

Mining Gene-Disease Associations with Open Targets

Webinar

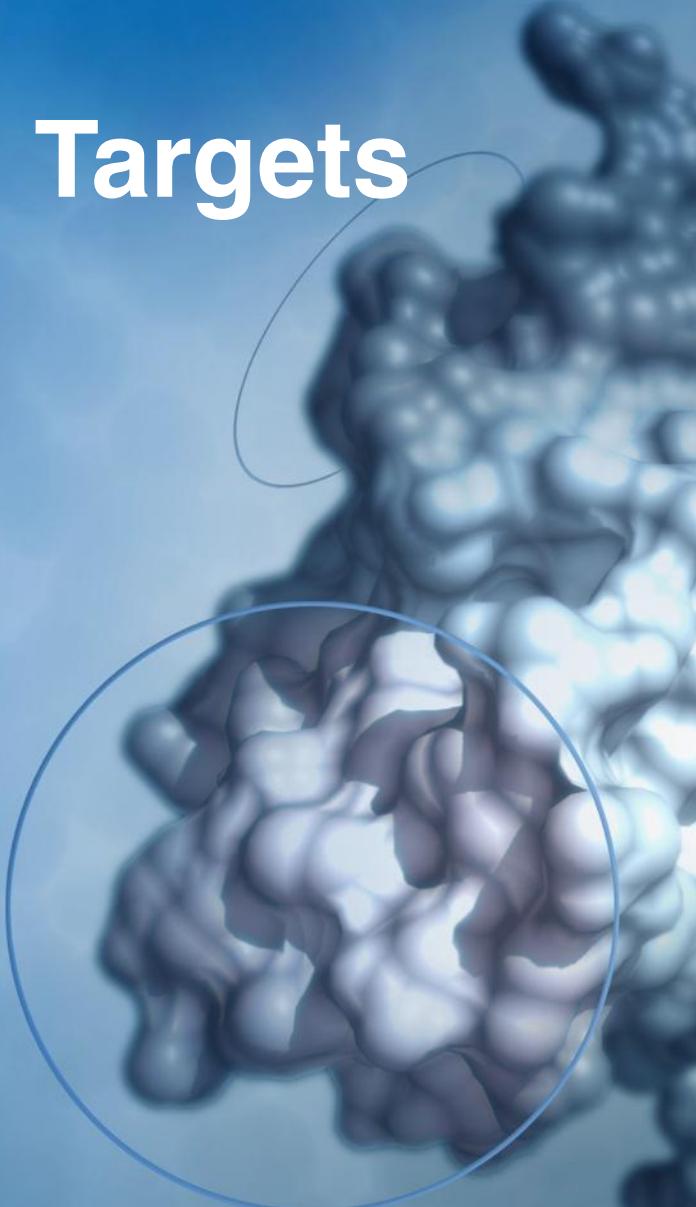
Biogen

Denise Carvalho-Silva

Wellcome Genome Campus, United Kingdom
Open Targets Consortium
Core Bioinformatics team

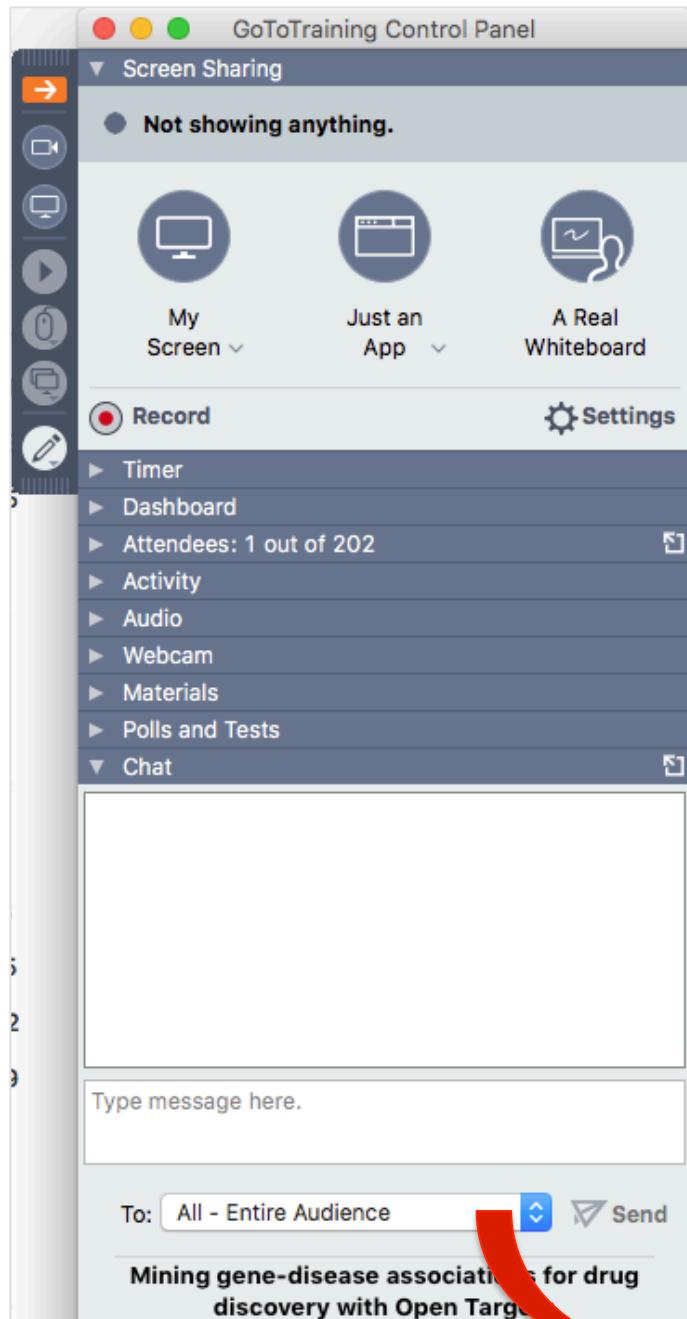


Open Targets

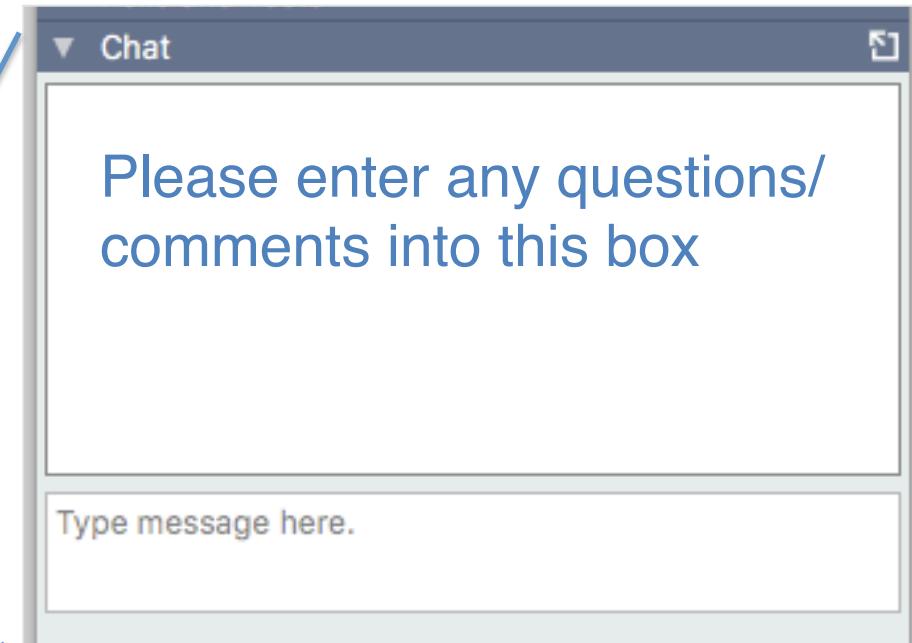


The logistics for the next hour

- Slides uploaded in the ‘materials’ section as presentation.pdf
- exercises.pdf: anytime, at your own pace
Email support@targetvalidation.org for further discussions
- All microphones muted
- Take questions at the end: look for the chat box!



Chat box



Address chat to 'entire audience'.

Burning question
during the webinar?



Outline

- Drug Discovery and its the challenges
- The Open Targets Consortium
- The Open Targets Platform
- Quick live demo
- Get in touch

Webinar's objectives

What is the Open Targets Consortium?

What is the Open Targets Platform?

How to navigate the Platform?

How to connect with the team



Open Targets

Drug discovery path: timeline

1. DISCOVERY



IDEA



BASIC RESEARCH

The majority of the research at this stage is publicly funded at universities, colleges and independent research institutions in every state.

2. DEVELOPMENT



CLINICAL TRIALS

Once a disease target is identified, drugs are designed and tested. Both public and privately funded research are involved.

PHASE I PHASE II PHASE III



REGULATORY APPROVAL

Human trials are completed. FDA approval. Industry is responsible for bringing a drug to market. Safety and evaluation continue after approvals.

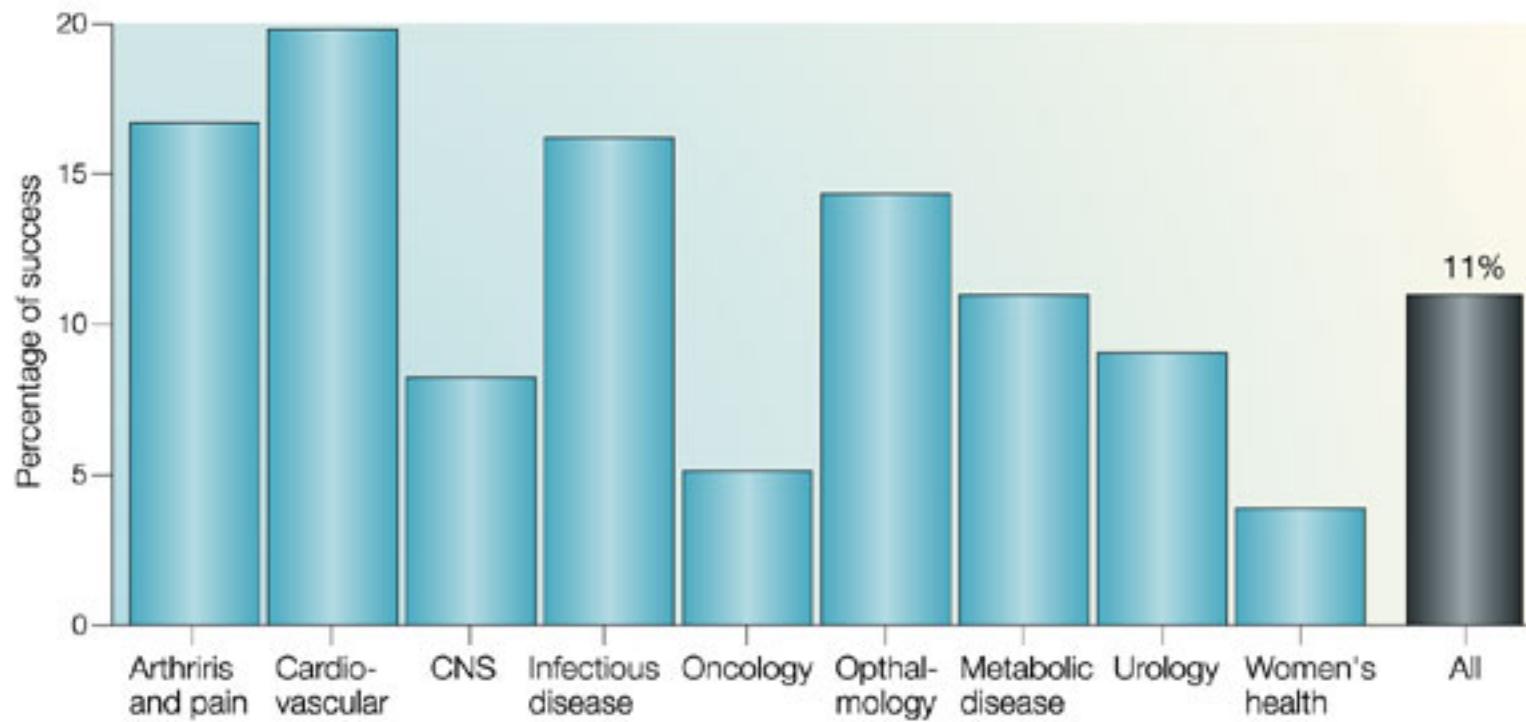


PATIENT CARE



Open Targets

Drug discovery: the challenges

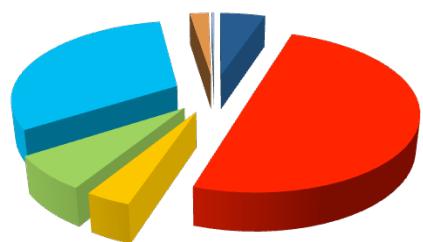


Lengthy, costly, low success rate, **HIGH ATTRITION RATES**

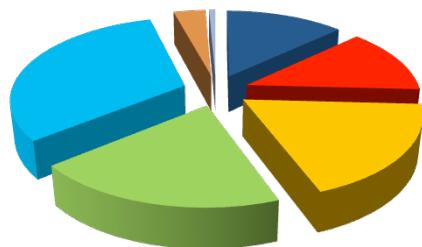
Source: doi:10.1038/nrd1470

What are the causes for the attrition?

Pre-clinical



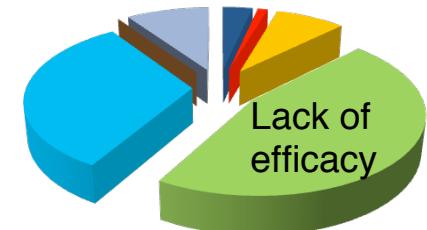
Phase I



Phase II



Phase III



- Pharmacokinetics/bioavailability
- Clinical safety
- Commercial
- Regulatory

- Non-clinical toxicology
- Efficacy
- Technical



*Professor Sir
Mike Stratton
Director, Sanger Institute*

Can we improve
target identification?



*Patrick Vallance, President
Pharmaceuticals R&D
GlaxoSmithKline*



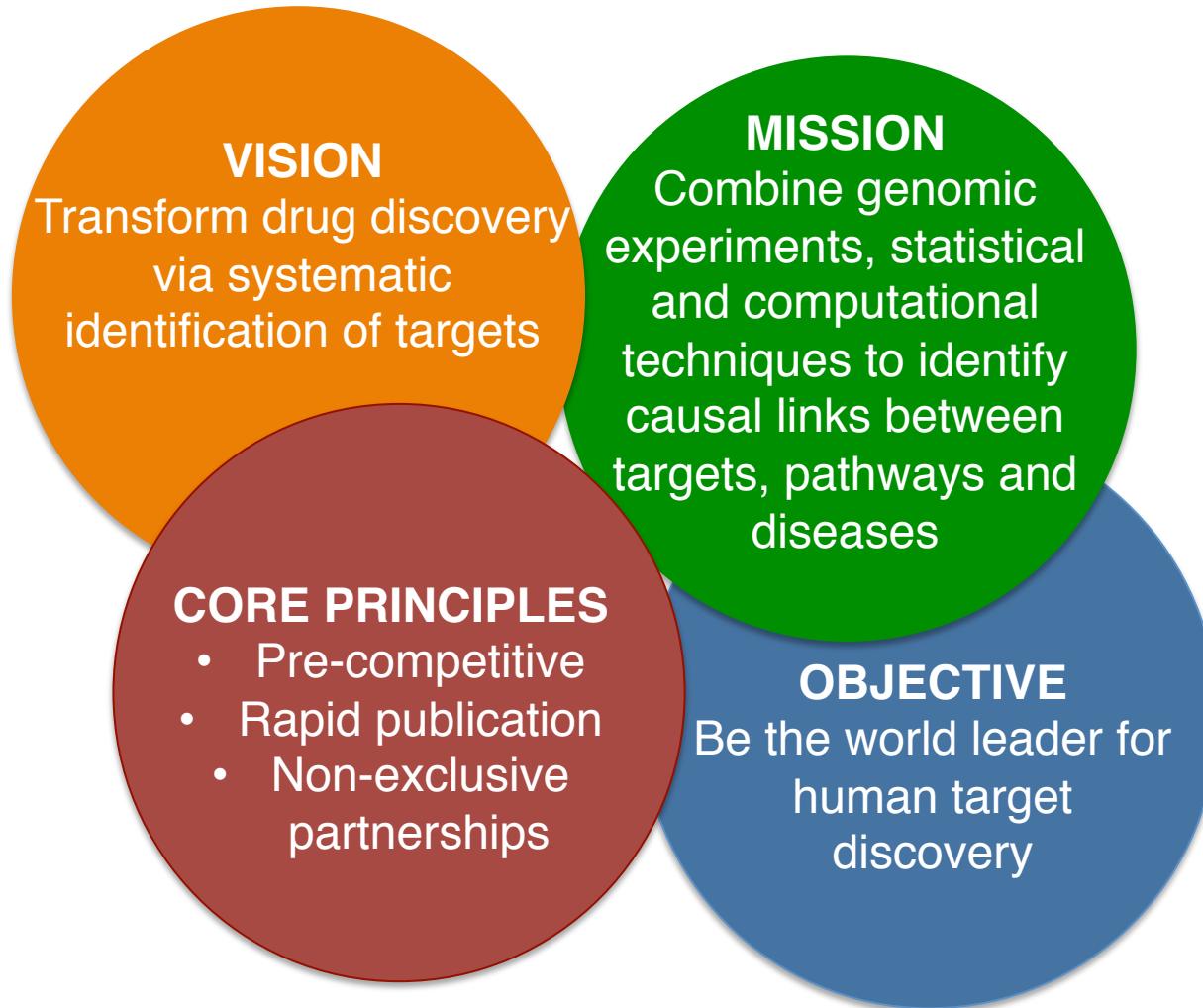
*Professor Dame
Janet Thornton
former Director, EMBL-EBI*

Yes, we can!
And we should.

But one institution
can not do it alone.



Open Targets Consortium*



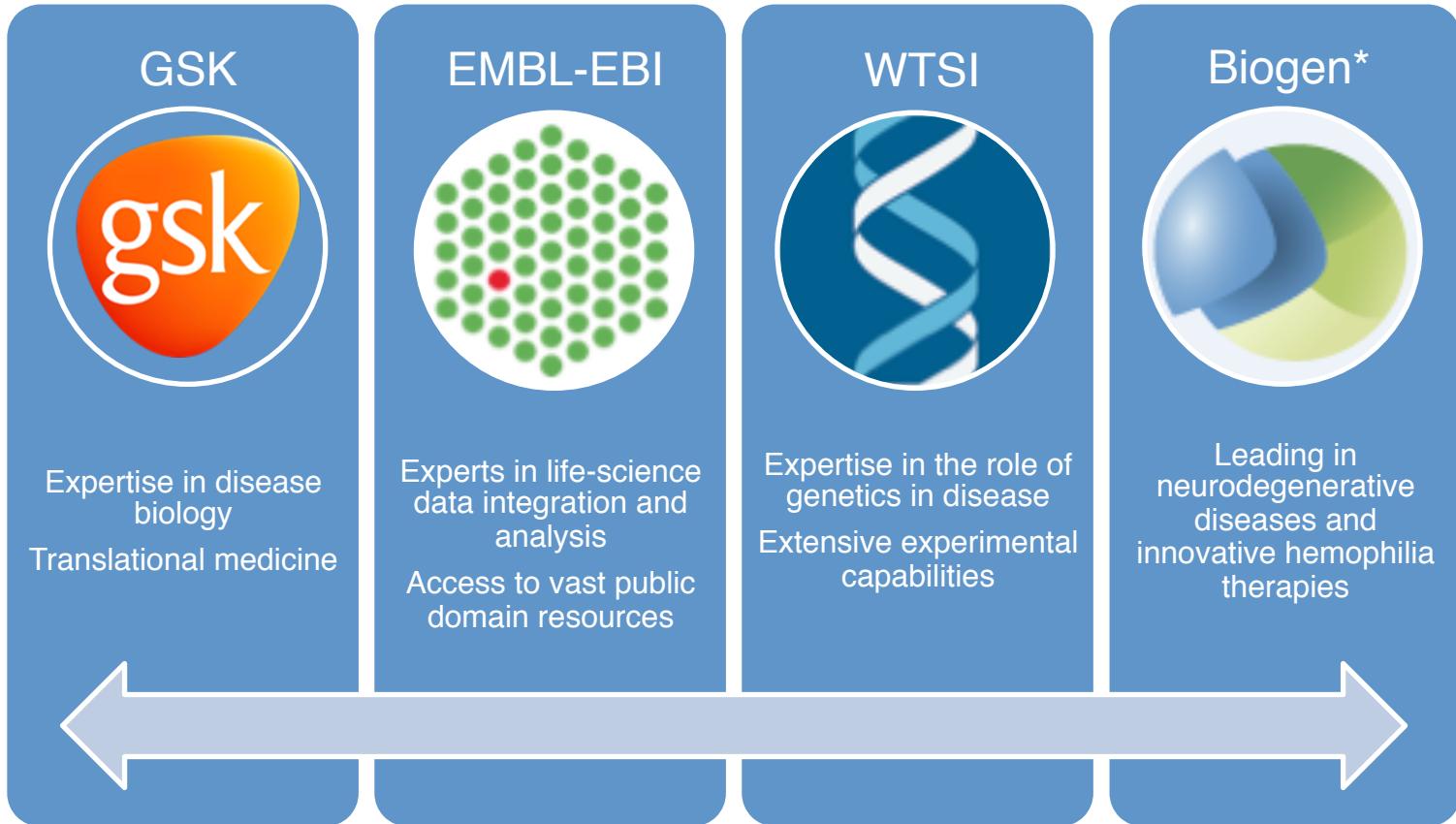
* Launched in March 2014
Three founding partners



EMBL-EBI



Who is Open Targets?



*Biogen joined the consortium in February 2016

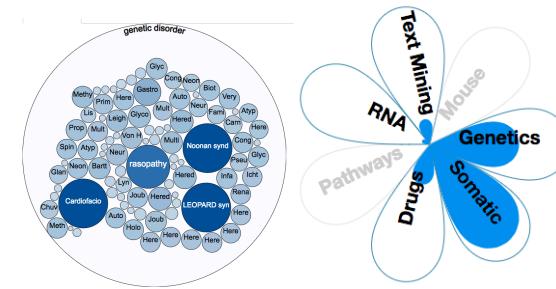
The two major areas of work* within Open Targets

Experimental projects



Generate new evidence
CRISPR
Organoids
Single cell RNASeq
Cell line fusion analyses
Metabolite GWAS

Core bioinformatics pipelines



Database for data integration
Web portal
REST API
Python client (fully supported)
R client (community)
Data dumps

* Concurrent
www.opentargets.org/projects

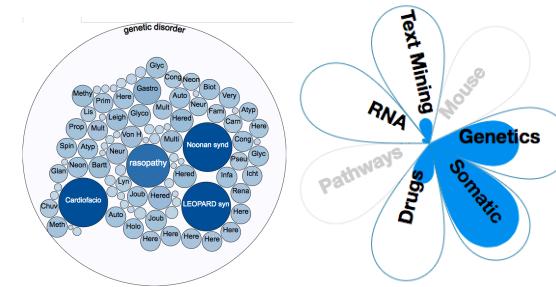
The two major areas of work* within Open Targets

Experimental projects



Generate new evidence
CRISPR
Organoids
Single cell RNASeq
Cell line fusion analyses
Metabolite GWAS

Core bioinformatics pipelines

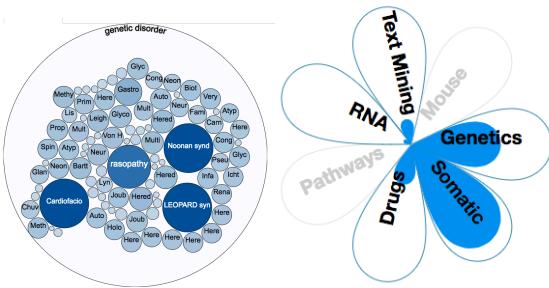


Database for data integration
Web portal
REST API
Python client (fully supported)
R client (community)
Data dumps

Open Targets Platform*

- Developed by the Core Bioinformatics team at EMBL-EBI
- Allow users to identify target–disease associations
- Improvements driven by you

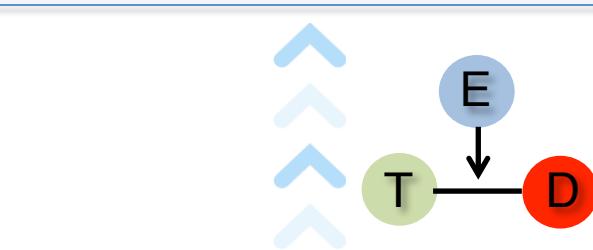
<https://www.targetvalidation.org/>



* First release: December 2015

Currently: Integration of existing data

Public Databases and Pipelines



Open Targets experimental data: NEW
Physiologically relevant and at scale

Oncology



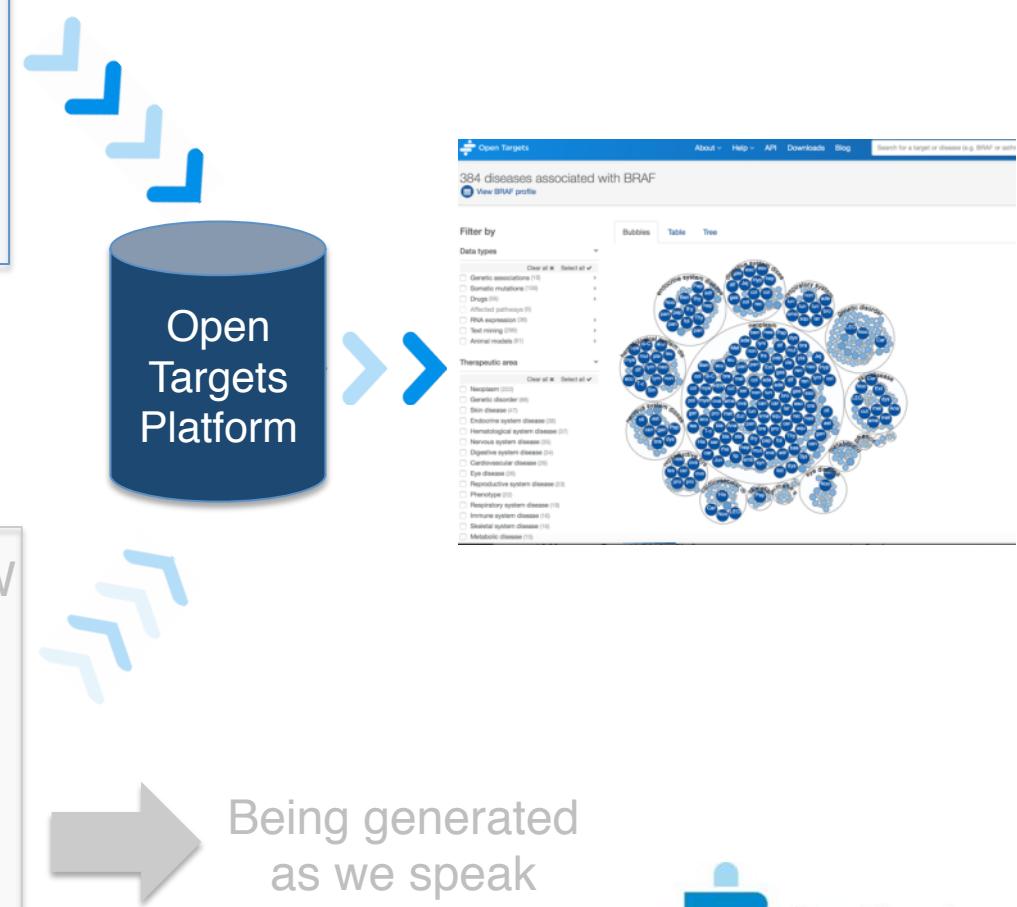
Immunity



Cross-Disease



Neuro



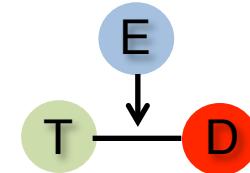
Being generated
as we speak



Open Targets

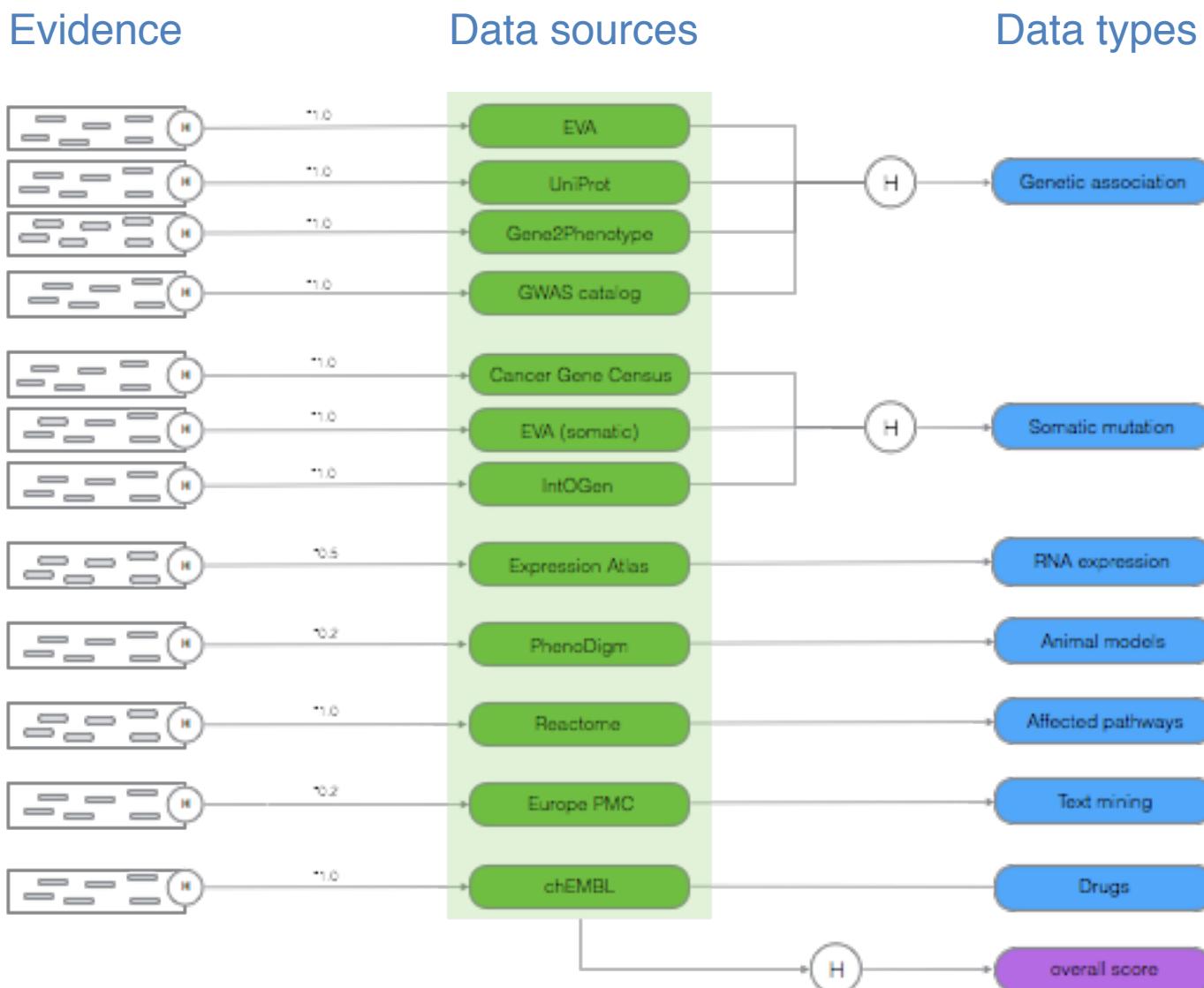
Evidence from publicly available data

- Similar data sources are grouped into data types

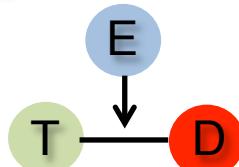


Data sources	Data types
GWAS catalog, UniProt, EVA, G2P	Genetic associations
Cancer Gene Census, EVA, IntOgen	Somatic mutations
Expression Atlas	RNA expression
ChEMBL	Drugs
Reactome	Affected pathways
Europe PMC	Text mining
PhenoDigm	Animal models
Your favourite data?	Let us know!

Confidence score



It's allow for replication and deflates the effect of large amounts of data



We support decision-making

A) Which targets are associated with a disease?

B) What evidence supports this target-disease association?

C) Are there FDA drugs for this association?

D) For a target, are there other diseases associated with it?

G) Can I find out about the mechanisms of the disease?

F) What else can I find out about my drug target?

E) If so, can I get associations for diseases from different therapeutic areas?



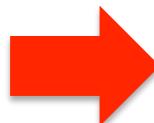
Which targets are associated with a disease?



Demo 1:

screenshots: next slides
coursebook: pages 9 - 15

The screenshot shows the homepage of the Open Targets Platform. At the top, there is a navigation bar with links for Survey, About, Help, API, Downloads, and Blog. Below the navigation is the platform's logo, "Open Targets Platform", and the tagline "Find new targets for drug discovery". A search bar contains the placeholder text "Search for a target or disease" with a magnifying glass icon. Below the search bar, there is a "Try:" section with links to BRAF, PTEN, Asthma, and Inflammatory bowel disease. On the far right, there are "Feedback" and "Follow us" buttons.



The screenshot shows the search results for "multiple sclero". The search bar at the top contains "multiple sclero". The first result is "multiple sclerosis" with "2697 targets associated". It is categorized as a "Disease". The description states: "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...". Below this, there is a "Targets" section listing "MBP myelin basic protein". Further down, there is a "Diseases" section listing "relapsing-remitting multiple sclerosis" and "autoimmune disease > multiple sclerosis > relapsing-remitting multiple ...". Another section lists "chronic progressive multiple sclerosis" and "autoimmune disease > multiple sclerosis > chronic progressive multiple...".

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

181 targets associated with multiple sclerosis

[View disease profile](#)

Filter the results

Filter by

Data types

- Clear all Select all
- Genetic associations (181)
 - GWAS catalog (178)
 - UniProt literature (3)
 - European Variation A... (1)
 - Somatic mutations (0)
 - Drugs (129)
 - Affected pathways (0)
 - RNA expression (1k)
 - Text mining (1k)
 - Animal models (4)

Pathway types

Target class

- Clear all Select all
- Enzyme (19)
 - Unclassified protein (8)
 - Membrane receptor (6)
 - Transcription factor (2)
 - Secreted protein (1)
 - Other membrane protein
 - Surface antigen (1)
 - Transporter (1)

Your target list

Choose File No file chosen

Upload your own list of targets

Data types
(Genetic Associations,
Drugs, etc)

Pathway types
(Metabolism,
Cell cycle)

Target class
(Receptor, Kinase)

CD24
IL12B
PTPRC
HLA-F
NR1H3
CYP27B1
IL7R
CD6
SP140

Showing 1 to 50 of 181 targets
Search:

Total number of targets associated with multiple sclerosis based on Genetic association only



Total number of targets associated with multiple sclerosis based on Genetic association only

association score
Genetic associations
Somatic mutations
Drugs
Affected pathways
RNA expression
Text mining
Animal models
Target name

TNFRSF1A
KCNB2
IL2RA
CD86
CD58
HLA-DQA1
TTC34
STAT3
CD24
IL12B
PTPRC
HLA-F
NR1H3
CYP27B1
IL7R
CD6
SP140
TNF receptor superfamily ...
potassium voltage-gated ...
interleukin 2 receptor sub...
CD86 molecule
CD58 molecule
butyrophilin like 2
C-type lectin domain fami...
major histocompatibility c...
major histocompatibility c...
T-cell activation RhoGTPa...
ecotropic viral integration ...
prostaglandin E receptor 4
major histocompatibility c...
tetrastricopeptide repeat d...
signal transducer and acti...
tumor necrosis factor sup...
Abelson helper integration...
C-X9-C motif containing 1
major histocompatibility c...

[http://www.targetvalidation.org/
disease/EFO_0003885/associations?
fcts=datatype:genetic_association](http://www.targetvalidation.org/disease/EFO_0003885/associations?fcts=datatype:genetic_association)



http://www.targetvalidation.org/disease/EFO_0003885

2697 targets associated with multiple sclerosis



[View disease profile](#)

Phenotypes

Drugs

Disease Classification

Drugs

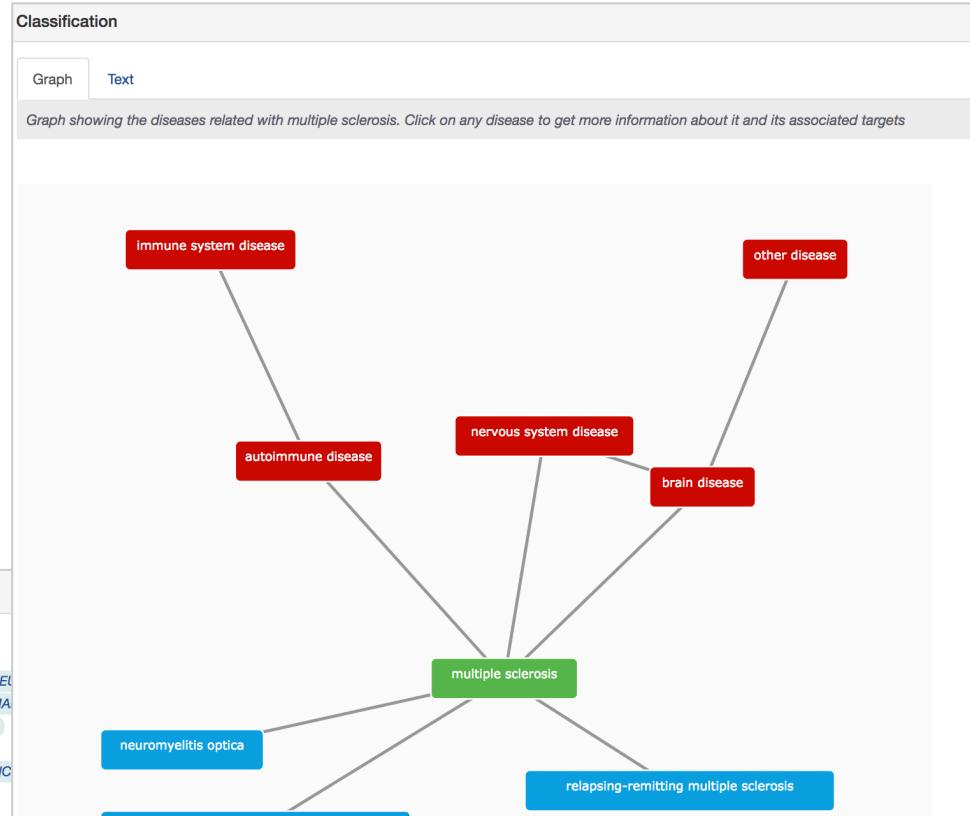
Source: ChEMBL

Found 40 unique drugs: ABATACEPT ALEMKTUZUMAB BACLOFEN BOTULINUM TOXIN TYPE A PURIFIED NEUROPEPTIDE Y CORTICOTROPIN DEXAMETHASONE DEXTRAMPHETAMINE DICLOFENAC DIMETHYL FUMARATE FINGOLIMOD GS1101 HYDROXYUREA INTERFERON BETA-1A INTERFERON BETA-1B LAMOTRIGINE METHYLPHENIDATE METHYLPREDNISOLONE MIRABEGRON MITOXANTRONE MYCOPHENOLATE MOFETIL OCRELIZUMAB OXCARBAZEPINE PEGINTERFERON BETA-1A PREDNISOLONE PREDNISOLONE PHOSPHORIC ACID RITUXIMAB SIMVASTATIN Siponimod TERIFLUONIMIDE

Showing 1 to 10 of 1,000 entries

Search:

Disease	Drug Information			Mechanism of action	Activity	Gene-Drug Evidence	
	Drug	Phase	Status			Target class	Evidence source
relapsing-remitting multiple sclerosis	INTERFERON BETA-1A	Phase IV	Completed	Protein Interferon alpha/beta receptor agonist DailyMed	agonist Membrane receptor	Curated from Clinical Trials Information	



Open Targets

Demo 2: Evidence supporting the *CD86* - multiple sclerosis association

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

Choose your favourite internet browser*

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

Demo 3: your list of targets for a disease

These are some of the genes associated with Alzheimer's disease from the medical literature: *HFE*, *PSEN1*, *TF*, *APOE*, *ADRB2*, *PSEN2*, and *A2M*.

- Which of these have the strongest association w/ Alzheimer's?
- Is there any membrane receptors in this list?
- If so, can you find which amino acids of this receptor (putative drug target) correspond to the extracellular domain (s) of the protein?

Exercises

Pages 25-27, 33-34

*Learn at a pace that works for you.
You can get in touch at any time!*



support@targetvalidation.org

Wrap up

Open Targets Platform:

For drug target ID and selection in drug discovery

Rank target-disease associations: disease prioritization

Integrated information on targets and diseases

Intuitive graphical interface

Oh Yes!
And all is 100% free
and open source



Open Targets

Alternative ways to access the data

Looking for our entire datasets?

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

- All target-disease associations: 215 MB
- All evidence: 4.35 GB

Looking for extracts of our datasets?

- API: REST calls, Python client
- R client: maintained by the community



Open Targets

How to cite us

Published online 8 December 2016

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

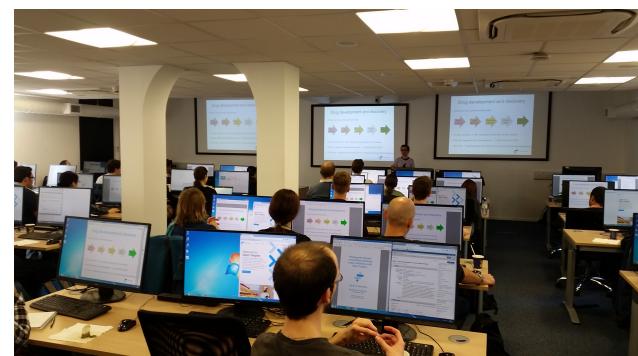
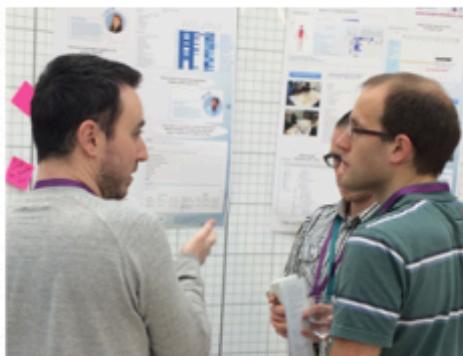
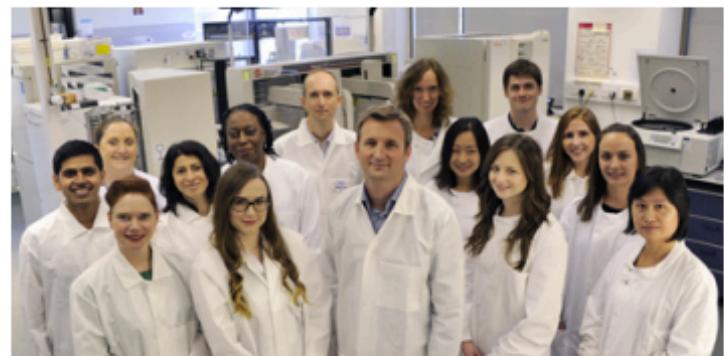
Open Targets: a platform for therapeutic target identification and validation

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

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

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

Acknowledgements



support@targetvalidation.org



Open Targets

Get in touch



@targetvalidate



support@targetvalidation.org



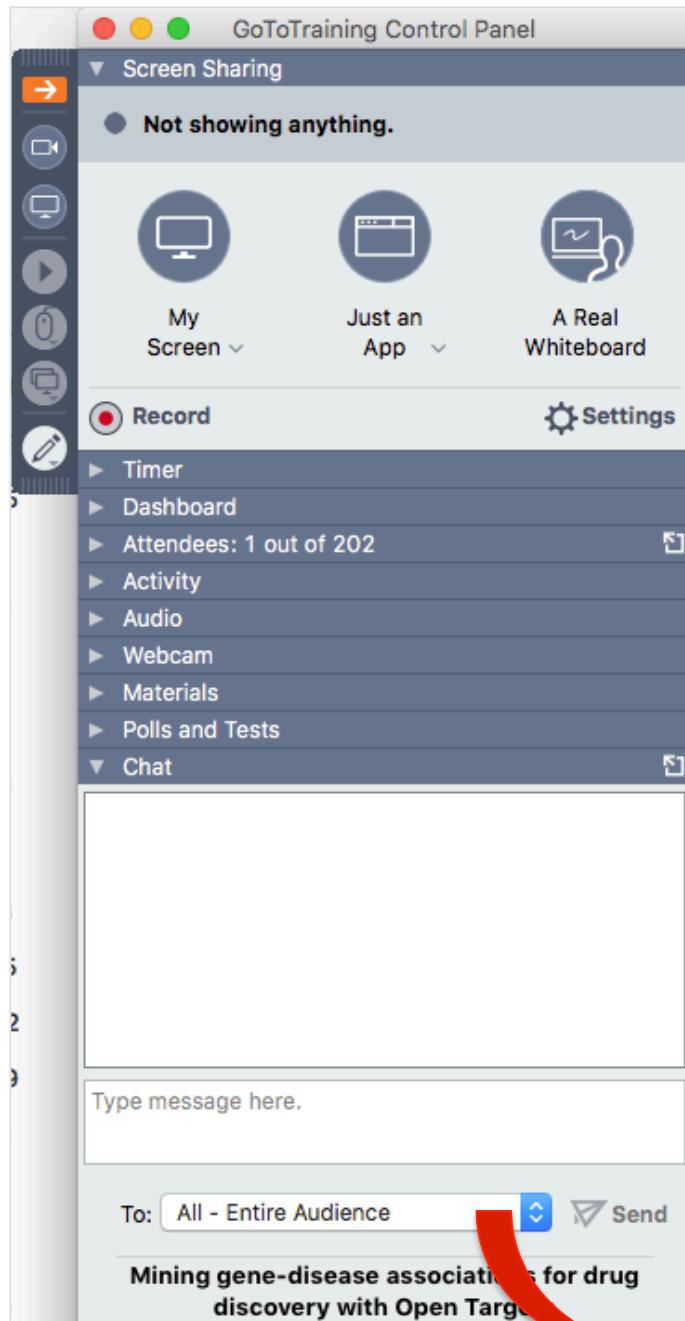
www.facebook.com/OpenTargets/



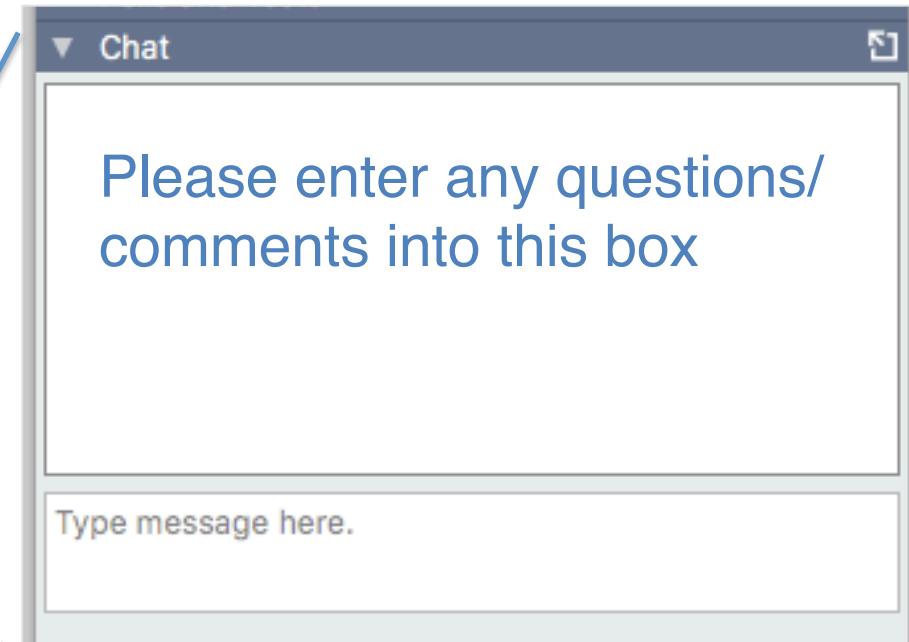
blog.opentargets.org/



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



Chat box



Address chat to 'entire audience'.



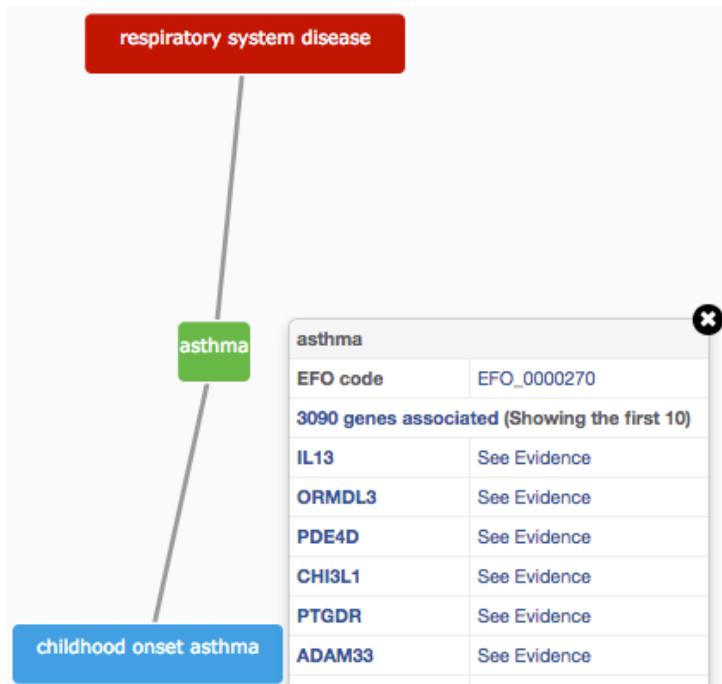
Extra Extra Extra

How do we associate diseases and phenotypes w/ targets?

- 1 ChEMBL, UniProt, EVA (w/ ClinVar) curate diseases and phenotypes
- 2 Map disease/phenotypes to an ontology using EFO and HPO terms
- 3 Use genes as proxies for our targets
- 4 Create target-disease evidence JSON objects
- 5 Calculate for each supporting evidence the likelihood of gene A being associated with disease B
- 6 Compute integrated target-disease scores at the levels of data source, data type and overall score

Experimental Factor Ontology* (EFO)

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



controlled vocabulary
+
hierarchy (relationship)

* <https://www.ebi.ac.uk/efo/>

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

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

Statistical integration, aggregation and scoring*

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

*https://github.com/opentargets/association_score_methods

Factors affecting the relative strength of an evidence

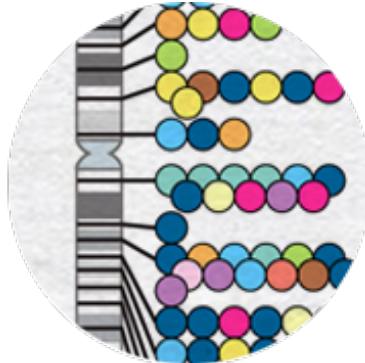
e.g. *GWAS Catalog*

$$S = f * s * c$$

f, relative occurrence of a target-disease evidence

s, strength of the effect described by the evidence

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



f= sample size (cases versus controls)

s = predicted functional consequence

c = *p*-value reported in the paper

Aggregating scores across the data

- Using a mathematical function, the harmonic sum*

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

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

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

* PMID: 19107201, PMID: 20118918

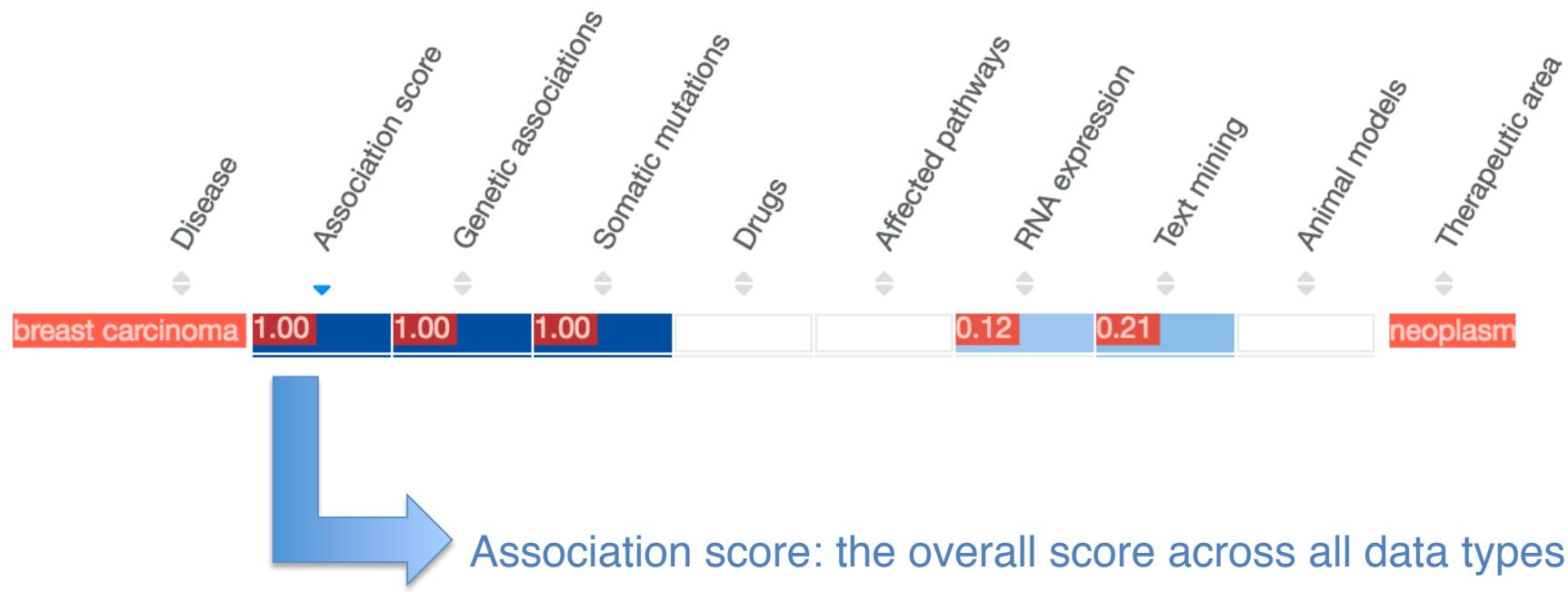
Disclaimer: score, dos and don'ts

- 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



- Based on the data sources
- Different weight applied:

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

REST API endpoints



public : Publicly supported stable API.

Open/Hide | List operations | Expand operations

GET /public/evidence

POST /public/evidence

GET /public/evidence/filter

POST /public/evidence/filter

GET /public/association

GET /public/association/filter

POST /public/association/filter

GET /public/search

GET /public/auth/request_token

GET /public/auth/validate_token

GET /public/utils/ping

GET /public/utils/version

GET /public/utils/stats

- Query association and evidence by gene identifiers and diseases
- Filter by type of evidence

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

GET

/public/association



Implementation notes

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

Parameters

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

Response messages

HTTP status code	Reason	Model
200	Successful response	

[Try it out!](#)[Hide response](#)

Request URL

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



Response body

```
{  
  "from": 0,  
  "facets": null,  
  "took": 6,  
  "therapeutic_areas": [],  
  "total": 1,  
  "data": [  
    {  
      "target": {  
        "gene_info": {  
          "symbol": "PTGS2",  
          "ensembl_id": "ENSG00000073756",  
          "name": "PTGS2",  
          "chromosome": 12, "start": 123456789, "end": 123456789},  
        "evidence": [{"source": "Ensembl", "score": 100, "type": "Gene-Disease"}, {"source": "OMIM", "score": 80, "type": "Gene-Disease"}],  
        "summary": "PTGS2 is associated with various diseases, including cardiovascular diseases and metabolic disorders."},  
      "disease": {  
        "name": "Cardiovascular Disease",  
        "efo_id": "EFO_0003767",  
        "description": "A group of diseases that affect the heart and blood vessels."},  
        "summary": "The association score for PTGS2 and Cardiovascular Disease is approximately 0.85."},  
      "score": 0.85  
    }  
  ]  
}
```

- Paste the URL in a location bar in a browser
- Use the terminal window (e.g. with CURL)
- Use one of our clients (i.e. R and Python)

Python and R clients for the REST API

opentargets
latest

Search docs

Tutorial
High Level API
Low Level API
Code Documentation
Changelog

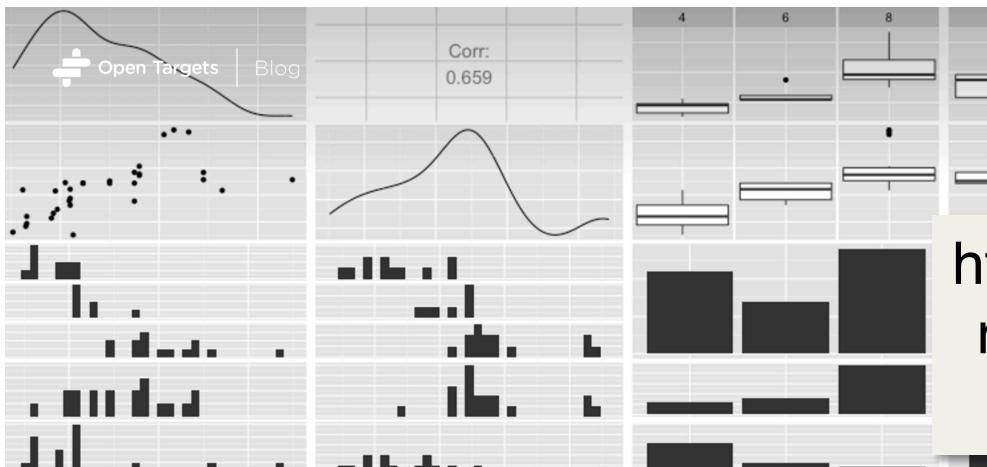
Docs » opentargets - Python client for targetvalidation.org

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