models for disease)

Concurrent
www.opentargets.org/projects

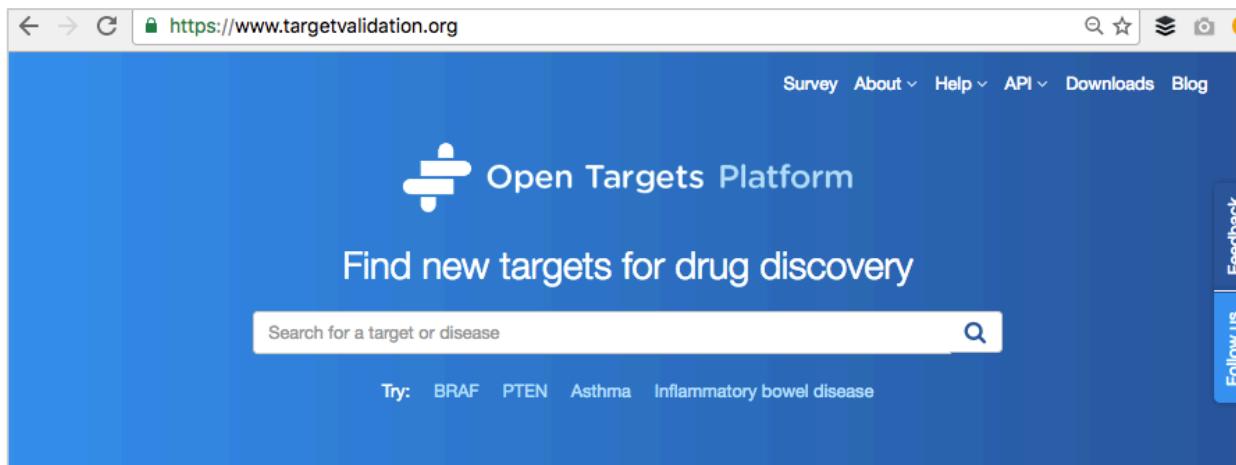
Core bioinformatics pipelines



Integration of available data
Web interface
REST API
Data dumps

Open Targets Platform*

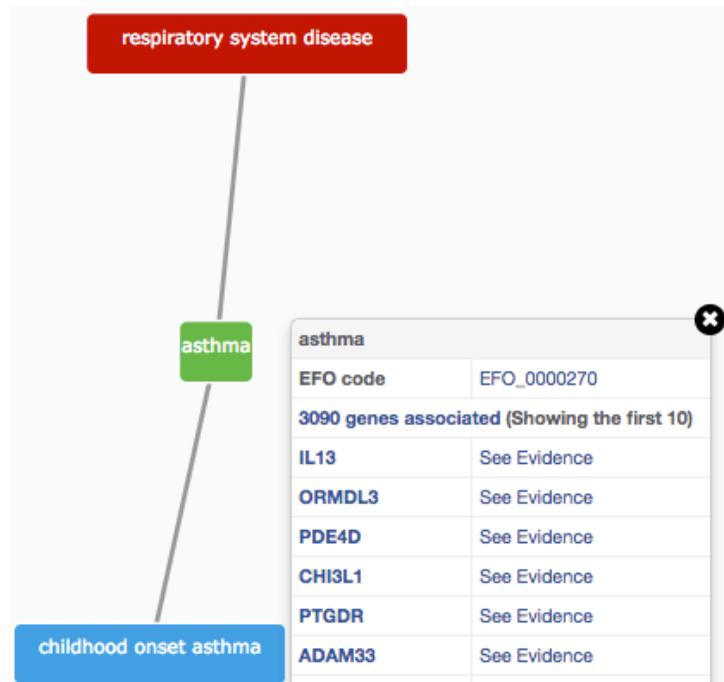
- Developed by the Core team at EMBL-EBI
- To allow users to identify target and disease associations
- Improvements driven by you



* First release: December 2015

Experimental Factor Ontology (EFO)

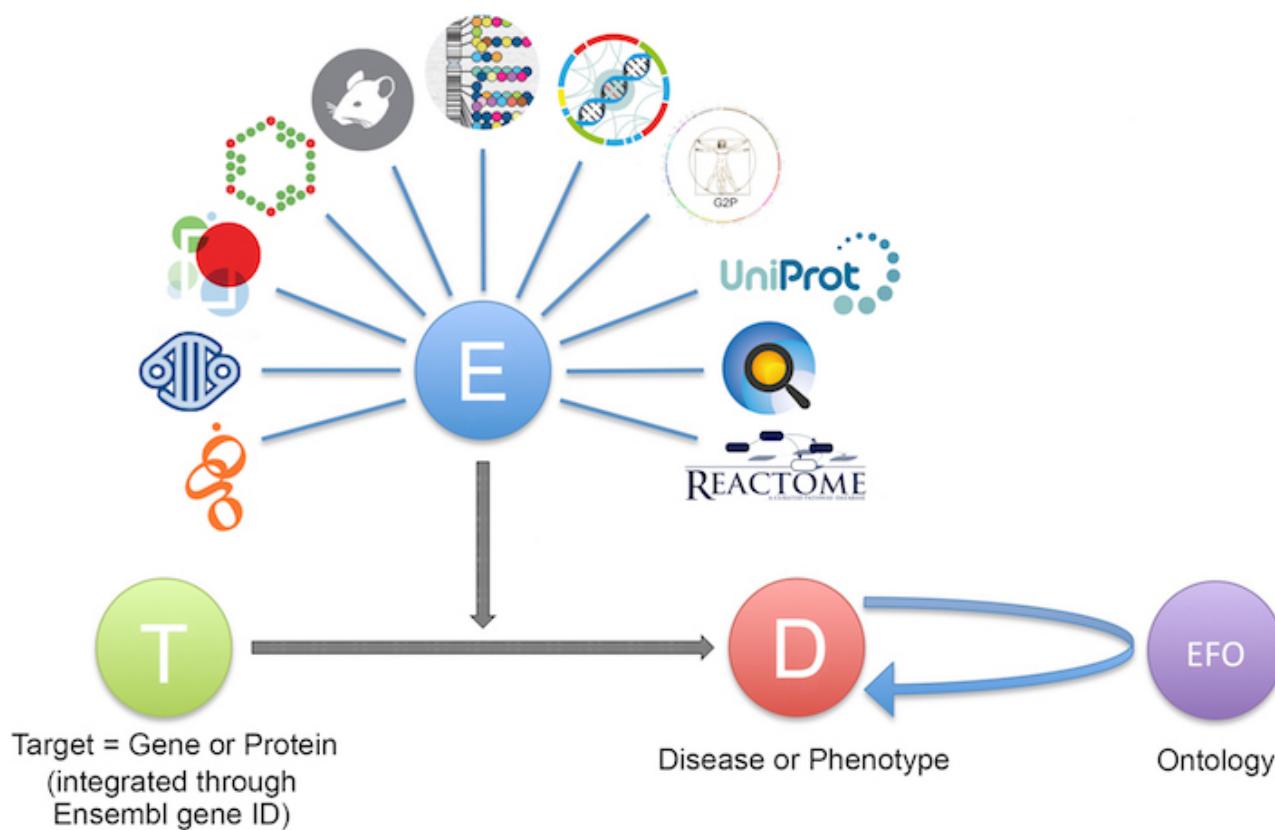
- Ontology: dictionary of relationships between entities
- EFO: way to organise experimental variables (e.g. diseases)



controlled vocabulary
+
hierarchy (relationship)

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

Evidence model



Original sources → more being added

Evidence from publicly available data

Open Targets Platform

Germline variants

EVA

Gene2Phenotype

Genomics England

GWAS

PheWAS

UniProt

Somatic mutations

Cancer Gene Census

EVA

IntOgen

UniProt

Drugs

ChEMBL

Affected pathways

Reactome

Differential RNA expression

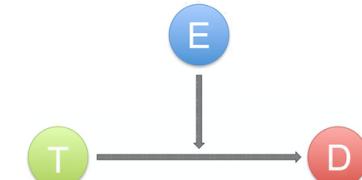
Expression Atlas

Animal models

PhenoDigm

Text mining

EuropePMC

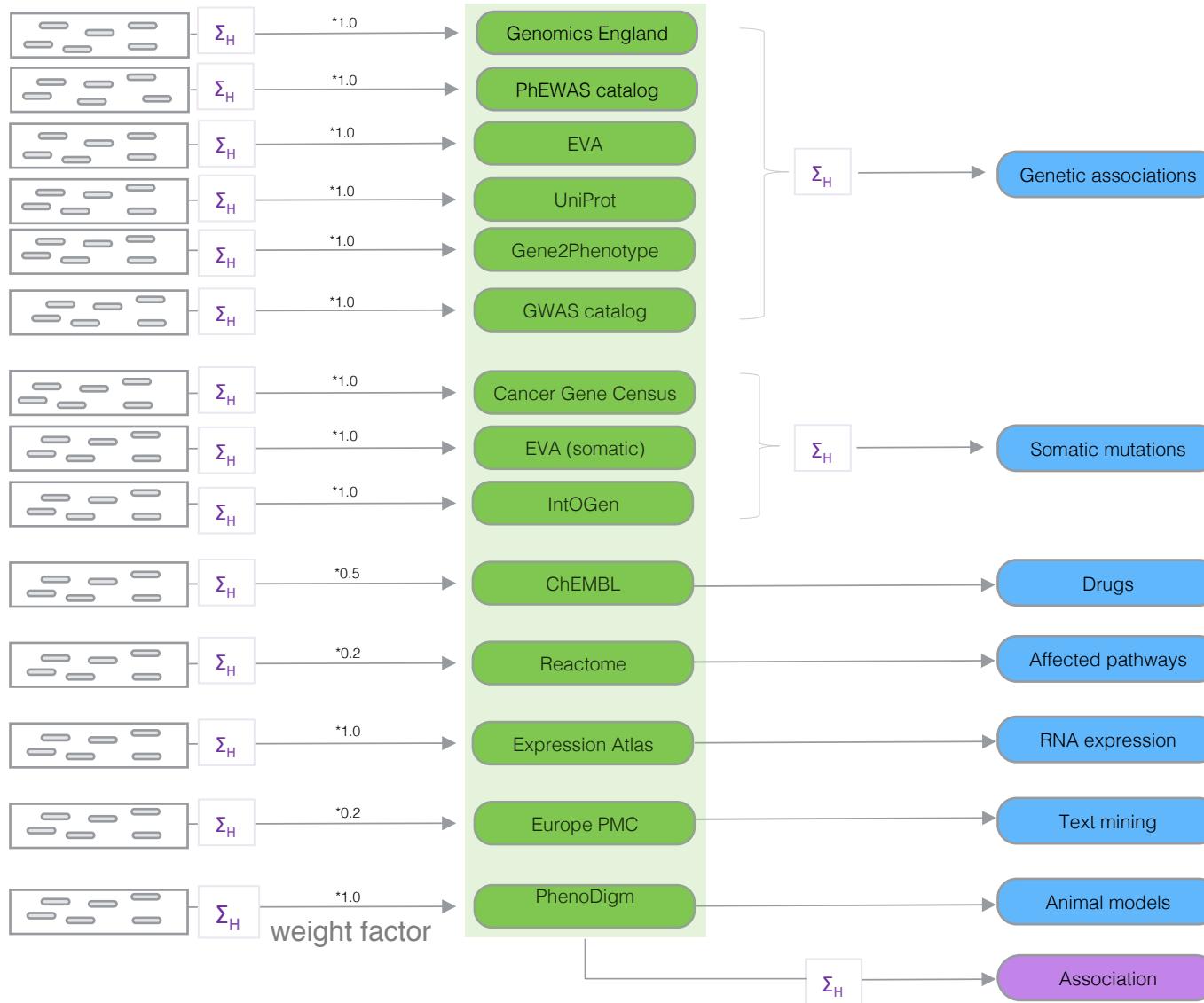


https://www.targetvalidation.org/data_sources

How do we associate diseases and phenotypes with targets?

- 1 Curate diseases and phenotypes
- 2 Map disease/phenotypes to an ontology using EFO, HPO, DO, OMIM
- 3 Use genes as proxies for our targets
- 4 Create target-disease evidence JSON objects
- 5 Calculate for each supporting evidence the likelihood of gene A being associated with disease B
- 6 Compute aggregated target-disease scores at the levels of data source, data type and overall score

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)

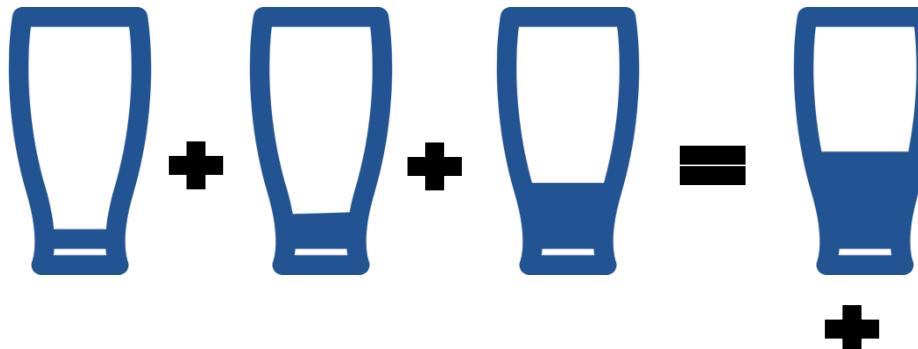


$$S_1 + S_2/2^2 + S_3/3^2 + S_4/4^2 + S_i/i^2$$

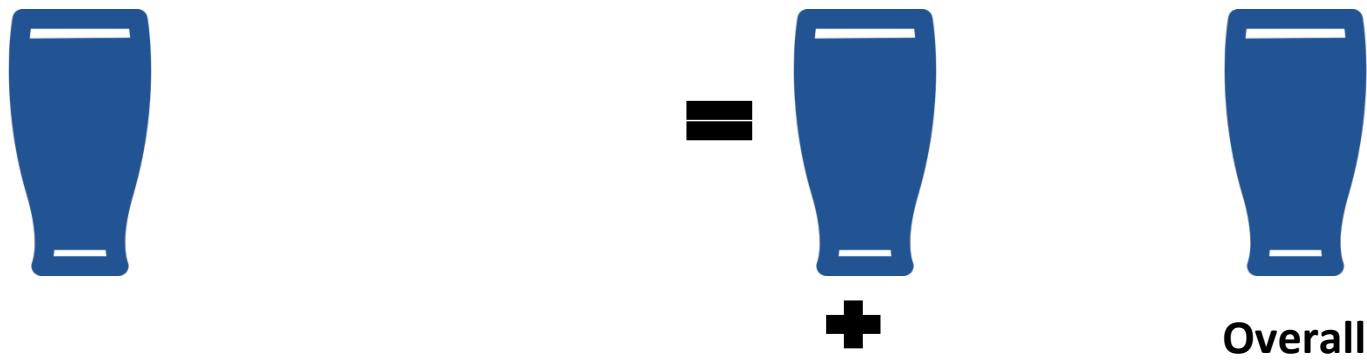
Note: Each data set has its own scoring and ranking scheme

Target-Disease Association Score

EuropePMC
(Text Mining)

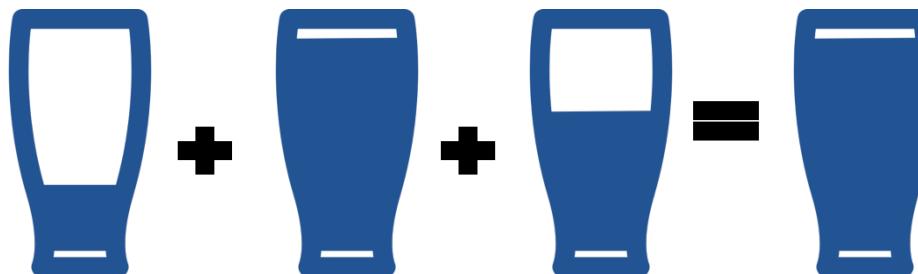


UniProt
(Manual Curation)



Overall

ChEMBL
(Manual Curation)



VERY simplified diagram

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

Today 13:00-16:00

- Introduction
- Live demos

14:30-14:45 break?

- Hands-on exercises
- Wrap up and feedback survey

Demo 1: Disease centric workflow

Which targets are associated with 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 ...

coursebook: pages 8-15

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

Choose your favourite internet browser*

*Supported ones: Internet Explorer 11 (and above), Chrome, Firefox and Safari

Demo 2: Evidence for an association

What is the evidence for the association between *CD86* and multiple sclerosis?



Open Targets Platform ≡ Q

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

The screenshot shows the Open Targets Platform interface. At the top is a blue header with the platform's logo and a search bar. Below the header, the main content area has a title "Evidence for CD86 in multiple sclerosis". To the left, there is a large, stylized flower-like diagram where each petal represents a different source of evidence: Text Mining, RNA, Pathways, Drugs, Genetics, Somatic, and Mouse. To the right of the diagram, there are two detailed pages: one for "CD86" (listing it as a CD86 molecule with synonyms B7.2, B7-2, and CD28LG2) and one for "multiple sclerosis" (listing its synonyms: MS (Multiple Sclerosis), MS, MULTIPLE SCLEROSIS ACUTE FULMINATING, Disseminated Sclerosis, Sclerosis...). Two blue callout boxes point to these pages: one labeled "Target profile page" pointing to the CD86 page, and another labeled "Disease profile page" pointing to the multiple sclerosis page.

Demo 2: Evidence for an association

- Which genetic evidence supports this association?
- Can you view this in a genome browser display?
- Are there any drugs in clinical trials for this disease?
- Which cell/tissue has the highest RNA expression(Illumina Body Map data)?
- Are there other diseases of the nervous system associated with this target? Can you export the table with this information? How strong is this association?

Demo 3: 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?

Alternative ways to access the data

The screenshot shows a web browser window with the URL <https://www.targetvalidation.org/download> in the address bar. The page itself has a blue header with the Open Targets Platform logo and navigation icons. The main content area is titled "Data Download" and contains text explaining that all data from targetvalidation.org is available for download as compressed JSON files. It describes the availability of associations and evidence objects via API methods. Below this, a section titled "2017 Feb (Latest)" lists two download links: "Association objects (2016-12-09, 215MB, md5sum)" and "Evidence objects (2016-12-09, 4.35Gb, md5sum)".

All data from targetvalidation.org is available for download as compressed JSON files.

We provide downloads of all associations between target and disease calculated by the platform, as well as all the evidence used in calculating each associations. These are the same objects returned by the corresponding [/public/associations](#) and [/public/evidence](#) API methods. Head to the API documentation for further details.

2017 Feb (Latest)

- Association objects (2016-12-09, 215MB, md5sum)
- Evidence objects (2016-12-09, 4.35Gb, md5sum)

Open Targets REST API



public : Publicly supported stable API.

Open/Hide | List operations | Expand operations

GET /public/evidence

POST /public/evidence

GET /public/evidence/filter

POST /public/evidence/filter

GET /public/association

GET /public/association/filter

POST /public/association/filter

GET /public/search

GET /public/auth/request_token

GET /public/auth/validate_token

GET /public/utils/ping

GET /public/utils/version

GET /public/utils/stats

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

Interactive API documentation

GET

/public/association

Implementation notes

After integrating all evidence connecting a target to a specific disease, we compute an association score by mean of an harmonic sum. This **association score** provides an indication of how strong the evidence behind each connection is and can be used to rank genes in order of likelihood as drug targets. The association id is constructed by using the ensembl id of the gene and the EFO id for the disease (eg. ENSG00000073756-EFO_0003767). The method returns an association object, which contain data and summary on each evidence type included in the calculation of the score, as well as the score itself.

Parameters

Parameter	Value	Description	Parameter type	Data type
id	ENSG00000073756-EFO_0003767	an association ID usually in the form of TARGET_ID-DISEASE_ID	query	string

Response messages

HTTP status code	Reason	Model
200	Successful response	

Try it out!

[Hide response](#)

Request URL

http://targetvalidation.org/api/latest/public/association?id=ENSG00000073756-EFO_0003767

Response body

```
{  
  "from": 0,  
  "took": 32,  
  "data_version": "17.04",  
  "query": {},  
  "total": 1,  
  "data": [  
    {  
      "target": {  
        "gene_info": {  
          "symbol": "PTGS2",  
          "name": "Prostaglandin-endoperoxide synthase 2",  
          "ensembl_id": "ENSG00000073756",  
          "uniprot_id": "P48331",  
          "ensembl_gene_id": "ENSG00000073756",  
          "chromosome": "19",  
          "start": 133000000, "end": 133000000, "strand": 1  
        }  
      }  
    }  
  ]  
}
```



REST API calls: some examples*

GET

/public/search

* http://targetvalidation.org/api/latest/public/search?q=EFO_0003767

* <http://targetvalidation.org/api/latest/public/search?q=asthma>

GET

/public/association/filter

[http://www.targetvalidation.org/api/latest/public/association/filter?
target=ENSG00000110324&direct=false&fields=is_direct&fields=disease.efo_info.lab
el&size=100](http://www.targetvalidation.org/api/latest/public/association/filter?target=ENSG00000110324&direct=false&fields=is_direct&fields=disease.efo_info.label&size=100)

GET

/public/evidence/filter

[https://targetvalidation.org/api/latest/public/evidence/filter?
target=ENSG00000141867&disease=EFO_0000565&datatype=expression_atl
as&size=100&format=json](https://targetvalidation.org/api/latest/public/evidence/filter?target=ENSG00000141867&disease=EFO_0000565&datatype=expression_atlas&size=100&format=json)

* blog.opentargets.org/tag/api/

How to run these REST endpoints

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

*<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

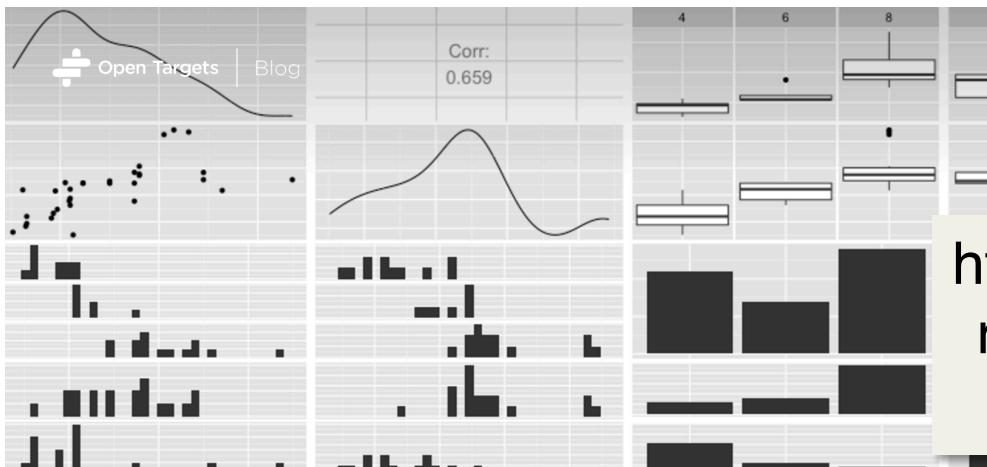
Docs » opentargets - Python client for targetvalidation.org

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

Today 13:00-16:00

- Introduction
- Live demos

14:30-14:45 break?

- Hands-on exercises
- Wrap up and feedback survey

Hands-on exercises

Pages 26-29

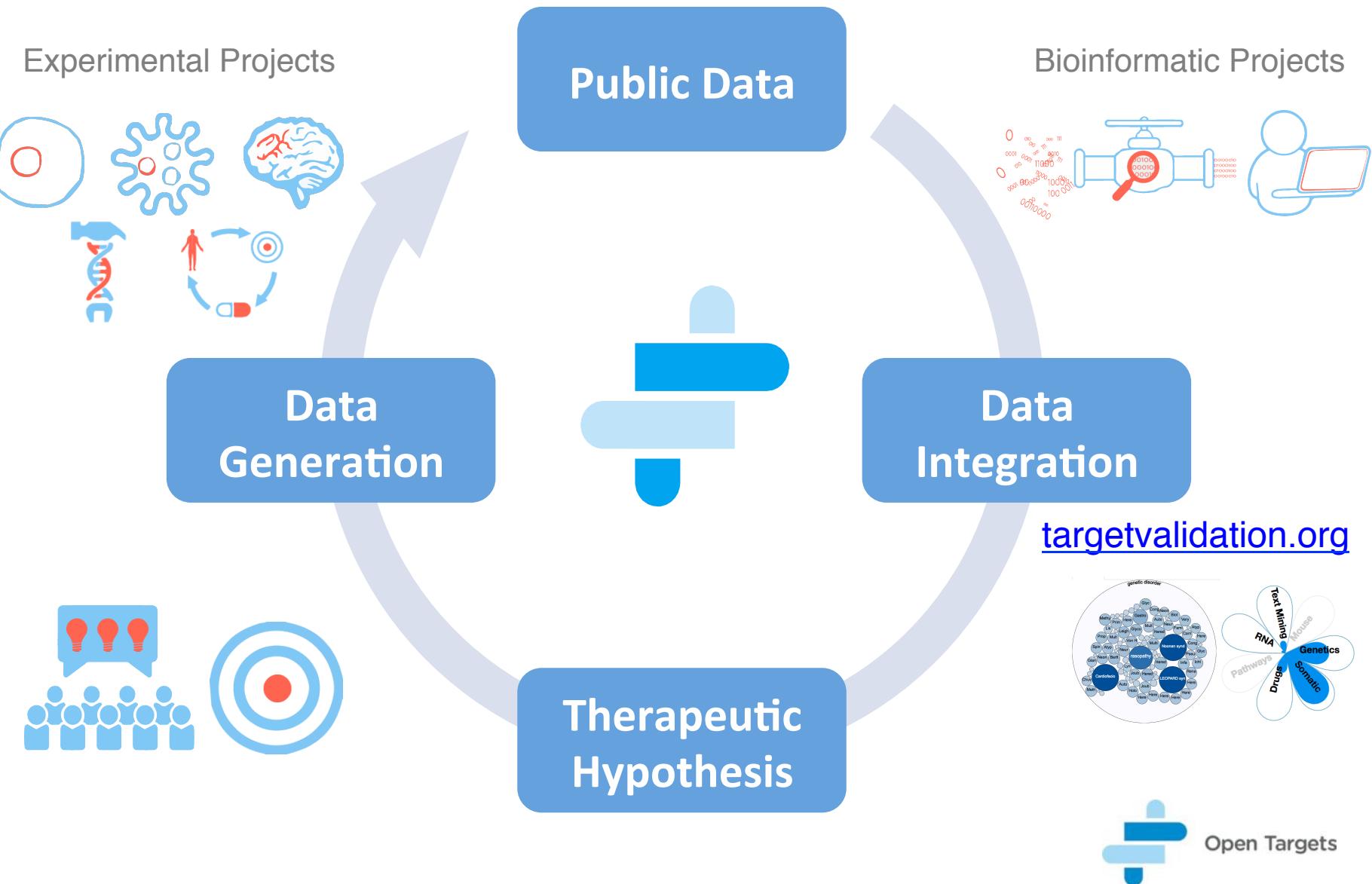
Today 13:00-16:00

- Introduction
- Live demos

14:30-14:45 break?

- Hands-on exercises
- Wrap up and feedback survey

Wrap up



Open Targets Platform

For drug target ID and selection in drug discovery

Rank target and disease associations by priority
using machine learning models and multiple data sources

Integrated information from multiple sources

Intuitive interface

Oh Yes!
And all is 100% free

26K targets

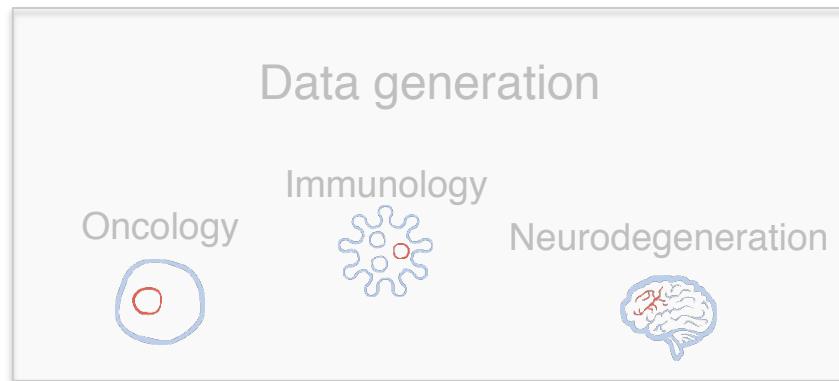
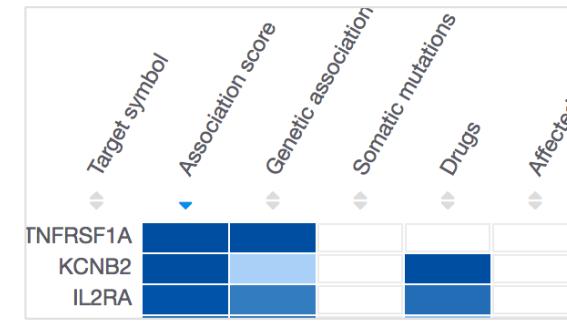
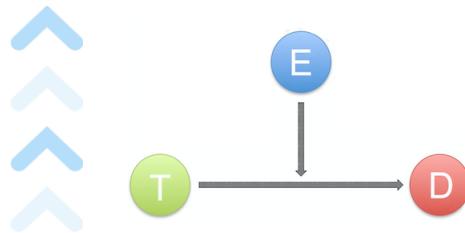
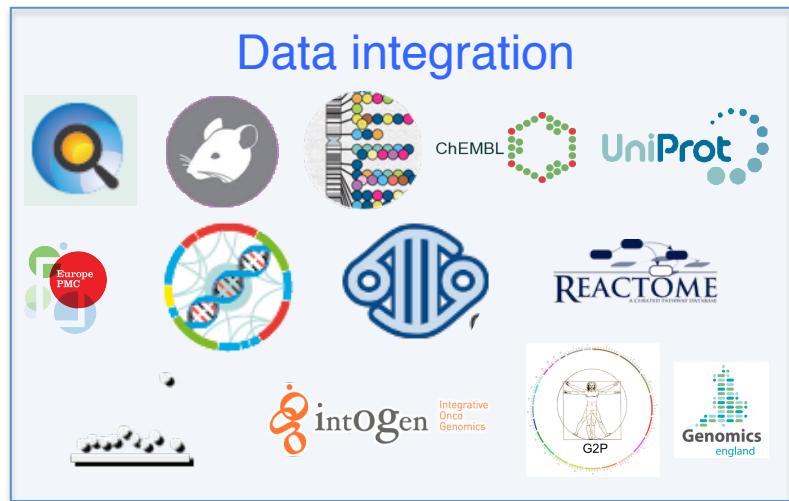
9.1K diseases

2.8 M associations

5.3 M evidence

June 2017 release

Present and Future



Generated
as I speak

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?

...



Open Targets

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



Support, dissemination, video tutorials



support@targetvalidation.org



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



[@targetvalidate](#)



blog.opentargets.org/



www.facebook.com/OpenTargets/

<https://vimeo.com/233638392>



<http://imgur.com/a/JIDCP>

<http://imgur.com/a/LKDhp>

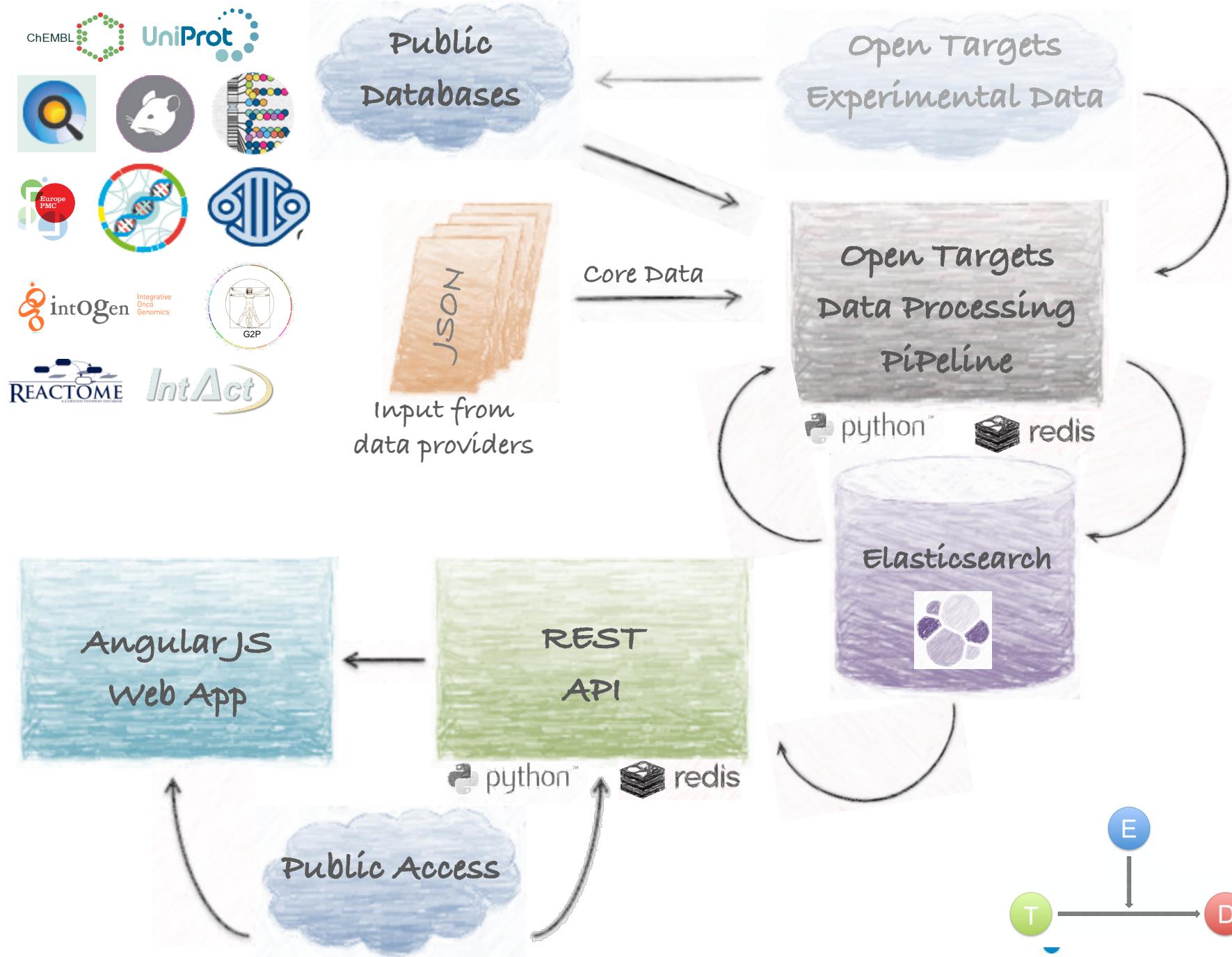


Open Targets

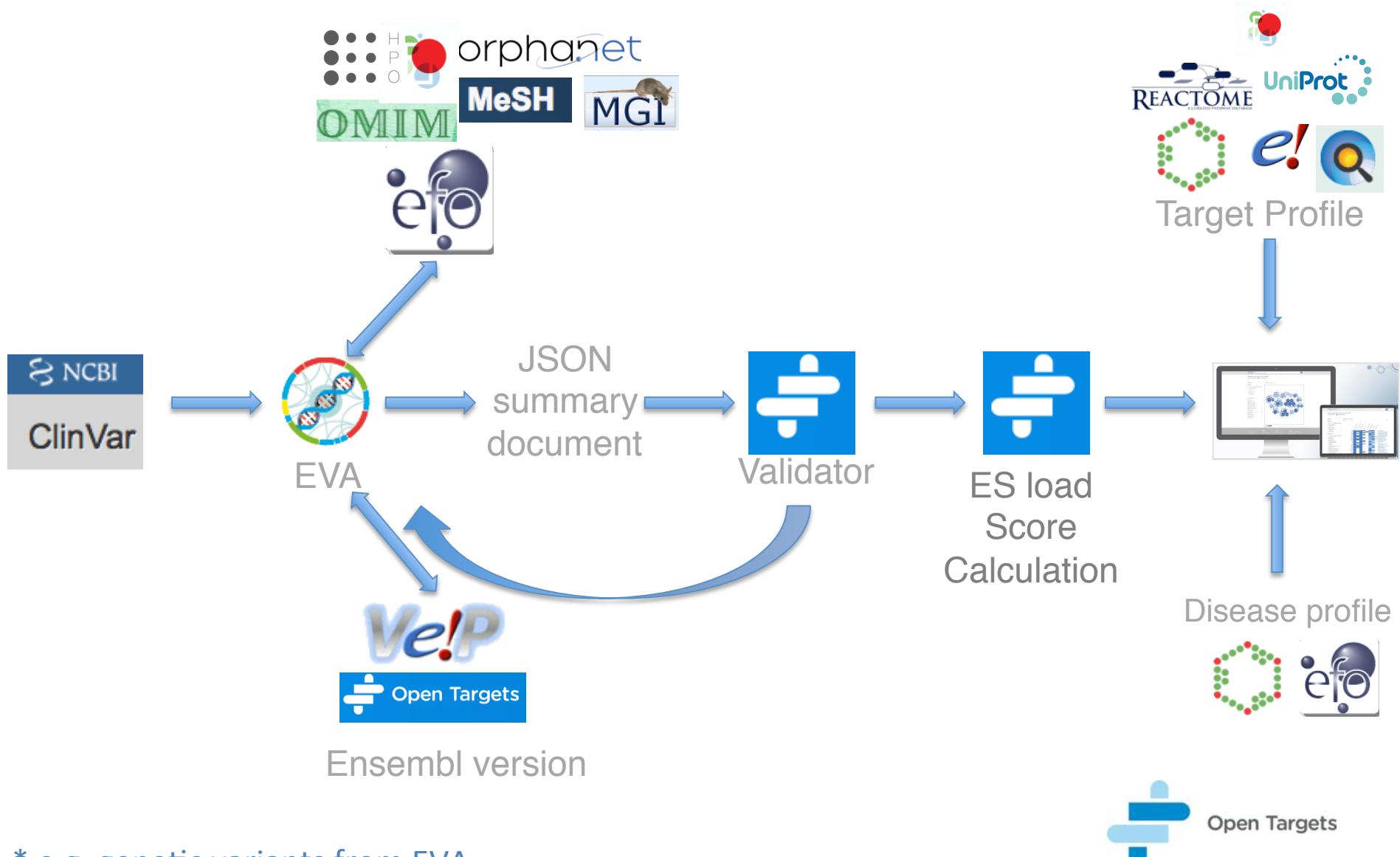
Feedback

<http://aruk-231017>

Extra Extra Extra (slides)



Data flow pipeline*

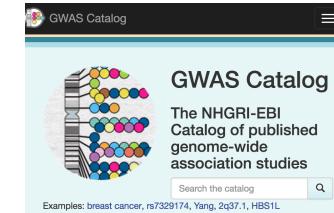
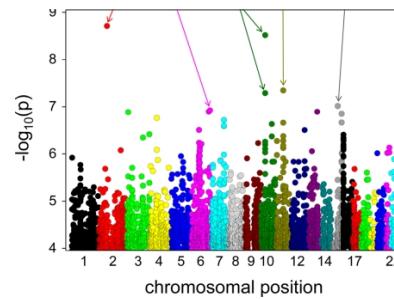
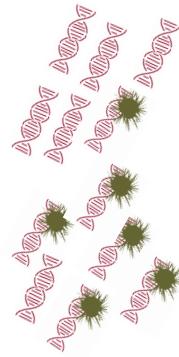
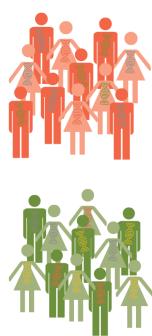


* e.g. genetic variants from EVA



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

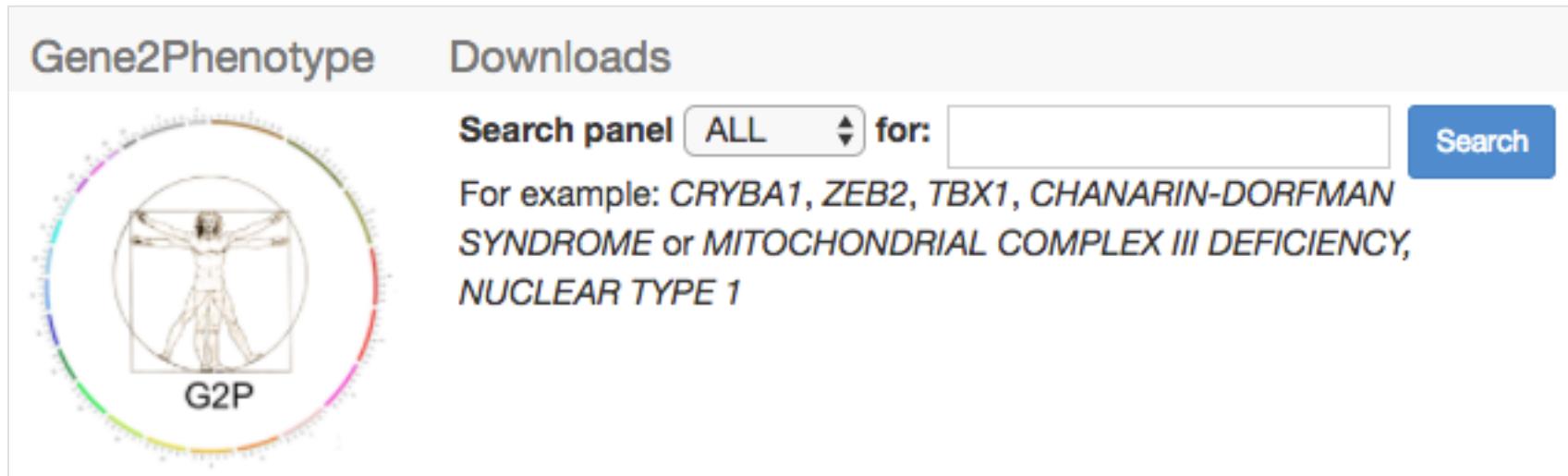


EMBL-EBI train online



Open Targets

Data sources: Gene2Phenotype



Gene2Phenotype Downloads

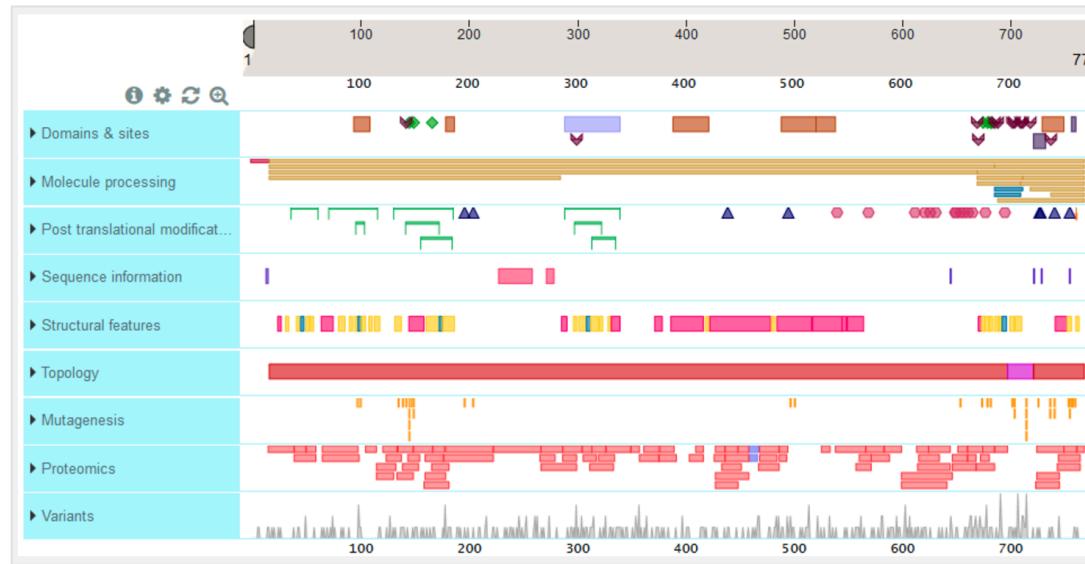
Search panel ALL for: **Search**

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



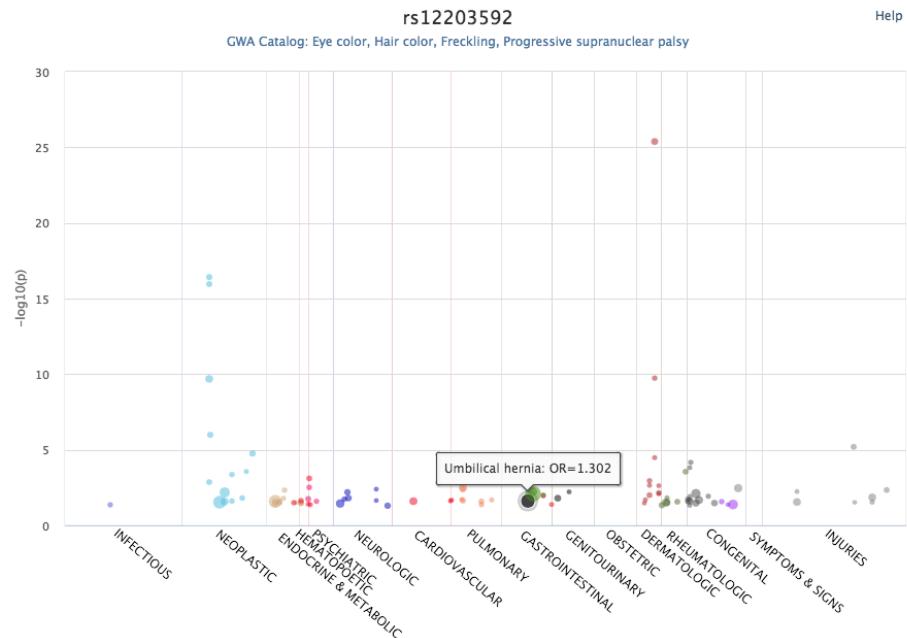
EMBL-EBI train online



Open Targets

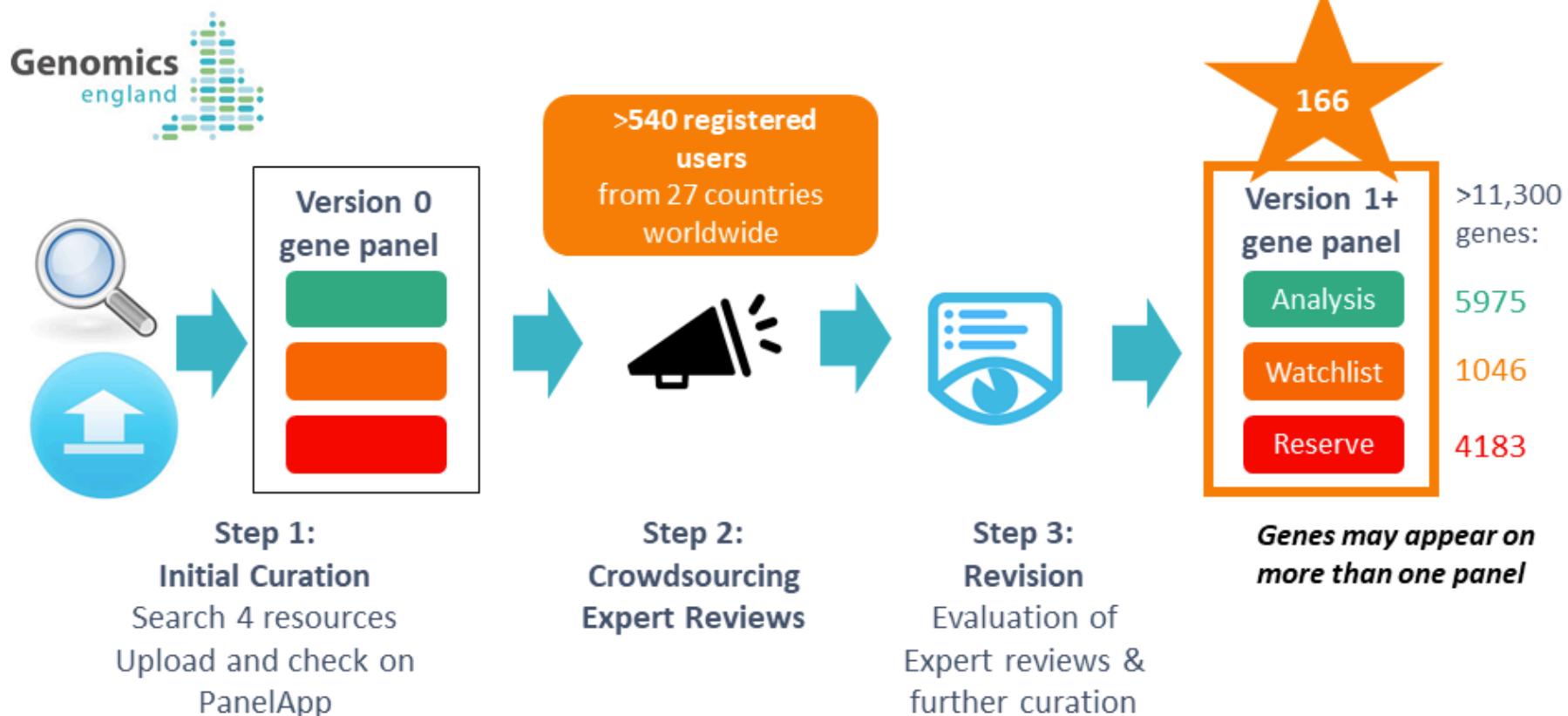
Data sources: PheWAS

- 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 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

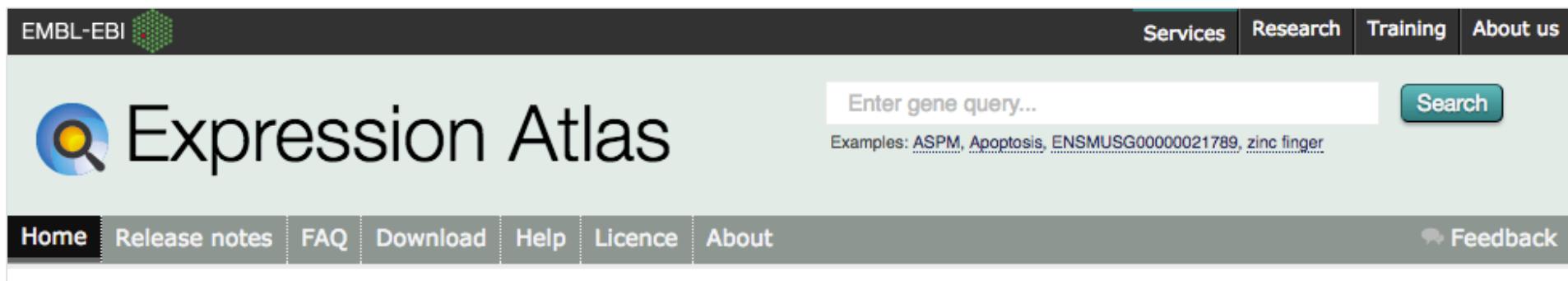


EMBL-EBI train online



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

JSON summary document

```
hort_name": "Franke A"}, {"last_name": "Alizadeh", "full_name": "Alizadeh Behrooz Z", "short_name": "Alizadeh BZ"}, {"last_r  
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_000040,  
EFO_0005140", "EFO_0003767"]], "efo_id": "http://www.ebi.ac.uk/efo/EFO_0003767", "label": "inflammatory bowel disease"  
_association_fields": {"pubmed_refs": "http://europepmc.org/abstract/MED/26192919", "object": "http://www.ebi.ac.uk/efo/EFO_0000405",  
ntifiers.org/dbsnp/rs11150589", "study_name": "cttv009_gwas_catalog", "sample_size": "96486", "gwas_panel_resolution": "9000000",  
: "http://identifiers.org/ensembl/ENSG00000005844"}, "evidence": {"variant2disease": {"gwas_sample_size": 96486, "unique_experiments":  
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": "2017-03-23T03:44:36+00:00"}, "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": [{"type": "pvalue", "id": "http://purl.obolibrary.org/obo/ECO_0000205", "label": "curator inference"}], "date asserted":  
"2017-03-23T03:44:36+00:00"}, "evidence_codes_info": [{"{"eco_id": "GWAS", "label": "Genome-wide association study evidence_pipeline", "type": "CTTV-custom annotation pipeline"}, {"{"eco_id": "ECO_0000205", "label": "curator inference", "type": "upstream_gene_variant"}]}, {"{"label": "CTTV-custom annotation pipeline"}], [{"{"eco_id": "ECO_0000205", "label": "curator inference", "type": "functional_consequence"}], {"{"label": "Functional consequence"}], "gene2variant": {"functional_consequence": "http://purl.obolibrary.org/obo/SO_0001631", "provenance": [{"lit_act": "2017-03-23T03:44:36+00:00"}]}, "true", "statement": "Primary submitter of data"}, "database": {"dbxref": {"version": "2017-03-23T03:44:36+00:00", "id": "http://identifiers.org/gwascatalog"}, "id": "GWAS Catalog", "version": "2017-03-23T03:44:36+00:00"}, "is_associated": true, "resource_score": {"type": "probabilistic", "id": "http://purl.obolibrary.org/obo/ECO_0000205", "label": "curator inference"}, "date asserted": "2017-03-23T03:44:36+00:00"}, "evidence_codes": [{"type": "probabilistic", "id": "http://purl.obolibrary.org/obo/ECO_0000205", "label": "curator inference"}, {"type": "evidence", "id": "http://identifiers.org/eco/cttv_mapping_pipeline"}, "date asserted": "2017-03-23T03:44:36+00:00"}, "evidence_codes": [{"type": "evidence", "id": "http://identifiers.org/eco/cttv_mapping_pipeline"}, {"type": "probabilistic", "id": "http://purl.obolibrary.org/obo/ECO_0000205", "label": "curator inference"}, {"type": "probabilistic", "id": "http://identifiers.org/eco/cttv_mapping_pipeline", "label": "curator inference"}], "validated_against_schema_version": "1.2.5", "res": {"association_score": 0.24183029962242697}, "type": "genetic_association", "id": "f8aa5612c7f01940f3958914fc6074ba"}  
loads denise$
```

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

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

Factors affecting the relative strength of an evidence

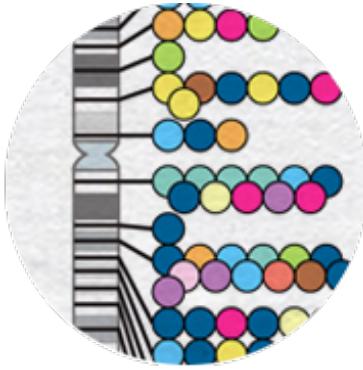
e.g. *GWAS Catalog*

$$S = f * s * c$$

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