

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

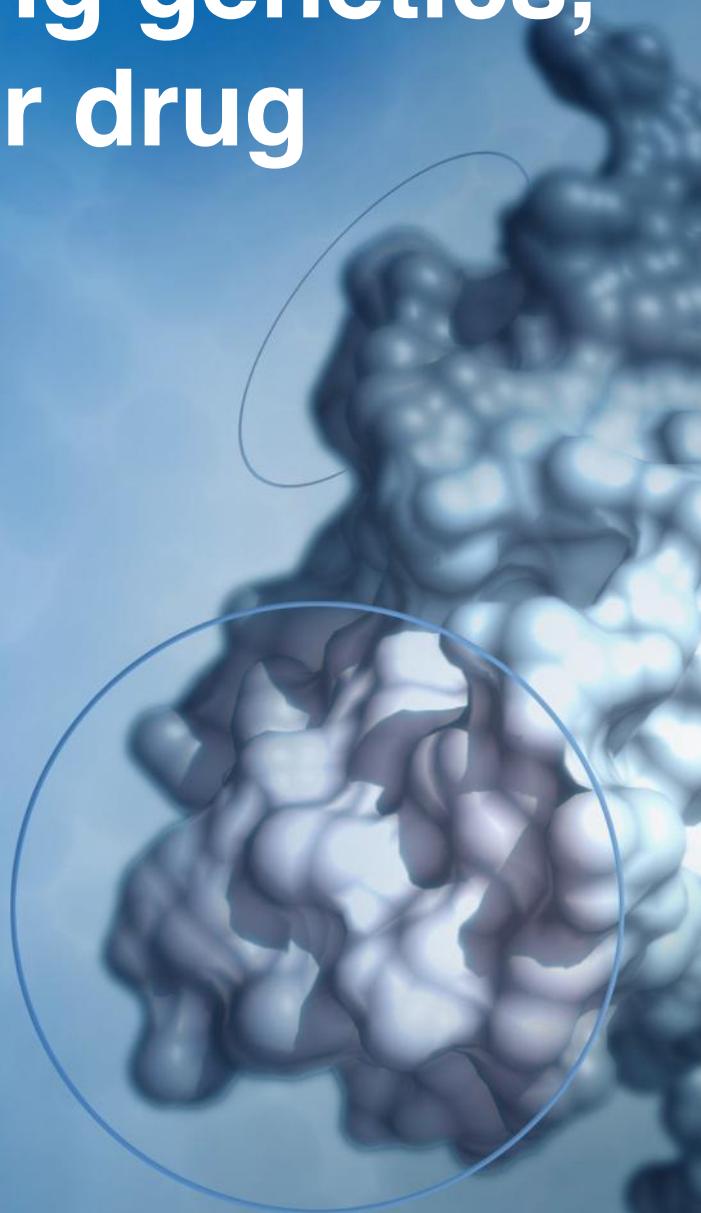
University of Oxford  
September 26<sup>th</sup> 2018

**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 use the  
Open Targets Platform for?

Other ways to  
access the data

How can you get in  
touch with the team?

# Materials

- Slides



- Coursebook



- Input files

SD-file
HFE
PSEN1
PR01557
APOE
ADRB2
PSEN2
CPAMD5
BACE1

[www.github.com/deniseOme/training](https://www.github.com/deniseOme/training)

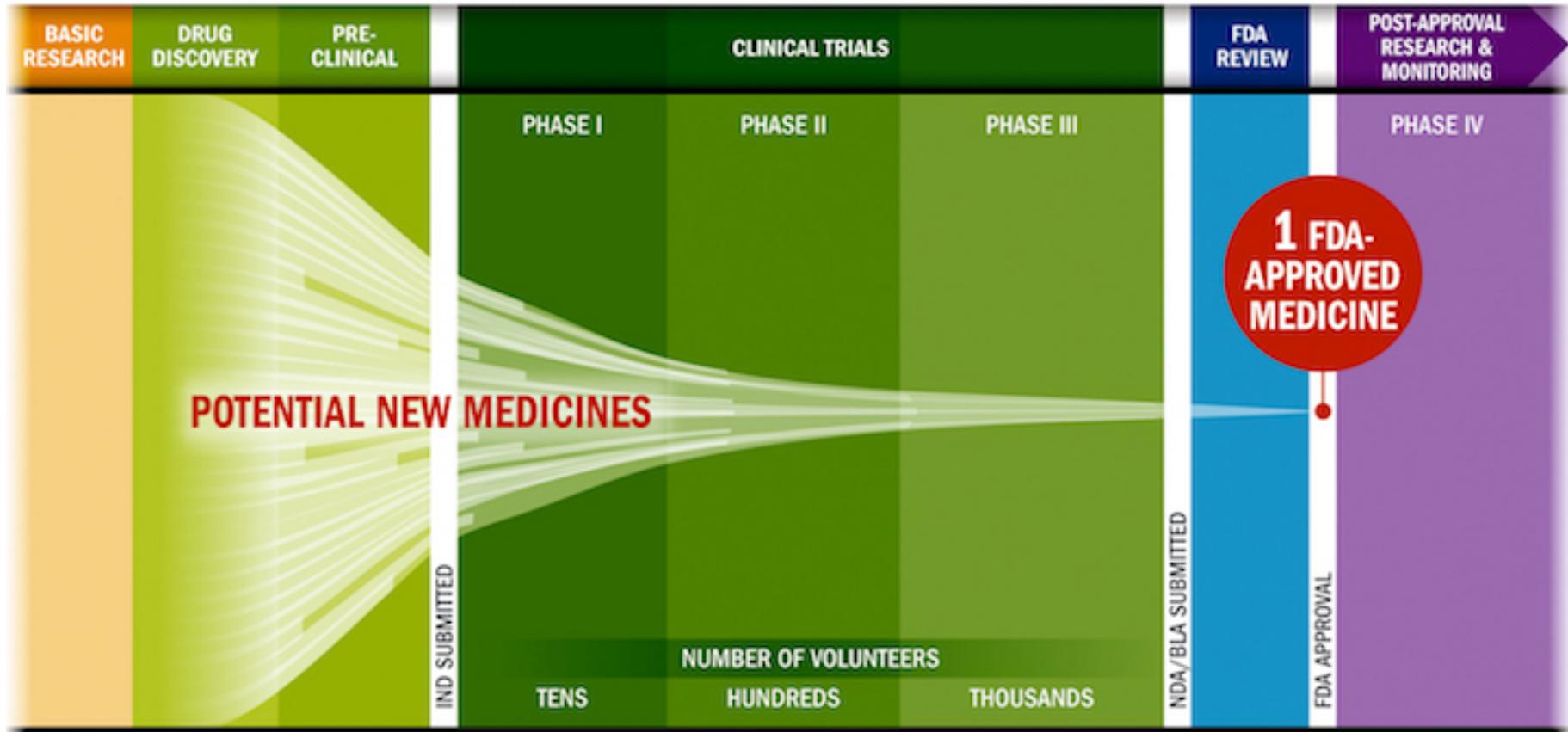
# In these ~3.0 hours

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

# In these ~3.0 hours

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

# Drug discovery: some challenges



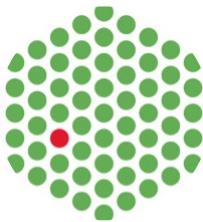
Lengthy, costly, low success rate, **HIGH ATTRITION RATES**

Source: PhRMA adaptation based on Tufts CSDD & School of Medicine, and FDA

# Open Targets

<https://www.opentargets.org>

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



2014

2016

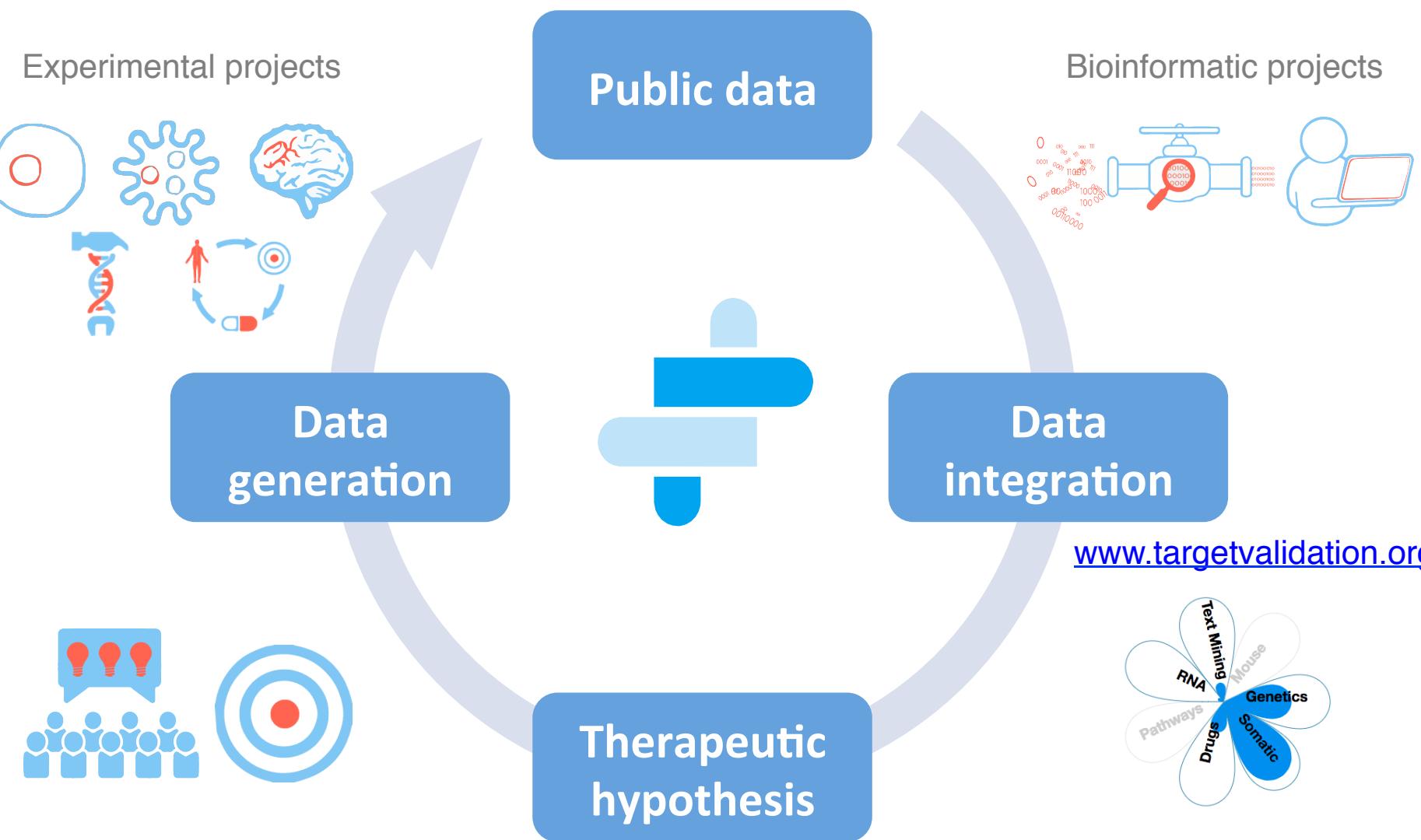
2017

2018

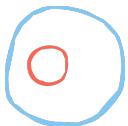


Open Targets

# Virtuous cycle in Open Targets



# Open Targets generates data



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



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



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



Open Targets

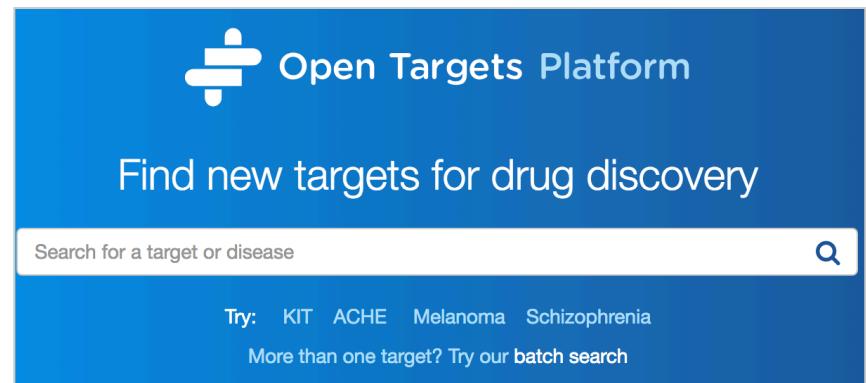
# Open Targets integrates data\*



Target annotations (cv)

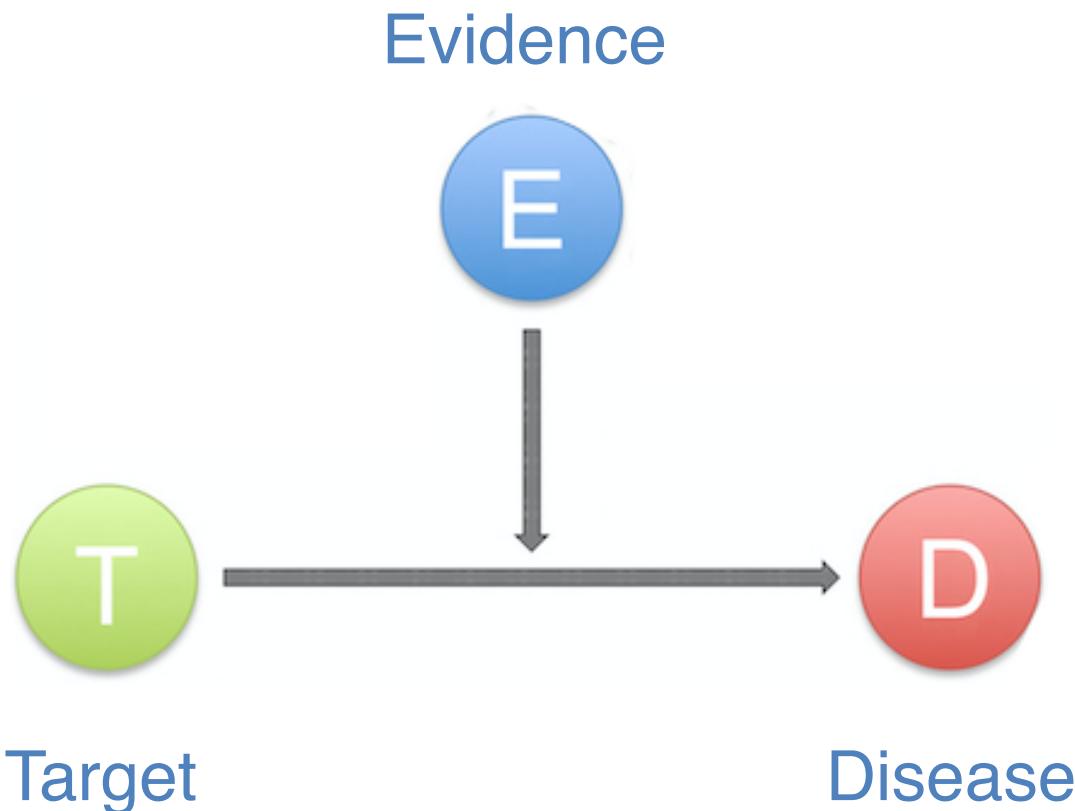
Disease annotations

Evidence for T-D associations



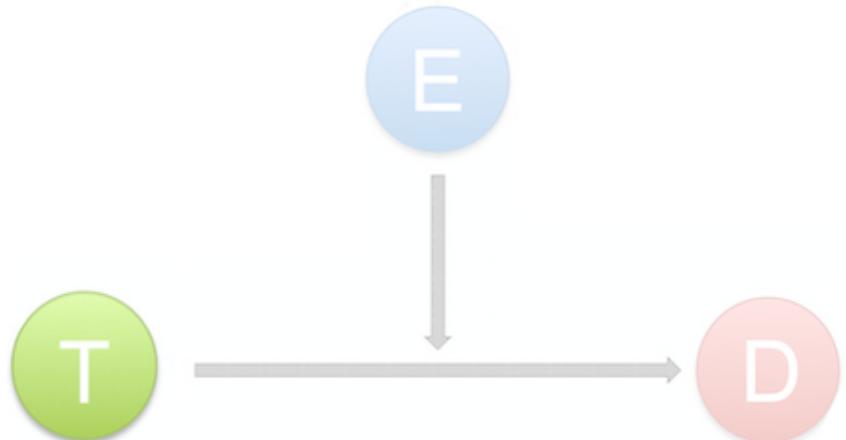
The image shows a screenshot of the Open Targets Platform search interface. At the top, there is a logo consisting of a white minus sign inside a blue square, followed by the text "Open Targets Platform". Below this, a large button with the text "Find new targets for drug discovery" is centered. A search bar with the placeholder "Search for a target or disease" and a magnifying glass icon is located below the button. At the bottom, there is a row of text with the word "Try:" followed by a list of diseases: KIT, ACHE, Melanoma, and Schizophrenia. Below this list, the text "More than one target? Try our batch search" is displayed.

# Open Targets Platform



# 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

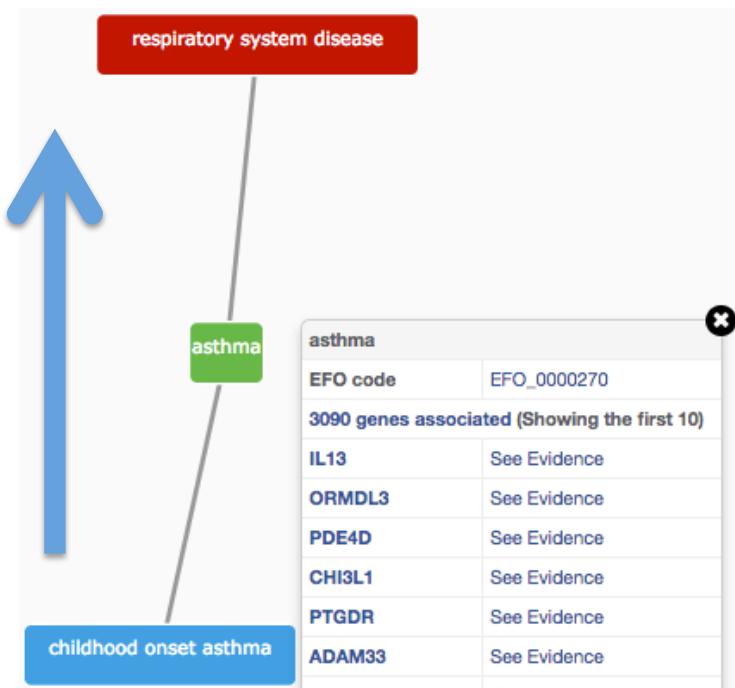
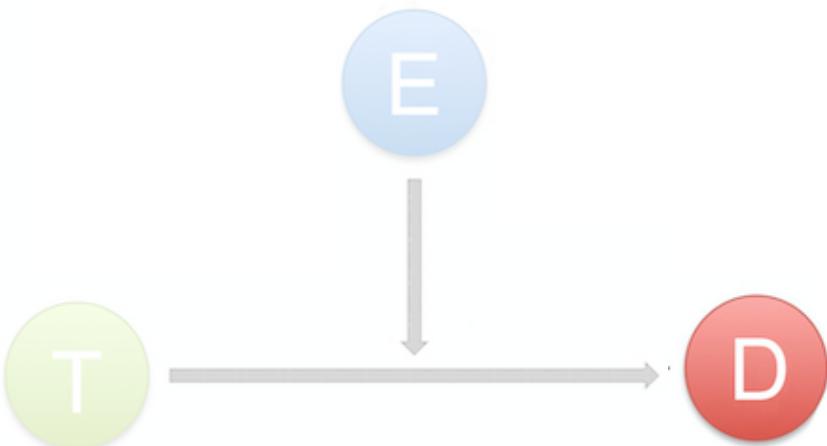


<https://docs.targetvalidation.org/faq/what-is-a-target>

# Diseases, and phenotype, traits

- Modified version of 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



# What can you use the Platform for?

Target annotations

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

Disease annotations

[http://www.targetvalidation.org/diseaset/EFO\\_0000228](http://www.targetvalidation.org/diseaset/EFO_0000228)

Evidence for T-D associations

[https://www.targetvalidation.org/evidence/ENSG00000141510/EFO\\_0000228](https://www.targetvalidation.org/evidence/ENSG00000141510/EFO_0000228)



# What can you use the Platform for?

Target annotations

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



Disease annotations

[http://www.targetvalidation.org/diseaset/EFO\\_0000228](http://www.targetvalidation.org/diseaset/EFO_0000228)

Evidence for T-D associations

[https://www.targetvalidation.org/evidence/ENSG00000141510/EFO\\_0000228](https://www.targetvalidation.org/evidence/ENSG00000141510/EFO_0000228)

# Target annotations

Drugs

Protein Information

Pathways

Similar targets (based on diseases in common)

Variants, isoforms and genomic context

Protein interactions

RNA and protein baseline expression

Mouse phenotypes

Protein Structure

Gene Ontology

Gene tree

Bibliography

Cancer hallmarks

Cancer biomarkers

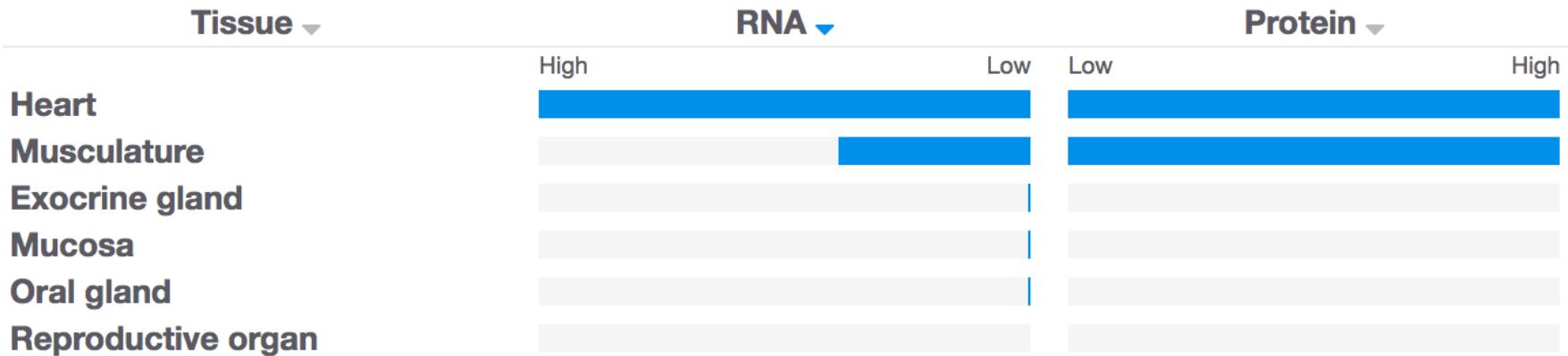


# Target annotations

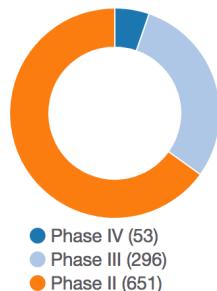
Group by

Organs

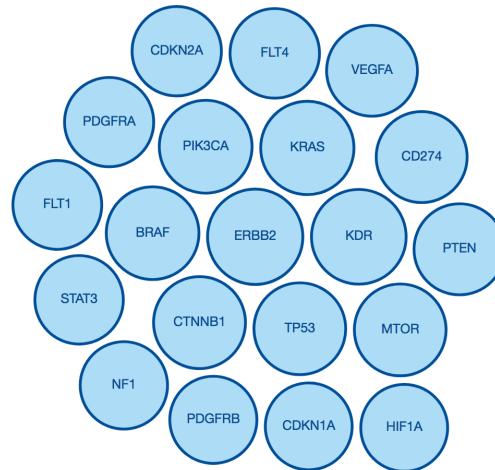
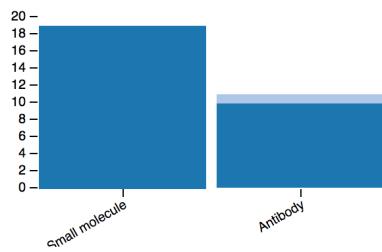
Anatomical Systems



Trials



Type vs Activity



Synthesis and biological evaluation of 2-styryl-4,6-disubstituted 2-(4-(dimethylamino)thiosemicarbazide, p-hydroxybenzaldehyde, ethylcyanide and their derivatives against cancer cell lines (HepG2 and HCT116). Furthermore, interpretation of the results showed clearly that the derivative with the highest cytotoxicity was the one containing the styryl group.

Magda A-A El-Sayed, Walaa M El-Husseiny, Naglaa I Abd

Journal of enzyme inhibition and medicinal chemistry 201

Hide abstract Show similar

## Abstract

A new series of 4,6-disubstituted 2-(4-(dimethylamino)thiosemicarbazide, p-hydroxybenzaldehyde, ethylcyanide and their derivatives against cancer cell lines (HepG2 and HCT116). Furthermore, interpretation of the results showed clearly that the derivative with the highest cytotoxicity was the one containing the styryl group.

Gene

Disease

Drug

Target and disease

# What can you use the Platform for?

Target annotations

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



Disease annotations

[http://www.targetvalidation.org/diseaset/EFO\\_0000228](http://www.targetvalidation.org/diseaset/EFO_0000228)

Evidence for T-D associations

[https://www.targetvalidation.org/evidence/ENSG00000141510/EFO\\_0000228](https://www.targetvalidation.org/evidence/ENSG00000141510/EFO_0000228)

# Disease annotations

Similar diseases (based on targets in common)



Phenotypes



Drugs

Bibliography



Classification

<https://docs.targetvalidation.org/getting-started/getting-started/disease-profile>

# What can you use the Platform for?

Target annotations

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



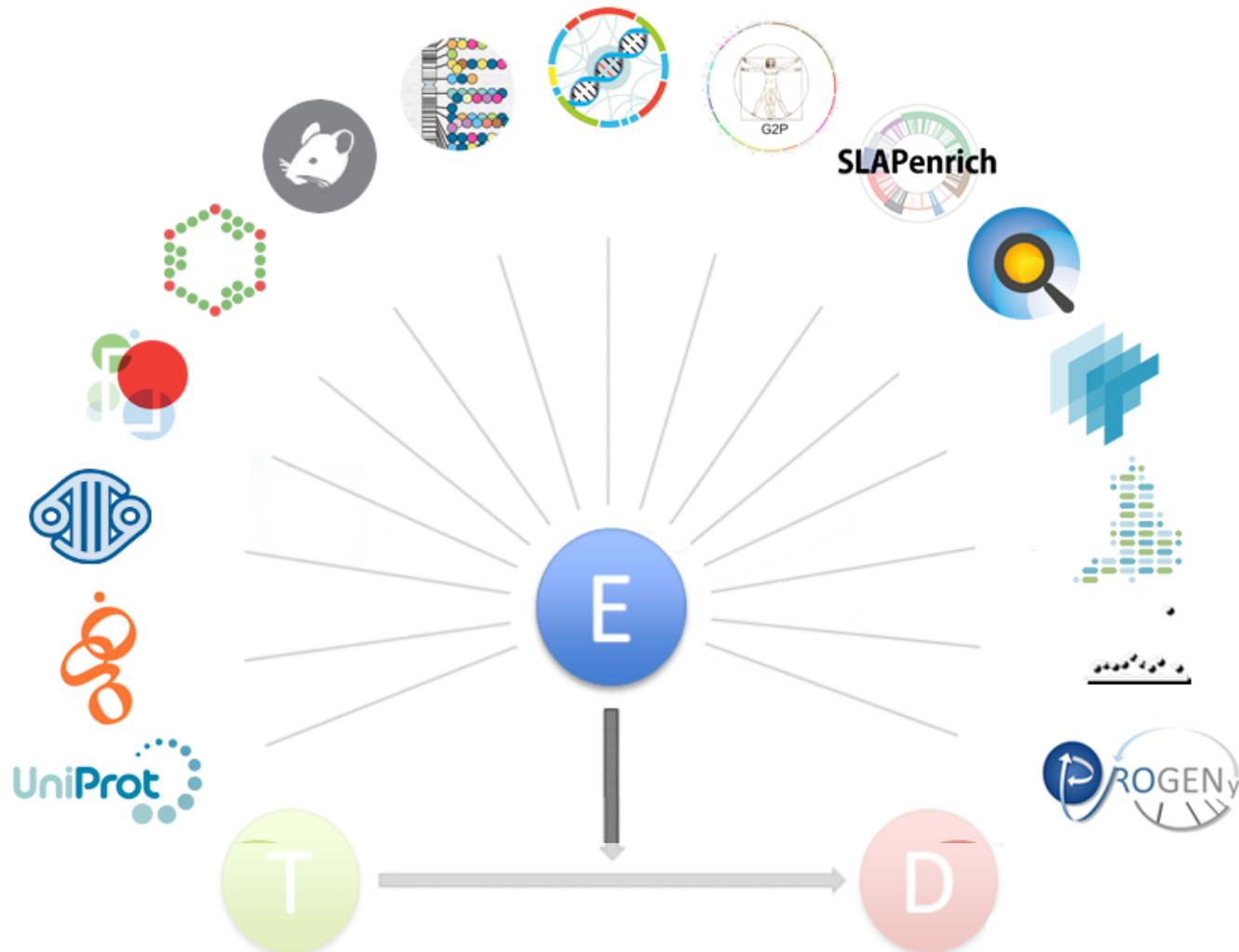
Disease annotations

[http://www.targetvalidation.org/diseaset/EFO\\_0000228](http://www.targetvalidation.org/diseaset/EFO_0000228)

Evidence for T-D associations

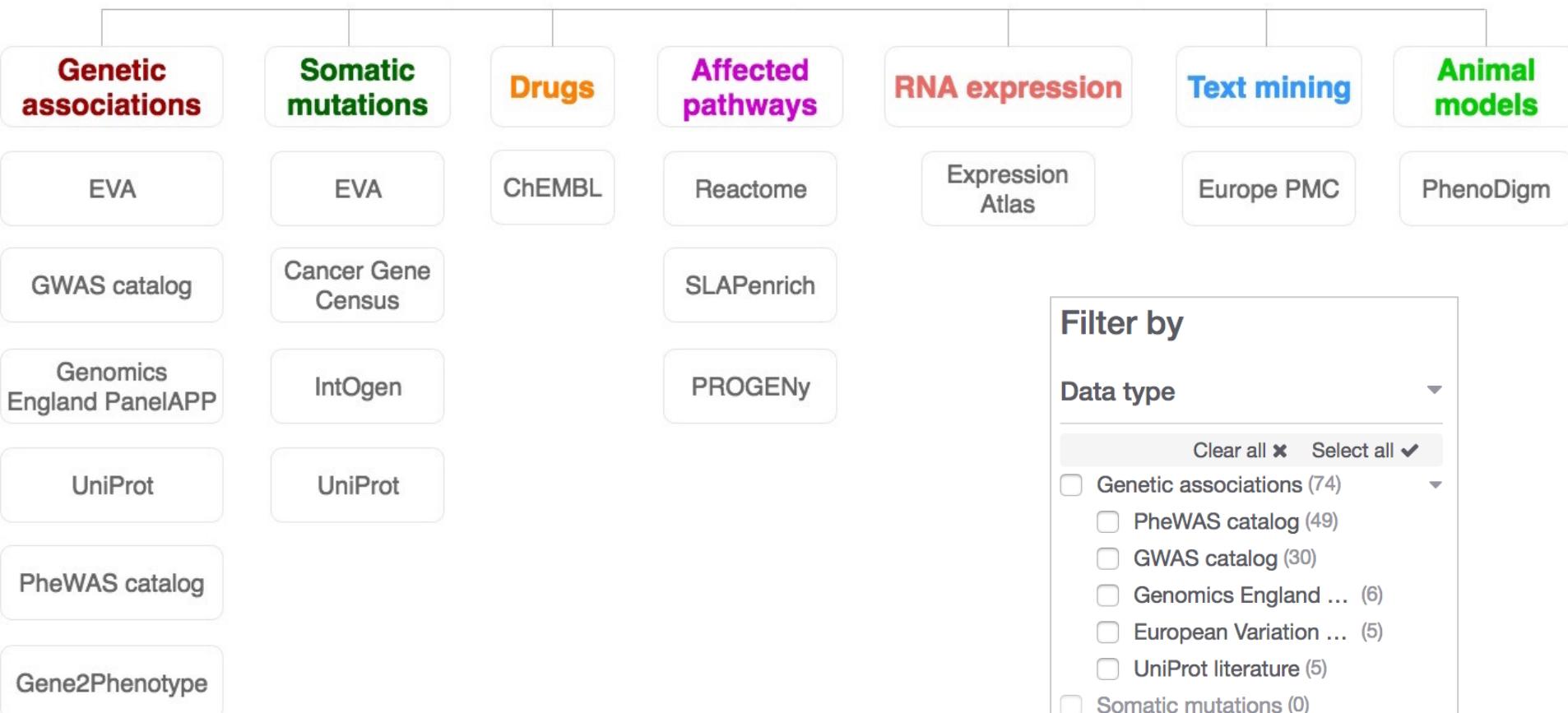
[https://www.targetvalidation.org/evidence/ENSG00000141510/EFO\\_0000228](https://www.targetvalidation.org/evidence/ENSG00000141510/EFO_0000228)

# Evidence for our T-D associations



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

# Data sources grouped into data types



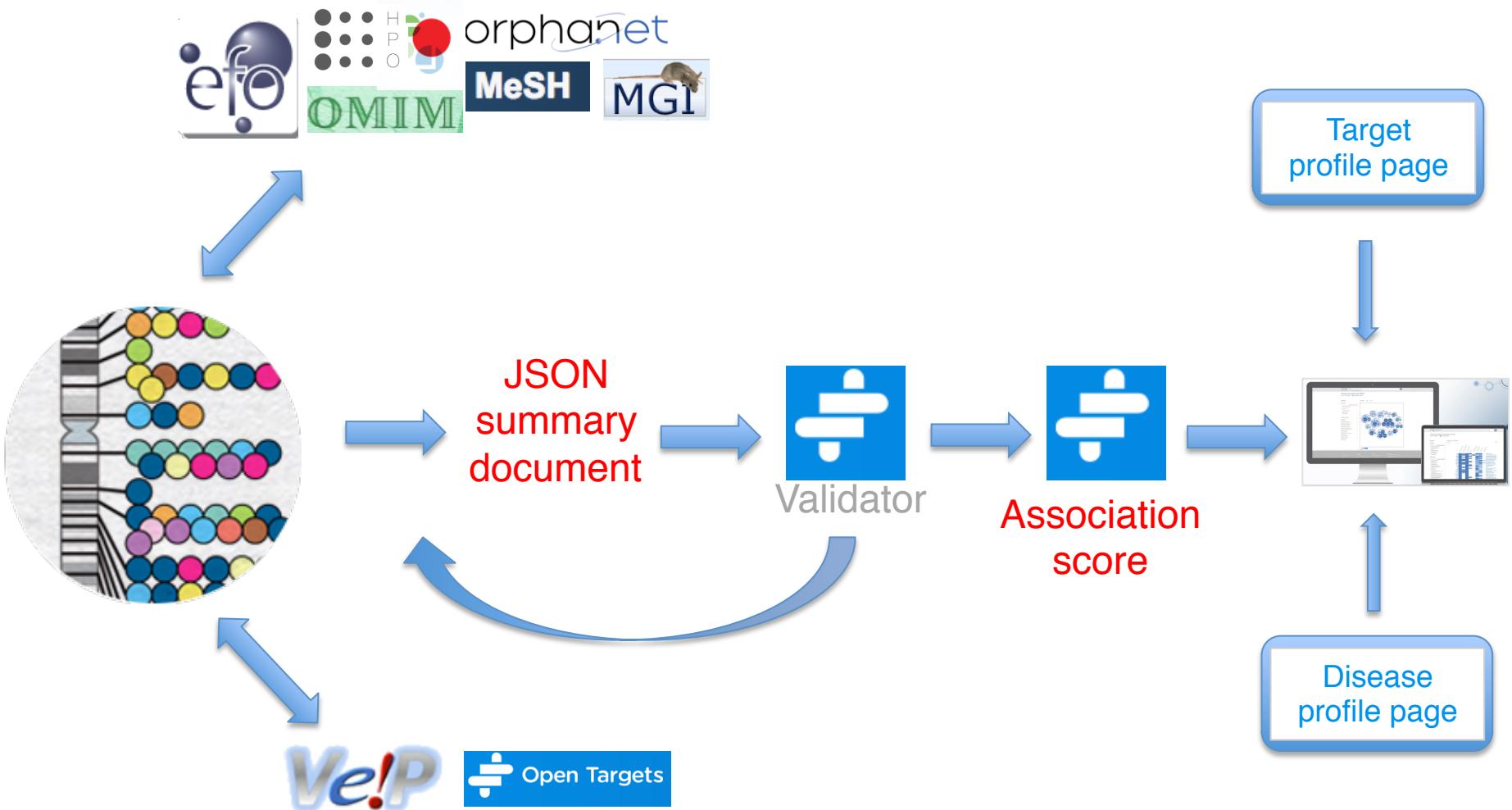
## Filter by

### 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 the data\* flows



\* e.g. germline variants from NHGRI-GWAS Catalog

# JSON summary document

```
hort_name": "Franke A"}, {"last_name": "Alizadeh", "full_name": "Alizadeh Behrooz Z", "short_name": "Alizadeh BZ"}, {"last_name": "Parkes 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_labels": ["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", "study_name": "cttv009_gwas_catalog", "sample_size": "96486", "gwas_panel_resolution": "9000000", "id": "http://identifiers.org/ensembl/ENSG00000005844"}, "evidence": {"variant2disease": {"gwas_sample_size": 96486, "unique_experiments": "http://europepmc.org/abstract/MED/26192919", "gwas_panel_resolution": 9000000, "provenance_type": {"literature": {"references": [{"lit_id": "http://purl.obolibrary.org/obo/ECO_0000205", "label": "Genome-wide association study evidence_pipeline", "label": "CTTV-custom annotation pipeline"}], "label": "CTTV-curated annotation pipeline", "label": "curator inference"}, "upstream_gene_variant": [{"functional_consequence": "http://purl.obolibrary.org/obo/SO_0001631", "provenance": 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", "codes": [{"label": "http://purl.obolibrary.org/obo/ECO_0000205", "label": "http://identifiers.org/eco/cttv_mapping_pipeline"}, "date_asserted": "2017-03-23T03:44:36+00:00", "evidence_codes": ["GWAS", "cttv_mapping_pipeline", "ECO_0000205", "SO_0001631"]}, "validated_against_schema_version": "1.2.5", "res": {"association_score": 0.24183029962242697}, "type": "genetic_association", "id": "f8aa5612c7f01940f3958914fc6074ba"}]  
loads denise$
```

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



# Association score



Which targets have more evidence for an association?



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

# Four-tier scoring framework

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

# Aggregating data → harmonic sum

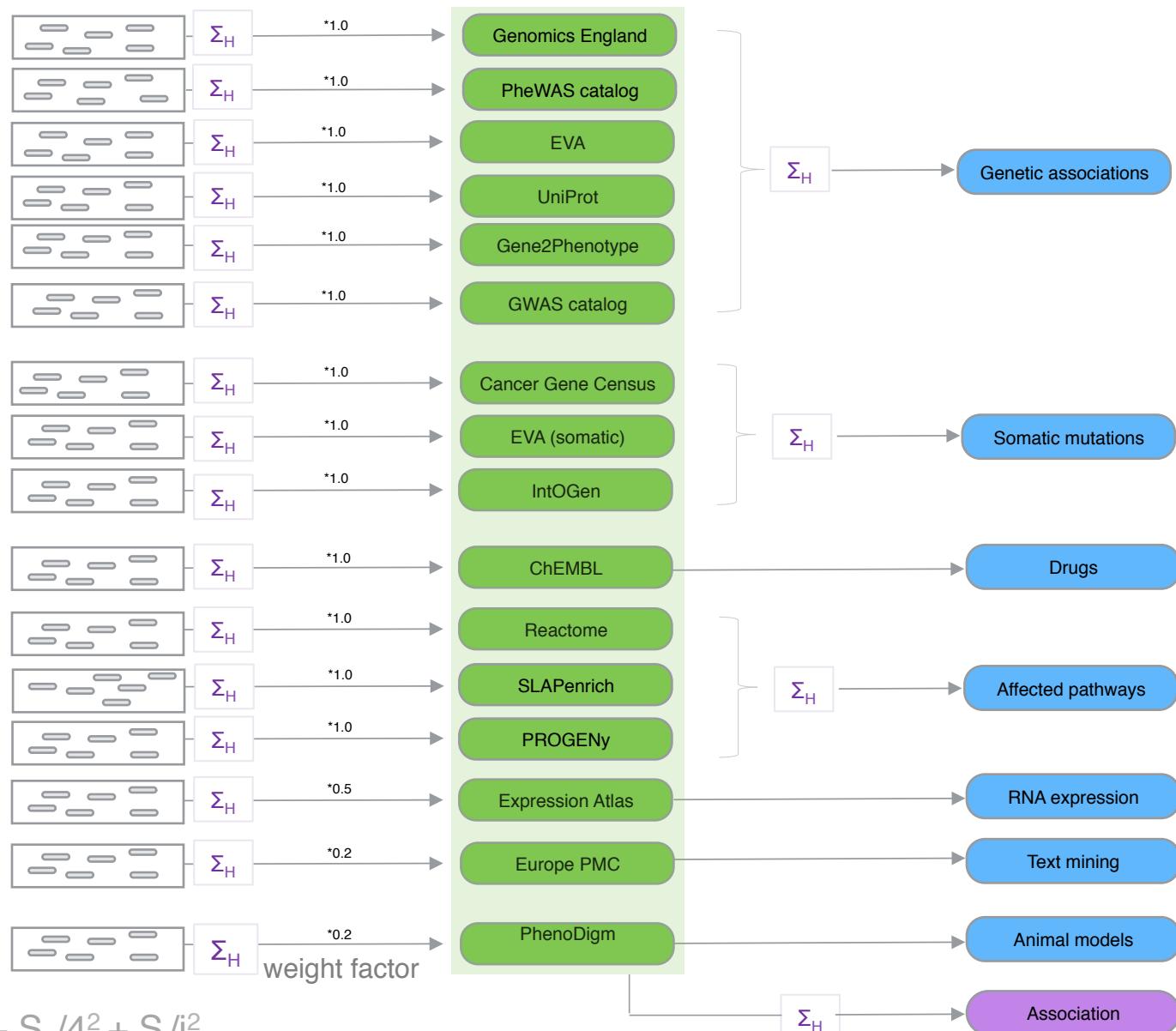
Score: 0 to 1 (max)

Calculated at 4 levels:

- Evidence
- Data source
- Data type
- Overall

Aggregation with  $\Sigma_H$   
(harmonic sum)

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



$$\Sigma_H$$

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

# Factors affecting the relative strength of an evidence

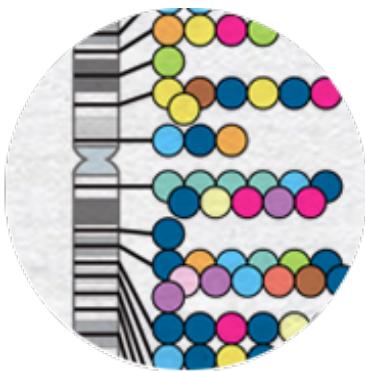
e.g. *GWAS Catalog*

$$S = f * s * c$$

f, relative occurrence of a target-disease evidence

s, strength of the effect described by the evidence

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

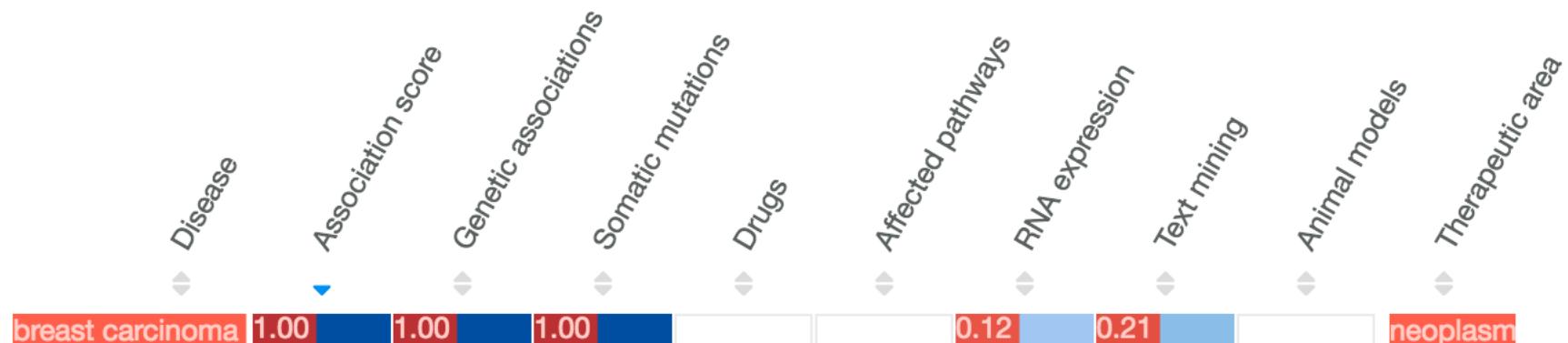


f = sample size (cases versus controls)

s = predicted functional consequence (VEP)

c = *p* value reported in the paper

# Ranking 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

# In these ~3.0 hours

- Introduction
- Open Targets Platform: live demos
- Exercises
- Other modes of data access
- Feedback 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 - 30

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

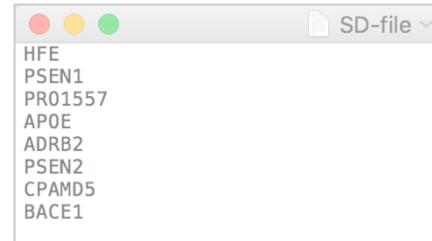


# In these 3.0 hours

- Introduction
- Open Targets Platform: live demos
- **Exercises**
- Other modes of data access
- Feedback and wrap up

# Hands-on exercises

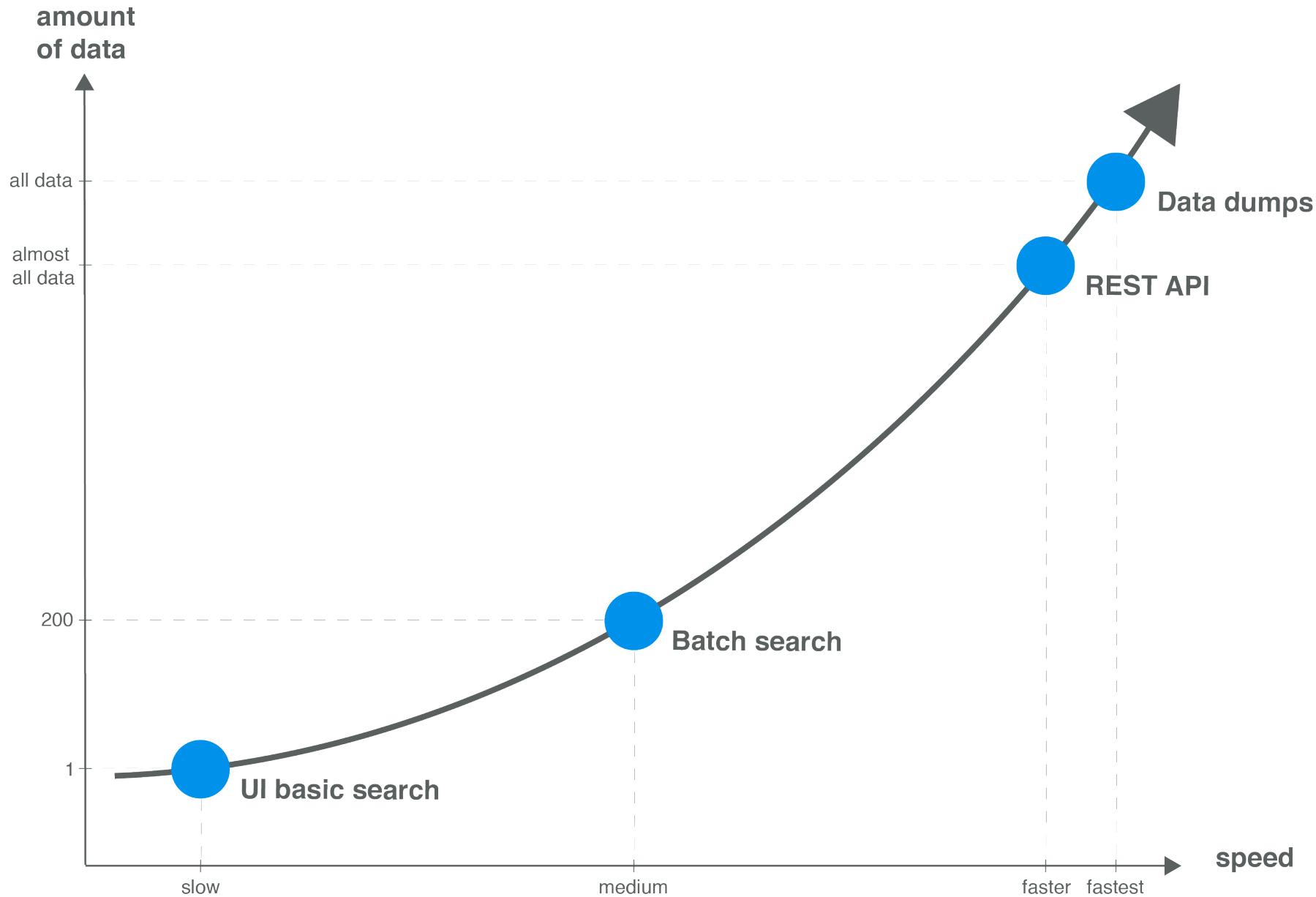
Pages 31-35



# In these ~3.0 hours

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

# Open Targets Platform data: access



# Demo 2: Batch search



We have a list of nine human targets (in cancer) for which there is already a TEP from the SGC.

Are these targets represented in other diseases than cancer?

Which pathways and GO terms are represented in this set of targets?



# 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

POST

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

GET

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

POST

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

Private: methods used by the UI to serve external data. Subject to change without notice

# REST API calls: some examples\*

GET

/public/search

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

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

GET

/public/association/filter

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

GET

/public/evidence/filter

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

# Breaking down the URLs

`https://api.opentargets.io/v3/platform/public/association/filter?`

`target=ENSG00000163914&size=10000&fields=target.id&fields=disease.id`

Server

Endpoint parameters

Parameters

`https://api.opentargets.io/v3/platform/ public/association/filter`

`?target=ENSG00000163914&size=10000&fields=target.id&fields=disease.id`

# REST API: some use cases

How to search

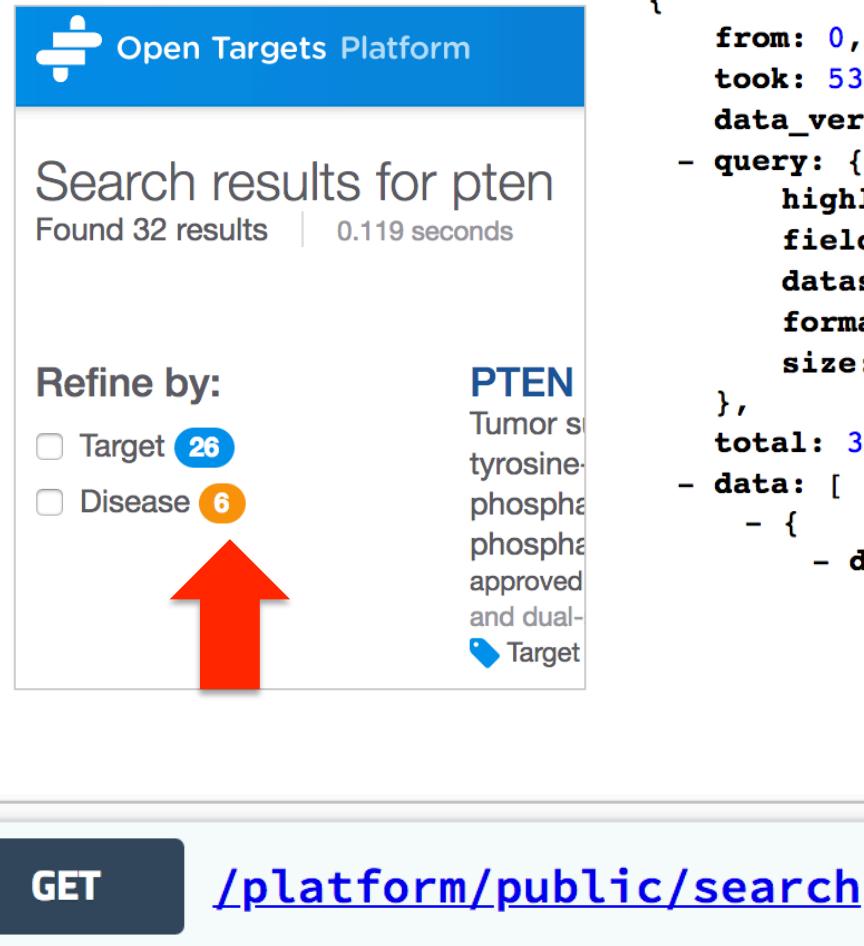


How to get all diseases  
associated with a target

How to get the association score  
for a target – disease pair

How to get the evidence for a  
target – disease association

# How to search



Open Targets Platform

Search results for pten

Found 32 results | 0.119 seconds

Refine by:

- Target 26
- Disease 6

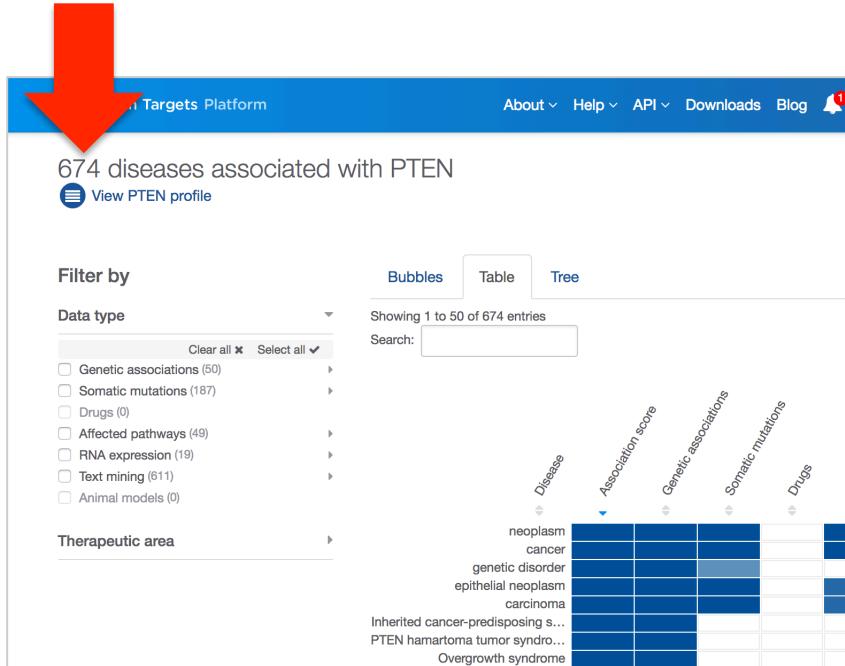
PTEN  
Tumor suppressor, tyrosine-phosphatase, phosphatase and dual-specificity kinase, Target

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 all diseases associated with a target



Open Targets Platform

674 diseases associated with PTEN

[View PTEN profile](#)

Filter by

Data type

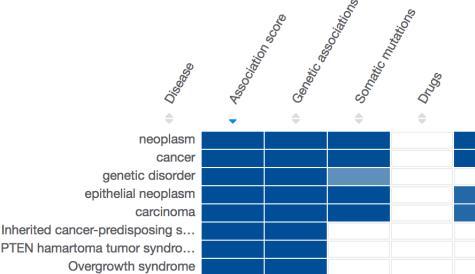
- Genetic associations (50)
- Somatic mutations (187)
- Drugs (0)
- Affected pathways (49)
- RNA expression (19)
- Text mining (611)
- Animal models (0)

Therapeutic area

Bubbles Table Tree

Showing 1 to 50 of 674 entries

Search:



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,  
                    "europemc": 0.3241324475302135,  
                    "pubmed": 0  
                }  
            }  
        }  
    ]  
}
```

# How to get the score for an association

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

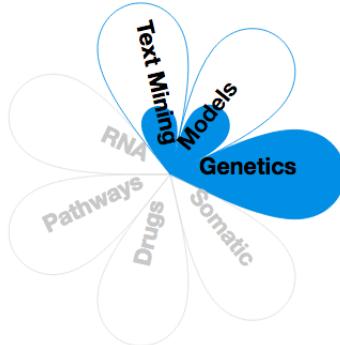
/platform/public/association

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,  
            }  
            - datasources: {  
                slapenrich: 0.8172153264,  
                expression_atlas: 0,  
                gene2phenotype: 0,  
            }  
        }  
    ]  
}
```

# How to get the evidence for an association

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](http://api.opentargets.io/v3/platform/public/evidence/filter)

**DMD**  
dystrophin  
Synonyms: BMD, DDX142, DDX164, DDX206, DDX230, DDX239, DDX268, DDX269, DDX270, DDX272

Anchors the extracellular matrix to the cytoskeleton via F-actin. Ligand for dystroglycan. Component of the dystrophin-associated glycoprotein complex which accumulates at the neuromuscular junction (...)

**Duchenne muscular dystrophy**  
Synonyms: Severe dystrophinopathy, Duchenne type, DMD

```
],  
  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)

# REST API: webinar

YouTube GB

Search

## getAssociationFilter



```
{  
    "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,  
                    "europeme": 0.3241324475302135,  
                    "ccle": 0  
                }  
            }  
        }  
    ]  
}
```

http://api.opentargets.io/v3/platform/public/association/filter?

target=ENSG00000171862

&select=28:56 / 40:43

<https://youtu.be/KQbfhwpeEvc>

# How to run our REST endpoints

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

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

# Paste the URL in the a location bar

The screenshot shows a web browser window with the following details:

- Address Bar:** https://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000167207
- Content:** A JSON object representing a search result for the target gene ENSG00000167207. The JSON is very long and contains numerous fields related to gene expression, association scores, and various databases.

```
{"from": 0, "took": 27, "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": 0}], "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}, "total": 503, "data": [{"target": {"gene_info": {"symbol": "NOD2", "name": "nucleotide binding oligomerization domain containing 2"}}, "id": "ENSG00000167207"}, {"association_score": {"datatypes": [{"literature": 0.14497512075102087, "rna_expression": 0.0006625794025179989, "genetic_association": 1.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.29744967878526557, "affected_pathway": 0.0}, {"overall": 1.0, "datasources": [{"slapenrich": 0.0, "expression_atlas": 0.0006625794025179989, "europepmc": 0.14497512075102087, "twentythreeandme": 0.0, "uniprot_literature": 1, "phenodigm": 0.29744967878526557, "eva": 0.9245864351309815, "gene2phenotype": 0.0, "gwas_catalog": 1, "intogen": 0.0, "genomics_england": 0.0, "uniprot": 1, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.001632417506314012}], "disease": {"efo_info": {"therapeutic_area": {"labels": [], "codes": []}, "path": [{"EFO_0000540"}], "label": "immune system disease"}, "id": "EFO_0000540"}, {"is_direct": false, "evidence_count": {"datatypes": {"literature": 1463.0, "rna_expression": 32.0, "genetic_association": 133.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 192.0, "affected_pathway": 0.0}, "total": 1820.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 32.0, "europepmc": 1463.0, "twentythreeandme": 0.0, "uniprot_literature": 3.0, "phenodigm": 192.0, "eva": 11.0, "gene2phenotype": 0.0, "gwas_catalog": 17.0, "intogen": 0.0, "genomics_england": 0.0, "uniprot": 93.0, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 9.0}, "id": "ENSG00000167207-EFO_0000540"}, {"target": {"gene_info": {"symbol": "NOD2", "name": "nucleotide binding oligomerization domain containing 2"}}, "id": "ENSG00000167207"}, {"association_score": {"datatypes": [{"literature": 0.09408949039475947, "rna_expression": 0.0, "genetic_association": 1.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.29621591672500813, "affected_pathway": 0.0}, {"overall": 1.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 0.0, "europepmc": 0.09408949039475947, "twentythreeandme": 0.0, "uniprot_literature": 0.98, "phenodigm": 0.29621591672500813, "eva": 0.900426255952381, "gene2phenotype": 0.0, "gwas_catalog": 0.0, "intogen": 0.0, "genomics_england": 1.0, "uniprot": 1, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.0013905322}], "disease": {"efo_info": {"therapeutic_area": {"labels": [], "codes": []}, "path": [{"EFO_0000508"}], "label": "genetic disorder"}, "id": "EFO_0000508"}, {"is_direct": true, "evidence_count": {"datatypes": []}}]}]}]
```

```
{  
    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

# Command line e.g. CURL -X GET

```
denise-m1:~ denise$ curl -X GET https://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000167207&direct=true&size=310
[[1] 45611
[2] 45612
[2]+ Done direct=true
denise-m1:~ denise$ {"from": 0, "took": 13, "next": [1.3668802, "ENSG00000167207-0", "ENSG00000167207-101988"], "data_version": "17.09", "query": {"sort": ["harmonic-sum_overall"]}, "search": null, "rna_expression_level": 0, "protein_expression_tissue": [], "score_value_types": ["overall"], "datatype": [], "fields": null, "format": "json", "facet_size": null, "disease": [], "protein_expression_level": 0, "datastructure": "dense", "facets": "false", "rna_expression_tissue": [], "target": ["ENSG00000167207"], "target_class": [], "cap_scores": true, "pathway": [], "size": 10}, "total": 503, "data": [{"target": {"gene_info": {"symbol": "NOD2", "name": "nucleotide binding oligomerization domain containing 2"}}, "id": "ENSG00000167207"}, {"association_score": 0.14497512075102087, "datatypes": {"literature": 0.14497512075102087, "rna_expression": 0.0006625794025179989, "genetic_association": 1.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.29744967878526557, "affected_pathway": 0.0}, "overall": 1.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 0.0006625794025179989, "europewPMC": 0.14497512075102087, "twentythreeandme": 0.0, "uniprot_literature": 1, "phenodigm": 0.29744967878526557, "eva": 0.9245864351309815, "gene2phenotype": 0.0, "gwas_catalog": 1, "intogen": 0.0, "genomics_england": 0.0, "uniprot": 1, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.01632417506314012}, "disease": {"efo_info": {"therapeutic_area": {"labels": [], "codes": []}}, "path": [{"EFO_0000540"}], "label": "immune system disease"}, "id": "EFO_0000540"}, {"is_direct": false, "evidence_count": {"datatypes": {"literature": 1463.0, "rna_expression": 32.0, "genetic_association": 133.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 192.0, "affected_pathway": 0.0}, "total": 1820.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 32.0, "europewPMC": 1463.0, "twentythreeandme": 0.0, "uniprot_literature": 3.0, "phenodigm": 19.2, "eva": 11.0, "gene2phenotype": 0.0, "gwas_catalog": 17.0, "intogen": 0.0, "genomics_england": 0.0, "uniprot": 93.0, "chembl": 0.0, "cancer_gene_census": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 9.0}, "id": "ENSG00000167207-EFO_0000540"}, {"target": {"gene_info": {"symbol": "NOD2", "name": "nucleotide binding oligomerization domain containing 2"}}, "id": "ENSG00000167207"}, {"association_score": {"datatypes": {"literature": 0.09408949039475947, "rna_expression": 0.0, "genetic_association": 1.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.29621591672500813, "affected_pathway": 0.0}, "overall": 1.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 0.0, "europewPMC": 0.09408949039475947, "twentythreeandme": 0.0, "uniprot_literature": 0.98, "phenodigm": 0.29621591672500813, "eva": 0.900426255952381, "gene2phenotype": 0.0, "gwas_catalog": 0.0, "intogen": 0.0, "genomics_england": 1.0, "uniprot": 1, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.0013905322}, "disease": {"efo_info": {"therapeutic_area": {"labels": [], "codes": []}}, "path": [{"EFO_0000508"}], "label": "genetic disorder"}, "id": "EFO_0000508"}, {"is_direct": true, "evidence_count": {"datatypes": {"literature": 138.0, "rna_expression": 0.0, "genetic_association": 25.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 186.0, "affected_pathway": 0.0}, "total": 349.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 0.0, "europewPMC": 138.0, "twentythreeandme": 0.0, "uniprot_literature": 1.0, "phenodigm": 186.0, "eva": 7.0, "gene2phenotype": 0.0, "gwas_catalog": 0.0, "intogen": 0.0, "genomics_england": 1.0, "uniprot": 15.0, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 1.0}, "id": "ENSG00000167207-EFO_0000508"}, {"target": {"gene_info": {"symbol": "NOD2", "name": "nucleotide binding oligomerization domain containing 2"}}, "id": "ENSG00000167207"}, {"association_score": {"datatypes": {"literature": 0.1448254072034677, "rna_expression": 0.0, "genetic_association": 1.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.25615753501943783, "affected_pathway": 0.0}, "overall": 1.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 0.0, "europewPMC": 0.1448254072034677, "twentythreeandme": 0.0, "uniprot_literature": 1, "phenodigm": 0.25615753501943783, "eva": 0.697569444444445, "gene2phenotype": 0.0, "gwas_catalog": 1, "intogen": 0.0, "genomics_england": 0.0, "uniprot": 1, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.0016307895945663268}, "disease": {"efo_info": {"therapeutic_area": {"labels": {"immune system disease": {"codes": [{"EFO_0000540"}]}}, "path": [{"EFO_0000540", "EFO_0005140"}], "label": "autoimmune disease"}, "id": "EFO_0005140"}, {"is_direct": true, "evidence_count": {"datatypes": {"literature": 1286.0, "rna_expression": 4.0, "genetic_association": 108.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 11.0, "affected_pathway": 0.0}, "total": 1409.0, "datasources": {"slapenrich": 0.0, "expression_atlas": 4.0, "europewPMC": 1286.0, "twentythreeandme": 4.0, "uniprot_literature": 2.0, "phenodigm": 11.0, "eva": 4.0, "gene2phenotype": 0.0, "gwas_catalog": 17.0, "intogen": 0.0, "genomics_england": 0.0, "uniprot": 78.0, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 7.0}, "id": "ENSG00000167207-EFO_0005140"}, {"target": {"gene_info": {"symbol": "NOD2", "name": "nucleotide binding oligomerization domain containing 2"}}, "id": "ENSG00000167207"}, {"association_score": {"datatypes": {"literature": 0.1456119020355387, "rna_expression": 7.3235e-06, "genetic_association": 1.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 0.2425402006966178, "affected_pathway": 0.625}, "overall": 1.0, "datasources": {"slapenrich": 0.625, "expression_atlas": 7.3235e-06, "europewPMC": 0.1456119020355387, "twentythreeandme": 0.0, "uniprot_literature": 1, "phenodigm": 0.2425402006966178, "eva": 0.0, "gene2phenotype": 0.0, "gwas_catalog": 1, "intogen": 0.0, "genomics_england": 0.0, "uniprot": 1, "chembl": 0.0, "cancer_gene_census": 0.0, "reactome": 0.0, "uniprot_somatic": 0.0, "eva_somatic": 0.0, "phewas_catalog": 0.001519999744292431}, "disease": {"efo_info": {"therapeutic_area": {"labels": {"digestive system disease": {"codes": [{"EFO_0000405"}]}}, "path": [{"EFO_0000405"}], "label": "digestive system disease"}, "id": "EFO_0000405"}, {"is_direct": true, "evidence_count": {"datatypes": {"literature": 1367.0, "rna_expression": 10.0, "genetic_association": 110.0, "somatic_mutation": 0.0, "known_drug": 0.0, "animal_model": 12.0, "affected_pathway": 2.0}, "total": 1501.0, "datasources": {"slapenrich": 2.0, "expression_atlas": 10.0, "europewPMC": 1367.0, "twentythreeandme": 0.0, "uniprot_literature": 2.0, "phenodigm": 12.0}}}}
```



Can we change the way the associations are scored? Perhaps increase the weight given to text mining data?



Yes, you can with the Open Targets Python client!

A screenshot of the opentargets documentation homepage. The header is blue with the text "opentargets" and "latest". Below the header is a search bar labeled "Search docs". On the left side of the main content area, there is a sidebar with links to "Tutorial", "High Level API", "Low Level API", "Code Documentation", and "Changelog".

[Docs](#) » opentargets - Python client for targetvalidation.org

## opentargets - Python client for targetvalidation.org

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

# 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)

# In these 3.0 hours

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

Your feedback is important

<https://tinyurl.com/sgc-260918>

[genetics.opentargets.org](https://genetics.opentargets.org)

# Open Targets Genetics coming soon!

Our newest product, Open Targets Genetics, is almost ready. We will be launching it on **Thursday 18 October 2018** during an interactive workshop at the ASHG conference in San Diego. To be the first to know when it goes live, sign up for updates using the form below.

Enter email...

Subscribe

# Open Targets Platform

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

21K  
targets

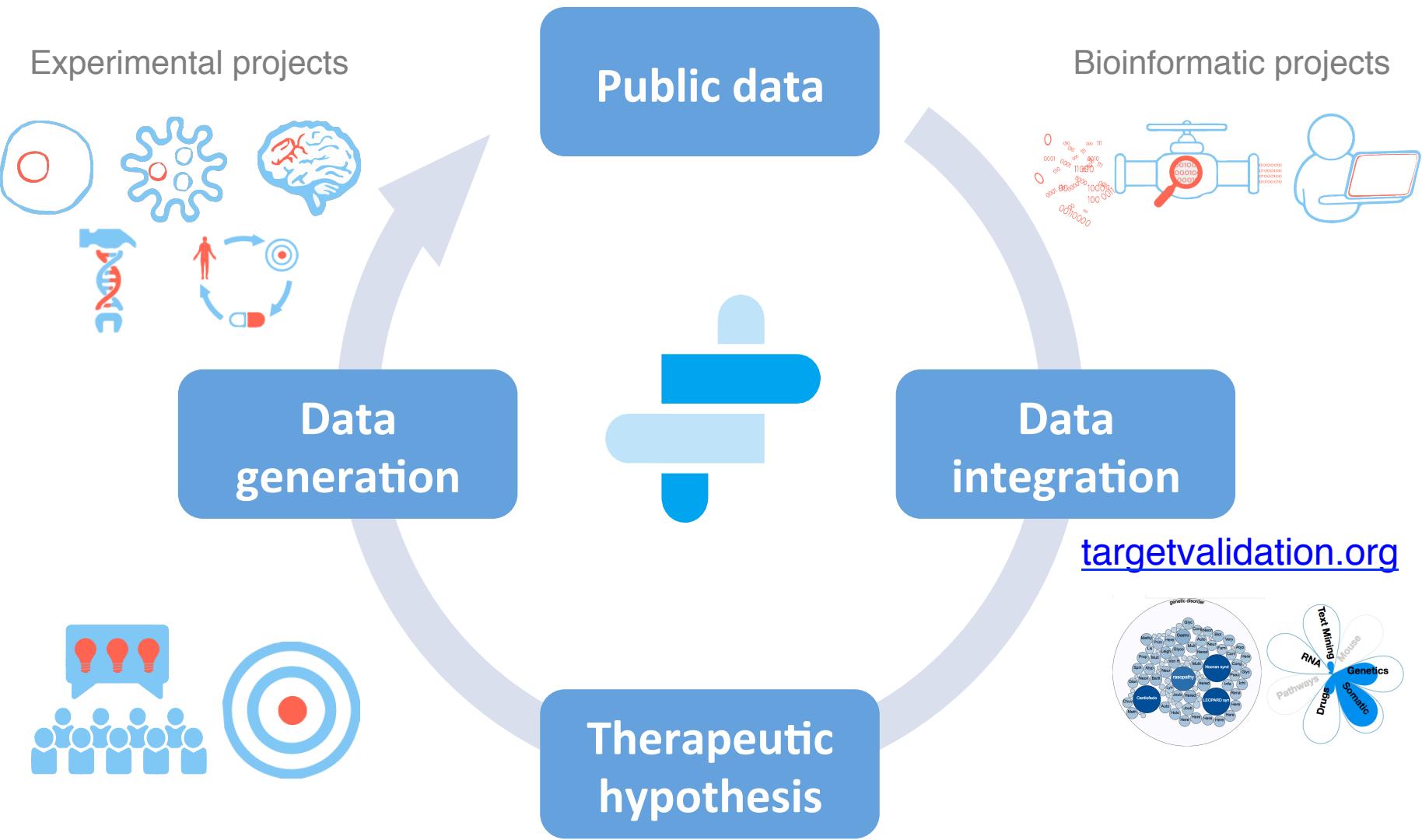
10K  
diseases

2.9 M  
associations

6.5 M  
evidence

August 2018 release

# Virtuous cycle in Open Targets



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

# We support decision-making

Which targets are associated with a disease?

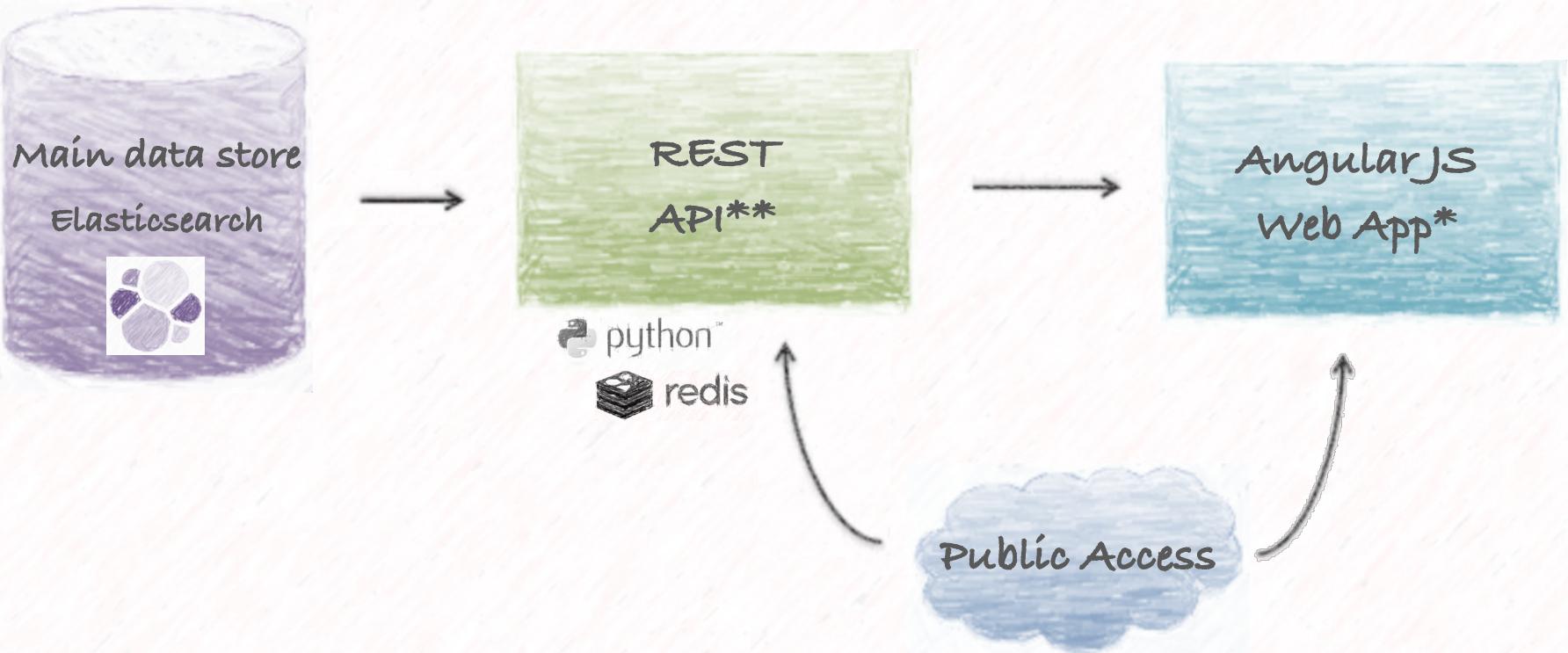
Can I find out about the mechanisms of the disease?

Are there FDA drugs for this association?



...

# How to access the Platform



# Our breakthrough paper

Published online 8 December 2016

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

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

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

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

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

# Get in touch



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



[support@targetvalidation.org](mailto:support@targetvalidation.org)



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



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



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



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

# Acknowledgements



Open Targets

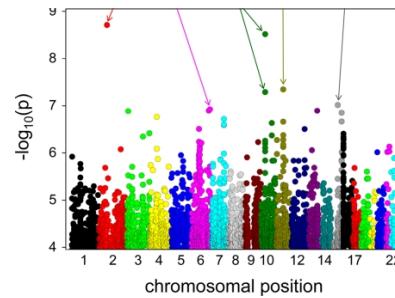
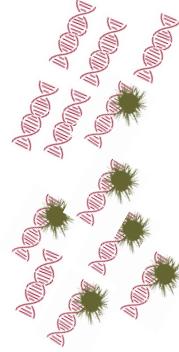
# Extra extra extra

Details on data sources to associate  
targets and diseases

Other Open Targets resources

# Data sources: GWAS catalog

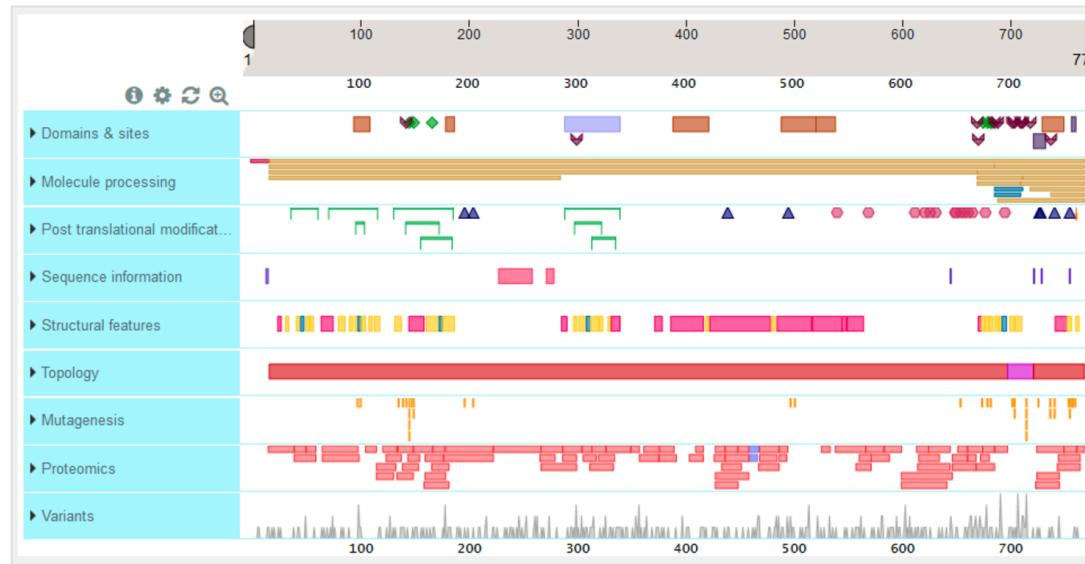
- Genome Wide Association Studies
- Array-based chips → genotyping 100,000 SNPs genomewide



Open Targets

# Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients

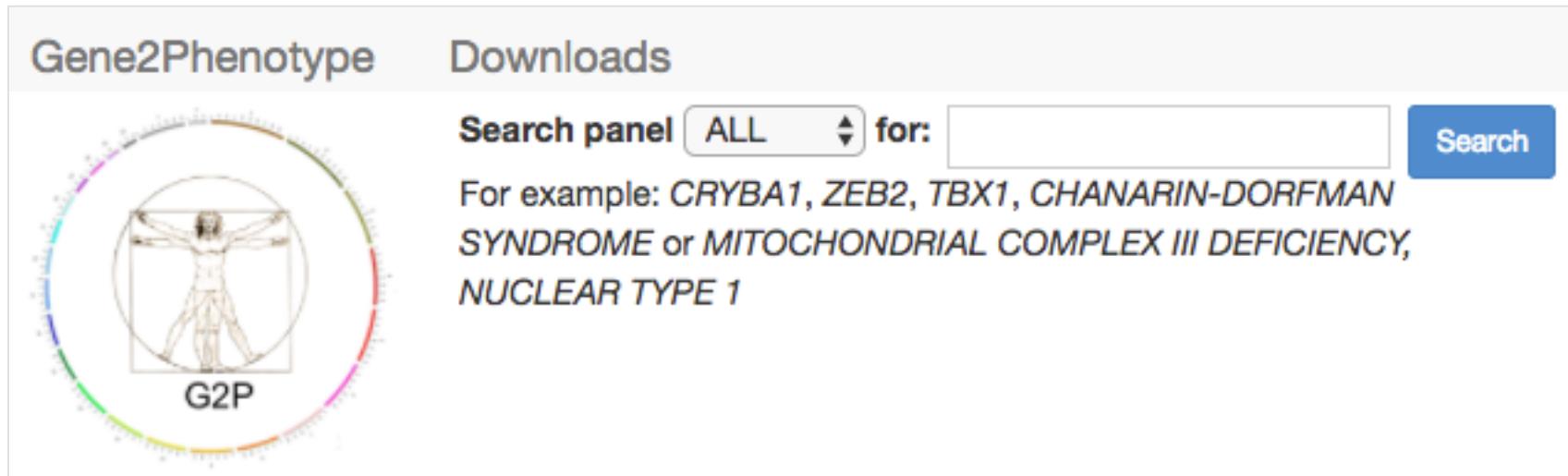


EMBL-EBI train online



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



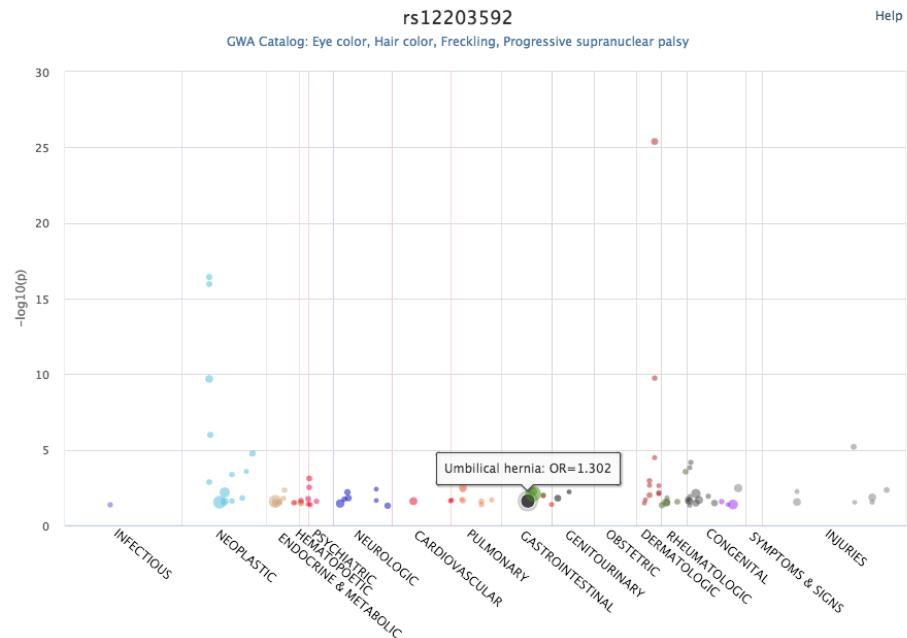
EMBL-EBI train online



Open Targets

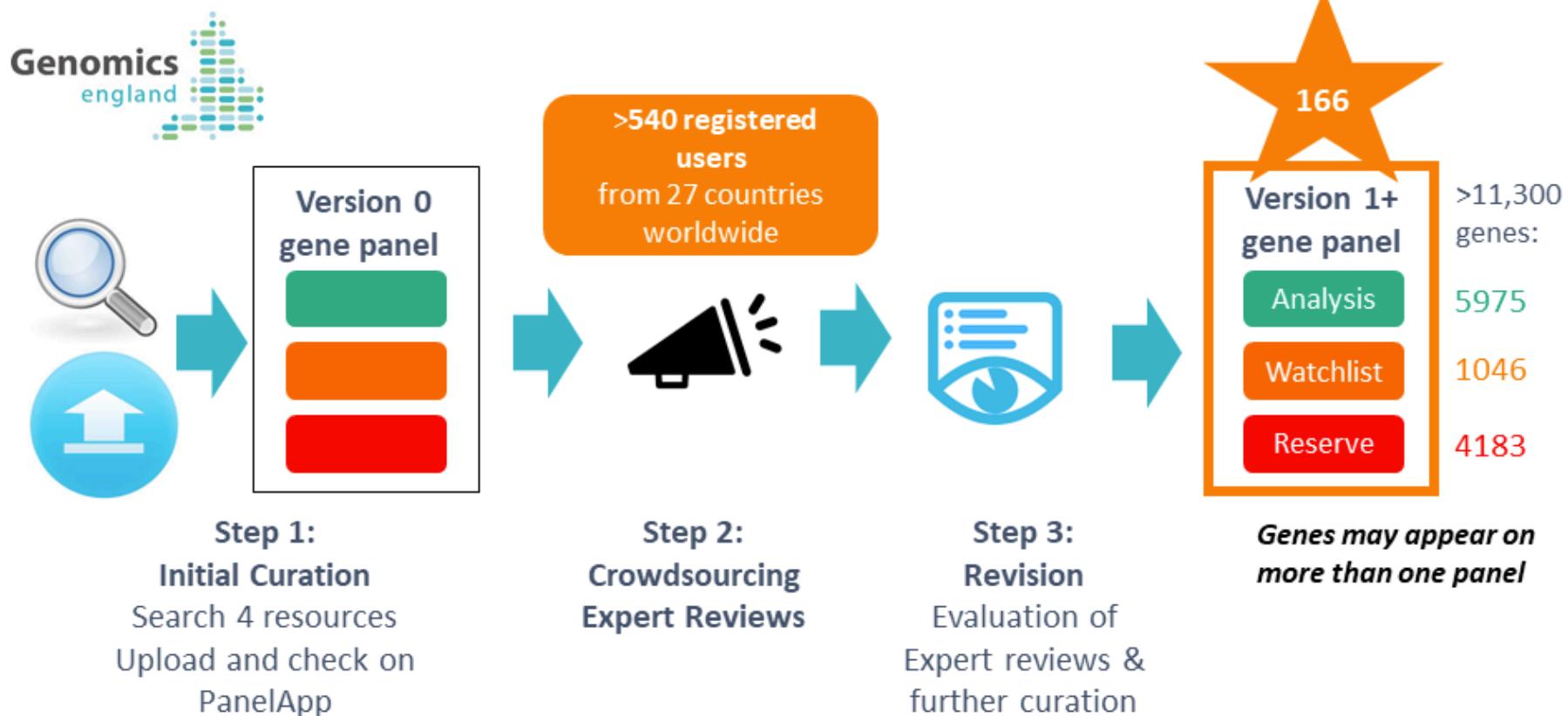
# Data sources: PheWAS

- 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: SLAPenrich

## METHOD

### Dissecting the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich

Francesco Iorio<sup>1,5†</sup>, Luz Garcia-Alonso<sup>1,5</sup>, Jonathan Brammell<sup>2</sup>, Iñigo Martincorena<sup>2</sup>, David R Wille<sup>3,5</sup>, Ultan McDermott<sup>2,5</sup> and Julio Saez-Rodriguez<sup>1,4,5\*†</sup>

- 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

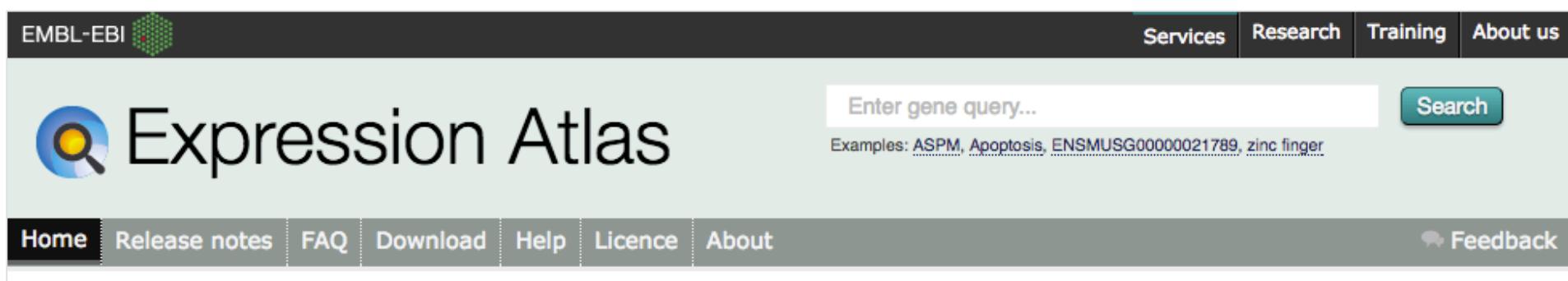
OPEN

## Perturbation-response genes reveal signaling footprints in cancer gene expression

Michael Schubert<sup>1</sup>, Bertram Klinger<sup>2,3</sup>, Martina Klünemann<sup>2,3</sup>, Anja Sieber<sup>2,3</sup>, Florian Uhlitz<sup>2,3</sup>, Sascha Sauer<sup>4</sup>, Mathew J. Garnett<sup>5</sup>, Nils Blüthgen<sup>2,3</sup> & Julio Saez-Rodriguez<sup>2,3</sup>

- 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 $\alpha$ , 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



EMBL-EBI train online



# 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



Open Targets

# Data sources: PhenoDigm

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

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

Diseases Tissue phenotype associations Secondary phenotypes

- Semantic approach to associate mouse models with diseases

# Aggregating scores across the data

- Using a mathematical function, the harmonic sum\*

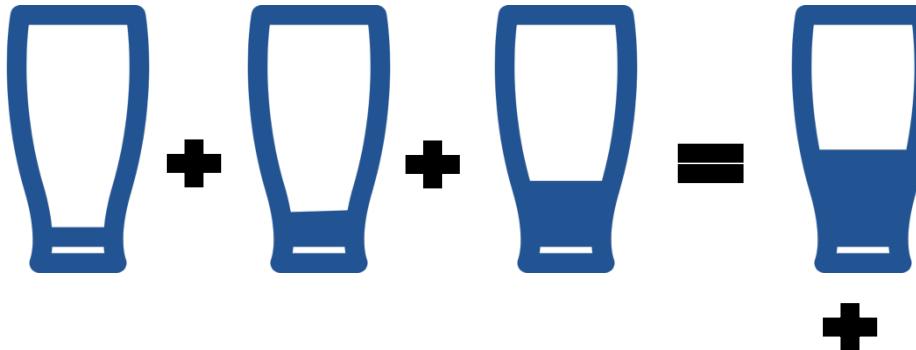
$$S_{1..i} = S_1 + \frac{S_2}{2^2} + \frac{S_3}{3^2} + \frac{S_4}{4^2} \dots + \frac{S_i}{i^2}$$

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

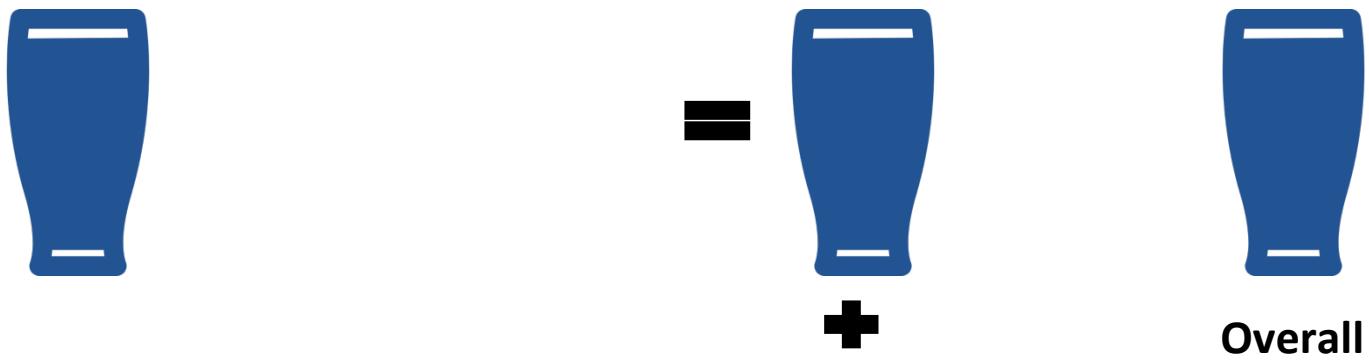
- Advantages:
  - A) account for replication
  - B) deflate the effect of large amounts of data e.g. text mining

# Target-Disease Association Score

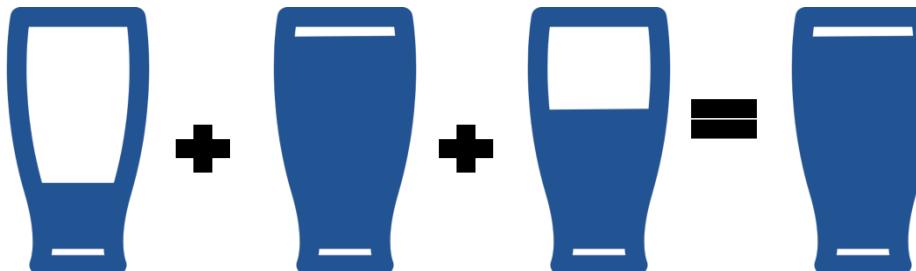
EuropePMC  
(Text Mining)



UniProt  
(Manual Curation)



ChEMBL  
(Manual Curation)

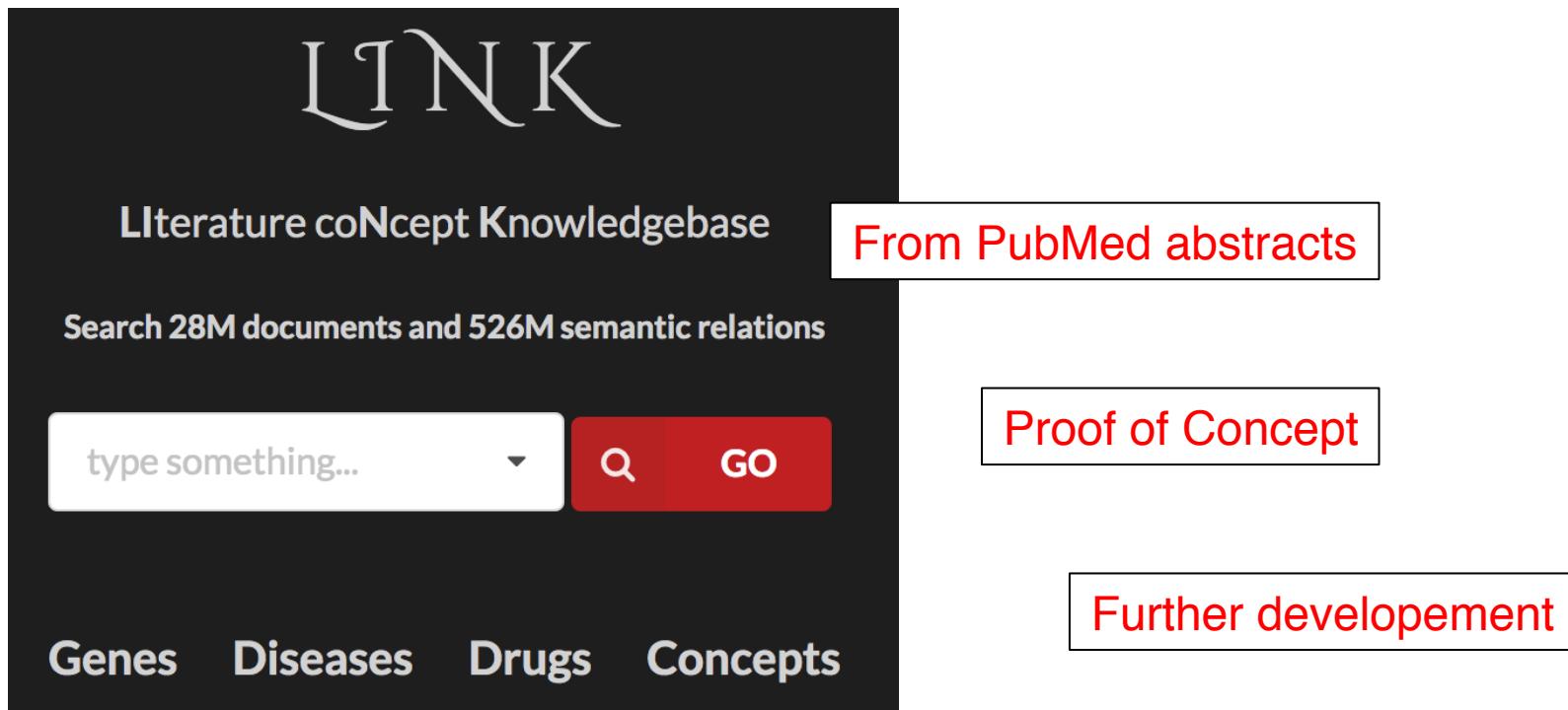


VERY simplified diagram

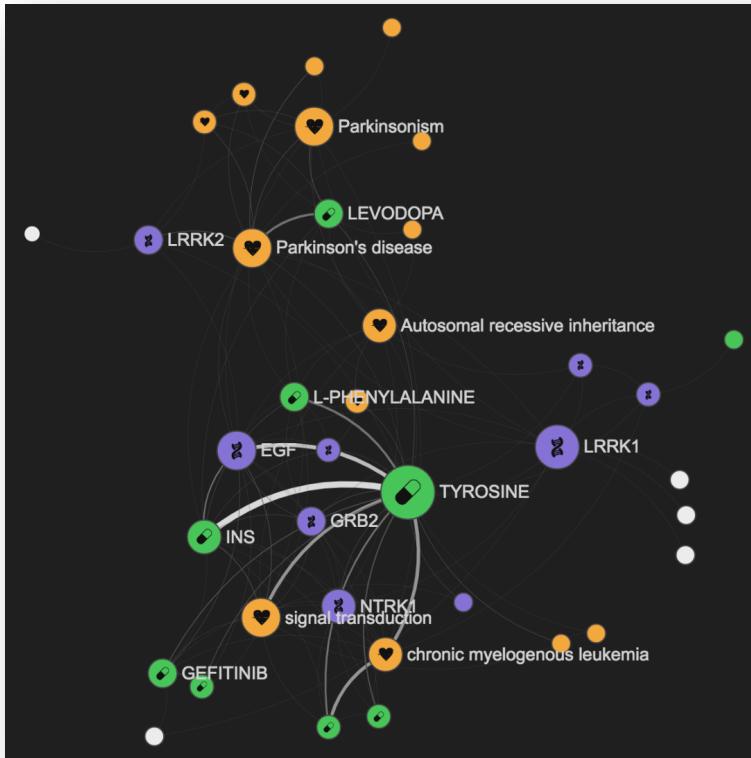
# Open Targets LINK

<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

# Open Targets toolkit: 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 x

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

