

Mining Gene-Disease Associations with Open Targets

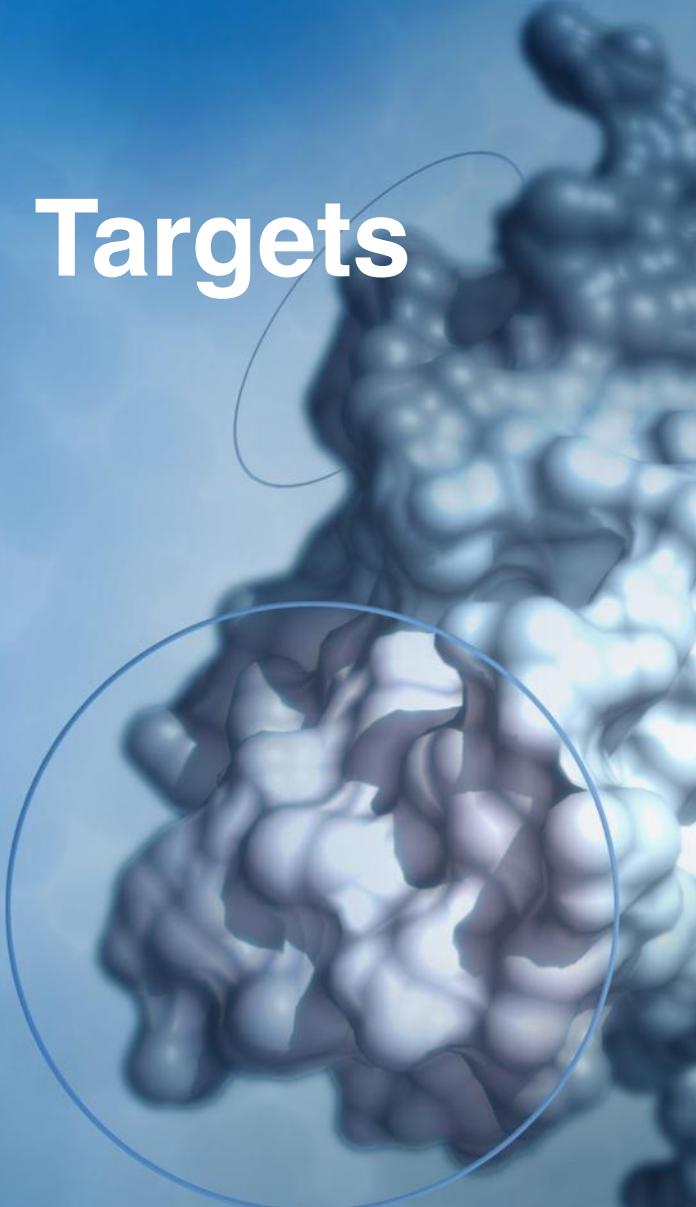
Part of “Bioinformatics for Discovery”

Denise Carvalho-Silva

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



Open Targets



Today 13:30-14:30

- Open Targets Platform:

Live demos

Hands-on exercises

- Course wrap up

Objectives for this hour session

What is the Open Targets Consortium?

What is the Open Targets Platform?

How to **navigate** the Platform?

How to **connect** with us



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?

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 of the protein?

Exercises

Pages 25-27, 33-34

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

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

<http://nar.oxfordjournals.org/content/45/D1/D985.long>



...And the 2017 Breakthrough Articles Award goes to...

#OpenTargets, as well as @MonarchInit & @denovodb! @NAR_Open buff.ly/2iGMXlc

<http://www.narbreakthrough.com/>



RETWEETS
10

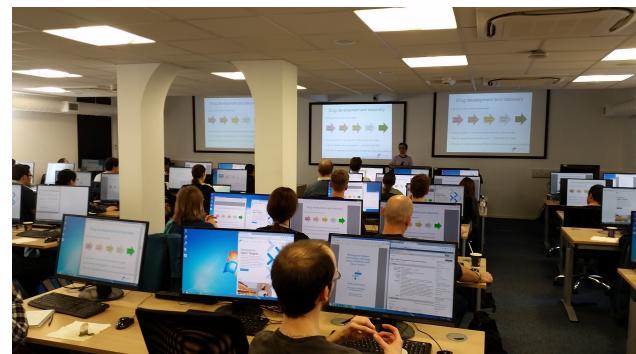
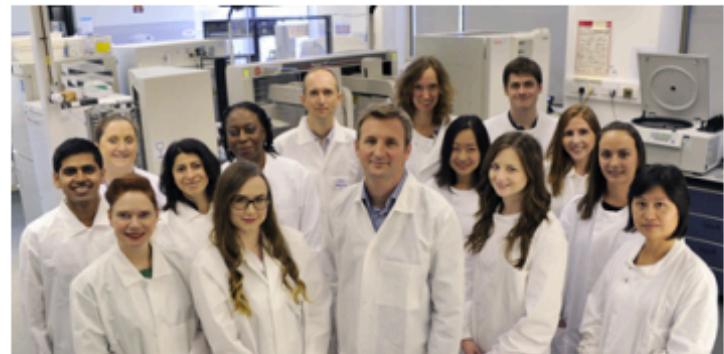
LIKES
8





- Search for many genes at once
- Search for PubMed, SNPs IDs, pathways
- Provide D-D and T-T associations
- Workshops ARUK (Cambridge) and DDU (Dundee)

Acknowledgements



support@targetvalidation.org

Get in touch



@targetvalidate



support@targetvalidation.org



www.facebook.com/OpenTargets/



blog.opentargets.org/



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

Bioinformatics forums (fora)



LATEST OPEN RNA-SEQ CHIP-SEQ SNP ASSEMBLY TUTORIALS TOOLS JOBS

Denise - Open Targets • 3.6k | Logout

Community Messages Votes My Posts My Tags Following

Biostars

BIOINFORMATICS EXPLAINED

Tutorial: Open Targets and programmatic access



If you are working in disease biology, drug discovery and/or validation, the chances are you have seen our previous post on the [Open Targets](#) project.

8

Perhaps you have even used our [Open Targets Platform](#) to search for target-disease associations on a gene by gene (or disease by disease) basis.



But if you are a bioinformatician working in disease biology, drug discovery and/or drug validation, the chances are you want to access and retrieve data on several genes or several diseases. All at the same time. To do this, you would access Open Targets in a programmatic way, wouldn't you?

Our public REST ([REpresentation State Transfer](#)) API is what you would be looking for. Easy access to our data via one of the following:

- Internet browser (using HTTPS)
- Command line (e.g. CURL)
- With any programming language (e.g. Python, R)

Note : Not a bioinformatician? Not to worry! You don't need to be one to access larger chunks of data from our API. If you are web-lab scientists with no programming skills, you can give the API a go too, using the internet browser (see below).

But how would you go about doing that?

- With a browser, you can simply copy and paste the URL in the URL bar in your favourite internet browser (e.g. Firefox, Google Chrome, Safari).



5 months ago by
[Denise - Open Targets](#)
• 3.6k
UK, Hinxton, EMBL-EBI

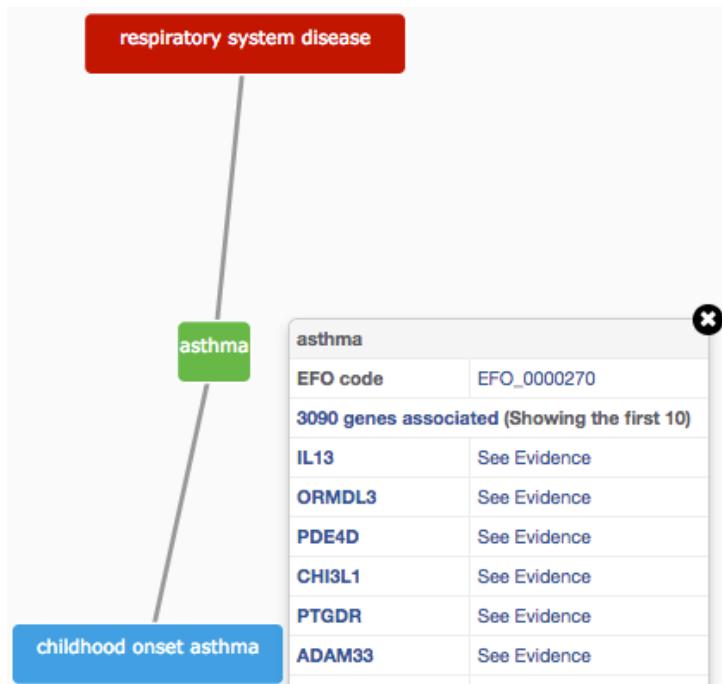
Extra Extra Extra

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

Factors affecting the relative strength of an evidence

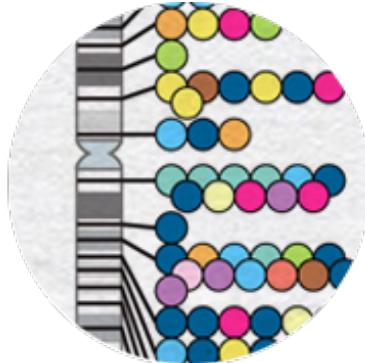
e.g. *GWAS Catalog*

$$S = f * s * c$$

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

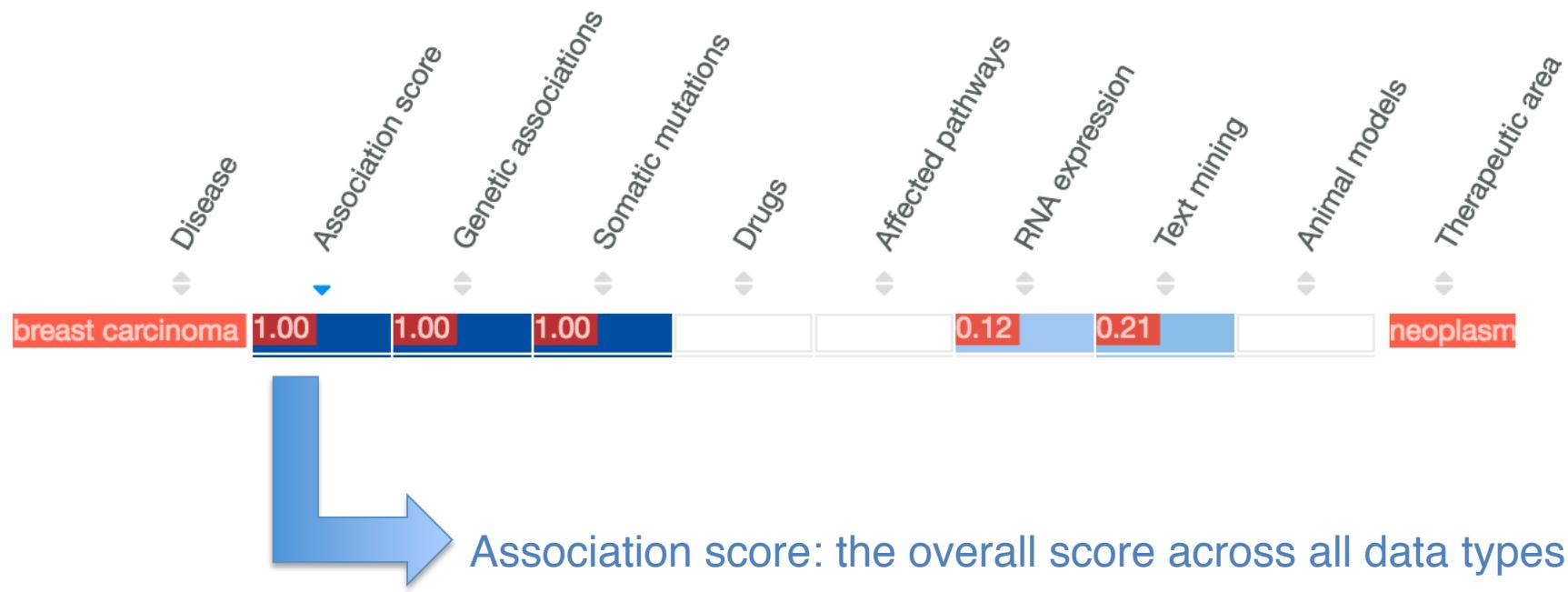
Disclaimer: score, dos and don'ts

- 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

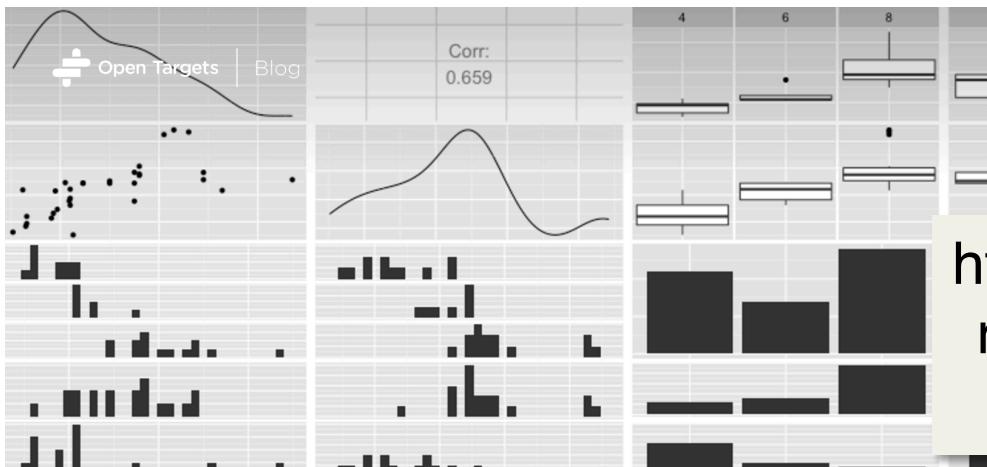
Docs » opentargets - Python client for targetvalidation.org

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