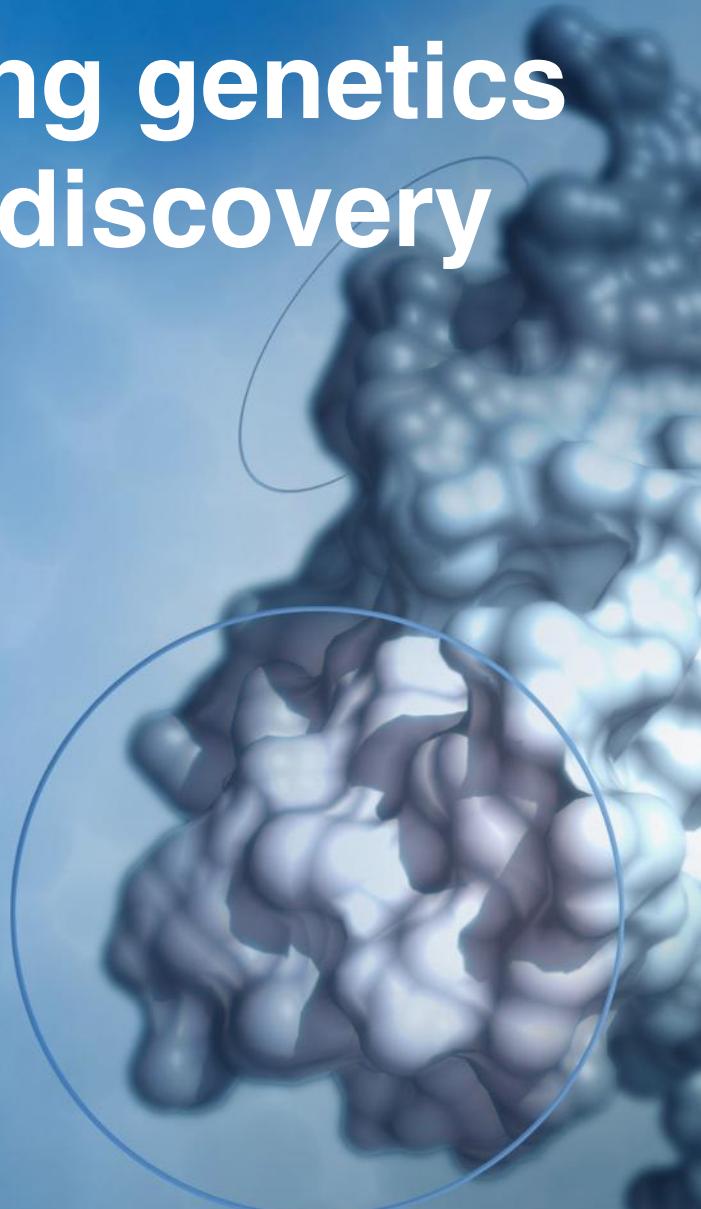


Open Targets: integrating genetics and genomics for drug discovery

BioData West Pre-conference Workshop
March 12th 2018, NASA Ames

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

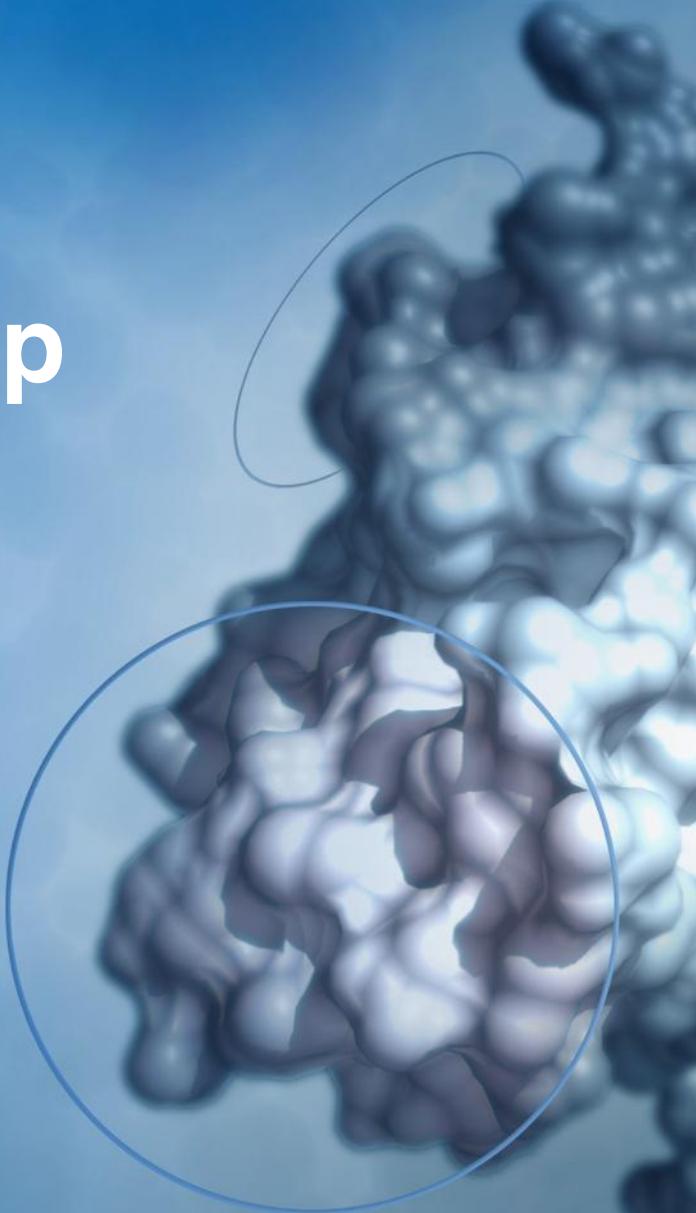


User Interface Workshop

... not datathon



Open Targets



Aims



Open Targets Partnership

How to navigate the
Open Targets Platform

How to query the Open
Targets REST server

How to get help

In these ~ 3.5 hours

- Introduction
- Live demos and exercises
 - User Interface (demos 1 and 2)
- REST API and simple calls
- Wrap up



In these ~ 3.5 hours

- Introduction
- Live demos and exercises
 - User Interface (demos 1 and 2)
- REST API and some examples
- Wrap up

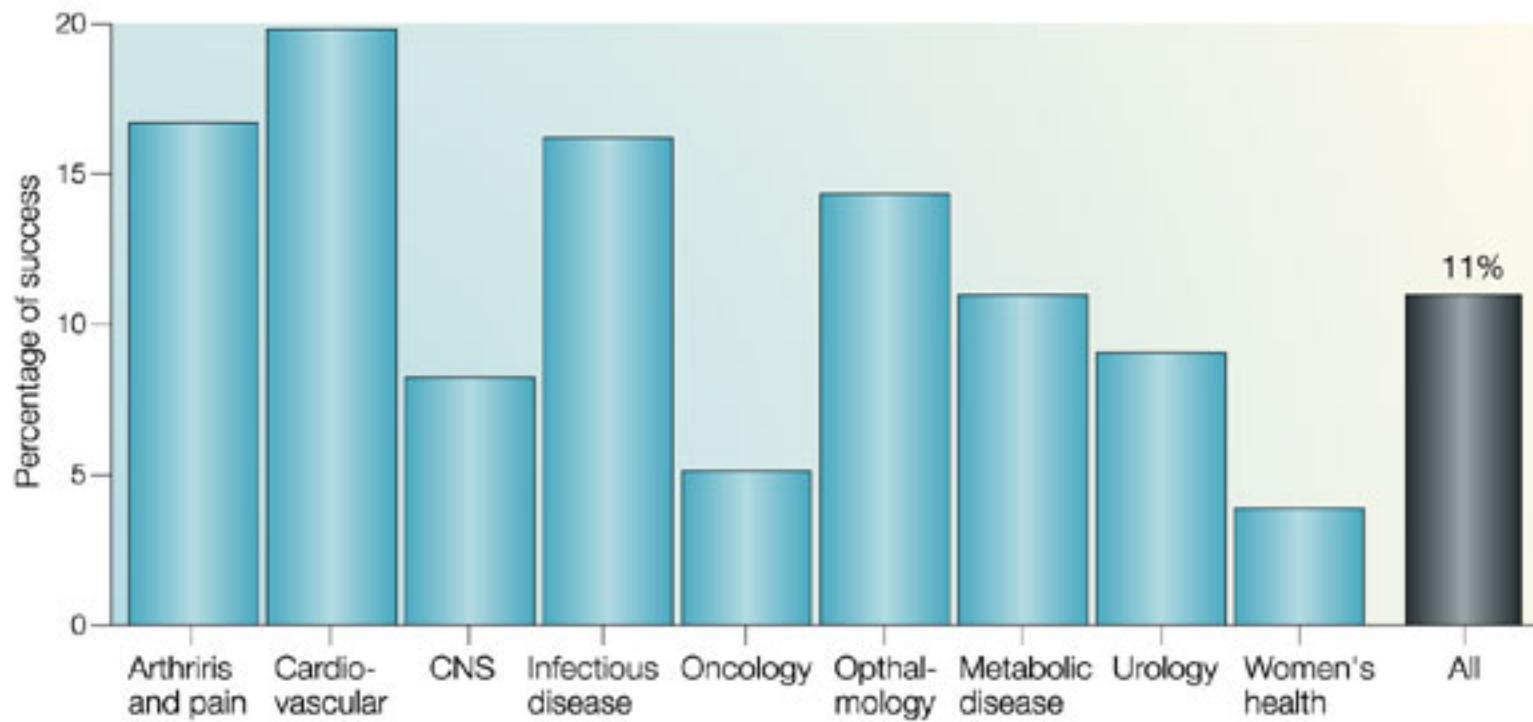


In these ~ 3.5 hours

- Introduction
- Live demos and exercises
 - User Interface (demos 1 and 2)
- REST API and some examples
- Wrap up



Drug discovery: some challenges



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

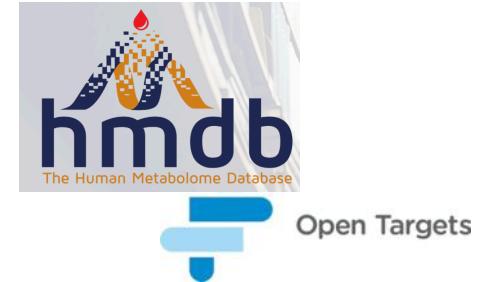
Source: doi:10.1038/nrd1470

Public databases for drug discovery

- EMBL-EBI (European Bioinformatics Institute)



- Elsewhere



Fit everything together



- Time consuming
- Possible lack of resources or expertise
- ...



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



Open Targets is all you need!

I know. If you're like me, you want to find as much data as possible, plus analyses and links to **the original source** for my own assessment.

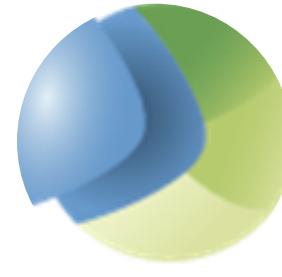
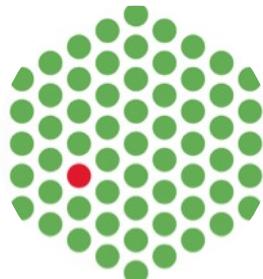
A resource that is **comprehensive, trustworthy, up-to-date, sustainable, easy-to-use** and free.



Open Targets

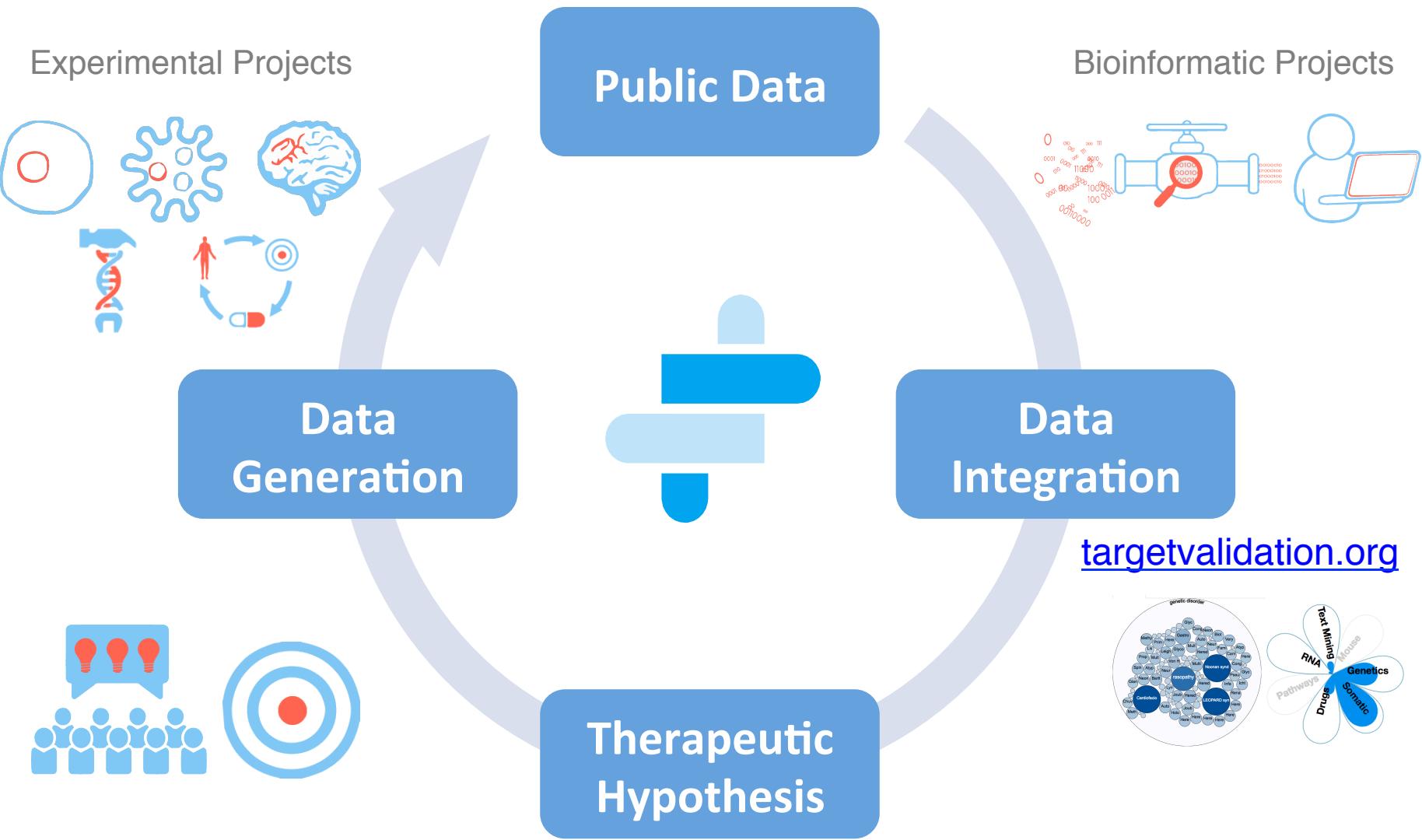
Our Vision

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



Open Targets

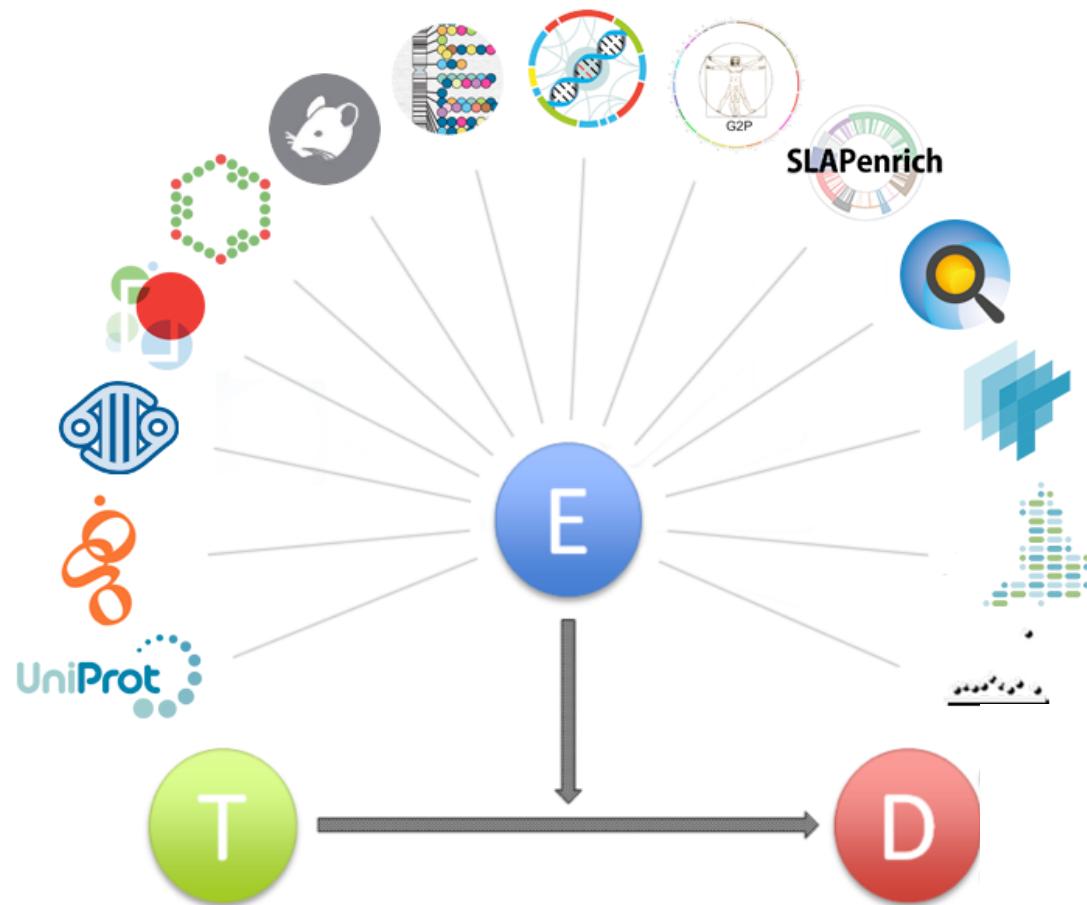
Virtuous cycle in Open Targets



Concurrent
www.opentargets.org/projects

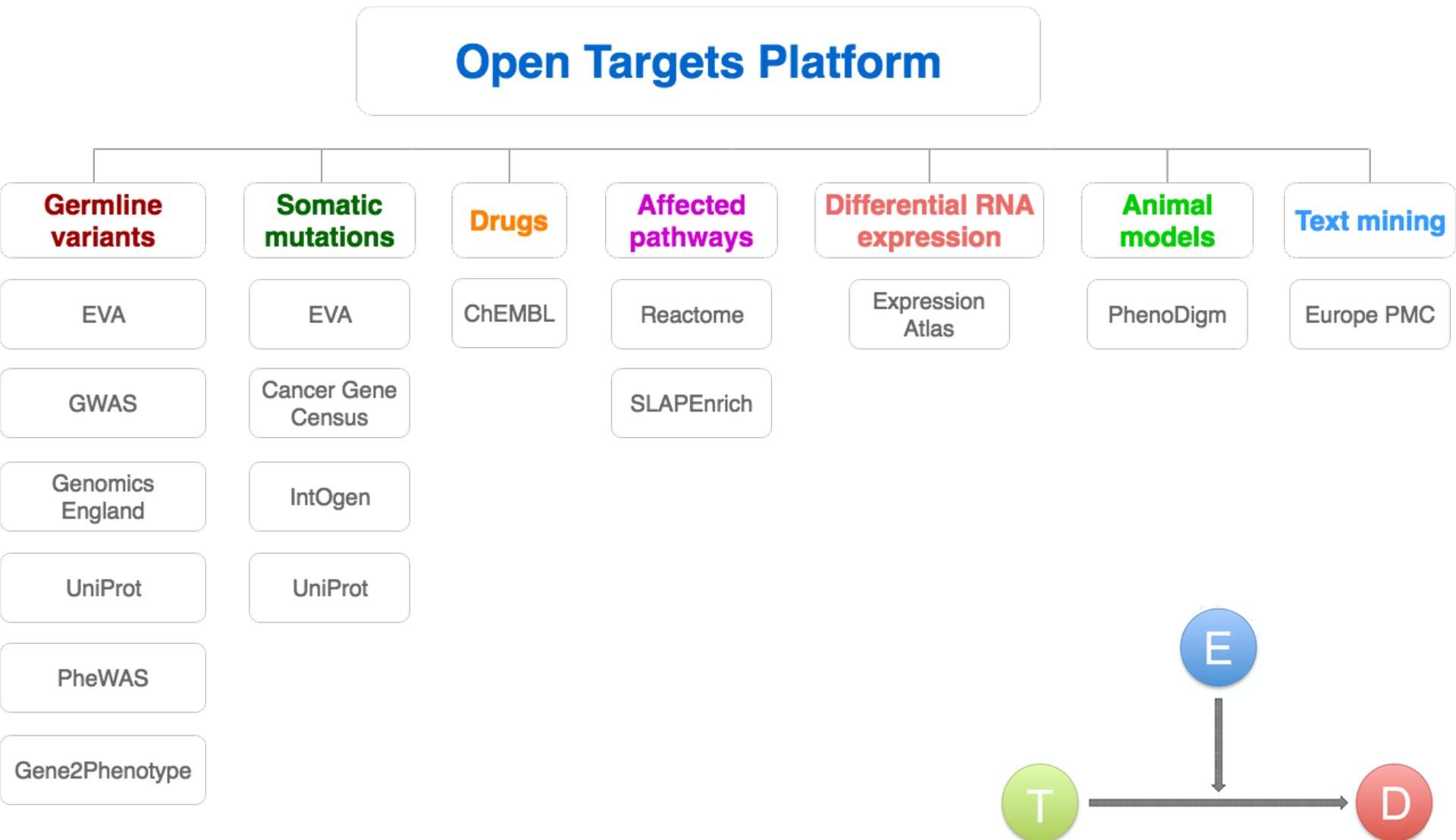
Integration of data

- 100 GSK biologists beta-tested
- 10 data sources then
- Additional sources*
- Future: crowdsourcing



www.targetvalidation.org/data_sources

Data Sources → Data types

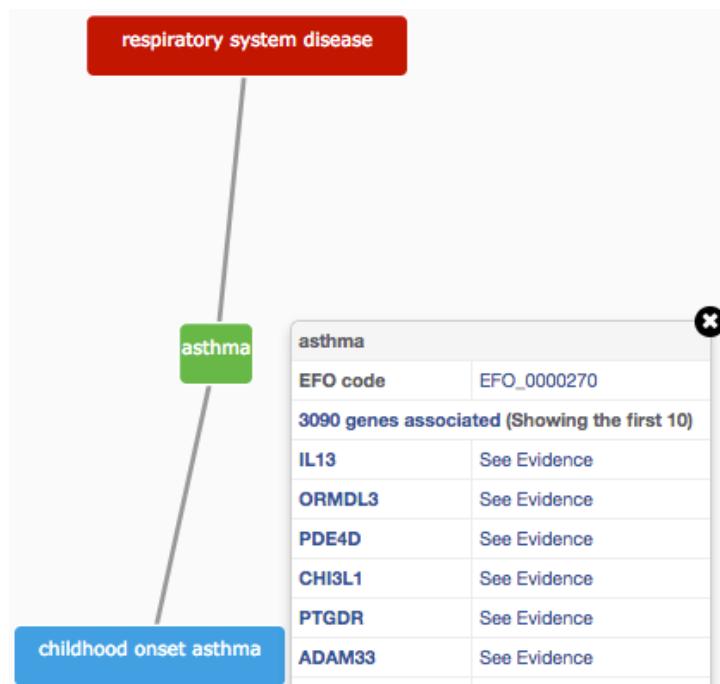


Which data to integrate?

- Suggest a causal link between the target and disease
- Aid an existing target hypothesis
- Be open, free and of easy access
- Be sustainable and likely maintained in the future
- Can be easily mapped (disease IDs, target IDs, GRCh38)

Disease ontology

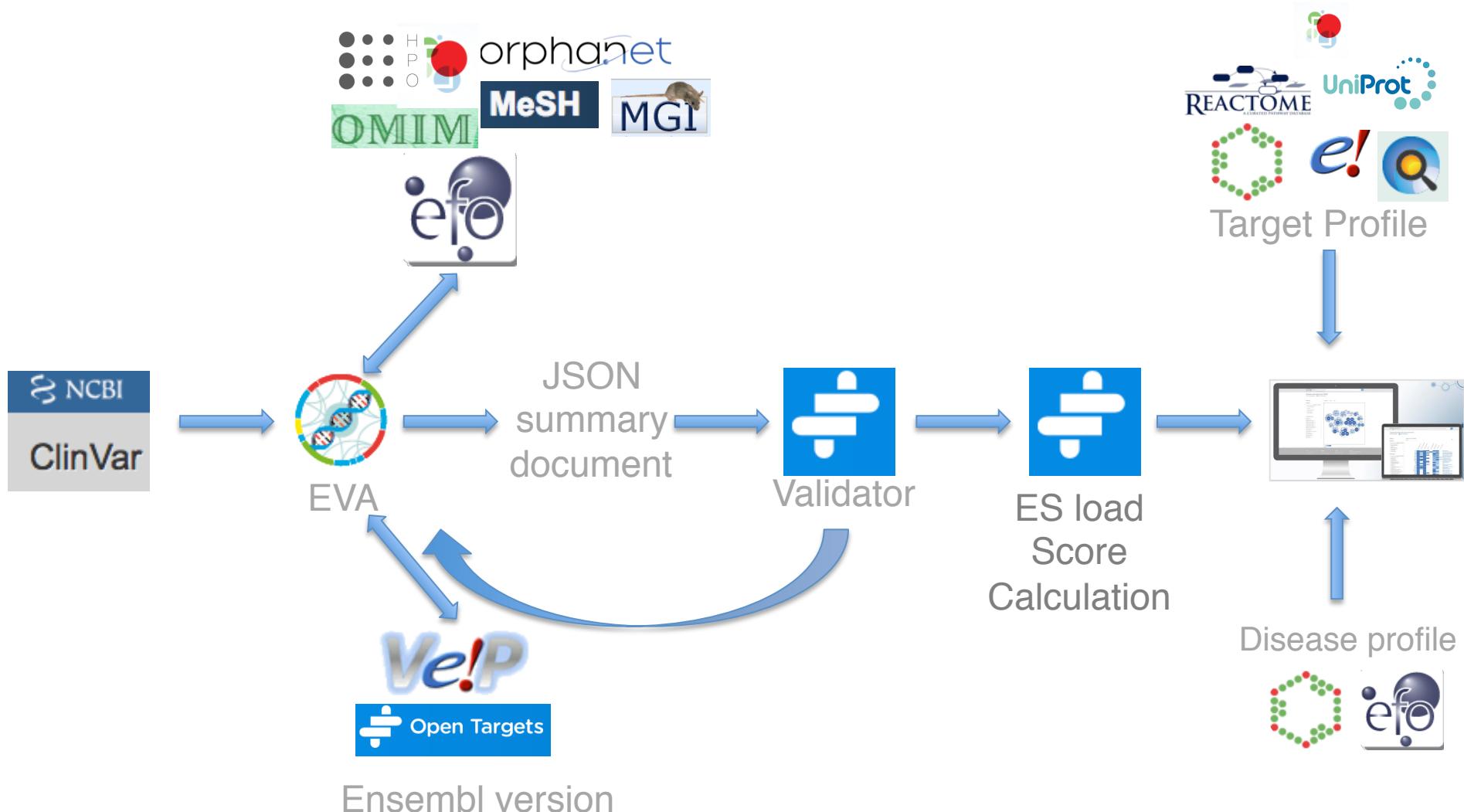
- Ontology: dictionary of relationships between entities
- EFO: way to organise experimental variables (e.g. diseases)



controlled vocabulary
(alzheimers AND alzheimer's)
+
hierarchy (relationship)

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

Integration – the data* flow



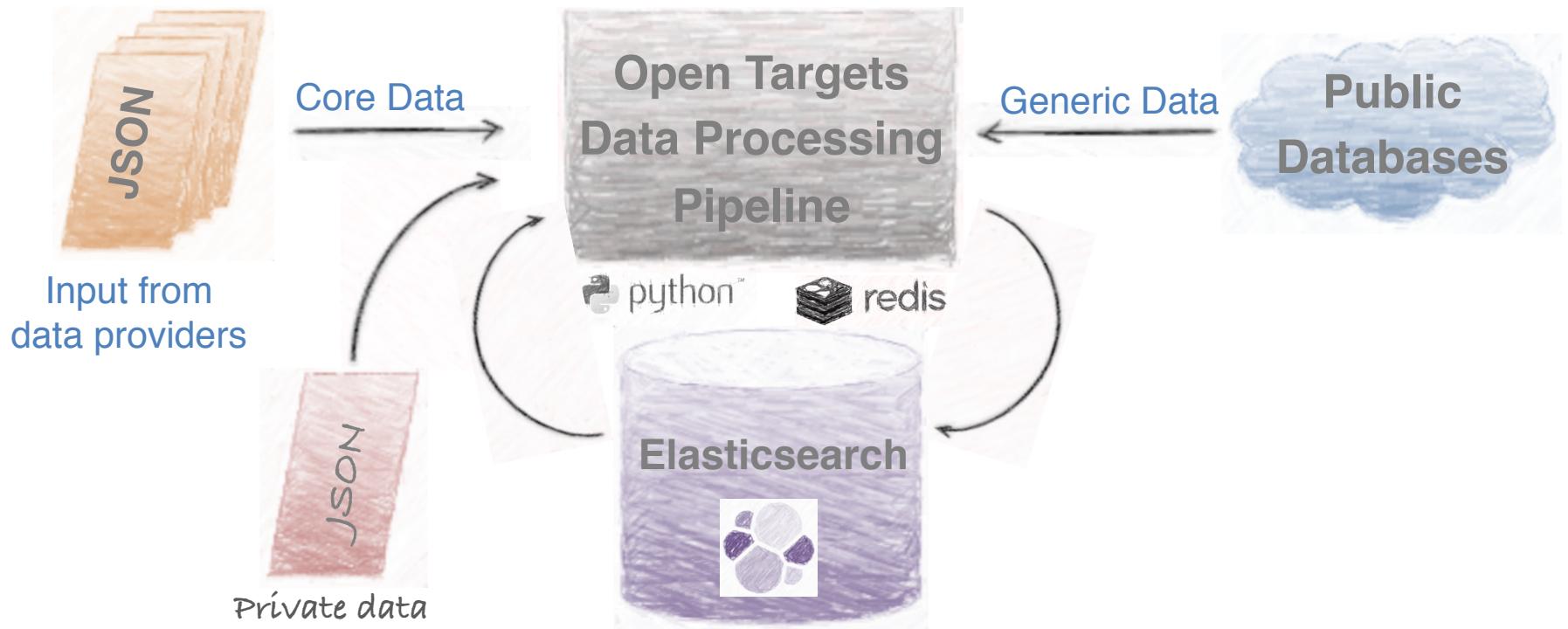
* e.g. genetic variants from EVA

JSON summary document

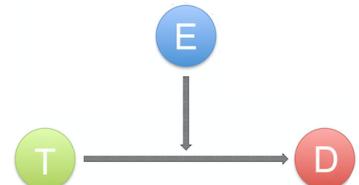
```
hort_name": "Franke A"}, {"last_name": "Alizadeh", "full_name": "Alizadeh Behrooz Z", "short_name": "Alizadeh BZ"}, {"last_name": "Parkes Miles", "short_name": "Parkes M"}, {"last_name": "B K", "full_name": "B K Thelma", "short_name": "B K T"}, {"last_name": "K J", "short_name": "Daly MJ"}, {"last_name": "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_bolbs": ["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_ntifiers.org/dbsnp/rs11150589", "study_name": "cttv009_gwas_catalog", "sample_size": "96486", "gwas_panel_resolution": "9000",  
: "http://identifiers.org/ensembl/ENSG00000005844"}, "evidence": {"variant2disease": {"gwas_sample_size": 96486, "unique_expepmc.org/abstract/MED/26192919", "gwas_panel_resolution": 9000000, "provenance_type": {"literature": {"references": [{"lit_act/MED/26192919"]}}}, "expert": {"status": true, "statement": "Primary submitter of data"}, "database": {"dbxref": {"version": "id": "http://identifiers.org/gwascatalog"}, "id": "GWAS Catalog", "version": "2017-03-23T03:44:36+00:00"}, "is_associated": {"pvalue", "method": {"description": "pvalue for the snp to disease association."}, "value": 9e-07}, "evidence_codes": [{"http://purl.obolibrary.org/obo/ECO_0000205", "http://purl.obolibrary.org/obo/ECO_0000205", "http://identifiers.org/eco/cttverted": "2017-03-23T03:44:36+00:00"}, "evidence_codes_info": [{"{"eco_id": "GWAS", "label": "Genome-wide association study ev:ing_pipeline", "label": "CTTV-custom annotation pipeline"}, {"{"eco_id": "ECO_0000205", "label": "curator inference"}], [{"{"upstream_gene_variant"}]}, "gene2variant": {"functional_consequence": "http://purl.obolibrary.org/obo/SO_0001631", "provenantrue, "statement": "Primary submitter of data"}, "database": {"dbxref": {"version": "2017-03-23T03:44:36+00:00", "id": "http://identifiers.org/cttv_mapping_pipeline"}, "id": "GWAS Catalog", "version": "2017-03-23T03:44:36+00:00"}, "is_associated": true, "resource_score": {"type": "probab:codes": [{"http://purl.obolibrary.org/obo/ECO_0000205", "http://identifiers.org/eco/cttv_mapping_pipeline"}, "date_asserted": {"evidence_codes": [{"GWAS", "cttv_mapping_pipeline", "ECO_0000205", "SO_0001631"}], "validated_against_schema_version": "1.2.9", "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

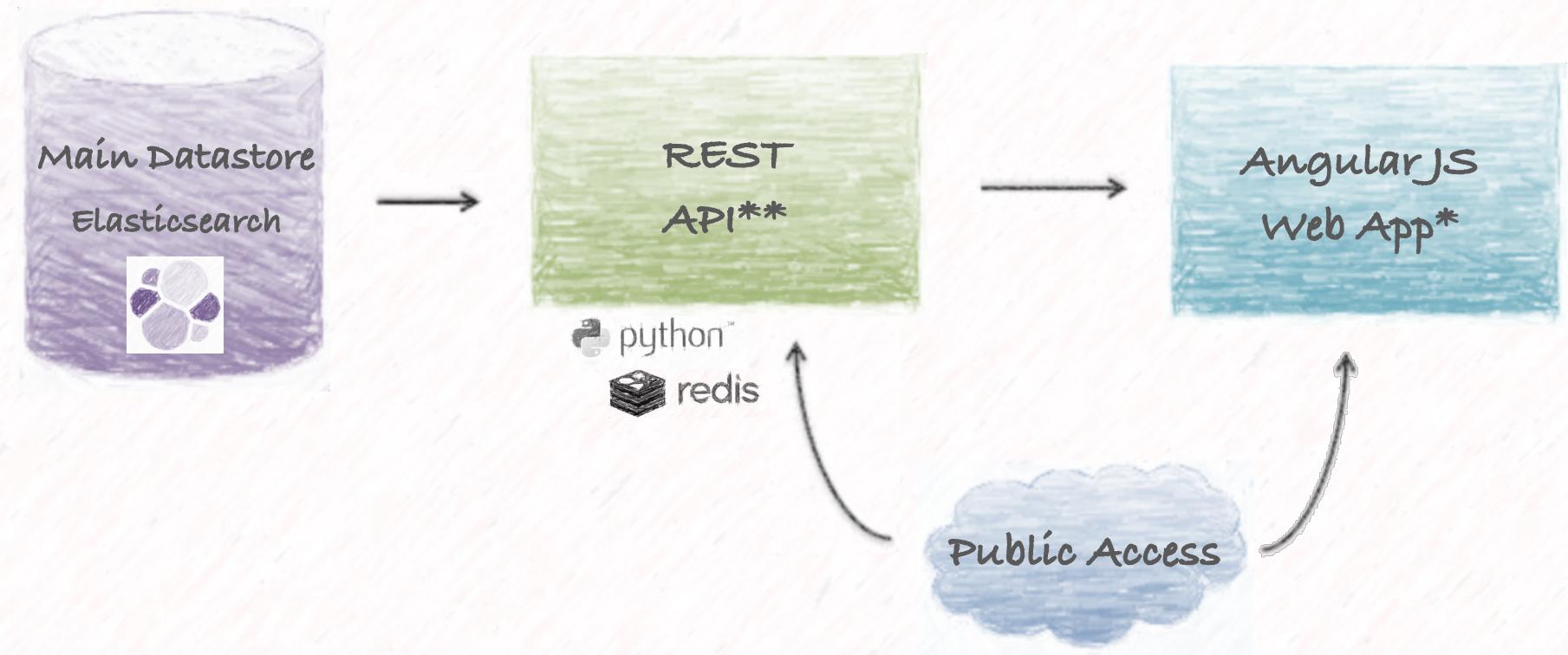
Components



1. Gather generic data from public resources (target, disease, pathways, etc...)
2. Validate Input
3. Analyse data and fit it to data model
4. Partners add private data



User access



*UI: first released in December 2015

<https://www.targetvalidation.org>

** API first release in April 2016

<https://api.opentargets.io>

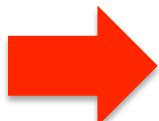
In these ~ 3.5 hours

- Introduction
- Live demos and exercises
 - User Interface (demos 1 and 2)
- REST API and some examples
- Wrap up



Demo 1: Disease centric workflow

Which targets are associated with a disease?



What is the evidence for the association between a target and a disease?

Find new targets for drug discovery

multiple sclero 🔍

multiple sclerosis
2697 targets associated

💡 Disease

An autoimmune disorder mainly affecting young adults and characterized by destruction of myelin in the central nervous system. Pathologic findings include multiple sharply demarcated areas of demyelination throughout the white matter of the central nervous system. Clinical manifestations include vis...

Targets
MBP myelin basic protein

Diseases
relapsing-remitting **multiple sclerosis**
autoimmune disease > multiple sclerosis > relapsing-remitting multiple ...

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

Choose your favourite internet browser*

*Supported ones: Internet Explorer 11 (not earlier versions), Chrome, Firefox, Safari

Demo 2: Several targets at once



We have a list of 26 possible targets for IBD (inflammatory bowel disease).

Are these targets represented in other diseases?

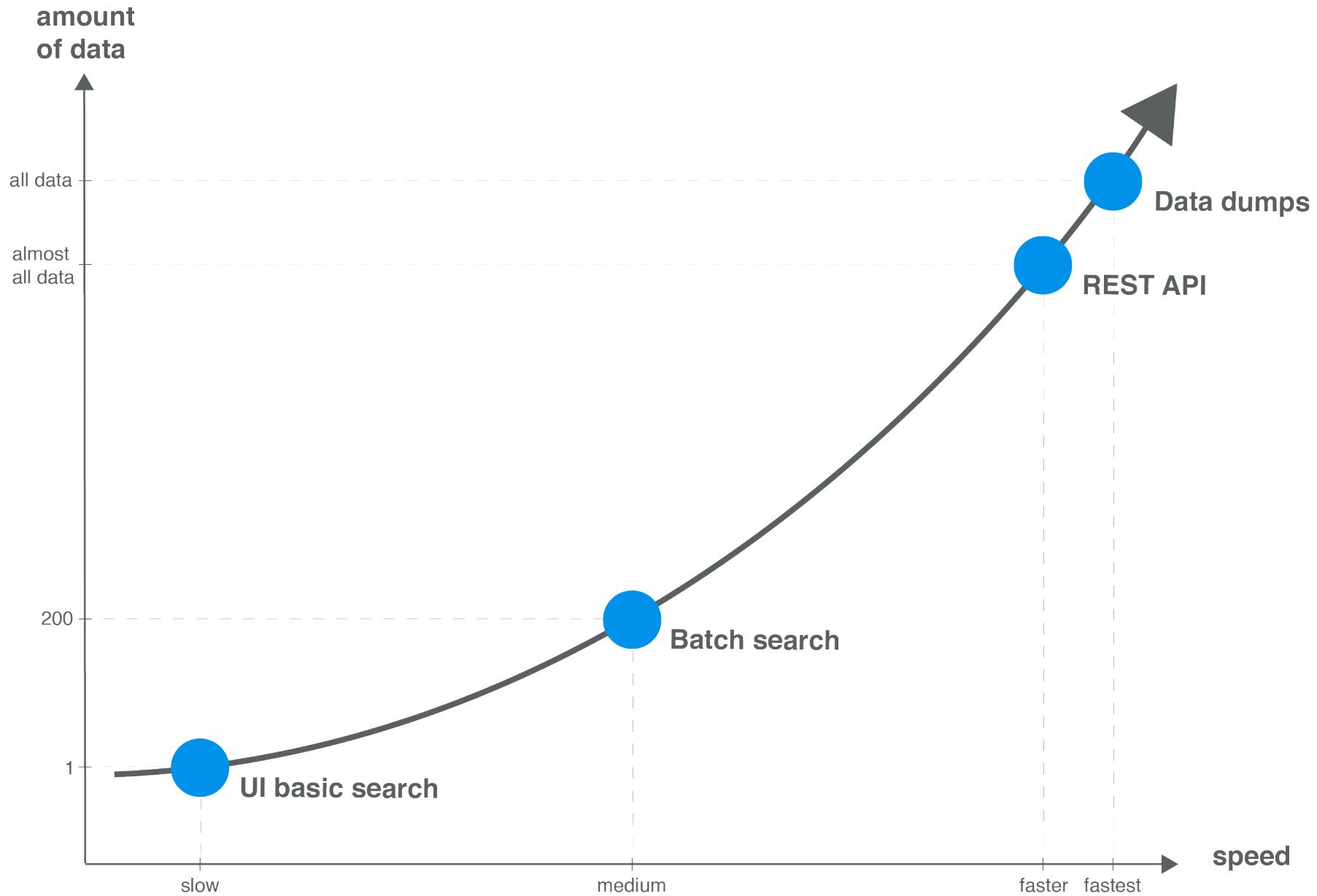
Which pathways are represented in this set of targets?

In these ~ 3.5 hours

- Introduction
- Live demos and exercises
 - User Interface (demos 1 and 2)
- REST API and some examples
- Wrap up



Other modes of data access



The documentation*

API ▾ Downloads Blog

API documentation

Python client

1



Open Targets

Search

PUBLIC

- `GET` getAssociationById
- `GET` getAssociationFilter
- `POST` postAssociationFilter
- `GET` getAuthToken
- `GET` ValidateToken
- `GET` getEvidenceById
- `POST` postEvidenceById
- `GET` getEvidenceFilter
- `POST` postEvidenceFilter
- `GET` getSearch

Public

Publicly supported stable API.

getAssociationById

Once we integrate all evidence connecting a target to a specific disease, we compute an association score by the means of an harmonic sum. This *association score* provides an indication of how strong the evidence behind each connection is and can be used to rank genes in order of likelihood as drug targets. The association ID is constructed by using the Ensembl ID of the gene and the EFO ID for the disease (e.g. ENSG0000073756-EFO_0003767). The method returns an association object, which contains the data and summary on each evidence type included in the calculation of the score, as well as the score itself.

PARAMETERS

Query Parameters ⓘ

→ `id` string **Required**
An association ID usually in the form of `TARGET_ID-DISEASE_ID`.

GET /platform/public/association

Server URL
`http://api.opentargets.io/v3/platform/public/association`

POST /platform/public/association

REQUEST SAMPLES

```
"{ "target": ["target1", "target2"] }"
```

<http://api.opentargets.io/v3/platform/docs#>

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=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`

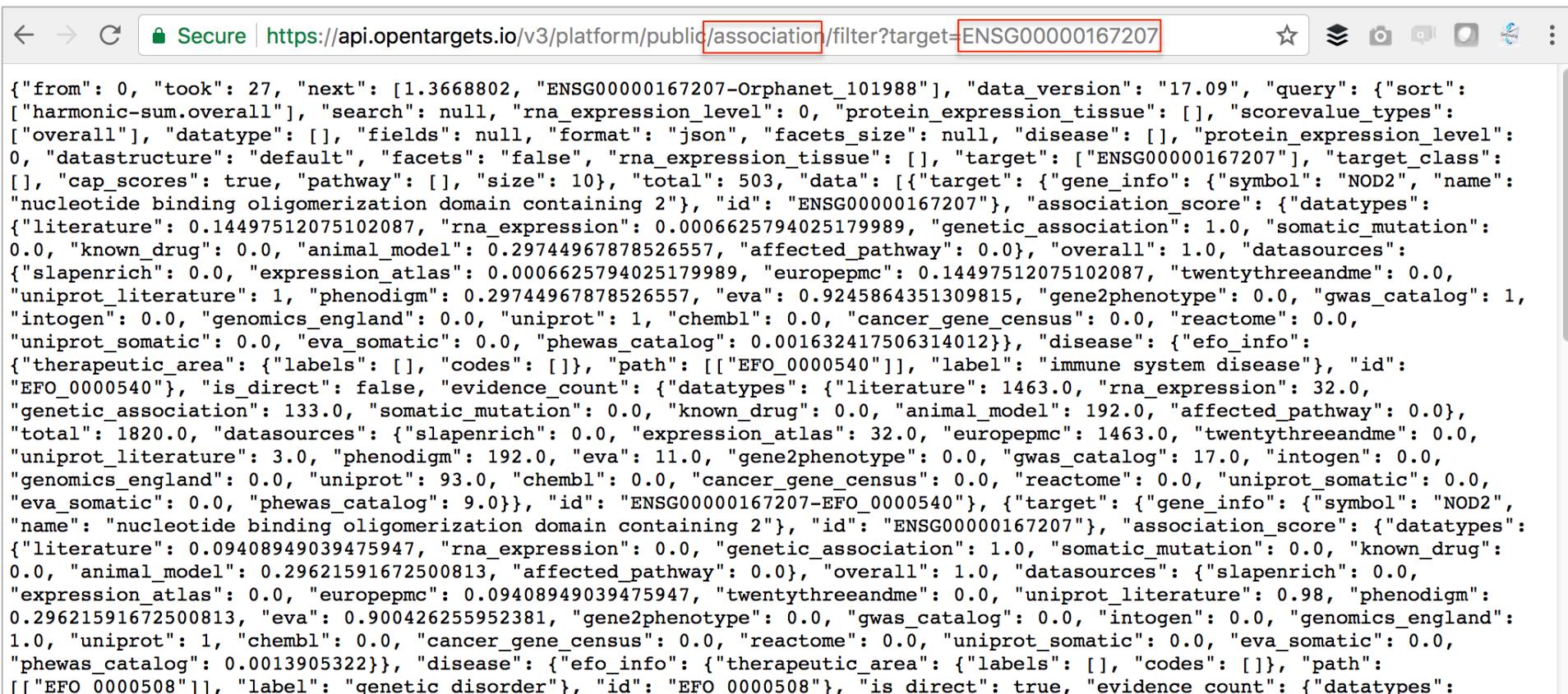
`?target=ENSG00000163914&size=10000&fields=target.id&fields=disease.id`

Several ways to run our REST endpoints

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

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

Paste the URL in the a location bar



The screenshot shows a browser window with the URL <https://api.opentargets.io/v3/platform/public/association/filter?target=ENSG00000167207>. The URL is highlighted with a red box. The page content is a JSON object representing a search result for the gene ENSG00000167207. The JSON includes fields such as 'from', 'took', 'next', 'data_version', 'query', 'target', 'association_score', and various association details like 'genetic_association', 'somatic_mutation', and 'known_drug' across different databases and sources.

```
{"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"], "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}}}}
```



Python and R clients for the REST API

opentargets
latest

Search docs

Tutorial
High Level API
Low Level API
Code Documentation
Changelog

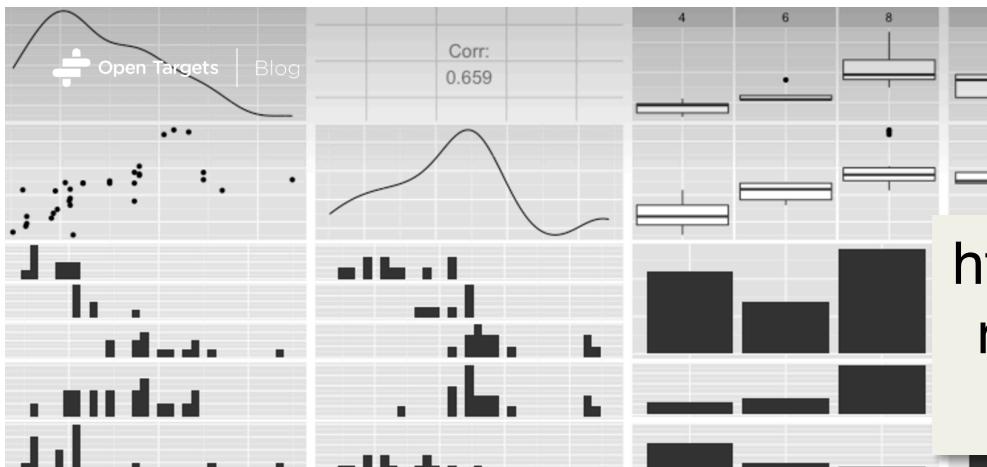
Docs » opentargets - Python client for targetvalidation.org

Edit on GitHub

opentargets - Python client for targetvalidation.org

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

<http://opentargets.readthedocs.io>



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

How to access Open Targets
with R

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 674 diseases
Tumor suppressor. Activates tyrosine-, serine- and phosphatase, removing phosphatidylinositol 3-kinase-activated kinase substrate. Approved symbol: PTEN and dual-specificity protein kinase.

 Target

```
{  
    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"  
                        }  
                    ]  
                }  
            }  
        }  
    ]  
}
```

GET

getSearch

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

How to get all diseases associated with a target

Open Targets Platform

About ▾ Help ▾ API ▾ Downloads Blog

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				

```
{  
    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,  
            }  
        }  
    ]  
}
```

GET getAssociationFilter

[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)

How to get the score for a target – disease pair

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

Showing 1 to 50 of 674 entr

Search:

Diseas...

neoplas...

GET

getAssociationById

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 tensin homolog",  
                    geneid: "ENSG00000171862"  
                },  
                id: "ENSG00000171862"  
            },  
            - association_score: {  
                - datatypes: {  
                    literature: 0.3241324475302135,  
                    rna_expression: 0,  
                    somatic_mutation: 1,  
                    genetic_association: 1,  
                    known_drug: 0,  
                    animal_model: 0,  
                    affected_pathway: 1  
                },  
                overall: 1,  
            }  
            - datasources: {  
                slapenrich: 0.817215326415924,  
                expression_atlas: 0,  
                gene2phenotype: 0,  
            }  
        }  
    ]  
}
```

How to get the evidence for an association

Evidence f



Genetic ass

Table E

Rare disease

Source: UniProt

Showing 1 to 5 o

```
],
  format: "json",
  fields: null,
  datastructure: "default",
  pathway: [ ],
  size: 10
},
  total: 5,
  data: [
    {
      + target: {...},
      validated_against_schema_version: "1.2.7",
      sourceID: "genomics_england",
      + disease: {...},
      + unique_association_fields: {...},
      - evidence: {
          is_associated: true,
          - urls: [
              {
                url: "https://panelapp.genomicsengland.co.uk/panels/597a33fc8f620307bbcc5f5e",
                nice_name: "Further details in the Genomics England PanelApp"
              }
            ],
          + provenance_type: {...},
          + evidence_codes_info: [...],
          + resource_score: {...},
          + evidence_codes: [...],
          date_asserted: "2017-11-27T17:48:00.516623"
        },
        access_level: "public",
        - scores: {
            association_score: 1
          },
          type: "genetic_association",
          id: "e6eba7e9d37146889828818546907903"
        }
      ]
    }
  ]
}
```

GET getEvidenceFilter

[http://api.opentargets.io/v3/platform/public/evidence/filter?
target=ENSG00000171862
&disease=Orphanet_2563](http://api.opentargets.io/v3/platform/public/evidence/filter?target=ENSG00000171862&disease=Orphanet_2563)

The documentation

<http://api.opentargets.io/v3/platform/docs#>



Open Targets

Search

- PUBLIC >
- PRIVATE >
- AUTH >
- UTILS >

Open Targets Platform REST API (v3)

Download OpenAPI specification: [Download](#)

Contact: support@targetvalidation.org License: Apache 2.0

The Open Targets Platform REST API

The Open Targets Platform API ('Application Programming Interface') allows you to interact with our platform via a series of REST services.

You can make calls to the latest version of our API using the base URL <https://api.opentargets.io/v3/platform>. Please make sure you use `http` calls, which we do not accept. Note that we only serve the latest version of the API. If you are querying an old version, please [get in touch](#) so that we can help.

We list below the methods available for you to query our data directly from the browser. These are generated from our Swagger UI. For every request you create, the interface will generate the code for you to copy and paste directly to a shell to obtain the same results without using an intermediate step.

Check our [API blog posts](#), for tutorials and additional information on how to use the API.

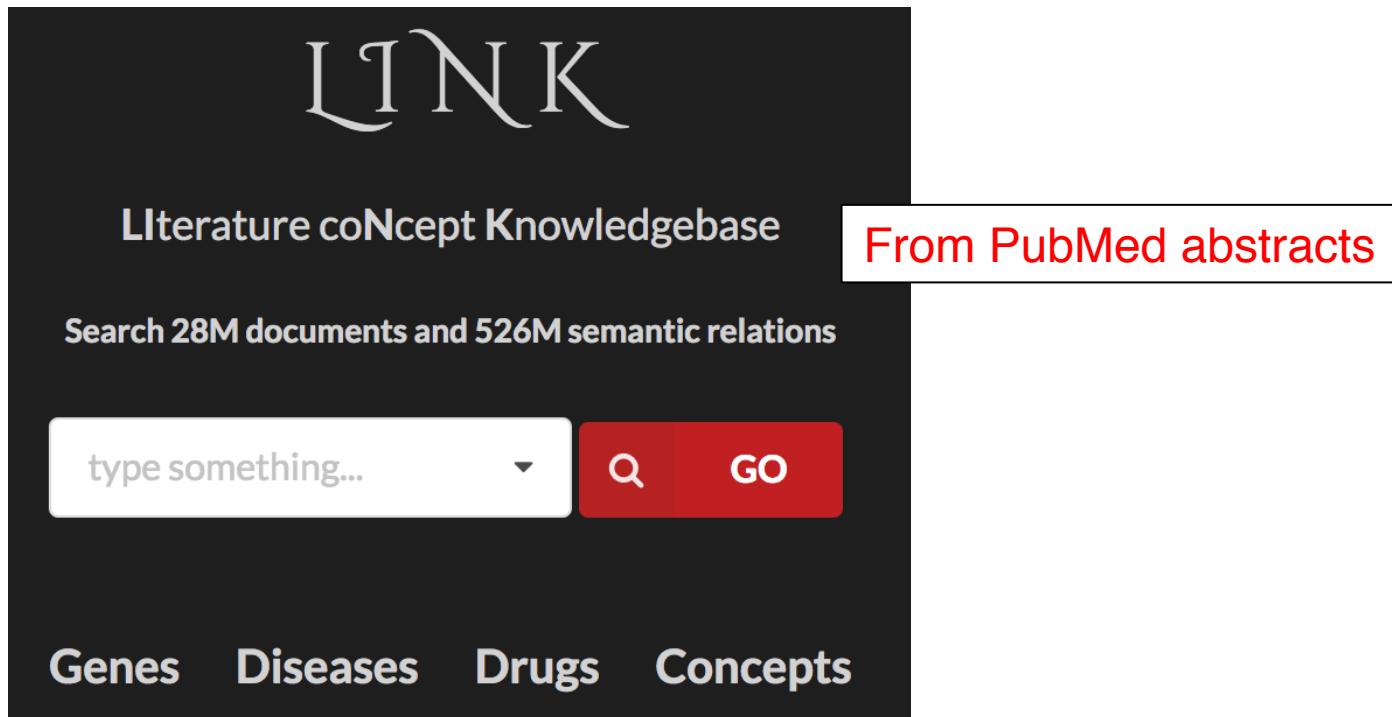
PRIVATE

- GET** `getApiDocs`
- GET** `getAutocomplete`
- POST** `postBestHitSearch`
- POST** `postDiseaseById`
- GET** `getDiseaseById`
- GET** `getECObyID`
- POST** `postEnrichmentTarget`
- GET** `getQuickSearch`
- POST** `postRelation`

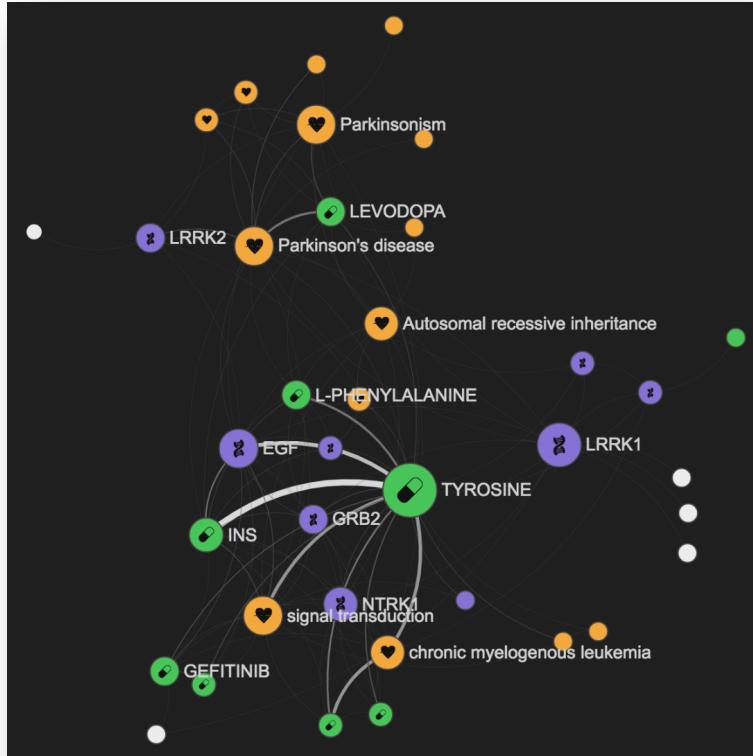
Private: methods used by the UI to serve external data.
Subject to change without notice

Beyond the Open Targets Platform

- LINK: Literature coNcept Knowledgebase



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



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

LRRK1 Irrk1 play in Parkinson's disease pd

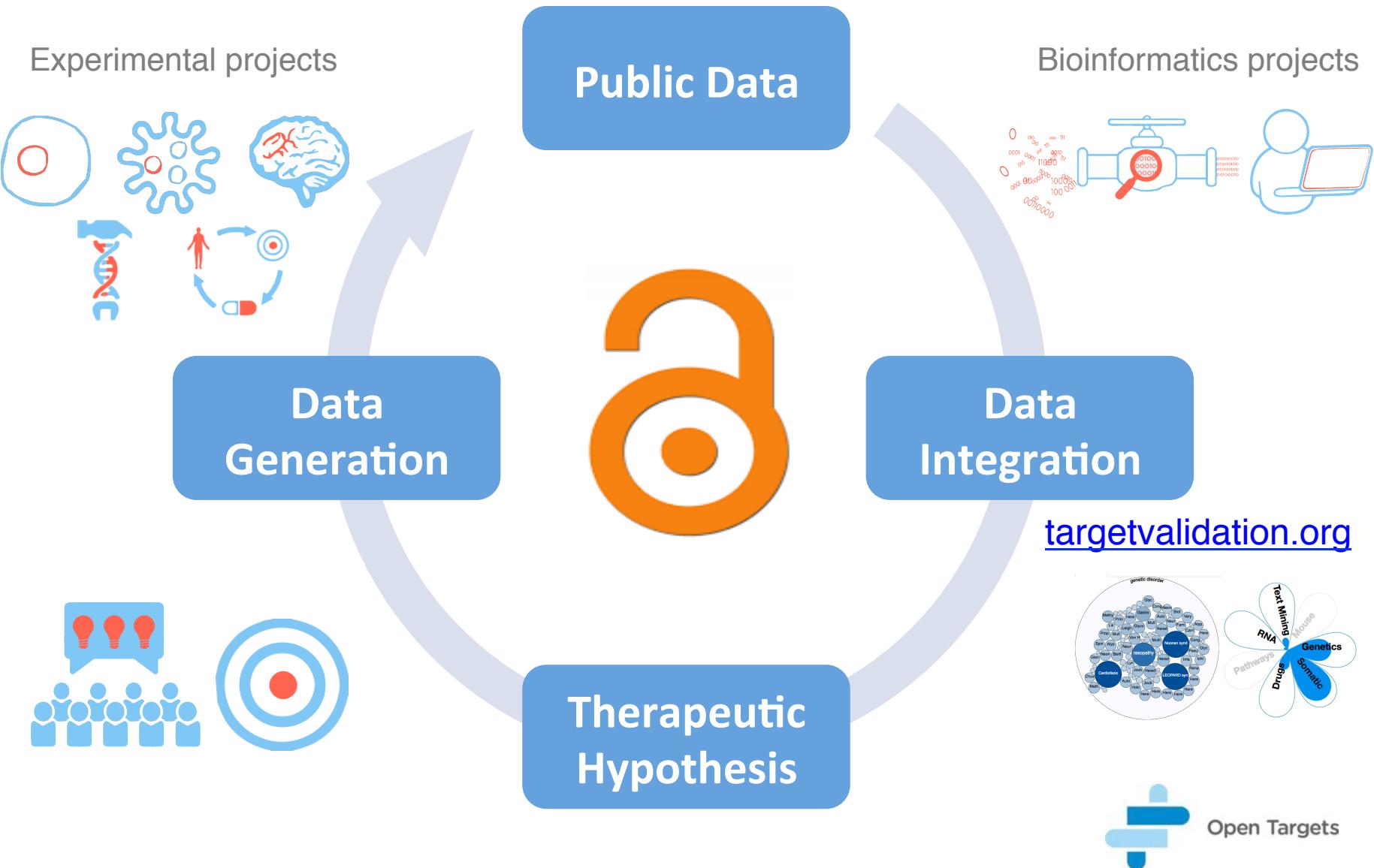
In contrast, LRRK1 GENE, the closest homologue to LRRK2, does not play any role in PD DISEASE CONCEPT.

PubMed: 28819229 2017-08-17

In these ~3.5 hours

- Introduction
- Live demos and exercises
 - User Interface (demos 1 and 2)
 - REST API and simple calls (demo 3)
- Wrap up

Open Targets Partnership



Open Targets Platform

- Resource of integrated data
 - Added new diseases
 - Graphics and reports easy to use
 - Drug discovery, translational medicine, disease biology
- Oh Yes!
And all is 100% free

20K
targets

9.7K
diseases

2.3 M
associations

5.9 M
evidence

February 2018 release

We support decision-making

Which targets are associated with a disease?

Can I find out about the mechanisms of the disease?

Are there FDA drugs for this association?



...

How to cite us

Published online 8 December 2016

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

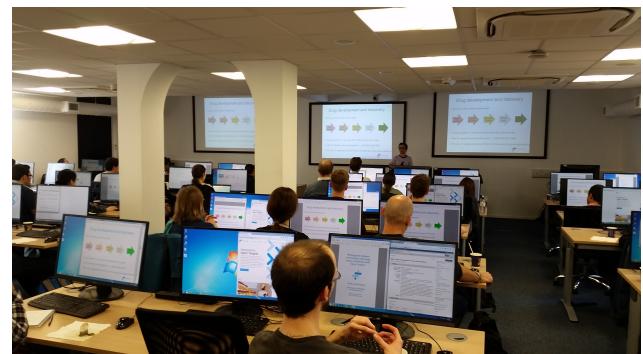
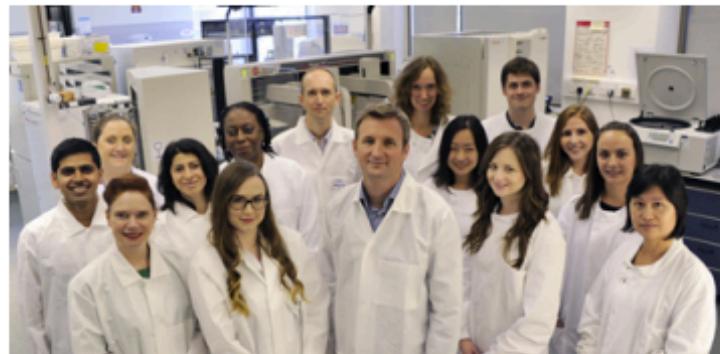
Open Targets: a platform for therapeutic target identification and validation

Gautier Koscielny^{1,2,*}, Peter An^{1,3}, Denise Carvalho-Silva^{1,4}, Jennifer A. Cham^{1,4}, Luca Fumis^{1,4}, Rippa Gasparyan^{1,3}, Samiul Hasan^{1,2}, Nikiforos Karamanis^{1,4}, Michael Maguire^{1,4}, Eliseo Papa^{1,3}, Andrea Pierleoni^{1,4}, Miguel Pignatelli^{1,4}, Theo Platt^{1,3}, Francis Rowland^{1,4}, Priyanka Wankar^{1,3}, A. Patrícia Bento^{1,4}, Tony Burdett^{1,4}, Antonio Fabregat^{1,4}, Simon Forbes^{1,5}, Anna Gaulton^{1,4}, Cristina Yenyxe Gonzalez^{1,4}, Henning Hermjakob^{1,4,6}, Anne Hersey^{1,4}, Steven Jupe^{1,4}, Şenay Kafkas^{1,4}, Maria Keays^{1,4}, Catherine Leroy^{1,4}, Francisco-Javier Lopez^{1,4}, Maria Paula Magarinos^{1,4}, James Malone^{1,4}, Johanna McEntyre^{1,4}, Alfonso Munoz-Pomer Fuentes^{1,4}, Claire O'Donovan^{1,4}, Irene Papatheodorou^{1,4}, Helen Parkinson^{1,4}, Barbara Palka^{1,4}, Justin Paschall^{1,4}, Robert Petryszak^{1,4}, Naruemon Pratanwanich^{1,4}, Sirarat Sarntivijal^{1,4}, Gary Saunders^{1,4}, Konstantinos Sidiropoulos^{1,4}, Thomas Smith^{1,4}, Zbyslaw Sondka^{1,5}, Oliver Stegle^{1,4}, Y. Amy Tang^{1,4}, Edward Turner^{1,4}, Brendan Vaughan^{1,4}, Olga Vrousou^{1,4}, Xavier Watkins^{1,4}, Maria-Jesus Martin^{1,4}, Philippe Sanseau^{1,2}, Jessica Vamathevan⁴, Ewan Birney^{1,4}, Jeffrey Barrett^{1,4,5} and Ian Dunham^{1,4,*}

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

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

Acknowledgements



Open Targets

Help!



support@targetvalidation.org



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



[@targetvalidate](#)



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

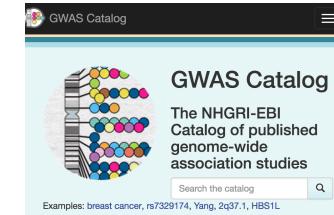
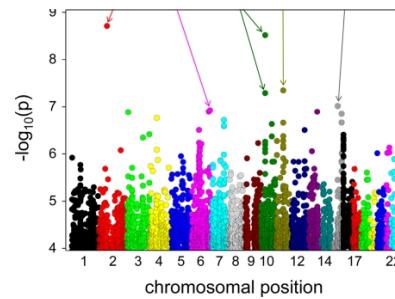
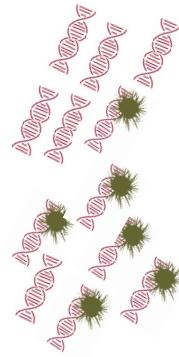
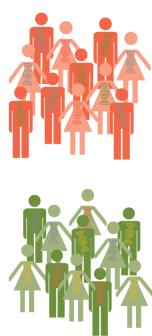


blog.opentargets.org/

Extra Extra Extra

Data sources: GWAS catalog

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



Open Targets

Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients

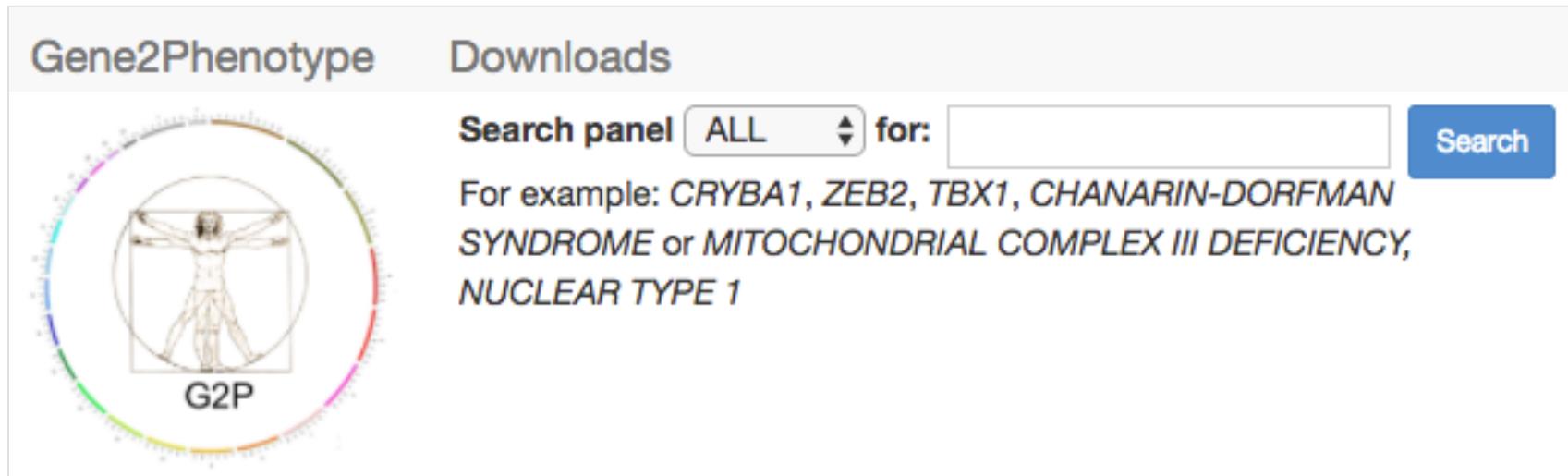


EMBL-EBI train online



Open Targets

Data sources: Gene2Phenotype



Gene2Phenotype Downloads

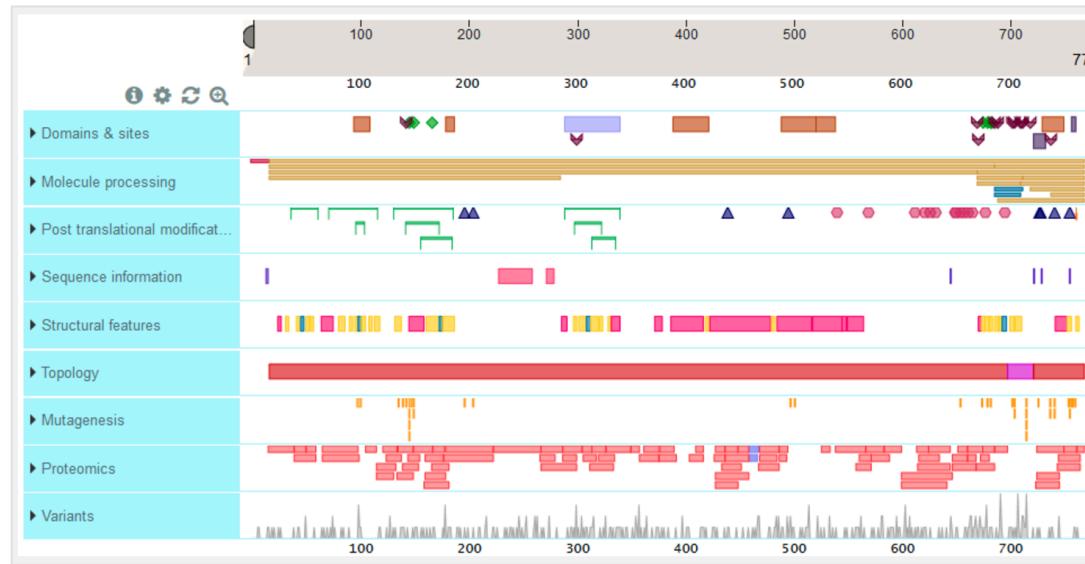
Search panel ALL for: **Search**

For example: **CRYBA1, ZEB2, TBX1, CHANARIN-DORFMAN SYNDROME or MITOCHONDRIAL COMPLEX III DEFICIENCY, NUCLEAR TYPE 1**

- Variants, genes, phenotypes in rare diseases
- Literature curation → consultant clinical geneticists in the UK

Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of coding variants in patients



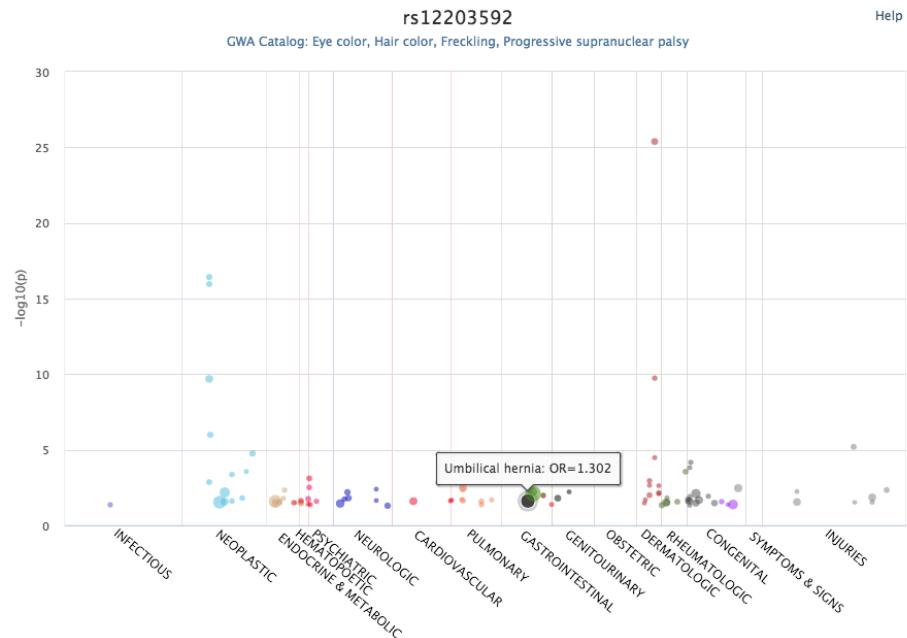
EMBL-EBI train online



Open Targets

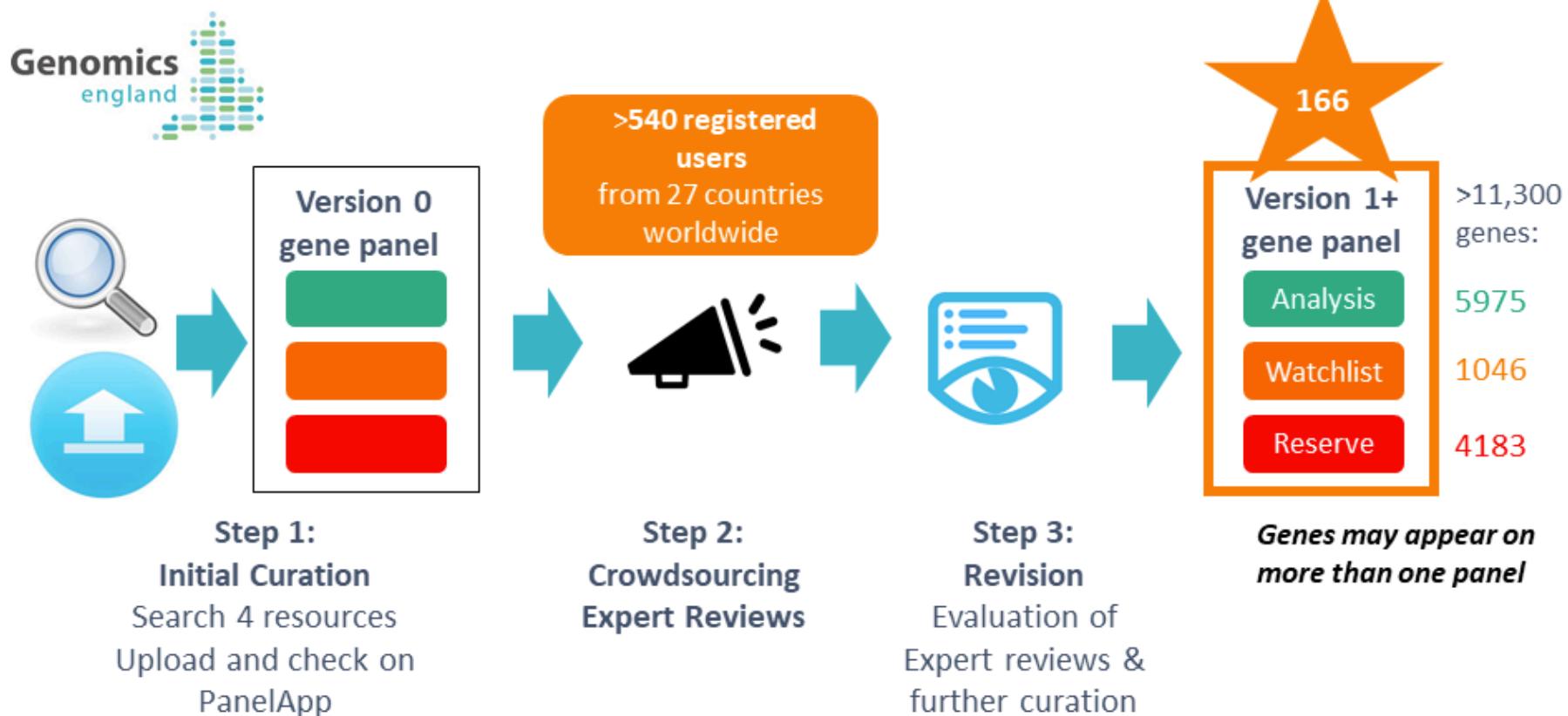
Data sources: PheWAS

- Phenome Wide Association Studies
- A variant associated with multiple phenotypes
- Clinical phenotypes derived from EMR-linked biobank BioVU
- ICD9 codes mapped to EFO



Data sources: GE PanelApp

- Aid clinical interpretation of genomes for the 100K project
- We include ‘green genes’ from version 1+ and phenotypes



Germline
variants

Somatic
mutations

Data sources: EVA

- With ClinVar information for rare diseases
- Clinical significance: pathogenic, protective

The screenshot shows the European Variation Archive (EVA) Clinical Browser interface. The top navigation bar includes links for Home, Submit Data, Study Browser, Variant Browser, Clinical Browser (which is highlighted in dark blue), GA4GH, API, FAQ, and Feedback. Below the navigation is a search bar with a magnifying glass icon and a "Filter" button. The main content area is titled "ClinVar Browser" with an information icon. It features a table with 960 records, showing columns for Position, Affecte..., Most Severe Consequence..., Trait, Clinical Significance, and ClinVar The first few rows of data are as follows:

...	Posi...	Affecte... i	A...	Most Severe Consequence...	Trait	Clinical Significance	ClinVar ...
2	480...	MSH6	T/G	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	G/A	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	C/T	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	C/T	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	G/T	5_prime_UTR...	Lynch synd...	Uncertain s...	RCV000...
2	480...	MSH6	G/T	5_prime_UTR...	Hereditary ...	conflicting ...	RCV000...



Data sources: The Cancer Gene Census

Census

Breakdown

Abbreviations

The cancer Gene Census is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer. The original census and analysis was published in [Nature Reviews Cancer](#) and supplemental analysis information related to the paper is also available.

- Genes with mutations causally implicated in cancer
- Gene associated with a cancer plus other cancers associated with that gene

Data sources: IntOGen

The screenshot shows the homepage of the intOGen website. At the top is a navigation bar with an orange gradient background. From left to right, it contains: the intOGen logo (a stylized orange 'i' icon followed by the word 'intOGen'), a 'Search' button with a magnifying glass icon, a 'Downloads' button with a download icon, an 'Analysis' button with a gear icon, an 'About' button with a speech bubble icon, and a 'Sign In' button with a user profile icon.

The main content area features the intOGen logo again, this time with the full name 'intOGen' in a large serif font next to a smaller orange 'i' icon. To the right of the name is the tagline 'Integrative Onco Genomics' in orange text.

- Genes and somatic (driver) mutations, 28 cancer types
- Involvement in cancer biology
- Rubio-Perez et al. 2015

Data sources: ChEMBL

EMBL-EBI

ChEMBL

EBI > Databases > Small Molecules > ChEMBL Database > Home

Search ChEMBL... Compounds Targets Assays

Ligand Search Target Search Browse Targets Browse Drugs Browse Drug Targets

- Known drugs linked to a disease and a known target
- FDA approved for clinical trials or marketing



EMBL-EBI train online



Open Targets

Affected
pathways

Data sources: Reactome



The image shows the Reactome homepage. At the top left, there is a purple box containing the text "Affected pathways". The main title "REACTOME" is displayed in large white letters, with "A CURATED PATHWAY DATABASE" in smaller text below it. To the right of the title is a decorative graphic featuring a red sphere, blue and green abstract shapes, and arrows. Below the title is a navigation bar with links: "About", "Content", "Documentation", "Tools", "Community", "Download", and "Contact". To the right of the navigation bar is a search bar containing the placeholder text "e.g. O95631, NTN1, signalin" and a "Search" button.

- Biochemical reactions and pathways
- Manual curation of pathways affected by mutations

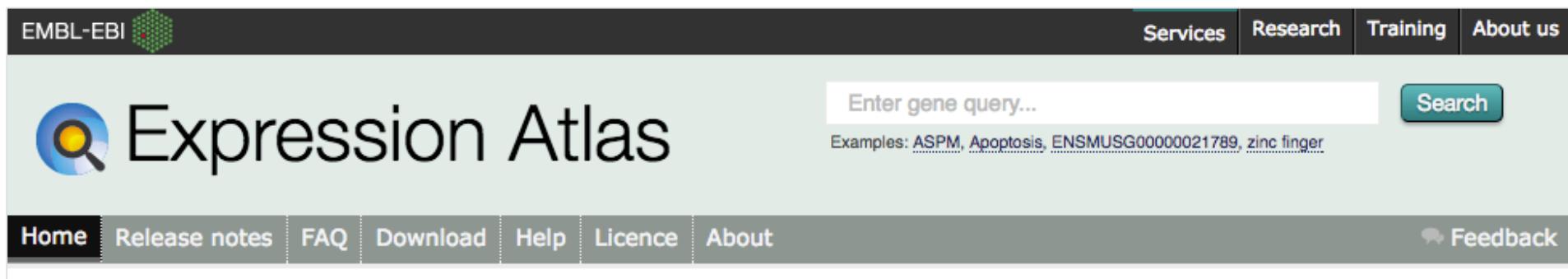


EMBL-EBI train online



Open Targets

Data sources: Expression Atlas



The screenshot shows the Expression Atlas website. At the top, there is a navigation bar with the EMBL-EBI logo, followed by links for Services, Research, Training, and About us. Below the navigation bar is a search bar with the placeholder "Enter gene query..." and a "Search" button. To the left of the search bar is a magnifying glass icon. The main title "Expression Atlas" is displayed prominently. Below the title is a horizontal menu bar with links for Home, Release notes, FAQ, Download, Help, Licence, and About. On the far right of this menu bar is a "Feedback" link.

- Baseline expression for human genes
- Differential mRNA expression (*healthy versus diseased*)



EMBL-EBI train online



Open Targets

Data sources: Europe PMC



Europe PMC

About

Tools

Developers

Help

Europe PMC plus

Search worldwide, life-sciences literature

- Mining titles, abstracts, full text in research articles
- Target and disease co-occurrence in the same sentence



EMBL-EBI train online



Open Targets

Data sources: PhenoDigm

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

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

Diseases Tissue phenotype associations Secondary phenotypes

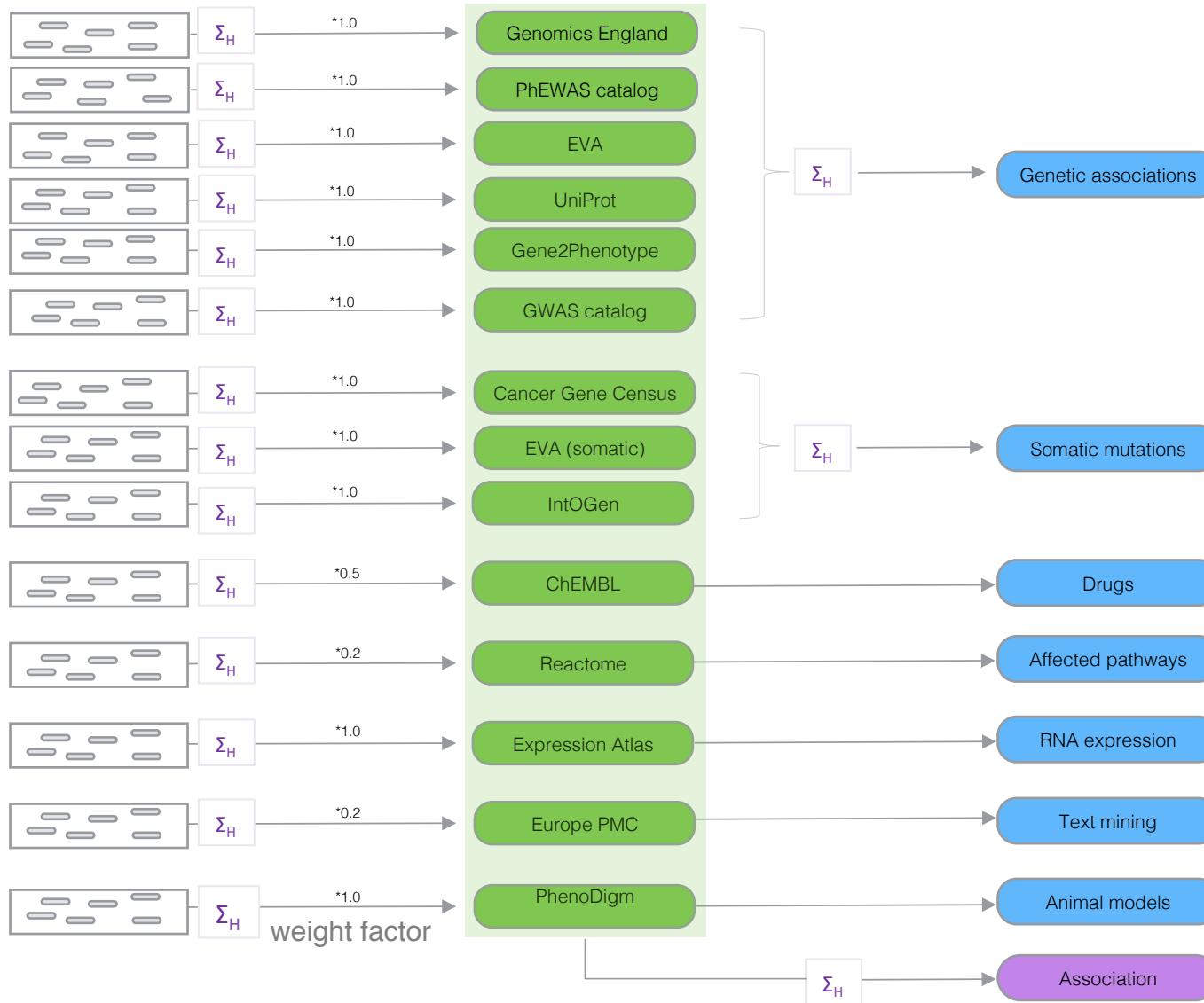
- Semantic approach to associate mouse models with diseases

How confident can you be of the target-disease associations in Open Targets?

Statistical integration, aggregation and scoring*

- A) per evidence (e.g. one SNP from a GWAS paper)
- B) per data source (e.g. SNPs from the GWAS catalog)
- C) per data type (e.g. Genetic associations)
- D) overall

How do we score the associations?



Score: 0 to 1 (max)

Calculated at 4 levels:

- Evidence
- Data source
- Data type
- Overall

Aggregation with (harmonic sum)



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

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

Factors affecting the relative strength of an evidence

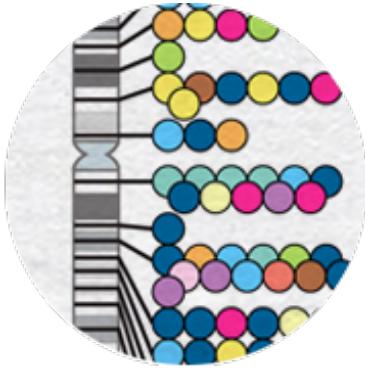
e.g. *GWAS Catalog*

$$S = f * s * c$$

f, relative occurrence of a target-disease evidence

s, strength of the effect described by the evidence

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



f = sample size (cases versus controls)

s = predicted functional consequence

c = *p* value reported in the paper



Open Targets

Aggregating scores across the data

- Using a mathematical function, the harmonic sum*

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

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

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

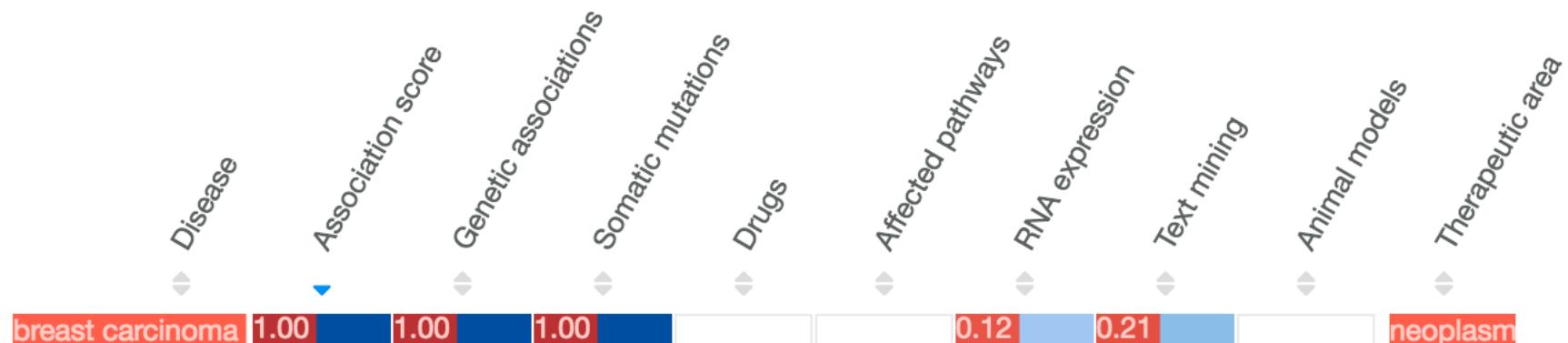
Disclaimer: score, dos and don'ts

- It's a ranking of target-disease associations
- It shows how confident we are in the association
- It's based on data sources, publicly available



- It can help you to design your null hypothesis
- It can help you to decide which target to pursue
- It is NOT sufficient on its own (use it in combination with...)

Ranking the target-disease association



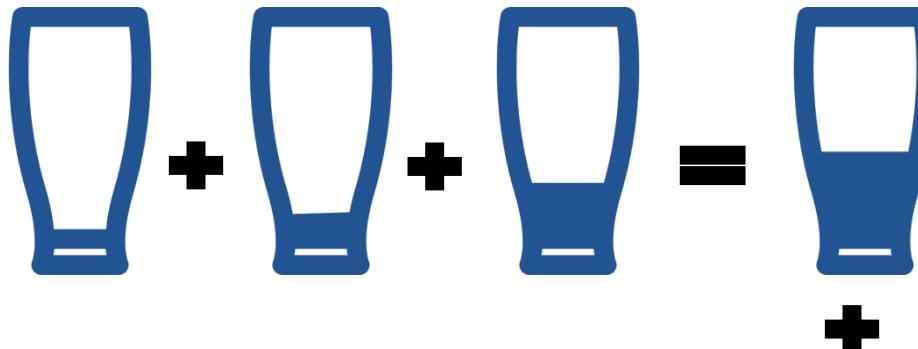
Association score: the overall score across all data types

- Based on the data sources
- Different weight applied:

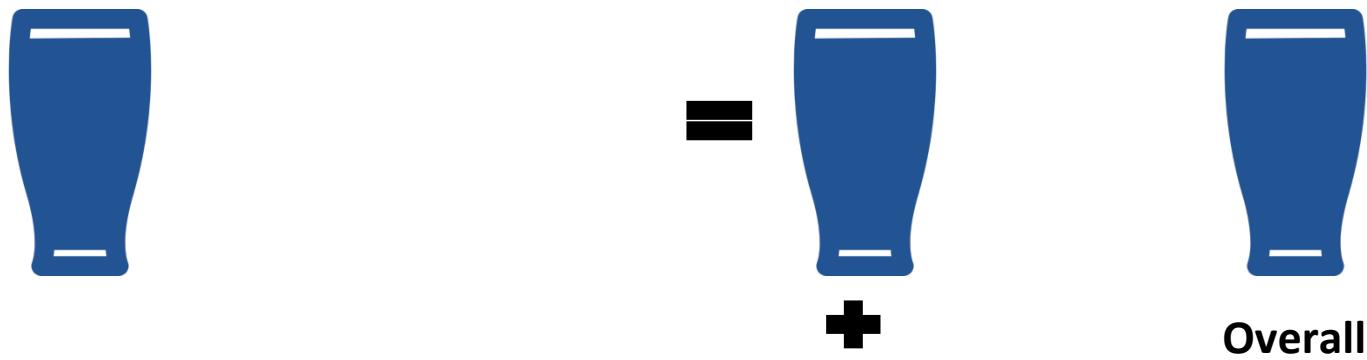
genetic association = drugs = mutations = pathways > RNA expression > animal models = text mining

Target-Disease Association Score

EuropePMC
(Text Mining)

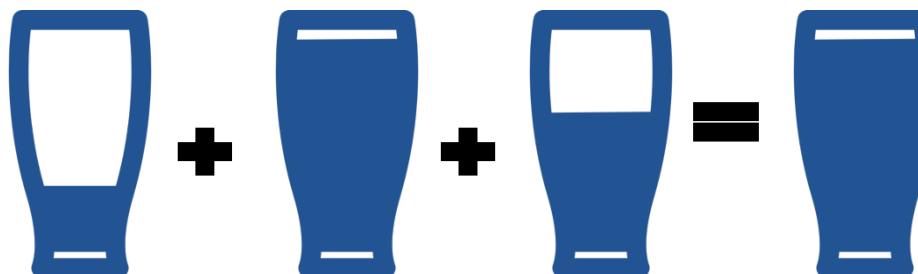


UniProt
(Manual Curation)



Overall

ChEMBL
(Manual Curation)



VERY simplified diagram

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