

GENE0006

You have two options:

- 1. Send an Email to
sms@textwall.co.uk**

**Type 4 things with a space in between
ONLY IN THE SUBJECT LINE:**

- 1) bs065 (*lower case only*)**
- 2) Student number**
- 3) Your Surname**
- 4) Unique word (*as listed in the example below*)**

e.g. bs065 1234567 SMITH Biology

OR

- 2. TEXT the same info to 0207 183 8329**
(Your standard text message rates will apply.)

ALL STUDENTS MUST RESPOND REGARDLESS OF VISA STATUS.

Materials

- Slides



- Coursebook



www.github.com/deniseOme/training/

Open Targets: integrating genetics, omics and chemistry for drug discovery

MSc Module

Understanding bioinformatics resources and their application

December 4th 2018

Dr Denise Carvalho-Silva

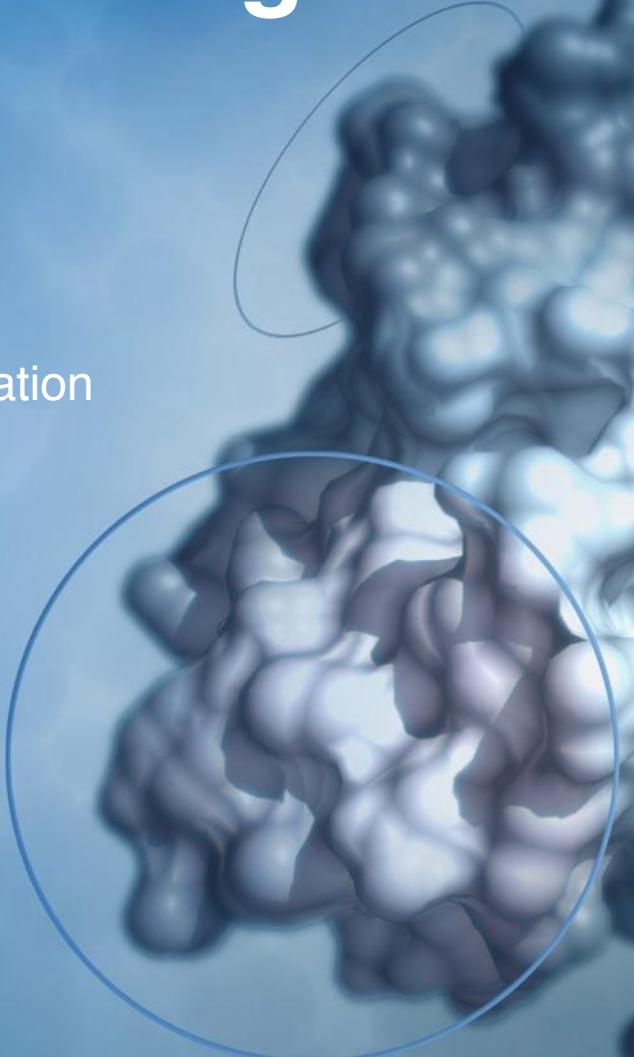
Open Targets / EMBL-EBI

Wellcome Genome Campus

United Kingdom



Open Targets



Aims of this workshop



What is Open Targets?

What can you do with the
Open Targets Platform?

Other ways to
access the data?

How to get
in touch?

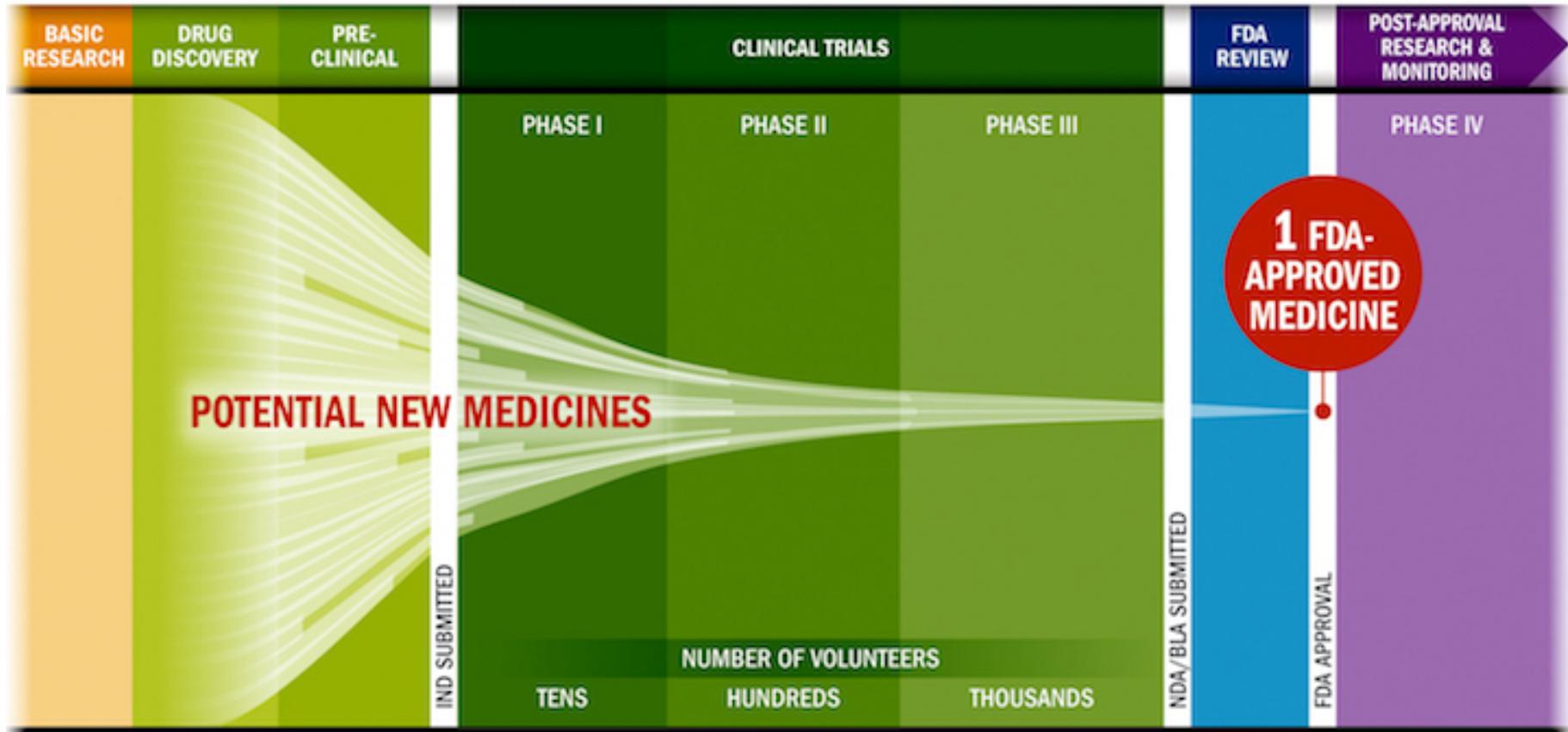
Today: 11:00-17:00

- Introduction
- Open Targets Platform: live demos
- Lunch – live demos- exercises
- Other modes of data access
- Feedback survey and wrap up

Today: 11:00-17:00

- Introduction
- Open Targets Platform: live demos
- Lunch – live demos- exercises
- Other modes of data access
- Feedback survey and wrap up

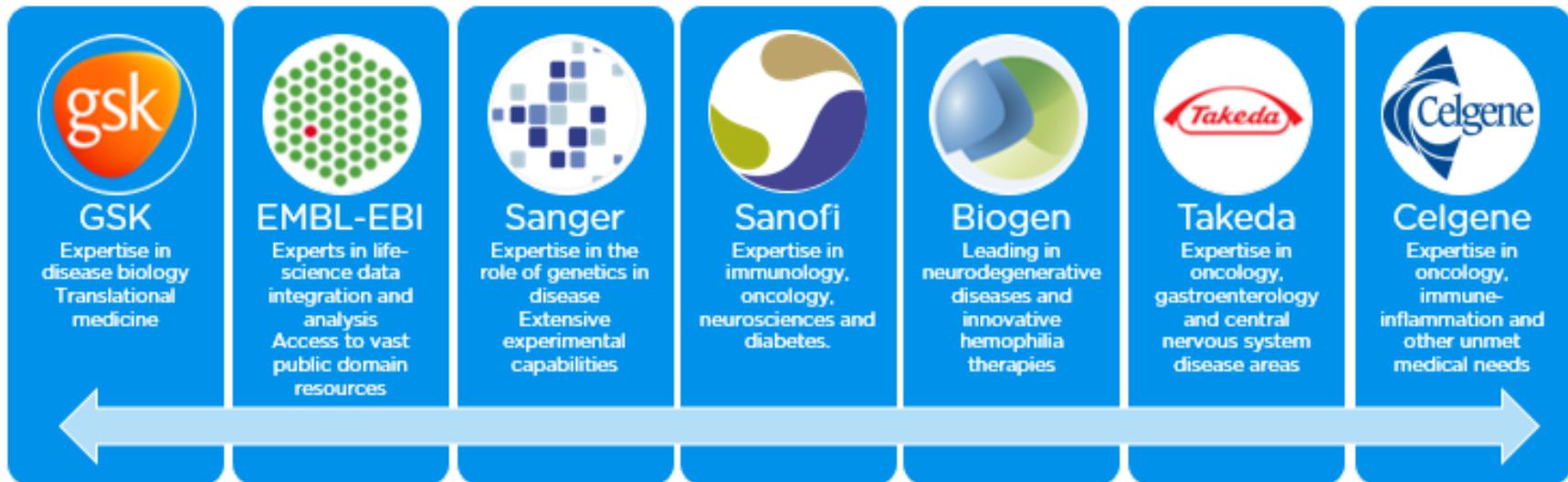
Drug discovery: some challenges



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

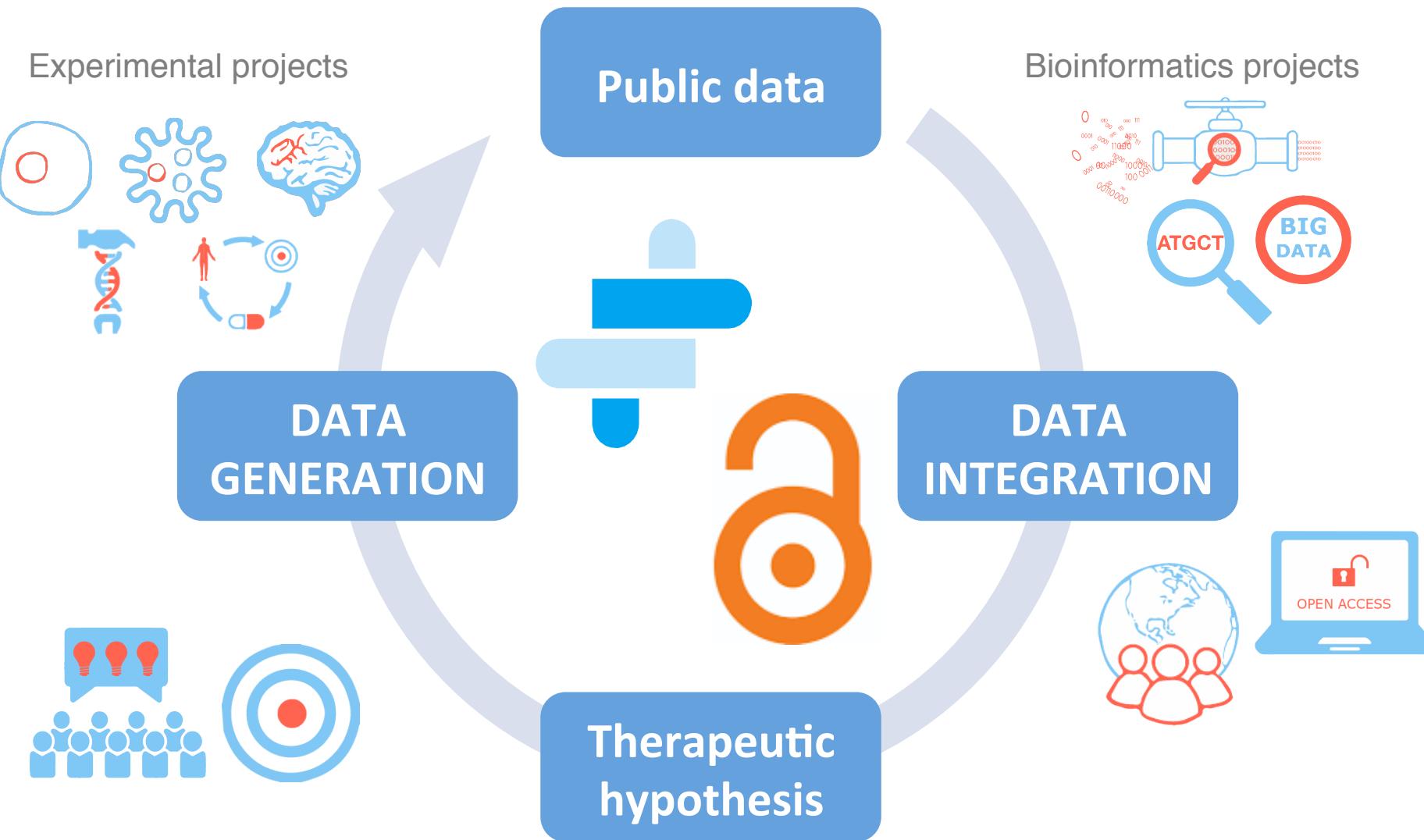
Open Targets

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

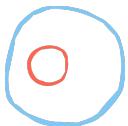


<https://www.opentargets.org>

Virtuous cycle in Open Targets



Data generation: Open Targets



- > 1,000 cancer cell lines + drug sensitivity data
- RNASeq, CRISPR/Cas9 screens
- Wellcome Sanger Institute, GSK and EMBL-EBI



- Genome wide knockouts in gut epithelium
- Organoids, metagenomics
- Wellcome Sanger Institute and GSK



- Alzheimer's and Parkinson's
- CRISPR/Cas9 screens, iPS cells
- Wellcome Sanger Institute, Biogen, Gurdon Institute



Open Targets

Data integration: Open Targets

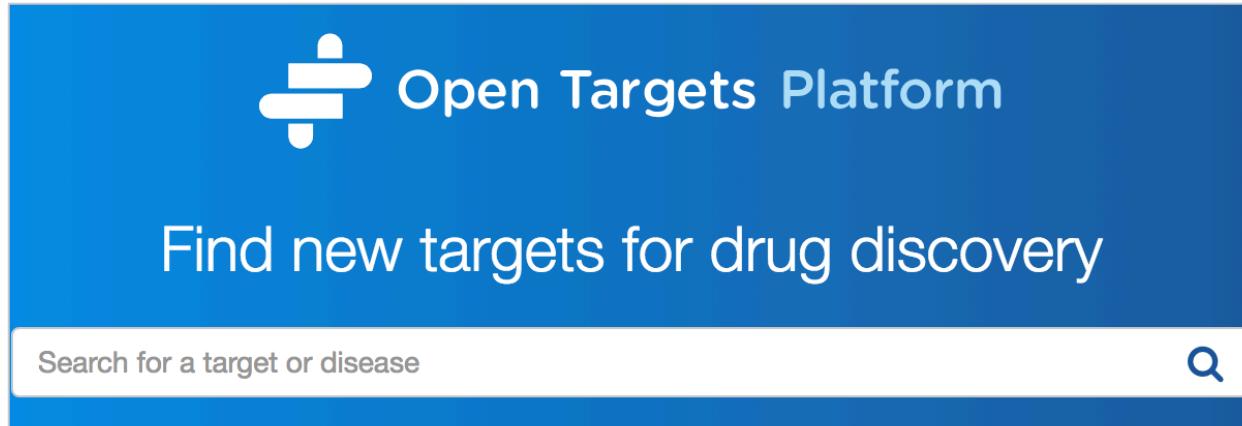
- Open Targets Platform
- Several databases in a single place
- Freely available: disease biology, drug discovery, translational



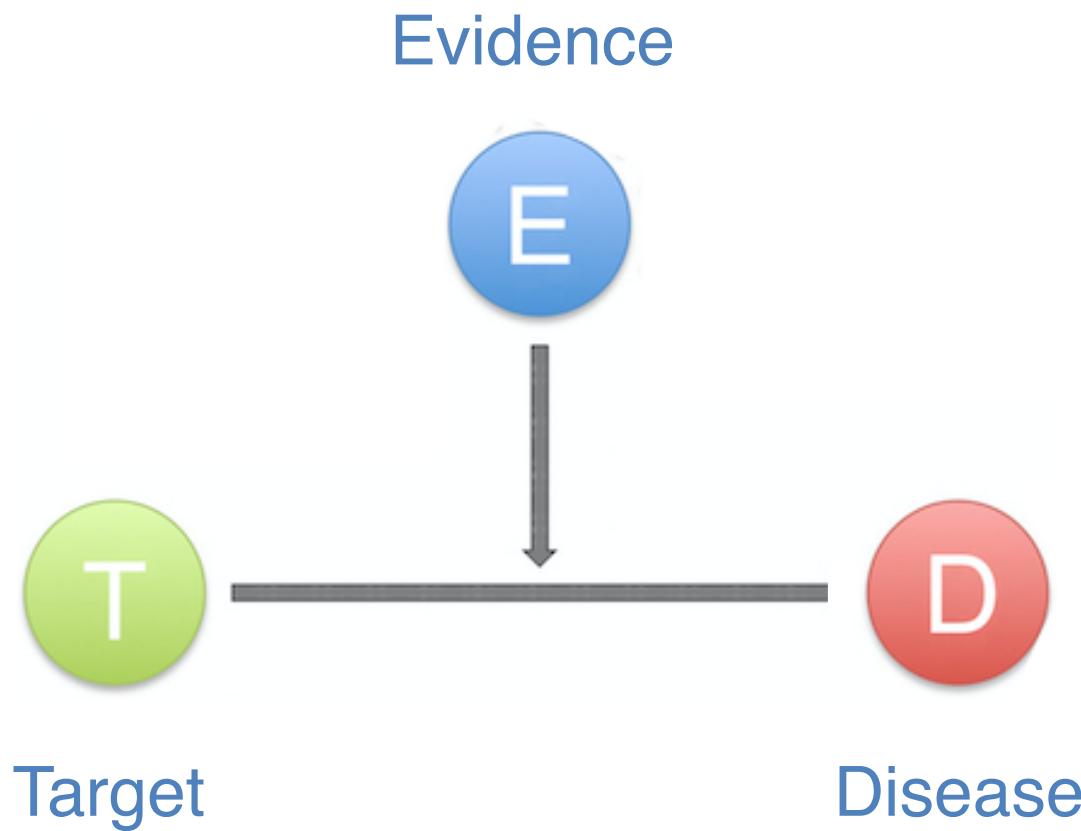
Data integration: reasons



- Data is everywhere
- Fit everything together: time
- Possible lack of resources or expertise



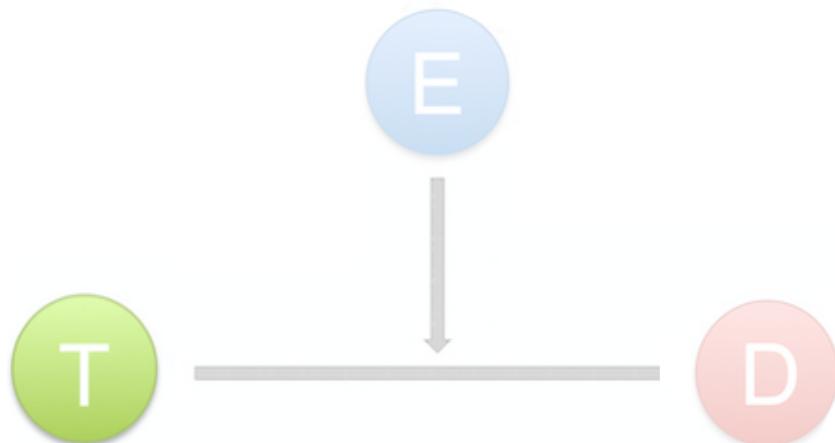
The image shows the Open Targets Platform homepage. At the top left is the platform's logo, which consists of a stylized white 'T' and 'O' icon followed by the text 'Open Targets Platform'. Below the logo is a blue header bar with the text 'Find new targets for drug discovery'. Underneath the header is a search bar containing the placeholder 'Search for a target or disease' and a magnifying glass icon.



Targets → genes (coding and not)

- HGNC names e.g. DMD, MIR100HG (synonyms)
- Ensembl Gene IDs e.g. ENSG00000105641
- UniProt IDs e.g. P15056

21K
targets



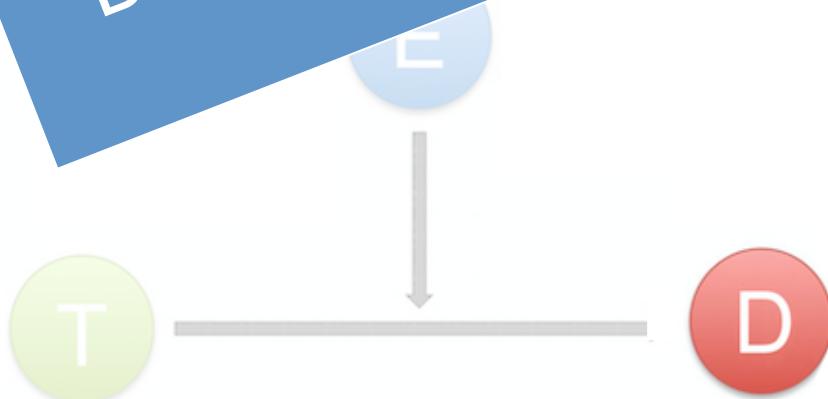
<https://docs.targetvalidation.org/faq/what-is-a-target>

Diseases and phenotypes

10K
diseases

- Rare diseases and common, complex diseases
- Modified version of Experimental Factor Ontology
- Controlled vocabulary
- Related terms

Direct evidence versus indirect evidence



Respiratory system disease

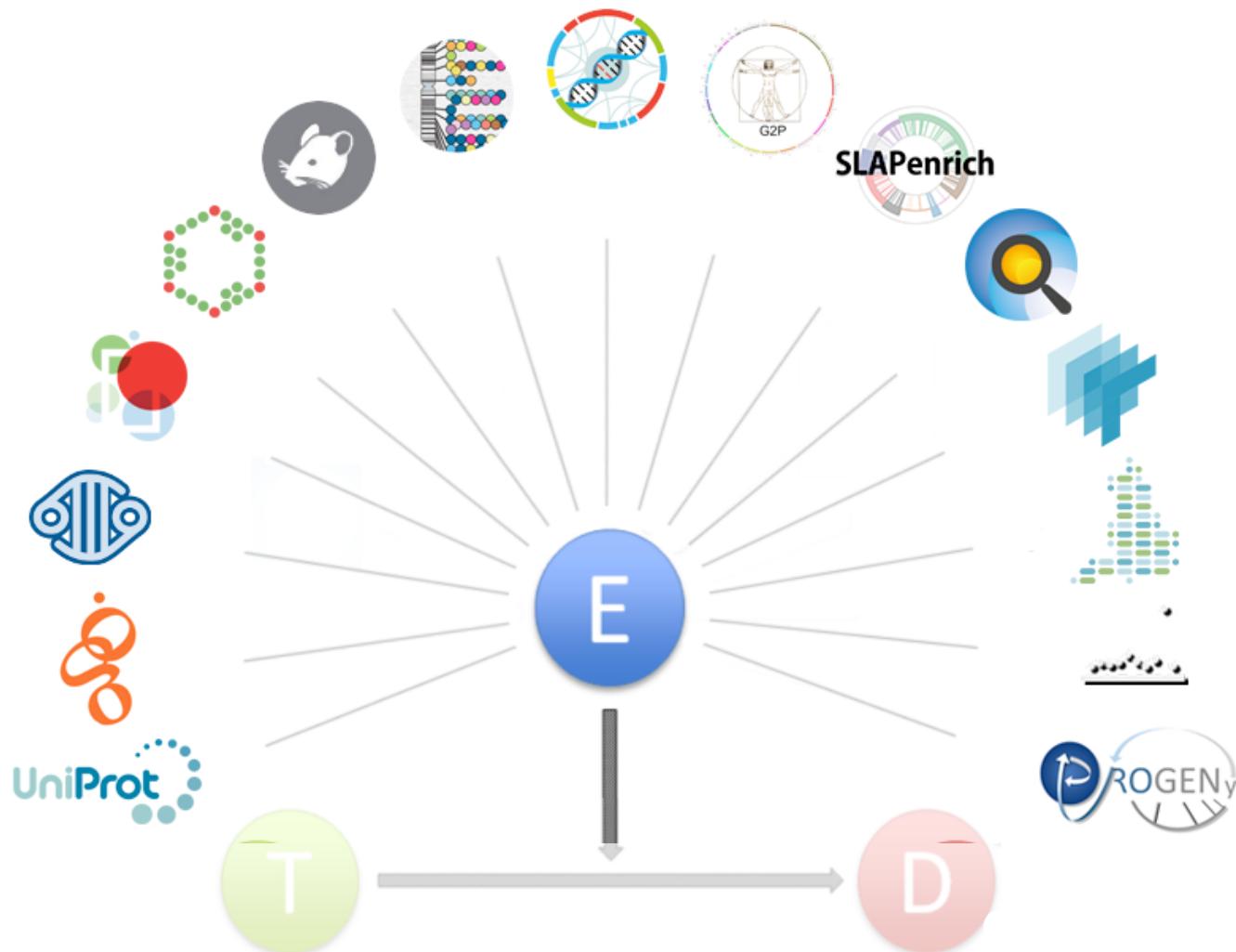
asthma

EFO code EFO_0000270

3090 genes associated (Showing the first 10)

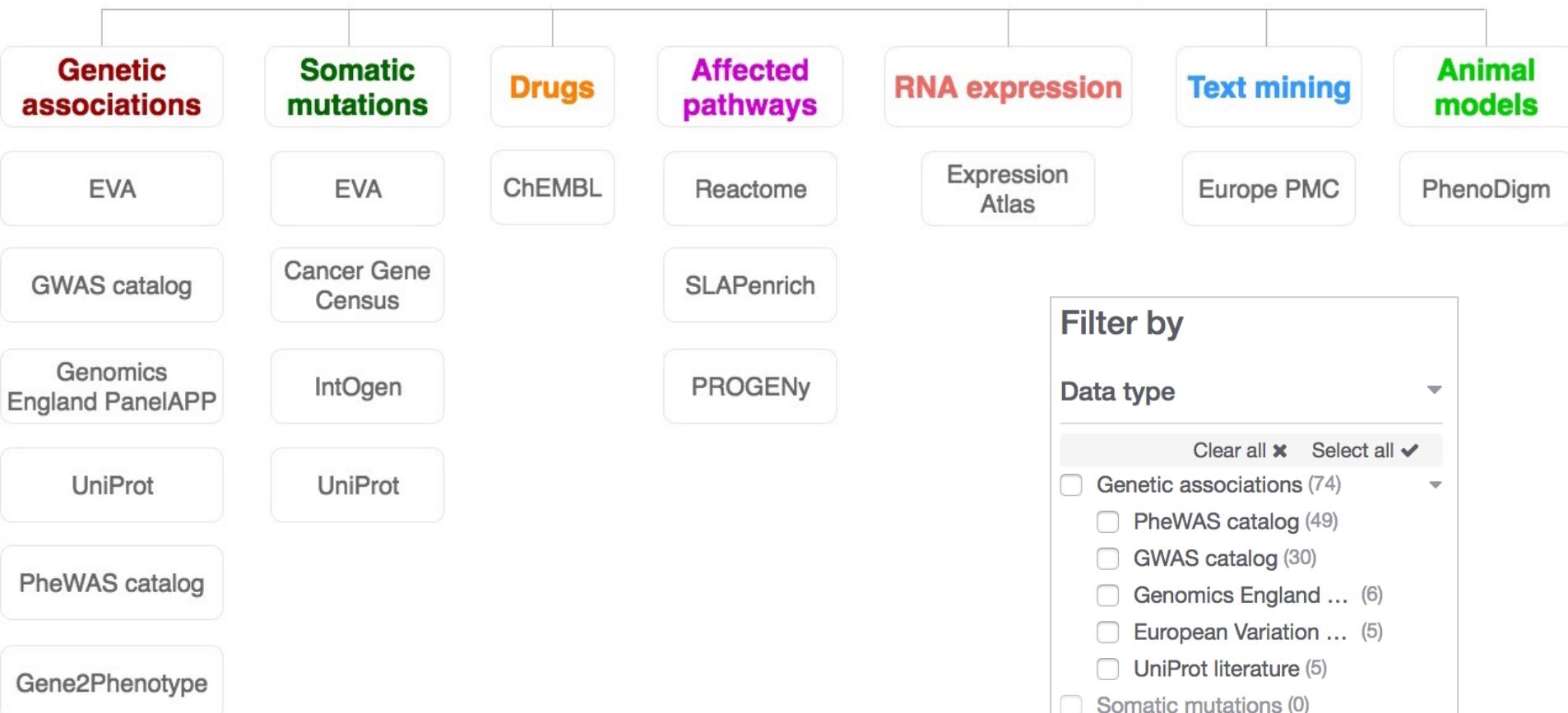
IL13	See Evidence
ORMDL3	See Evidence
PDE4D	See Evidence
CHI3L1	See Evidence
PTGDR	See Evidence
ADAM33	See Evidence

Evidence for our T-D associations



<https://docs.targetvalidation.org/data-sources/data-sources>

Data sources grouped into data types



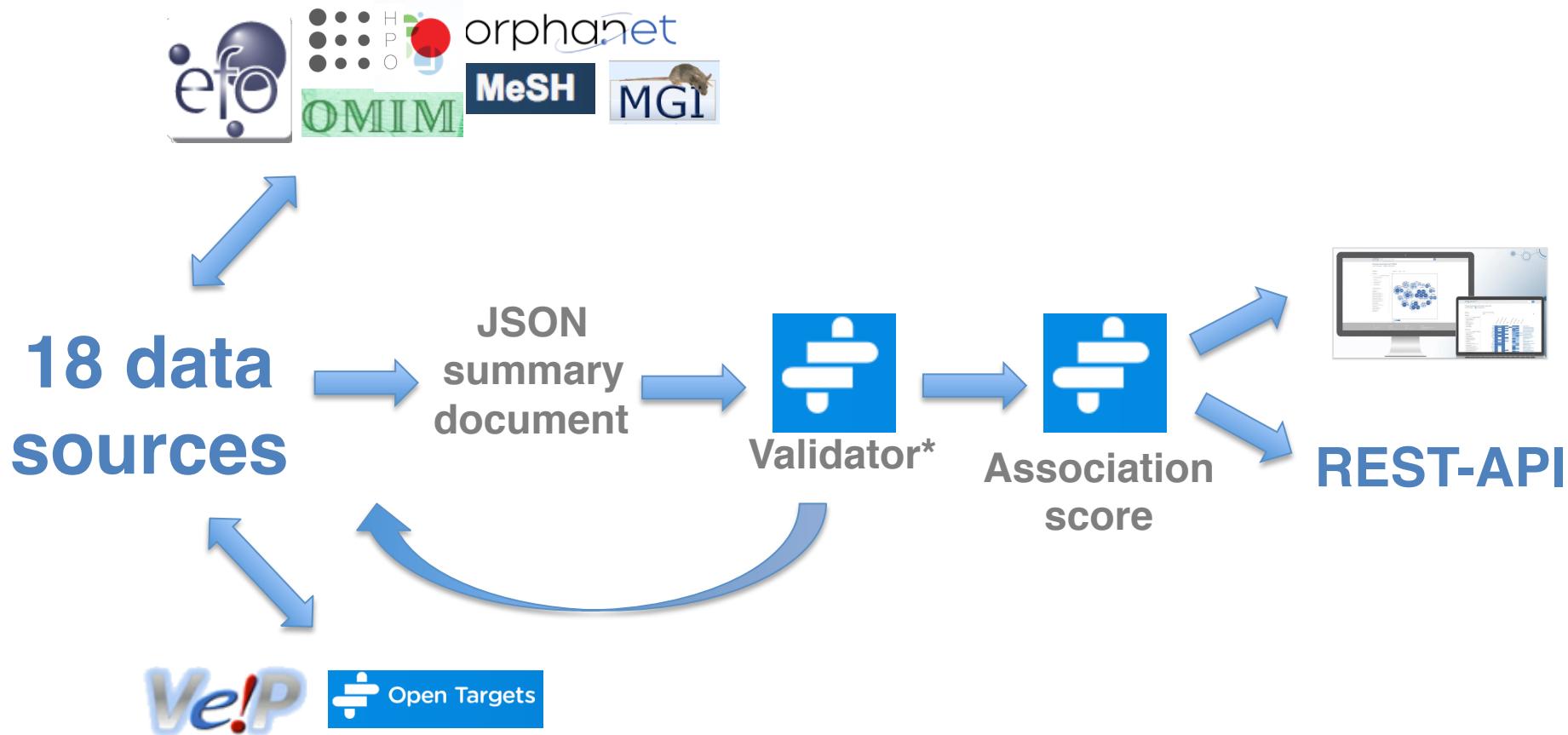
Filter by

Data type

Clear all Select all

- Genetic associations (74)
 - PheWAS catalog (49)
 - GWAS catalog (30)
 - Genomics England ... (6)
 - European Variation ... (5)
 - UniProt literature (5)
- Somatic mutations (0)
- Drugs (30)
- Affected pathways (21)
- RNA expression (4)
- Text mining (99)
- Animal models (45)

How does the data get integrated?



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

JSON schema
Expected structure, missing objects
Disease ID does not start as EFO, Orphanet

JSON summary document

```
"7","target":"http://identifiers.org/ensembl/ENSG00000012048","object":"http://www.ebi.ac.uk/efo/EF0_0004704","variant":"http://identifiers.org/dbsnp/rs1799949","study_name":"otar009_gwas_catalog","pvalue":"8e-11","confidence_interval":"0.1-0.18","odd_ratio":""}]},"evidence": {"variant2disease": {"gwas_sample_size": 69626, "unique_experiment_reference": "http://europepmc.org/abstract/MED/26414677", "gwas_panel_resolution": 2600000, "provenance_type": {"literature": {"references": [{"lit_id": "http://europepmc.org/abstract/MED/26414677"}]}}, "expert": {"status": true, "statement": "Primary submitter of data"}, "database": {"version": "2018-07-03T08:28:14+00:00", "id": "GWAS Catalog", "dbxref": {"version": "2018-07-03T08:28:14+00:00", "id": "http://identifiers.org/gwas_catalog"}}, "is_associated": true, "source_score": {"type": "pvalue", "method": {"description": "pvalue for the SNP to disease association."}, "value": 8e-11}, "evidence_codes": ["http://identifiers.org/eco/GWAS", "http://purl.obolibrary.org/obo/ECO_0000205"], "date_asserted": "2018-07-03T08:28:14+00:00"}, "genotype": {"variant": {"provenance_type": {"expert": {"status": true, "statement": "Primary submitter of data"}, "database": {"version": "2018-07-03T08:28:14+00:00", "id": "GWAS Catalog", "dbxref": {"version": "2018-07-03T08:28:14+00:00", "id": "http://identifiers.org/gwas_catalog"}}, "is_associated": true, "date_asserted": "2018-07-03T08:28:14+00:00", "evidence_codes": ["http://purl.obolibrary.org/obo/ECO_0000205", "http://identifiers.org/eco/cttv_mapping_pipeline"], "functional_consequence": "http://purl.obolibrary.org/obo/SO_0001819"}, "validated_against_schema_version": "1.2.8", "type": "genetic_association", "literature": {"references": [{"lit_id": "http://europepmc.org/abstract/MED/26414677"}]}}}]},"target": {"activity": "http://identifiers.org/cttv.activity/predicted_damaging", "id": "http://identifiers.org/ensembl/ENSG00000012048", "target_type": "http://identifiers.org/cttv.target/gene_evidence", "access_level": "public", "sourceID": "gwas_catalog", "variant": {"type": "SNP single", "id": "http://identifiers.org/dbsnp/rs1799949"}, "disease": {"id": "http://www.ebi.ac.uk/efo/EF0_0004703"}, "unique_association_fields": {"sample_size": 140426, "gwas_panel_resolution": 532488, "pubmed_refs": "http://europepmc.org/abstract/MED/29773799", "target": "http://identifiers.org/ensembl/ENSG00000012048", "object": "http://www.ebi.ac.uk/efo/EF0_0004703", "variant": "http://identifiers.org/dbsnp/rs1799949", "study_name": "otar009_gwas_catalog", "pvalue": "3e-8", "confidence_interval": "0.099-0.205", "odd_ratio": ""}, "evidence": {"variant2disease": {"gwas_sample_size": 140426, "unique_experiment_reference": "http://europepmc.org/abstract/MED/29773799", "gwas_panel_resolution": 532488, "provenance_type": {"literature": {"references": [{"lit_id": "http://europepmc.org/abstract/MED/29773799"}]}}, "expert": {"status": true, "statement": "Primary submitter of data"}, "database": {"version": "2018-07-03T08:50:28+00:00", "id": "GWAS Catalog", "dbxref": {"version": "2018-07-03T08:50:28+00:00", "id": "http://identifiers.org/gwas_catalog"}}, "is_associated": true, "source_score": {"type": "pvalue", "method": {"description": "pvalue for the SNP to disease association."}, "value": 3e-8}, "evidence_codes": ["http://identifiers.org/eco/GWAS", "http://purl.obolibrary.org/obo/ECO_0000205"], "date_asserted": "2018-07-03T08:50:28+00:00"}, "genotype": {"variant": {"provenance_type": {"expert": {"status": true, "statement": "Primary submitter of data"}, "database": {"version": "2018-07-03T08:50:28+00:00", "id": "GWAS Catalog", "dbxref": {"version": "2018-07-03T08:50:28+00:00", "id": "http://identifiers.org/gwas_catalog"}}, "is_associated": true, "date_asserted": "2018-07-03T08:50:28+00:00", "evidence_codes": ["http://purl.obolibrary.org/obo/ECO_0000205", "http://identifiers.org/eco/cttv_mapping_pipeline"], "functional_consequence": "http://purl.obolibrary.org/obo/SO_0001819"}, "validated_against_schema_version": "1.2.8", "type": "genetic_association", "literature": {"references": [{"lit_id": "http://europepmc.org/abstract/MED/26414677"}]}}}]}
```



* IDs (gene, variant, disease, papers) + evidence + source + stats for the score

Association score

Which targets have more evidence for an association with a disease?

What is the relative weight of the evidence for different targets?



Association score

4-tier framework

Statistical integration, aggregation and scoring

- 1) Evidence score: one single and unique evidence
- 2) Data source score: all evidence from one data source
- 3) Data type score: all sources from one data type
- 4) Overall association score: all evidence from all data sources

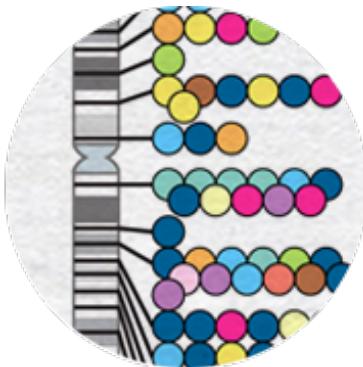
Step 1: evidence score (e.g. one SNP)

$$\text{score} = f * s * c$$

f, relative occurrence of a target-disease evidence

s, strength of the effect of the SNP

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



f = sample size (cases and controls)

s = predicted functional consequence (VEP)

c = *p* value reported in the paper

(Factors affecting the relative strength of GWAS Catalog evidence)

Factors affecting the relative strength of the evidence

Somatic mutations	Cancer Gene Census (functional consequence score of variants); European Variation Archive (functional consequence score of variants); IntOGen (binned score based on tumour type categories. If the gene has several signals of positive selections in the tumour, the score will be 0.25. If the gene is already described as a cancer gene and exhibits a signal of positive selection in a tumor type, the score will be 0.5. If in addition to a signal of positive selection, the gene is functionally connected to other genes in the same tumor type, the score will be 0.75)
Drugs	ChEMBL (Clinical trials phase binned score. Scores will be 0.09 for phase 0, 0.1 for phase I, 0.2 for Phase II, 0.7 for Phase III, and 1 for Phase IV drugs)
Affected pathways	Reactome (functional consequence of 1 for a pathway inferred by a curator). SLAPenrich evidence is scored according to Iorio F et al 2018 followed by quantifying, in large cohorts of cancer patients, the divergence of the total number of samples with genomic alterations in a Reactome-pathway from its expectation, accounting for mutational burdens and total exonic block lengths of genes in that pathway. PROGENy evidence is scored per sample and pathway following a modifications of the original implementation described by Schubert et al. 2016 . Further details can be found elsewhere .
RNA Expression	Expression Atlas score (normalised p-value, normalised expression fold change and normalised percentile rank)
Text mining	Europe PMC (weighting document sections, sentence locations and title for full text articles and abstracts (Kafkas et al 2016))
Animal models	PhenoDigm (similarity score between a mouse model and a human disease described by Smedley et al 2013)



Steps 2, 3, 4: Aggregating data

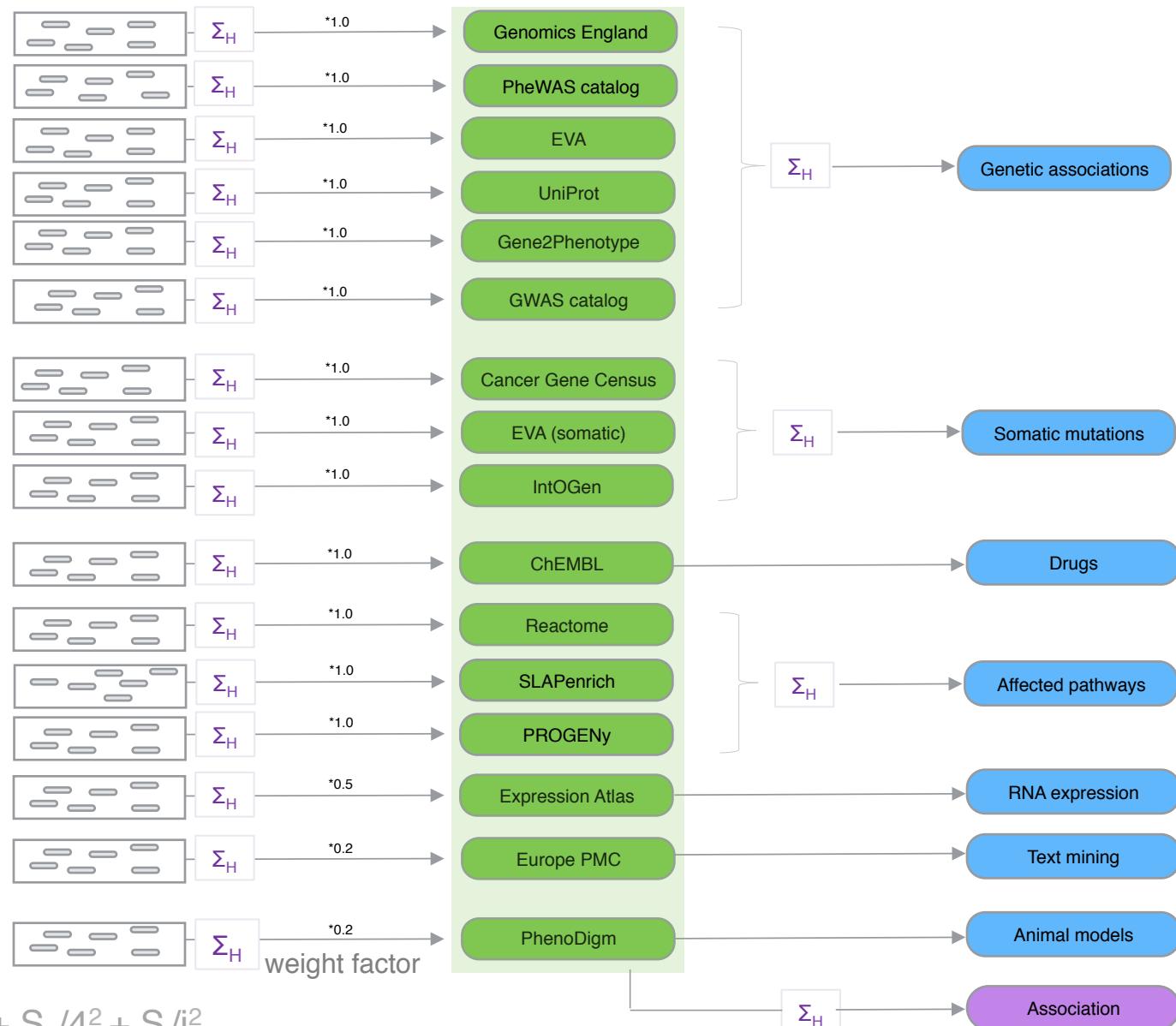
Score: 0 to 1 (max)

Calculated at four levels:

- Evidence
- Data source
- Data type
- Overall

Aggregation with Σ_H (harmonic sum)

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



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

What can you do with the Open Targets Platform?



- Target-disease associations (+ evidence + score)

https://www.targetvalidation.org/evidence/ENSG00000141510/EFO_0000228

- Disease annotations

http://www.targetvalidation.org/diseaset/EFO_0000228

- Target annotations

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

Today: 11:00-17:00

- Introduction
- Open Targets Platform: live demos
- Lunch – live demos - exercises
- Other modes of data access
- Feedback survey and 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 ...

Pages 7 - 17

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

Demo 2: Annotations of diseases/targets

What can I find about a disease?



What is the information available for my target?

Target profile page

CD86

CD86 molecule

Synonyms: B7.2, B7-2, CD28LG2

Receptor involved in the costimulatory signal essential for T-lymphocyte proliferation and interleukin-2 production, by binding CD28 or CTLA-4. May play a critical role in the early events of T-cell a...

Disease profile page

multiple sclerosis

Synonyms: MS (Multiple Sclerosis), MS, Multiple Sclerosis, Acute Fulminating, Disseminated Sclerosis, MULTIPLE...

Pages 17 - 25

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

Demo 3: Target centric workflow

Which diseases are associated with a target?



Open Targets Platform

CD86
CD86 molecule |  View associated diseases

Receptor involved in the costimulatory signal essential for T-lymphocyte proliferation and interleukin-2 production, by binding CD28 or CTLA-4. May play a critical role in the early events of T-cell activation and costimulation of naive T-cells, such as deciding between immunity and anergy that is mediated by CD28 and CTLA-4 respectively.

Open Targets Platform

Find new targets for drug discovery

cd86 

CD86
CD86 molecule
446 diseases associated
Target

Receptor involved in the costimulatory signal essential for T-lymphocyte proliferation and interleukin-2 production, by binding CD28 or CTLA-4. May play a critical role in the early events of T-cell activation and costimulation of naive T-cells, such as deciding between immunity and anergy that is mediated by CD28 and CTLA-4 respectively.

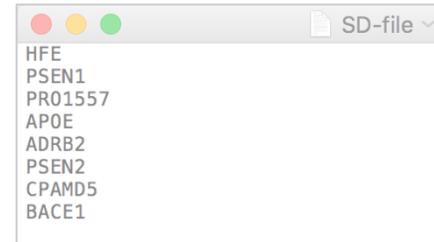


Today: 11:00-17:00

- Introduction
- Open Targets Platform: live demos
- **Lunch – live demos - exercises**
- Other modes of data access
- Feedback survey and wrap up

Hands-on exercises

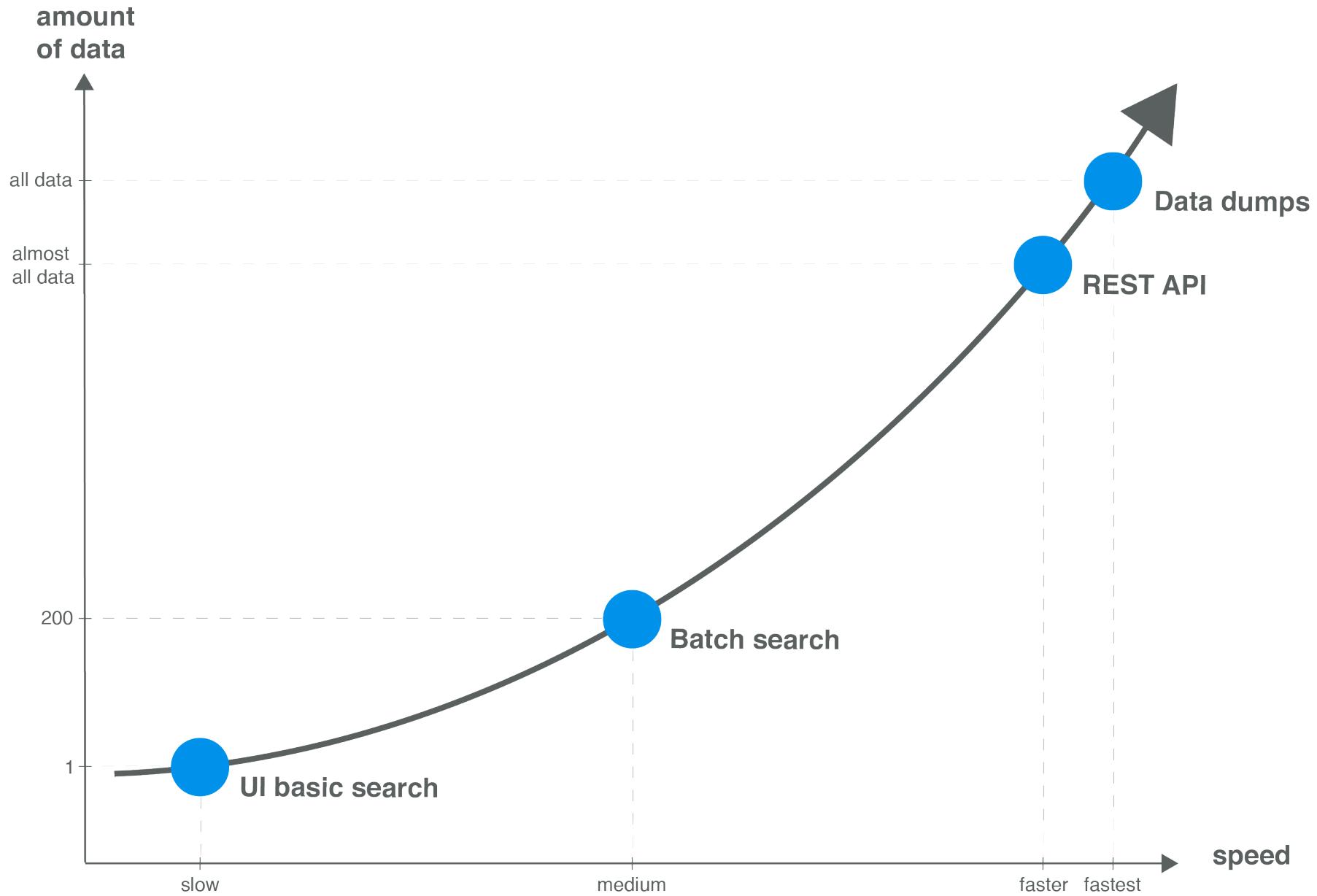
Pages 31-36



Today: 10:00-16:00

- Introduction
- Open Targets Platform: live demos
- Lunch – live demos - exercises
- Other modes of data access
- Feedback and wrap up

Modes of access → data volume



Demo 4: search of many targets at once



We have a list of 26 human genes that are potential targets for IBD

Are these targets also represented in other diseases?

Which pathways and GO terms are enriched in this set of genes?

Open Targets REST API

The video player shows a screenshot of the Open Targets Platform. The interface displays a list of diseases associated with the gene PTEN, with 674 entries. A search bar at the top right is shown. Below the search bar, there's a large blue header "getAssociationFilter". On the left, there's a sidebar titled "Filter by Data type" with checkboxes for various association types like "Genetic associations", "Somatic mutations", etc. The main content area shows a grid of disease names, with some highlighted in blue. At the bottom of the screenshot, there's a URL: "http://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000171862&offset=28:56 / 40:43". To the right of the screenshot, a large block of JSON code is displayed, representing the API response for the query.

```
{
  "from": 0,
  "took": 25,
  "next": [
    "ENSG00000171862-Orphanet_210548"
  ],
  "data_version": "17.12",
  "therapeutic_areas": [],
  "query": {
    "total": 674
  },
  "data": [
    {
      "target": {
        "gene_info": {
          "symbol": "PTEN",
          "name": "phosphatase and tensin homolog"
        },
        "id": "ENSG00000171862"
      },
      "association_score": {
        "datatypes": {
          "literature": 0.3241324475302135,
          "rna_expression": 0,
          "genetic_association": 1,
          "somatic_mutation": 1,
          "known_drug": 0,
          "animal_model": 0,
          "affected_pathway": 1
        },
        "overall": 1
      },
      "datasources": {
        "slapenrich": 0.817215326415924,
        "expression_atlas": 0,
        "europeme": 0.3241324475302135
      }
    }
  ]
}
```

<https://youtu.be/KQbfhwpeEvc>

Data downloads



Open Targets Platform

About ▾ Help ▾ API ▾ Downloads Blog

Data Download

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.

NOTE: the files below are useful only if you want to analyze the data. They are not a database dump and cannot be easily used to replicate the platform locally/somewhere else

2018 Aug

- [Association objects](#) (2018-06, 202MB)
- [Evidence objects](#) (2018-06, 2.44Gb)

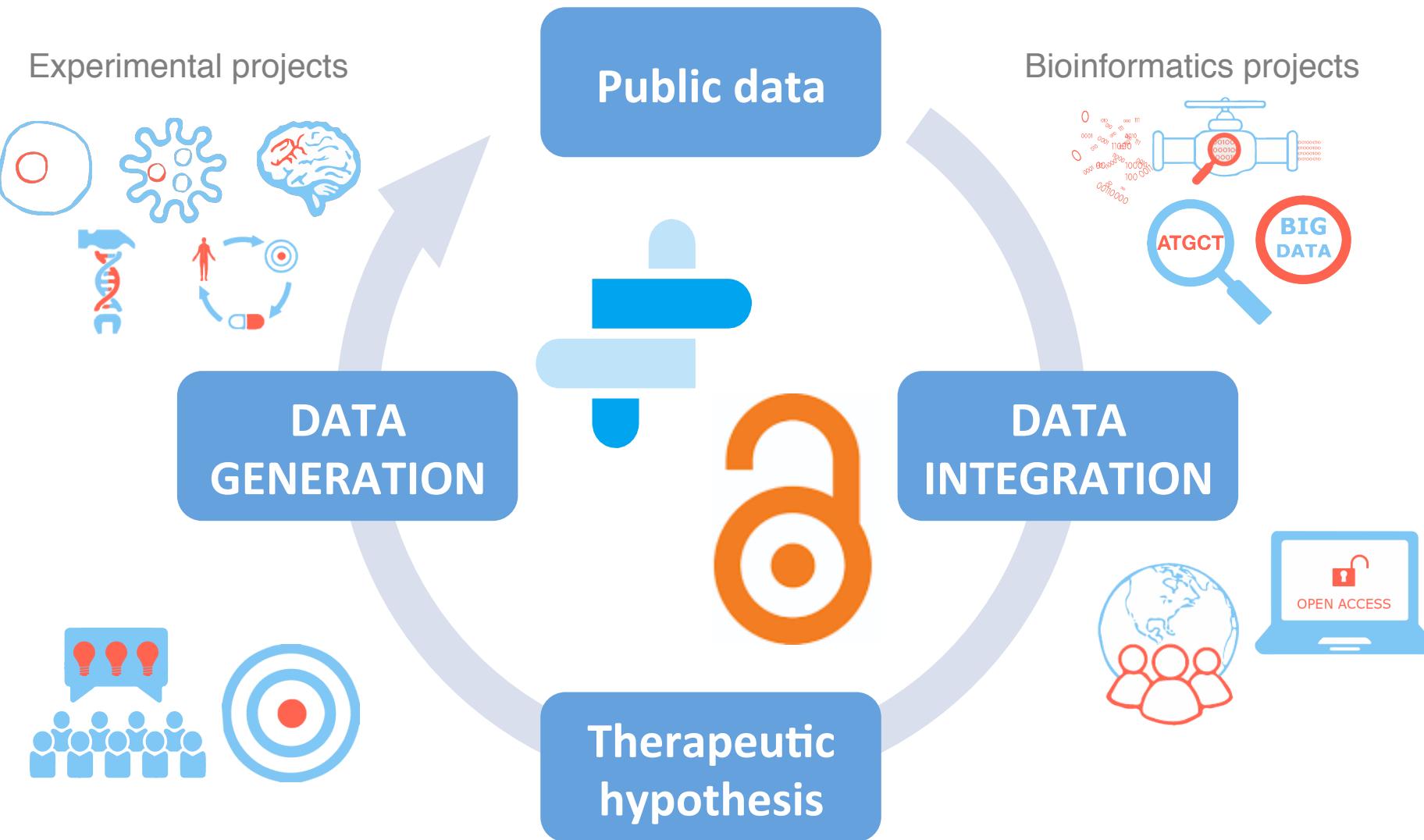
Today: 11:00-17:00

- Introduction
- Open Targets Platform: live demos
- Lunch – live demos - exercises
- Other modes of data access
- Feedback survey and wrap up

Your feedback is important

<https://tinyurl.com/UCL-031218>

Virtuous cycle in Open Targets



Open Targets Platform

- Resource of integrated multiomics data
- Added value (e.g. score) and links to original sources
- Graphical web interface: easy to use
- REST-API and data downloads

21K
targets

10K
diseases

3.1 M
associations

6.8 M
evidence

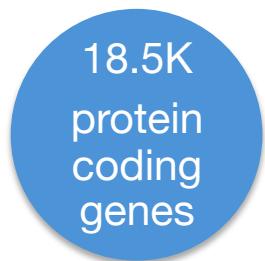
Association → causation?

The screenshot shows the homepage of the Open Targets Genetics website. The page has a white background with a blue hexagonal pattern border. At the top center is the Open Targets logo, which consists of three blue horizontal bars of increasing length from left to right, followed by the text "Open Targets Genetics". Below the logo is a sub-headline: "Explore **variant-gene-trait** associations from UK Biobank and GWAS Catalog". A search bar follows, containing the placeholder text "Search for a gene, variant or trait..." and a magnifying glass icon. Below the search bar are three rectangular buttons with rounded corners containing the text: "PCSK9", "1_154426264_C_T", and "rs4129267". Underneath these buttons is another button containing the text "Crohn's disease (de Lange KM et al. 2017)". At the bottom of the page is a link: "Subscribe to our newsletter".

<https://genetics.opentargets.org>

Open Targets Genetics

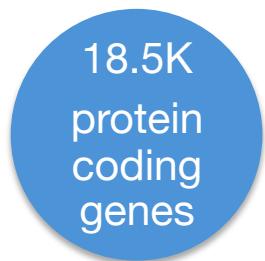
- Resource of integrated functional genomics data
- Statistical framework to guide target identification
- Map variants → genes → traits
- Graphical web interface and GraphQL schema and API



Release v1 – Oct 2018

Open Targets Genetics

- Resource of integrated functional genomics data
- Statistical framework to guide target identification
- Map variants → genes → traits
- Graphical web interface and GraphQL schema and API



Release v1 – Oct 2018

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

Nucleic Acids Research

<http://bit.ly/cite-us>

Issues Section browse ▾ Advance articles Submit ▾ Purchase About ▾



Nucleic Acids Research, 2018 **1**
doi: 10.1093/nar/gky1133

Open Targets Platform: new developments and updates two years on

Denise Carvalho-Silva^{1,2,*}, Andrea Pierleoni^{1,2}, Miguel Pignatelli^{1,2}, ChuangKee Ong^{1,2}, Luca Fumis^{1,2}, Nikiforos Karamanis^{1,2}, Miguel Carmona^{1,2}, Adam Faulconbridge^{1,2}, Andrew Hercules^{1,2}, Elaine McAuley^{1,2}, Alfredo Miranda^{1,2}, Gareth Peat^{1,2}, Michaela Spitzer^{1,2}, Jeffrey Barrett^{2,3}, David G. Hulcoop^{2,4}, Eliseo Papa^{2,5}, Gautier Koscielny^{2,4} and Ian Dunham^{1,2,*}

¹European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, ²Open Targets, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, ³Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK, ⁴GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK and ⁵Biogen, Cambridge, MA 02142, USA

Received September 14, 2018; Revised October 22, 2018; Editorial Decision October 23, 2018; Accepted October 26, 2018

Get in touch



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



support@targetvalidation.org



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



[@targetvalidate](https://twitter.com/targetvalidate)

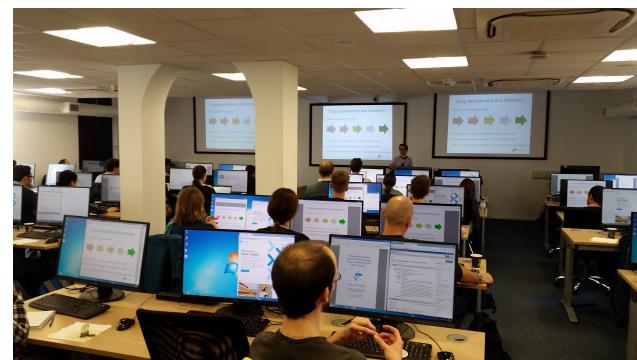
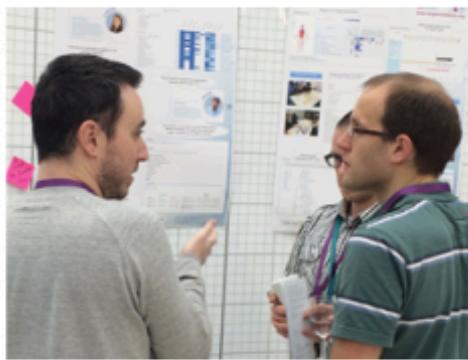
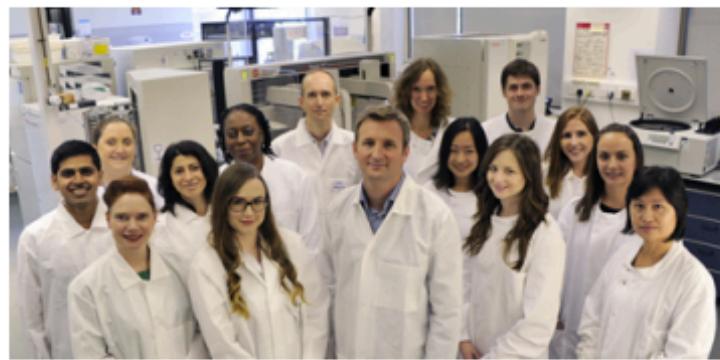


<http://tinyurl.com/opentargets-in>



blog.opentargets.org/

Acknowledgements



GENE0006

You have two options:

- 1. Send an Email to
sms@textwall.co.uk**

**Type 4 things with a space in between
ONLY IN THE SUBJECT LINE:**

- 1) bs065 (*lower case only*)**
- 2) Student number**
- 3) Your Surname**
- 4) Unique word (*as listed in the example below*)**

e.g. bs065 1234567 SMITH Biology

OR

- 2. TEXT the same info to 0207 183 8329**
(Your standard text message rates will apply.)

ALL STUDENTS MUST RESPOND REGARDLESS OF VISA STATUS.

**Extra extra extra
slides**

Target tractability

- Tractability: confidence that a modulator interacting with a target will produce a desirable biological effect
- Modified from Brown et al 2018 (PMID:30108951)
- 20,000 targets: small molecules (SM) and antibodies (Ab)

Target tractability buckets



- Clinical phases e.g. phase IV (bucket 1)
- Cellular localization e.g. plasma membrane (bucket 4)
- DrugEBIility – ensemble score e.g. > 0.7 (bucket 5)

Target tractability – small molecules

- Clinical precedence:



- Discovery precedence:



- Predictable tractable:



<https://www.targetvalidation.org/target/ENSG00000142192?view=sec:tractability>

<https://api.opentargets.io/v3/platform/private/target/ENSG00000142192>

Target tractability – antibodies



- Clinical precedence



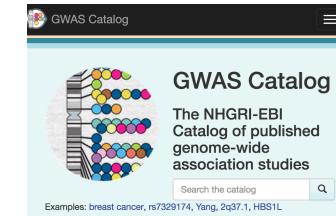
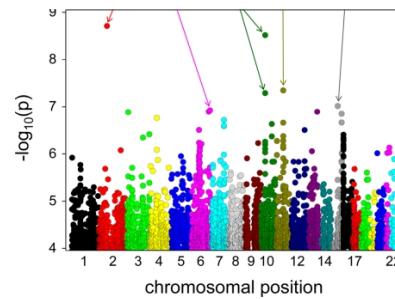
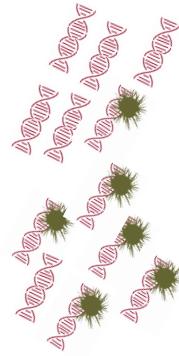
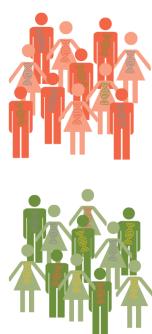
- Predictable tractable – high confidence



- Predictable tractable – medium low confidence

Data sources: GWAS catalog

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



Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients



Data sources: Gene2Phenotype

Gene2Phenotype

Downloads

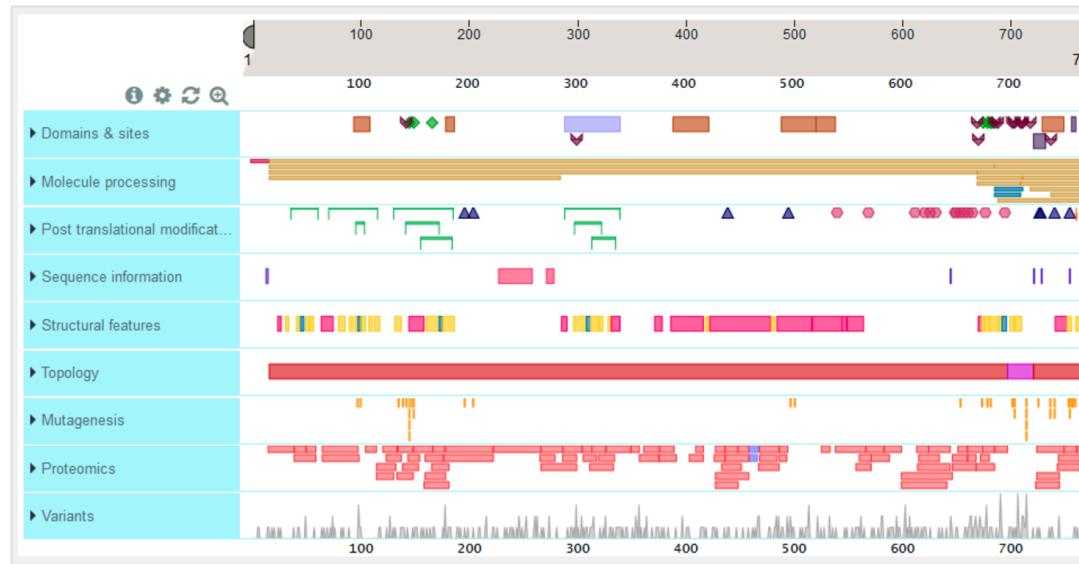
Search panel ALL for: Search

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

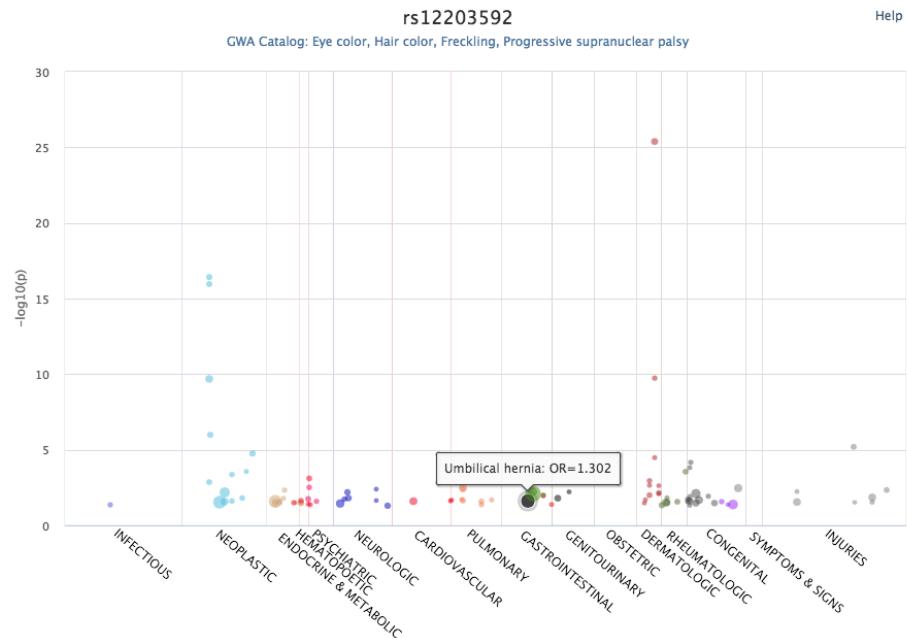


- Manual curation of coding variants in patients



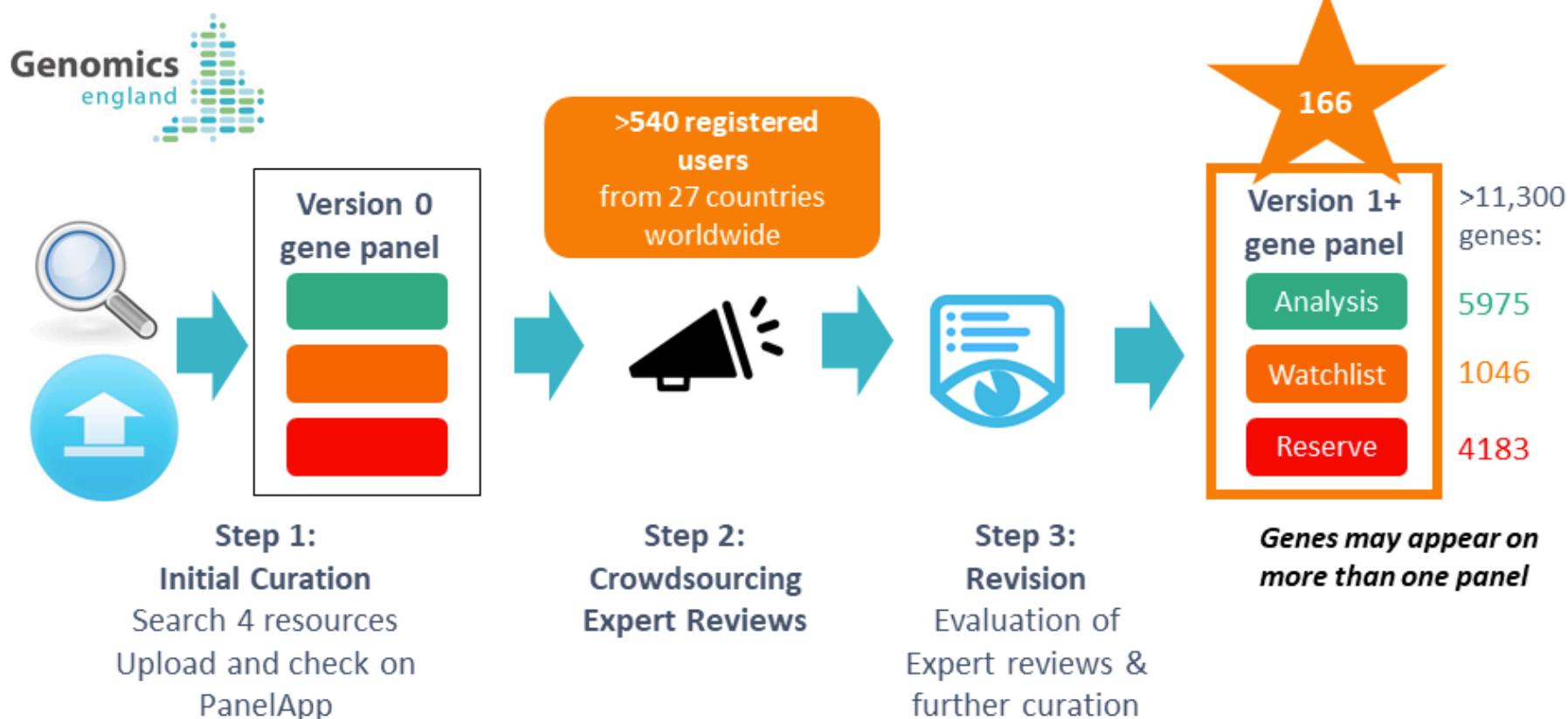
Data sources: PheWAS

- 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) website. The header features the EVA logo and the text "European Variation Archive". Below the header is a navigation bar with links: Home, Submit Data, Study Browser, Variant Browser, Clinical Browser (which is highlighted in black), GA4GH, API, FAQ, and Feedback. The main content area is titled "ClinVar Browser" with a help icon. On the left, there is a "Filter" section with buttons for "Reset" and "Sub...", and dropdown menus for "Position" (set to GRCh37), "Assembly" (set to GRCh37), "Filter By:" (set to Chromosomal), and a specific position entry "2:48000000-49000000". The main table displays 960 records, with the first 10 shown in detail:

...	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



- 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



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 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

Data sources: SLAPenrich

METHOD

Dissecting the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich

Francesco Iorio^{1,5†}, Luz Garcia-Alonso^{1,5}, Jonathan Brammell², Iñigo Martincorena², David R Wille^{3,5}, Ultan McDermott^{2,5} and Julio Saez-Rodriguez^{1,4,5*†}

- 374 pathways curated and mapped to cancer hallmarks
- Divergence of the total number of cancer samples with genomic alterations
- Mutational burden and total exonic block length of genes

Data sources: PROGENy

ARTICLE

DOI: 10.1038/s41467-017-02391-6

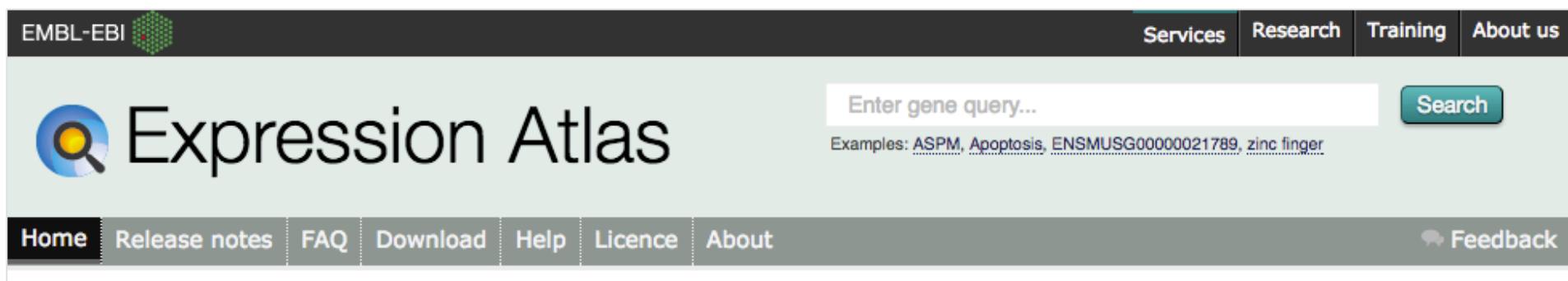
OPEN

Perturbation-response genes reveal signaling footprints in cancer gene expression

Michael Schubert¹, Bertram Klinger^{2,3}, Martina Klünemann^{2,3}, Anja Sieber^{2,3}, Florian Uhlitz^{2,3}, Sascha Sauer⁴, Mathew J. Garnett⁵, Nils Blüthgen^{2,3} & Julio Saez-Rodriguez^{2,3}

- Comparison of pathway activities between normal and primary samples from The Cancer Genome Atlas
- Inferred from RNA-seq: 9,250 tumour and 741 normal samples
- EGFR, hypoxia, JAK/STAT, MAPK, NFκB, PI3K, TGFb, TNF α , Trail, VEGF, and p53

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 on the left and navigation links for Services, Research, Training, and About us on the right. 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 prominently displayed in large black font next to the search bar. 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
 - target profile page
- Differential mRNA expression (healthy *versus* diseased):
 - target-disease associations



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
- Dictionary (not NLP)



EMBL-EBI train online

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. At the bottom of the page, there is a horizontal navigation 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*

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

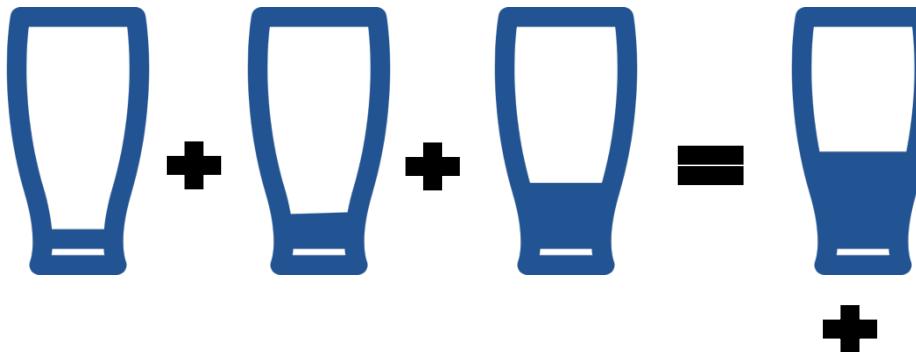
where S_1, S_2, \dots, S_i are the individual sorted evidence scores in descending order

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

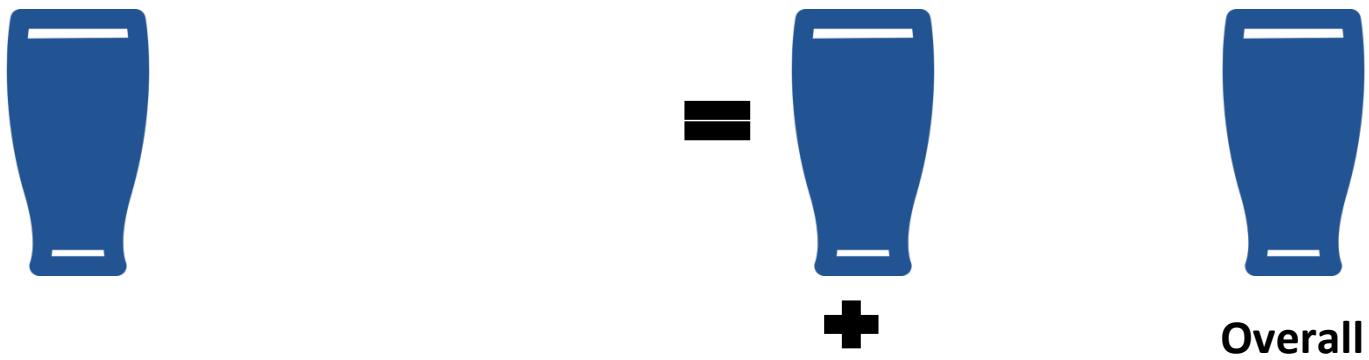
* PMID: 19107201, PMID: 20118918

Target-Disease Association Score

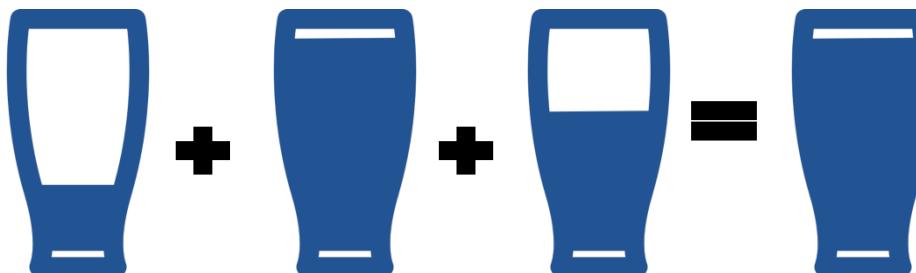
EuropePMC
(Text Mining)



UniProt
(Manual Curation)



ChEMBL
(Manual Curation)



Overall

VERY simplified diagram

Open Targets REST API

<https://api.opentargets.io/v3/platform/docs/swagger-ui>

filter Methods to filter the available evidence.

GET

[/platform/public/association/filter](#) Filter available associations

GET

[/platform/public/evidence/filter](#) Filter available evidence

<https://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000163914&fields=target.id&fields=disease.id&size=10000&format=csv>

Server

Endpoint

Parameters

How to run our REST endpoints (option 1)

Paste the URL in the location bar of your browser

The screenshot shows a browser window with the following details:

- Address Bar:** Secure | <https://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000167207>
- Content Area:** A large JSON object representing the search results for the target gene ENSG00000167207. The JSON is too long to show in full here, but it includes fields like "from", "took", "next", "data_version", "query", "target", "association_score", and various association types (e.g., "genetic_association", "somatic_mutation", "known_drug"). It also lists data sources and specific phenotypic associations.

```
{  
    from: 0,  
    took: 22,  
    ▼ next: [  
        1.3668802,  
        "ENSG00000167207-Orphanet_101988"  
    ],  
    data_version: "17.09",  
    ▼ query: {  
        ▼ sort: [  
            "harmonic-sum.overall"  
        ],  
        search: null,  
        rna_expression_level: 0,  
        protein_expression_tissue: [ ],  
        ▼ scorevalue_types: [  
            "overall"  
        ],  
        datatype: [ ],  
        fields: null,  
        format: "json",  
        facets_size: null,  
        disease: [ ],  
        protein_expression_level: 0,  
        datastructure: "default",  
        facets: "false",  
        rna_expression_tissue: [ ],  
        ▼ target: [  
            "ENSG00000167207"  
        ],  
        target_class: [ ],  
        cap_scores: true,  
        pathway: [ ],  
        size: 10  
    }  
}
```



JSONView

How to run our REST endpoints (option 2)

Command line e.g. CURL -X GET

```
denise-ml:~ denise$ curl -X GET https://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000167207&is_direct=true
[[1] 45611
[2] 45612
[2]+ Done
denise-ml:~ denise$ {"from": 0, "took": 13, "next": 1.3668802, "ENSG00000167207-Orphanet_101988": [{"target": {"id": "ENSG00000167207", "label": "nucleotide binding oligomerization domain containing 2"}, "association_score": {"datatypes": {"literature": 0.2974496, "slapenrich": 0.0, "expression_atlas": 0.0006625794025179989, "europemc": 0.14497512075102087, "reactome": 0.29744967878526557, "eva": 0.9245864351309815, "gene2phenotype": 0.0, "gwas_catalog": 1, "intogen": 0.0, "gene_census": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.0016}, "labels": [], "codes": []}, "path": [{"EFO_0000540"}], "label": "immune system disease"}, "id": "EFO_0000540", "literature": 1463.0, "rna_expression": 32.0, "genetic_association": 133.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.0, "total": 1820.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 32.0, "europemc": 1463.0, "twentythousand": 2.0, "eva": 11.0, "gene2phenotype": 0.0, "gwas_catalog": 17.0, "intogen": 0.0, "genomics_england": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 9.0}}, "id": "ENSG00000167207", "label": "nucleotide binding oligomerization domain containing 2"}, "id": "ENSG00000167207"}, "association_score": {"datatypes": {"literature": 0.0, "genetic_association": 1.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 0.0, "europemc": 0.09408949039475947, "twentythousand": 1591672500813, "eva": 0.900426255952381, "gene2phenotype": 0.0, "gwas_catalog": 0.0, "intogen": 0.0, "gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.00139053}}, "codes": [], "path": [{"EFO_0000508"}], "label": "genetic disorder"}, "id": "EFO_0000508"}, "is_direct": true}
```



How to run our REST endpoints (option 3)

Use our free Python client*



The screenshot shows a documentation page for the opentargets Python client. The left sidebar has a blue header with the 'opentargets' logo and 'latest' version. It includes links for 'Search docs', 'Tutorial', 'High Level API', 'Low Level API', 'Code Documentation', and 'Changelog'. The main content area has a white header with 'Docs' and the URL 'opentargets - Python client for targetvalidation.org'. A purple circular icon is in the top right. The main title is 'opentargets - Python client for targetvalidation.org'. Below it is a description: 'opentargets is the official python client for the Open Targets REST API at targetvalidation.org'.

Advantage: you can change the way the associations are scored e.g. increase the weight given to text mining data

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

Some use cases for the REST-API

How to search for a target (or disease)

How to find the associations for a target

How to get the underlying evidence

How to get the association score

How to get target annotations



How to search

Open Targets Platform

Search results for pten

Found 32 results | 0.119 seconds

Refine by:

- Target 26
- Disease 6

PTEN
Tumor suppressor, tyrosine-phosphatase, phosphatase and dual-specificity kinase, Target

GET /platform/public/search

```
{  
    from: 0,  
    took: 538,  
    data_version: "17.12",  
    - query: {  
        highlight: true,  
        fields: null,  
        datastructure: "default",  
        format: "json",  
        size: 675  
    },  
    total: 32, ←  
    - data: [  
        - {  
            - data: {  
                + ortholog: {...},  
                - top_associations: {  
                    - total: [  
                        - {  
                            score: 1,  
                            id: "ENSG00000171862-EFO_0000616"  
                        },  
                        - {  
                            score: 1,  
                            id: "ENSG00000171862-EFO_0000405"  
                        }  
                    ]  
                }  
            }  
        }  
    ]  
}
```

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

How to get the associations with a target

Open Targets Platform

674 diseases associated with PTEN

[View PTEN profile](#)

Filter by

Data type

- Genetic associations (50)
- Somatic mutations (187)
- Drugs (0)
- Affected pathways (49)
- RNA expression (19)
- Text mining (611)
- Animal models (0)

Therapeutic area

Bubbles Table Tree

Showing 1 to 50 of 674 entries

Search:

Disease Association score Genetic associations Somatic mutations Drugs

Disease	Association score	Genetic associations	Somatic mutations	Drugs
neoplasm				
cancer				
genetic disorder				
epithelial neoplasm				
carcinoma				
Inherited cancer-predisposing s...				
PTEN hamartoma tumor syndro...				
Overgrowth syndrome				

GET

[/platform/public/association/filter](#)

[http://api.opentargets.io/v3/platform/public/association/filter?
target=ENSG00000171862
&direct=true](http://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000171862&direct=true)

```
{  
    from: 0,  
    took: 25,  
    next: [  
        1.4207987,  
        "ENSG00000171862-Orphanet_210548"  
    ],  
    data_version: "17.12",  
    therapeutic_areas: [...],  
    query: {...},  
    total: 674,  
    data: [  
        {  
            target: {  
                gene_info: {  
                    symbol: "PTEN",  
                    name: "phosphatase and tensin homolog"  
                },  
                id: "ENSG00000171862"  
            },  
            association_score: {  
                datatypes: {  
                    literature: 0.3241324475302135,  
                    rna_expression: 0,  
                    genetic_association: 1,  
                    somatic_mutation: 1,  
                    known_drug: 0,  
                    animal_model: 0,  
                    affected_pathway: 1  
                },  
                overall: 1,  
            },  
            datasources: {  
                slapenrich: 0.817215326415924,  
                expression_atlas: 0,  
                europepmc: 0.3241324475302135,  
            }  
        }  
    ]  
}
```

How to get the the association score

674 diseases associated with PTEN

 View PTEN profile

Filter by

Data type

- Genetic associations (50) ▾
- Somatic mutations (187) ▾
- Drugs (0)
- Affected pathways (49) ▾
- RNA expression (19) ▾
- Text mining (611) ▾
- Animal models (0)

Therapeutic area

Bubbles Table Tree

Showing 1 to 50 of 674 entries

Search:



GET

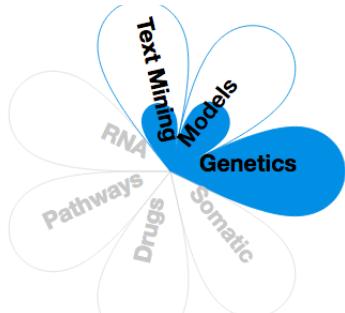
[/platform/public/association](#)

http://api.opentargets.io/v3/platform/public/association?id=ENSG00000171862-EFO_0000616

```
{  
    from: 0,  
    took: 0,  
    data_version: "17.12",  
    query: { },  
    total: 1,  
    - data: [  
        - {  
            - target: {  
                - gene_info: {  
                    symbol: "PTEN",  
                    name: "phosphatase and tu-  
                    geneid: "ENSG00000171862"  
                },  
                id: "ENSG00000171862"  
            },  
            - association_score: {  
                - datatypes: {  
                    literature: 0.3241324475:  
                    rna_expression: 0,  
                    somatic_mutation: 1,  
                    genetic_association: 1,  
                    known_drug: 0,  
                    animal_model: 0,  
                    affected_pathway: 1  
                },  
                overall: 1,  
            }  
        }  
    ]  
}
```

How to get the evidence

Evidence for DMD in Duchenne muscular dystrophy



Genetic associations

Table Browser

Rare diseases

Source: UniProt European Variation

Showing 1 to 10 of 297 entries

Search:

GET

[/platform/public/evidence/filter](#)

```
],
  data_version: "18.04",
- query: {
    - sort: [
        "scores.association_score"
    ],
    format: "json",
    fields: null,
    datastructure: "default",
    pathway: [ ],
    size: 10
},
  total: 297,
- data: [
    - {
        - target: {
            target_name: "DMD",
            id: "ENSG00000198947",
        - gene_info: {
            symbol: "DMD",

```

[http://api.opentargets.io/v3/platform/public/evidence/filter?
target=ENSG00000198947&disease=Orphanet_98896&datatype=genetic_association](http://api.opentargets.io/v3/platform/public/evidence/filter?target=ENSG00000198947&disease=Orphanet_98896&datatype=genetic_association)

How to get the target associations



Target tractability				
Summary of tractability information for APP for small molecule and antibody molecules				
Source: Open Targets				
Small molecule				
Clinical precedence		Discovery precedence		
Phase 4	Phase 2 or 3	Phase 0 or 1	PDB targets with ligands	Active compounds in ChEMBL
Antibody				
Clinical precedence		Predicted tractable - high confidence		
Phase 4	Phase 2 or 3	Phase 0 or 1	UniProt location - high confidence	GO cell component - high confidence



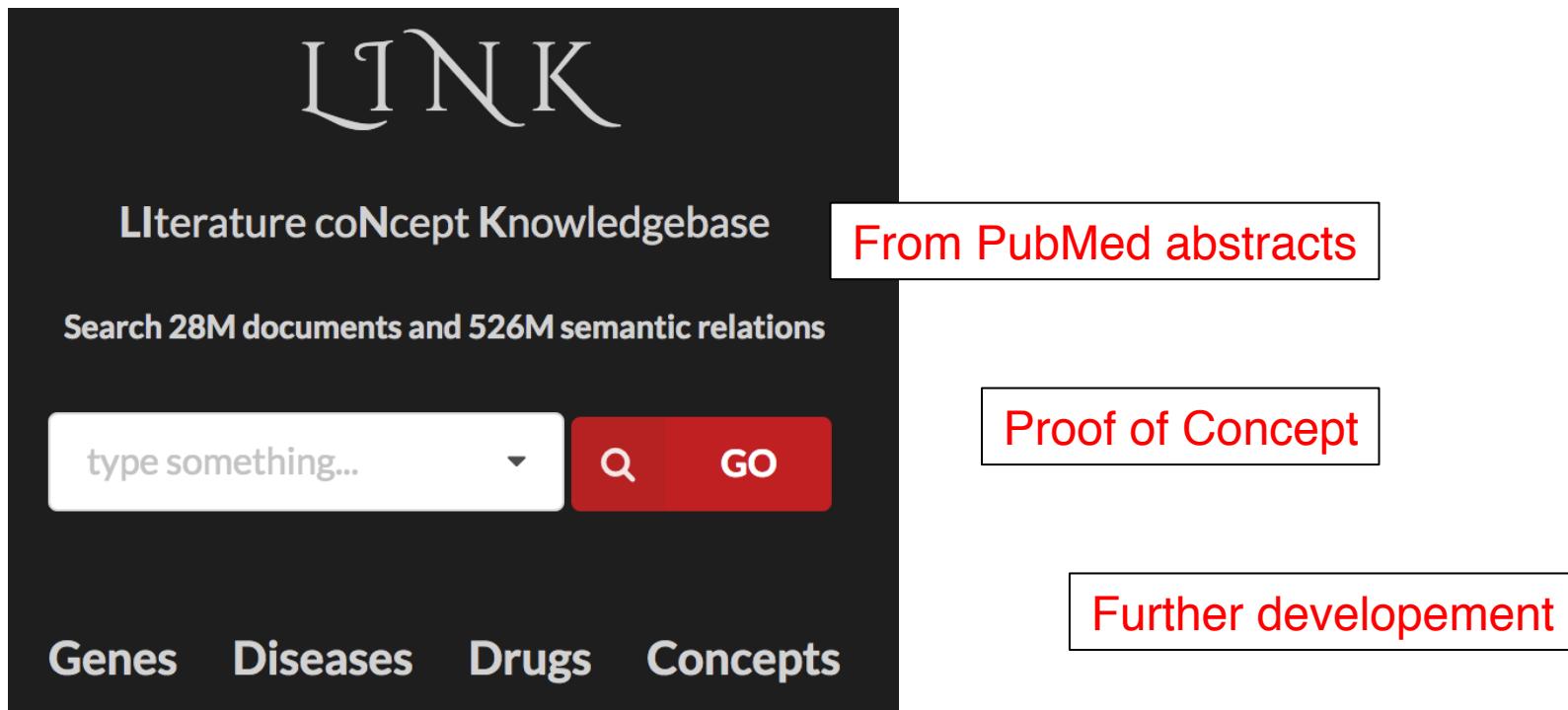
```
tractability:
  - smallmolecule: {
      top_category: "Discovery Precedence",
      small_molecule_genome_member: true,
      - buckets: [
          4,
          5,
          7,
          8
      ],
      high_quality_compounds: 285,
      ensemble: 0.76275614,
      - categories: {
          clinical_precedence: 0,
          predicted_tractable: 1,
          discovery_precedence: 1
      }
    },
  - antibody: {
      top_category: "Clinical Precedence",
      - buckets: [
          2,
          5,
          7
      ],
      - categories: {
          predicted_tractable_med_low_confidence: 0.25,
          clinical_precedence: 0.7000000000000001,
          predicted_tractable_high_confidence: 0.3000000
```

<http://api.opentargets.io/v3/platform/private/target/ENSG00000142192>

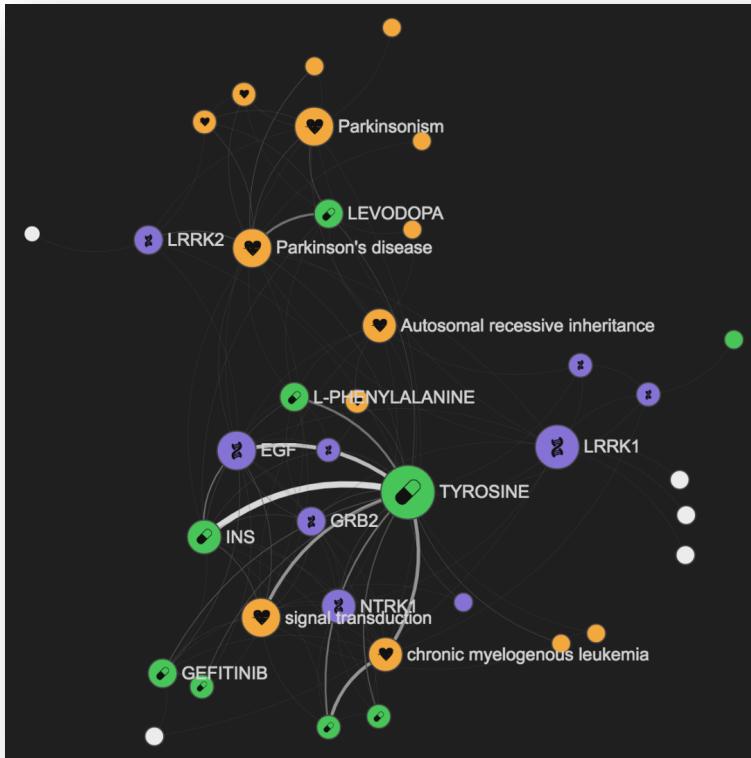
LINK

<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

