Mining Gene-Disease Associations with Open Targets

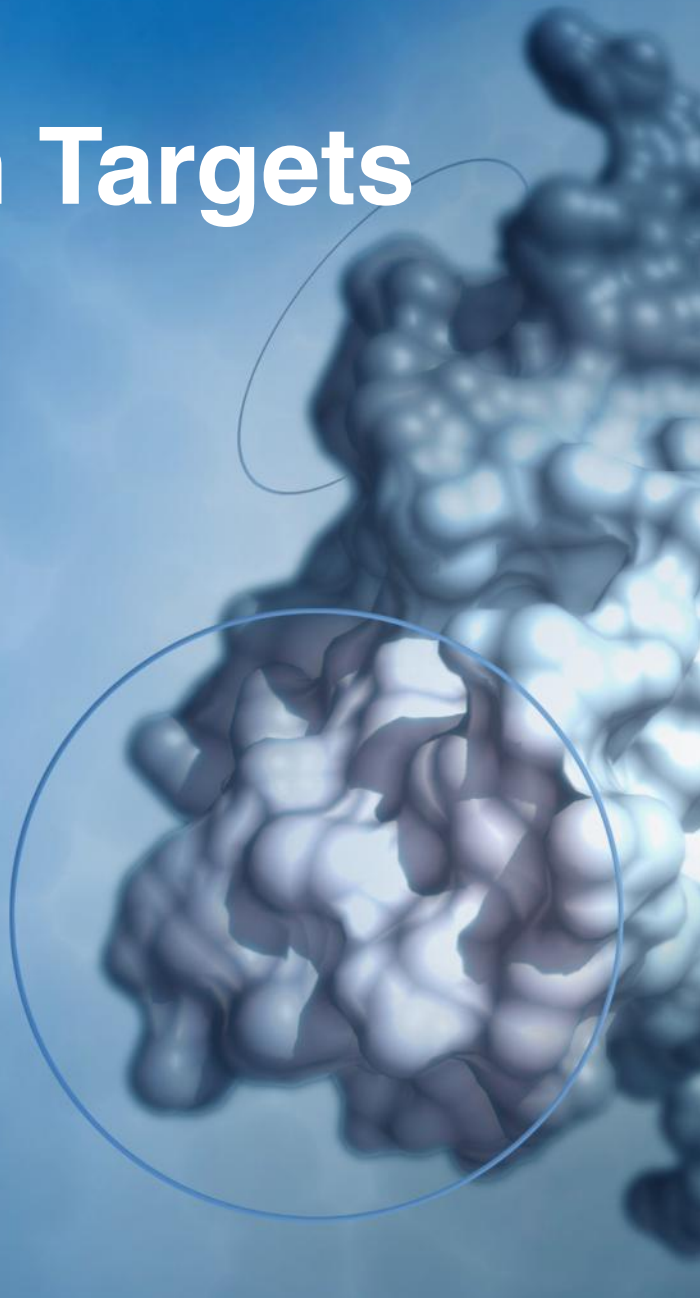
Webinar

Yale School of Medicine and friends

Denise Carvalho-Silva

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

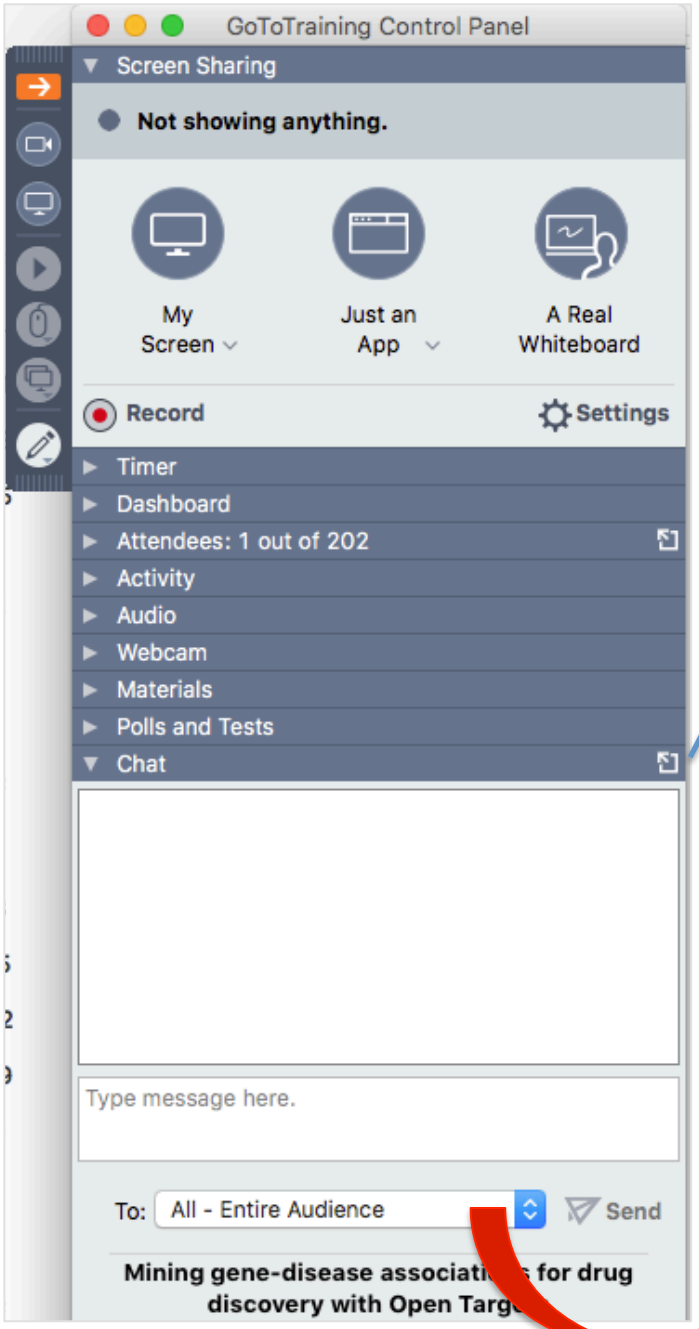
Open Targets

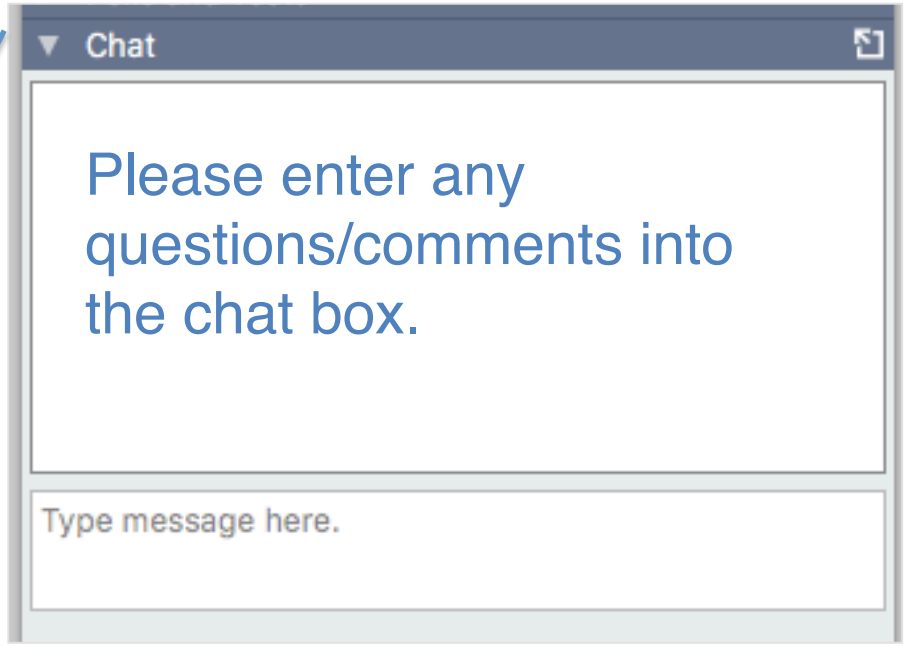


The logistics for the next hour

- Presentation is uploaded in the ‘materials’ section as slides.pdf
- Exercises: anytime, at your own pace (coursebook.pdf)
  Email support@targetvalidation.org for further discussions
- I’ll mute all microphones and take questions at the end
- Look out for the chat box

Chat box





Address chat to ‘entire audience’.

Outline

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

Open Targets

Webinar's objectives

What is the **Open Targets Consortium**?

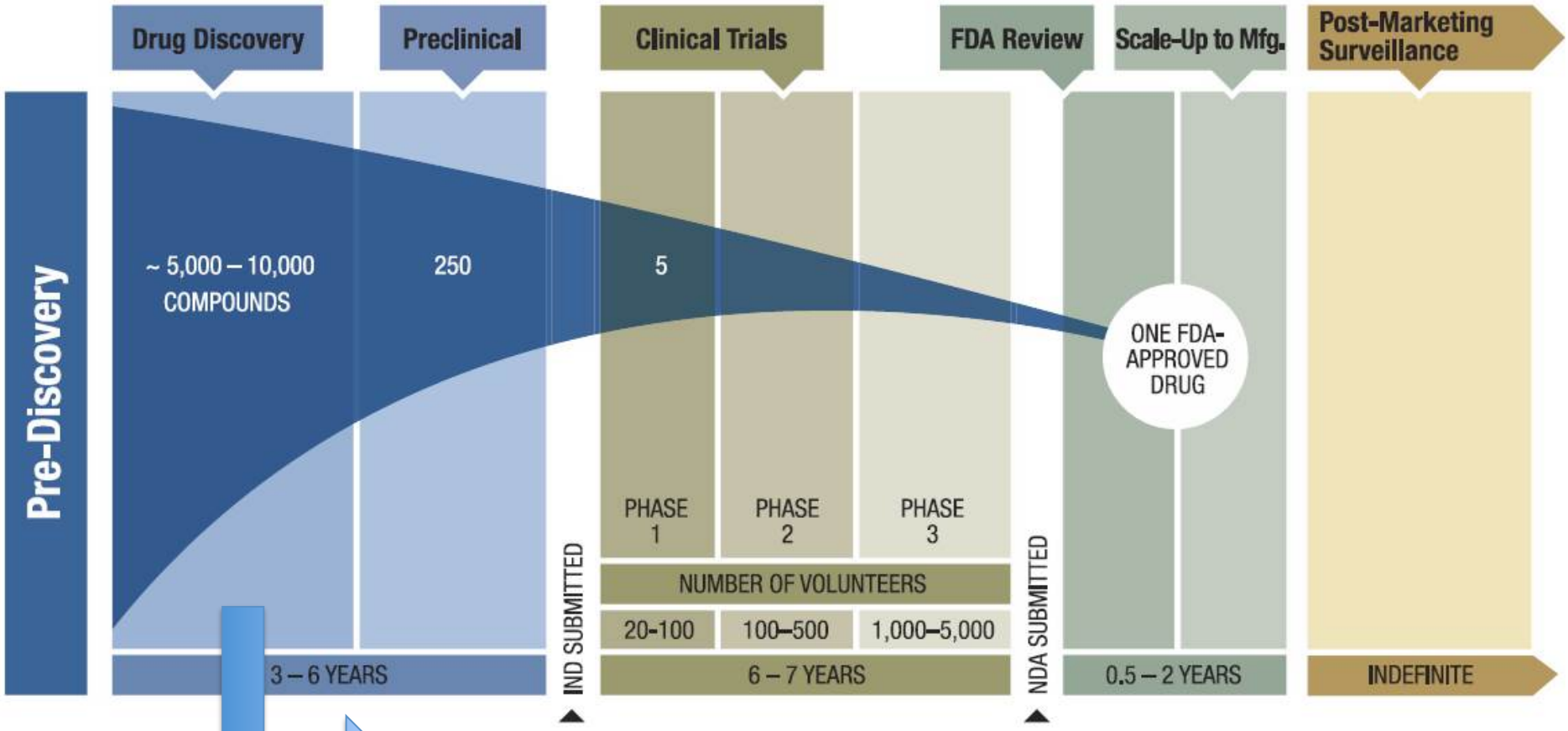
What is the **Open Targets Platform**?

How to **navigate** the Platform?

How to **connect** with us

Open Targets

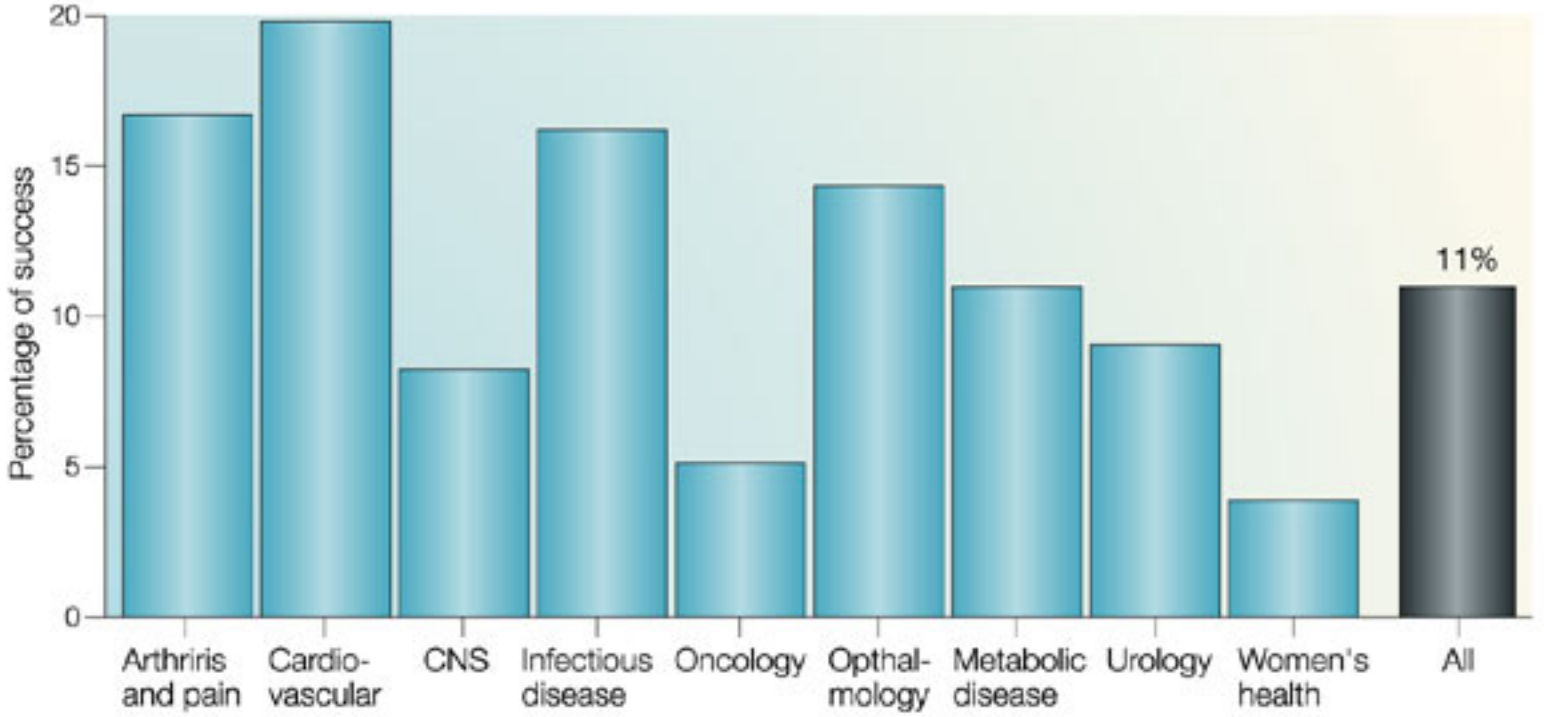
Drug discovery path: timeline



Target identification, assay development, lead generation

Source: http://www.focr.org/sites/default/files/feeds/AACR%20CPR.jpg

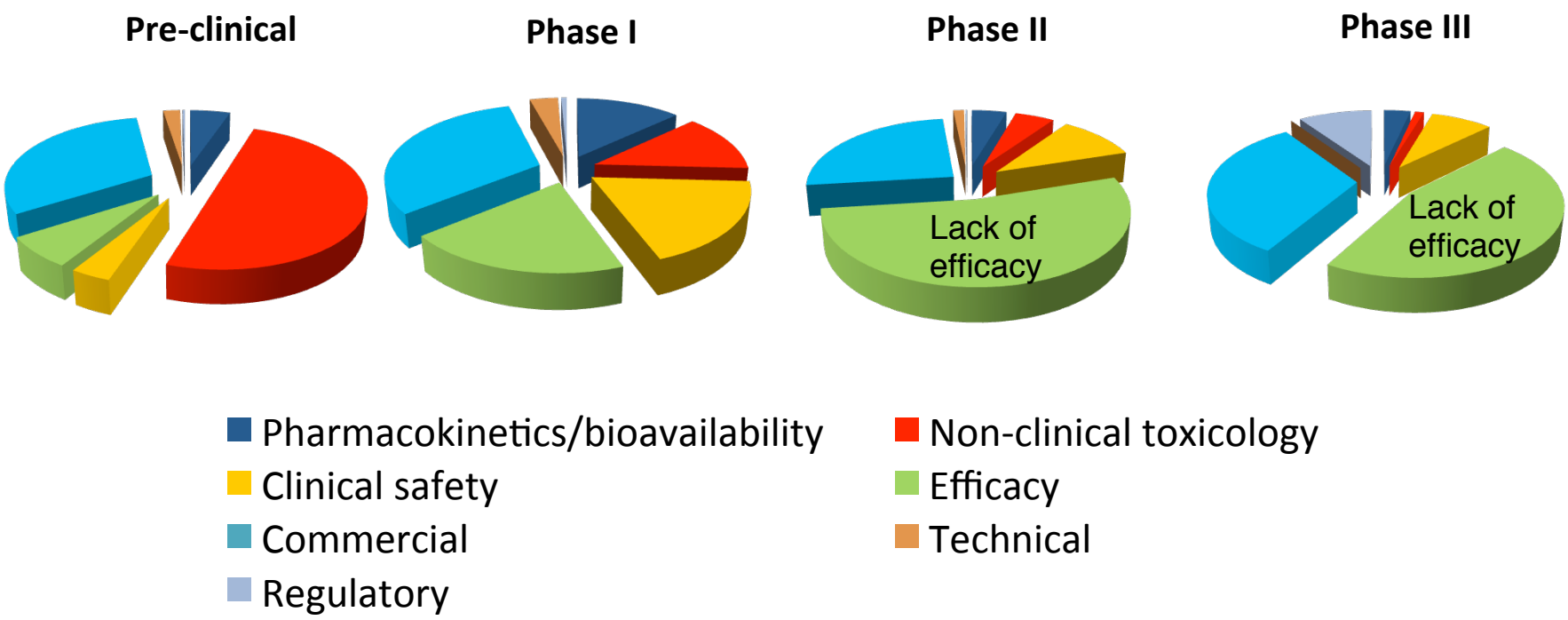
Drug discovery: the challenges



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

Source: doi:10.1038/nrd1470

What are the causes for the attrition?



THOMSON REUTERS

Professor Sir
Mike Stratton
Director, Sanger Institute

Can we improve target identification?

Yes, we can!
And we should.

Patrick Vallance, President
Pharmaceuticals R&D
GlaxoSmithKline

Professor Dame
Janet Thornton
former Director, EMBL-EBI

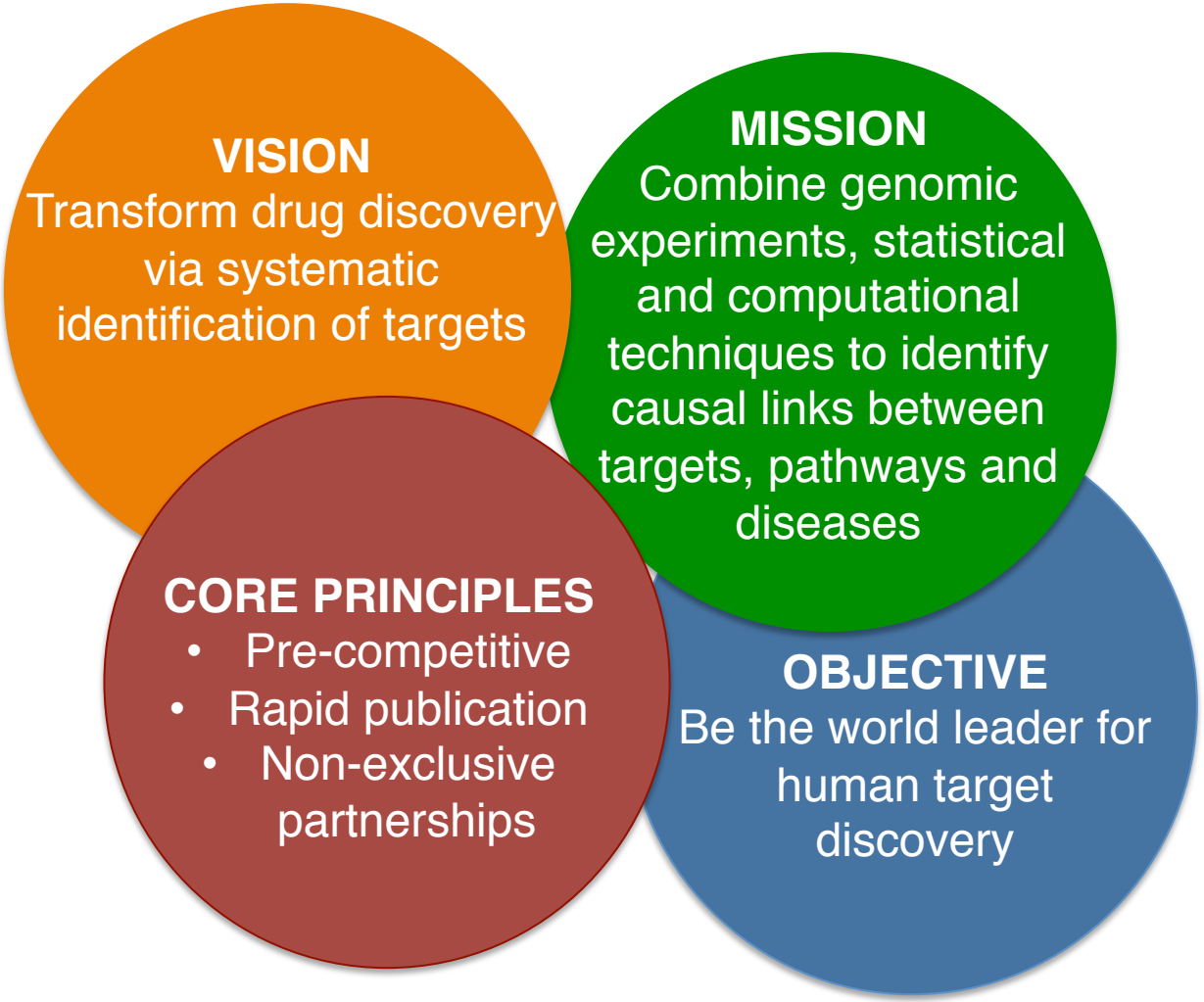
But one institution can not do it alone.

gsk

EMBL-EBI

wellcome trust sanger institute

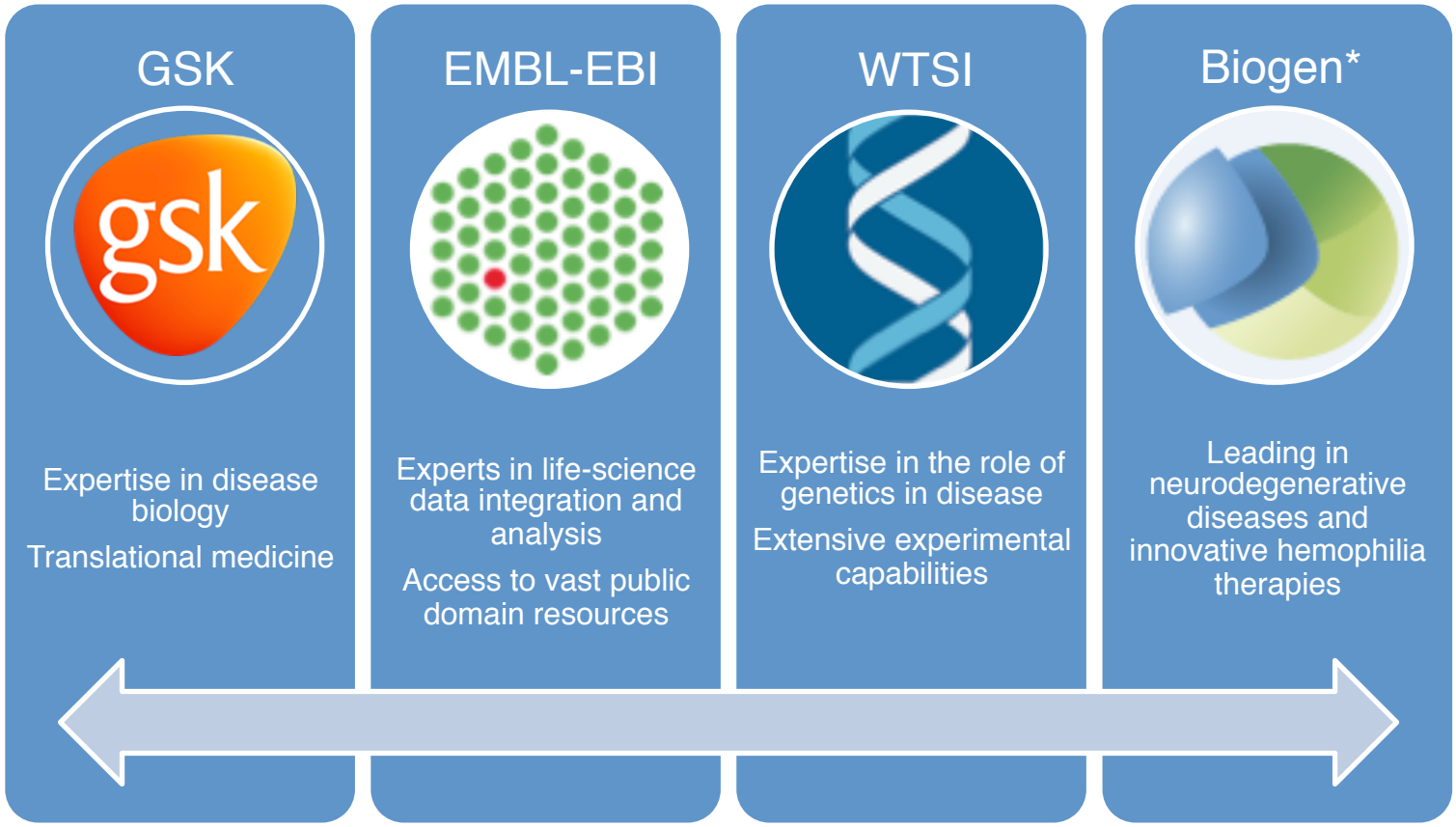
Open Targets Consortium*



* Launched in March 2014
Three founding partners

gsk wellcome trust sanger institute EMBL-EBI

Who is Open Targets?



*Biogen joined the consortium in February 2016

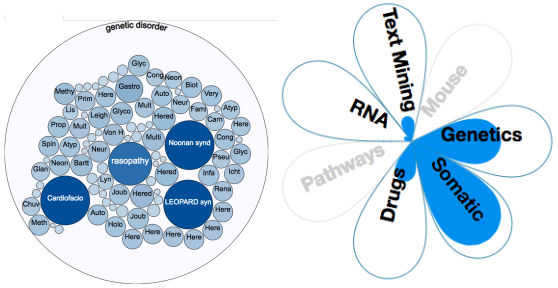
Open Targets

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



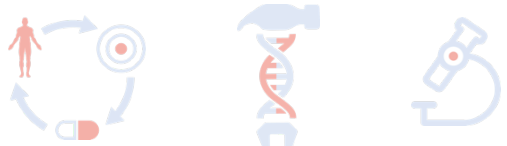
Database for data integration
Web portal
REST API
Python client (fully supported)
R client (community)
Data dumps

* Concurrent
www.opentargets.org/projects

Open Targets

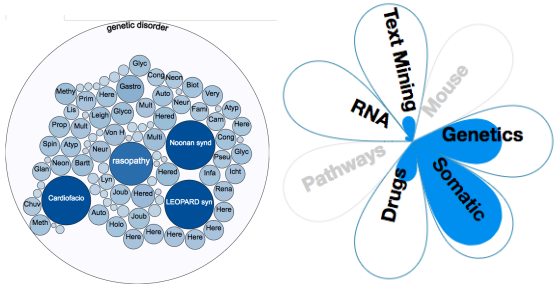
The two major areas of work* within Open Targets

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

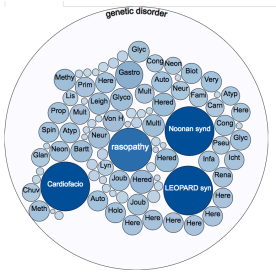
www.opentargets.org/projects

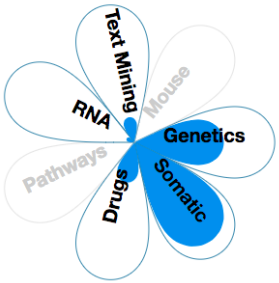
Open Targets

Open Targets Platform*

- Developed by the Core Bioinformatics team at EMBL-EBI
- Allow scientists to identify target–disease associations
- Frequent updates: new data, new web features
- Improvements driven by our user communities

https://www.targetvalidation.org/





* First release: December 2015

Open Targets

Currently: Integration of existing data

Public Databases and Pipelines

ChEMBL
UniProt
Europe PMC
REACTOME
IntAct
intOgen Integrative Onco Genomics
G2P

E
T
D

Open Targets Platform

www.targetvalidation.org

Graphical user interface

Open Targets experimental data: NEW
Physiologically relevant and at scale

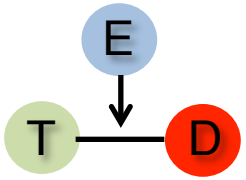
Oncology
Immunity
Neuro
Cross-Disease

Being generated as we speak.

Open Targets

Evidence from publicly available data

- Similar data sources are grouped into data types

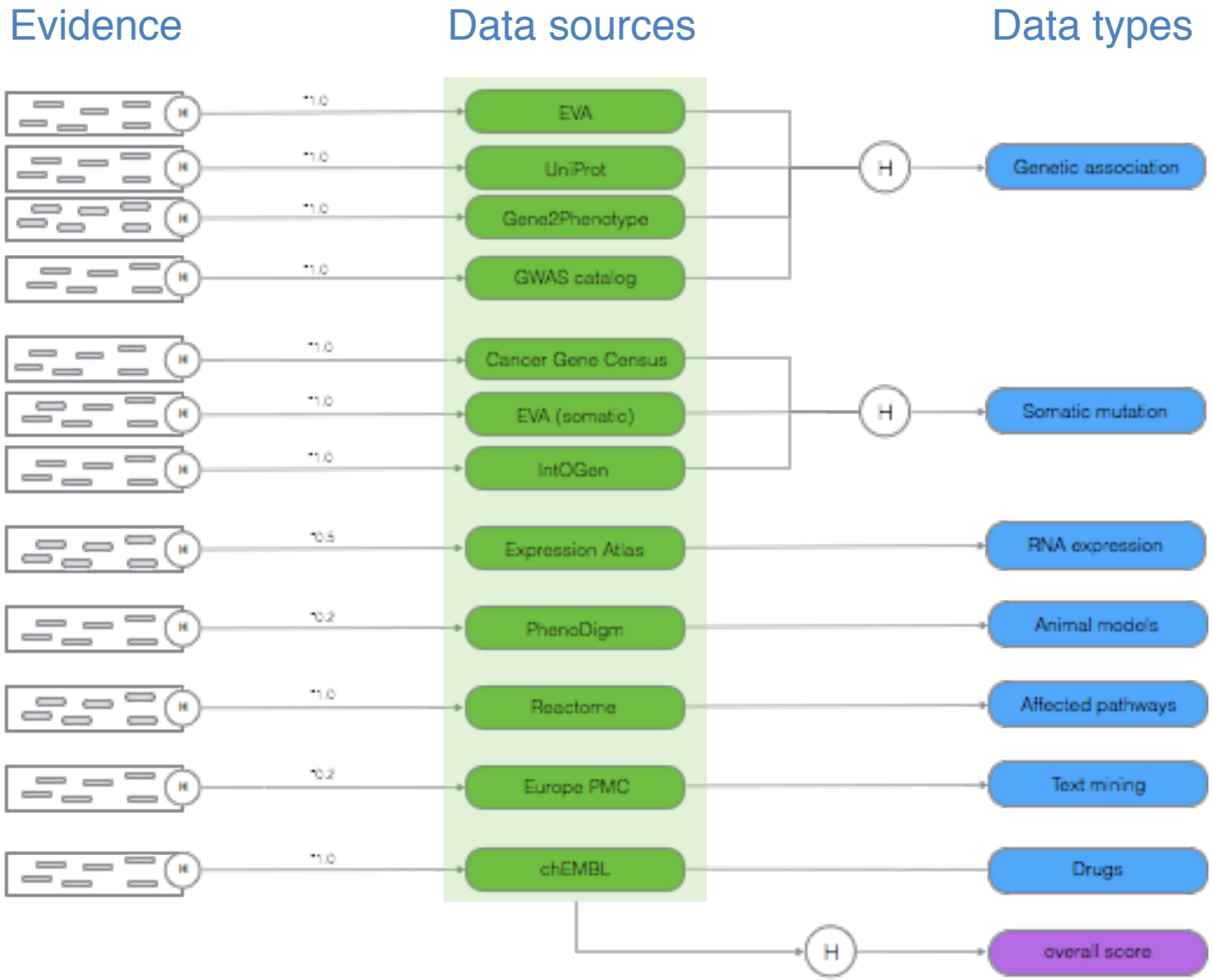


Data sources	Data types
GWAS catalog, UniProt, EVA, G2P	Genetic associations
Cancer Gene Census, EVA, IntOgen	Somatic mutations
Expression Atlas	RNA expression
ChEMBL	Drugs
Reactome	Affected pathways
Europe PMC	Text mining
PhenoDigm	Animal models
Your favourite data?	**Let us know!**

Open Targets

Confidence score for the T-D associations

Oliver Stegle's team (EMBL-EBI)



where H is a harmonic series defined by $S_{1..i} = S_1 + \frac{S_2}{2^2} + \frac{S_3}{3^2} + \frac{S_4}{4^2} .... + \frac{S_i}{i^2}$

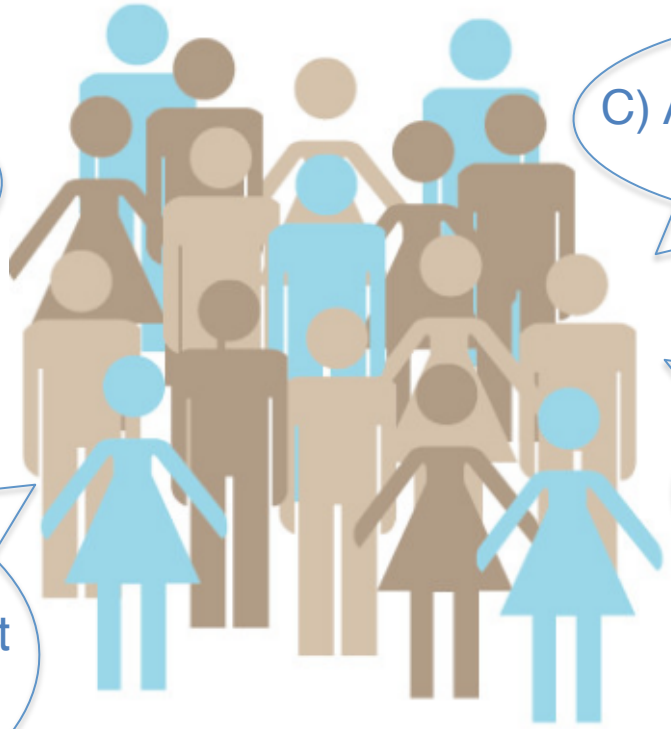
Its allow for replication and deflates the effect of large amounts of data

Open Targets

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?

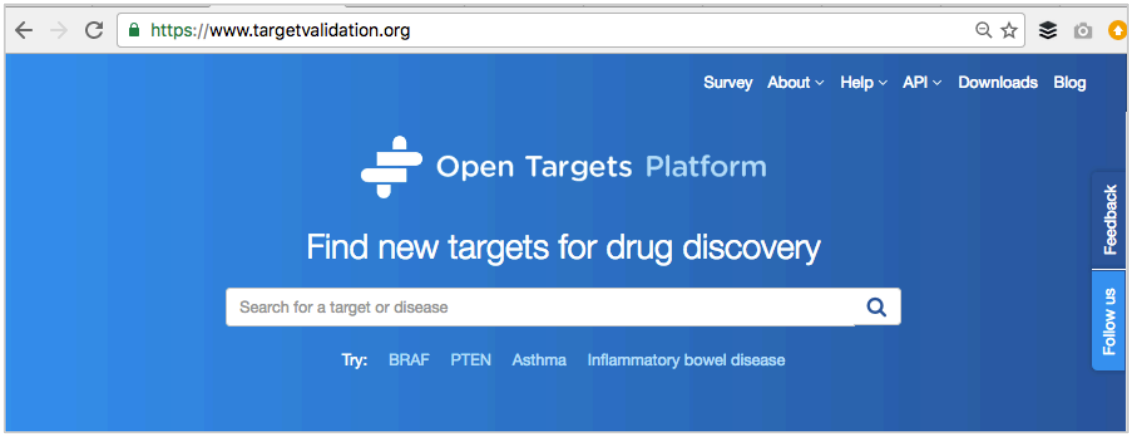
D) For a given target, are there other diseases associated with it?

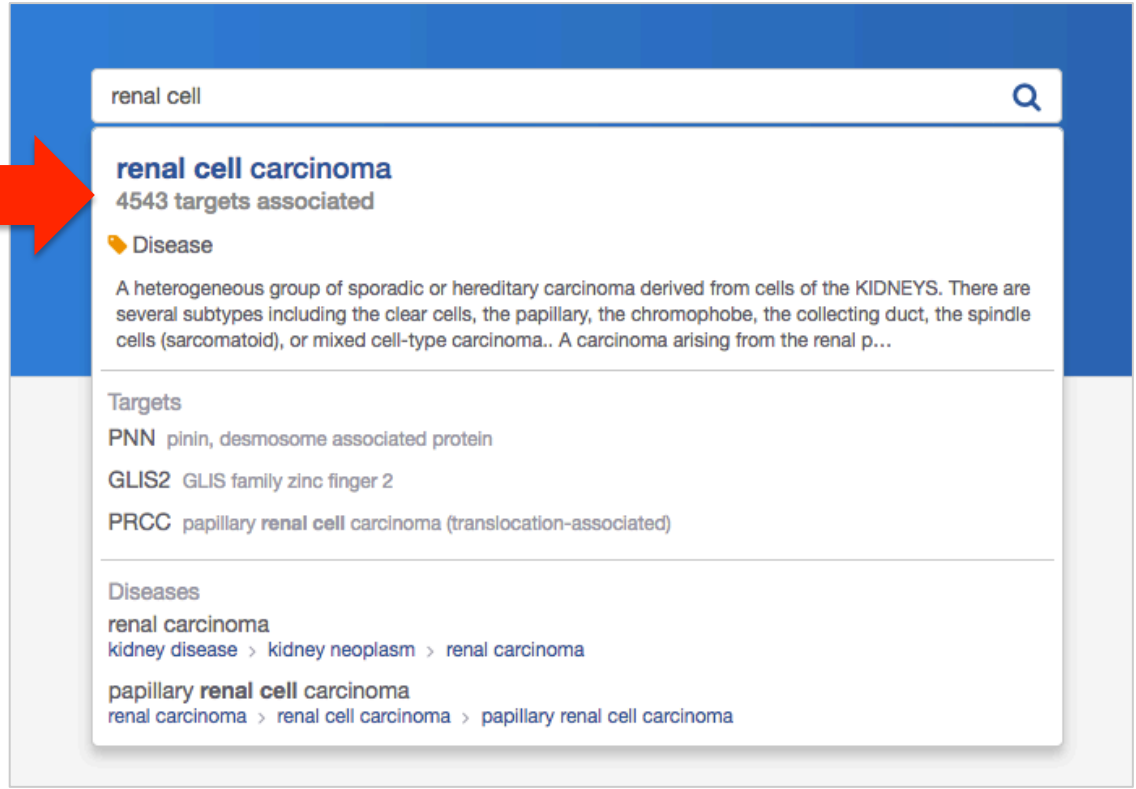
E) If this target is associated with other diseases, can I get the association for diseases from different therapeutic areas?

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

Open Targets

Which targets are associated with a disease?

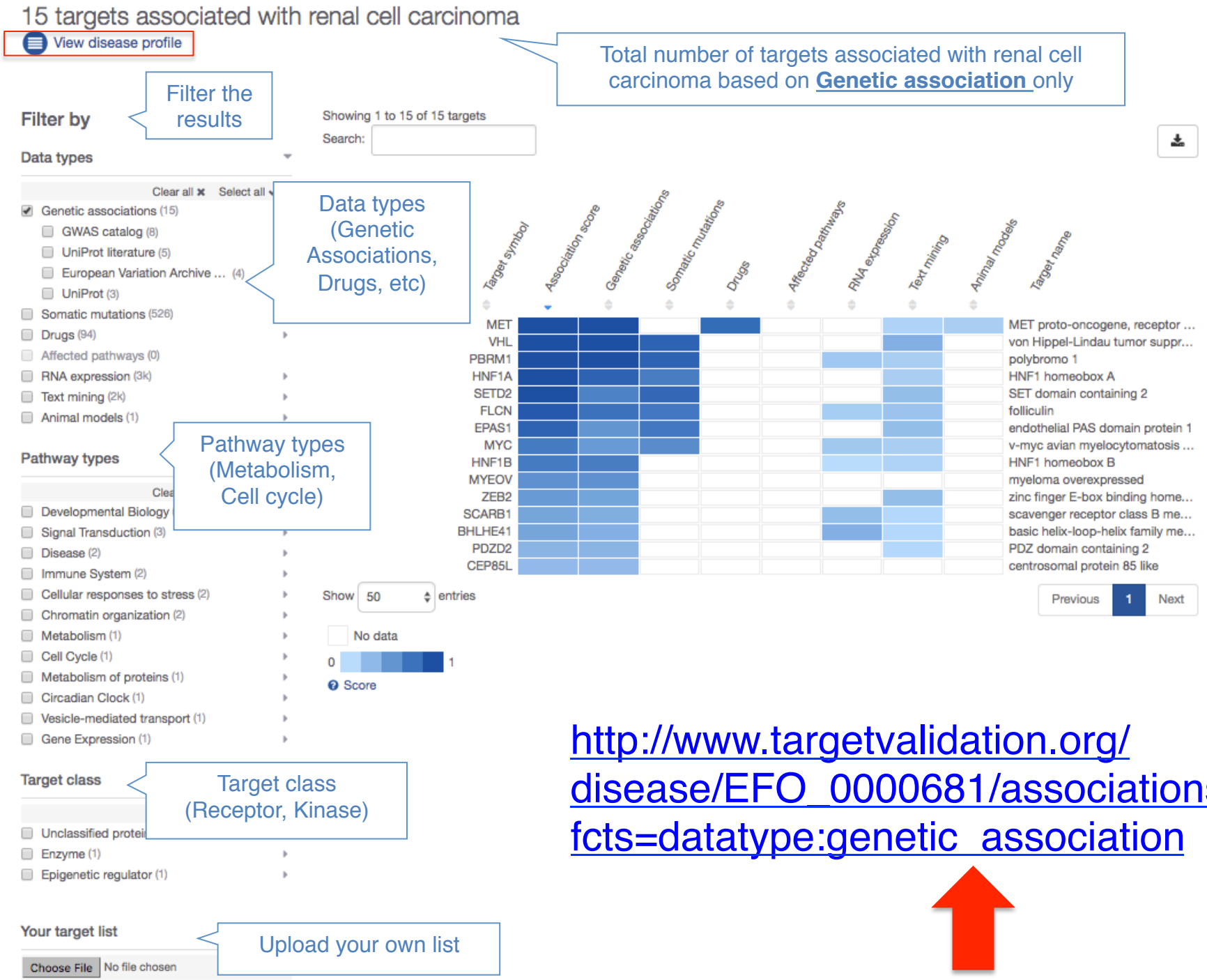




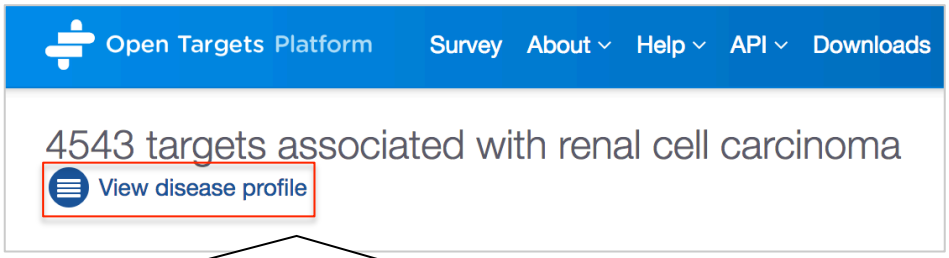
Demo 1:

screenshots: next slides
coursebook: pages 9 -15

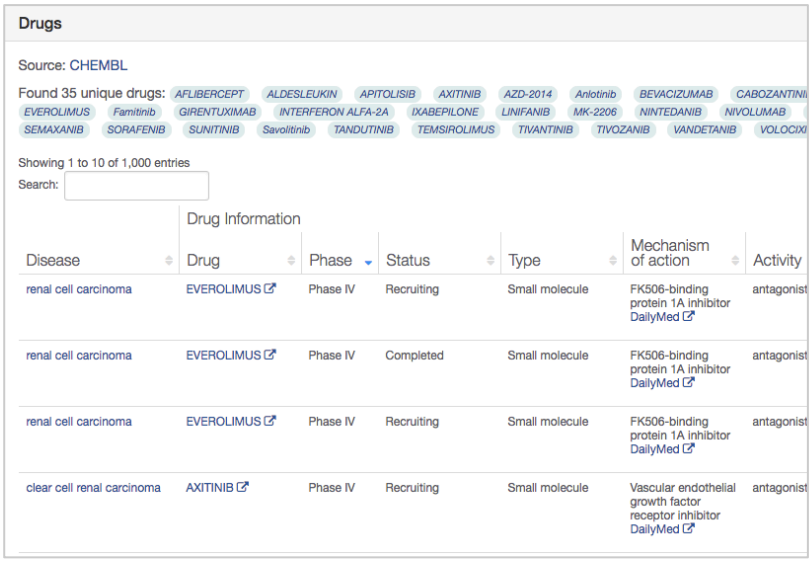
https://www.targetvalidation.org/

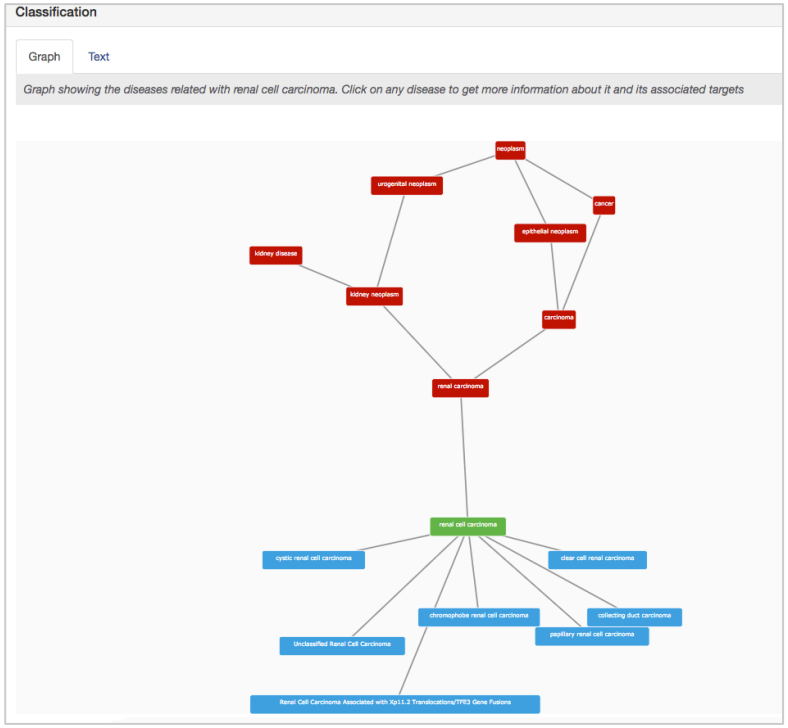

15 targets associated with renal cell carcinoma
View disease profile
Total number of targets associated with renal cell carcinoma based on Genetic association only
Filter by
Filter the results
Showing 1 to 15 of 15 targets
Search:
Data types
Clear all
Select all
Genetic associations (15)
GWAS catalog (8)
UniProt literature (5)
European Variation Archive ... (4)
UniProt (3)
Somatic mutations (526)
Drugs (94)
Affected pathways (0)
RNA expression (3k)
Text mining (2k)
Animal models (1)
Data types (Genetic Associations, Drugs, etc)
Pathway types
Developmental Biology
Signal Transduction (3)
Disease (2)
Immune System (2)
Cellular responses to stress (2)
Chromatin organization (2)
Metabolism (1)
Cell Cycle (1)
Metabolism of proteins (1)
Circadian Clock (1)
Vesicle-mediated transport (1)
Gene Expression (1)
Pathway types (Metabolism, Cell cycle)
Target class
Unclassified protei
Enzyme (1)
Epigenetic regulator (1)
Target class (Receptor, Kinase)
Your target list
Choose File
No file chosen
Upload your own list
Target symbol
Association score
Genetic associations
Somatic mutations
Drugs
Affected pathways
RNA expression
Text mining
Animal models
Target name
MET
VHL
PBRM1
HNF1A
SETD2
FLCN
EPAS1
MYC
HNF1B
MYEOV
ZEB2
SCARB1
BHLHE41
PDZD2
CEP85L
MET proto-oncogene, receptor ...
von Hippel-Lindau tumor suppr...
polybromo 1
HNF1 homeobox A
SET domain containing 2
folliculin
endothelial PAS domain protein 1
v-myc avian myelocytomatosis ...
HNF1 homeobox B
myeloma overexpressed
zinc finger E-box binding home...
scavenger receptor class B me...
basic helix-loop-helix family me...
PDZ domain containing 2
centrosomal protein 85 like
Show 50 entries
Previous
1
Next
No data
0
1
Score
http://www.targetvalidation.org/disease/EFO_0000681/associations?fcts=datatype:genetic_association

http://www.targetvalidation.org/disease/EFO_0000681



Phenotypes, Drugs, Disease Classification





A possible use case

Am J Hum Genet. 2006 Jun; 78(6): 1011–1025.
Published online 2006 Apr 25. doi: 10.1086/504300
PMCID: PMC1474084

Reconstruction of a Functional Human Gene Network, with an Application for Prioritizing Positional Candidate Genes

Lude Franke,[1] Harm van Bakel,[1] Like Fokkens,[1] Edwin D. de Jong,[2] Michael Egmont-Petersen,[3] and Cisca Wijmenga[1]

- Detect disease genes from a functional gene network
- 96 diseases (Mendelian and complex inheritance, cancer)
- 345 unique disease genes
- e.g. renal cell carcinoma (papillary) and *MET*

Demo 2: Evidence supporting the *MET*-papillary renal cell carcinoma 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?
- Is there a mouse model that mimics this disease?
- Which tissue has the highest RNA expression from GTEx?
- Are there other kidney diseases associated with this target? Can you export the table with this information? How strong is this association?

*coursebook: pages 16 -23

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

Franke et al (2006) described seven genes associated with Alzheimer's disease: ENSG00000091513, ENSG00000175899, ENSG00000143801, ENSG00000142192, ENSG00000130203, ENSG00000010704, ENSG00000080815.

- Which of these have the strongest association w/ Alzheimer's?
- Are there any targets, which are membrane receptors?
- Which amino acids of this membrane receptor (putative drug target) correspond to the extracellular domain?

*coursebook: pages 28-32

Exercises

Pages 24-27, 33-35

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

support@targetvalidation.org

Wrap up

Open Targets Platform is the place:

- For drug target ID and selection in drug discovery
- Rank target-disease associations: different sources
- Integrated information on target and diseases
- Intuitive graphical user interface

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

Breakthrough Articles from NUCLEIC ACIDS RESEARCH

http://www.narbreakthrough.com/

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 Vrousgou[1,4], Xavier Watkins[1,4], Maria-Jesus Martin[1,4], Philippe Sanseau[1,2], Jessica Vamathevan[4], Ewan Birney[1,4], Jeffrey Barrett[1,4,5] and Ian Dunham[1,4,*]

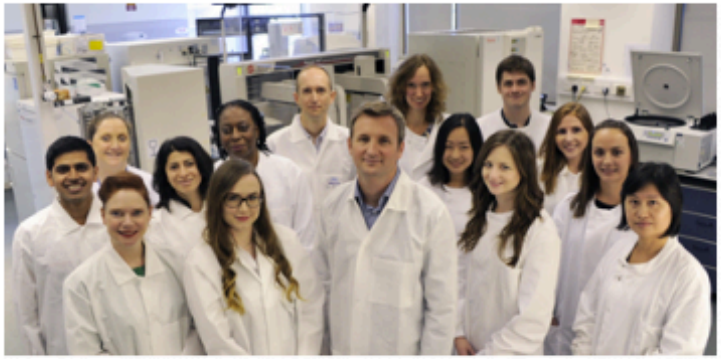
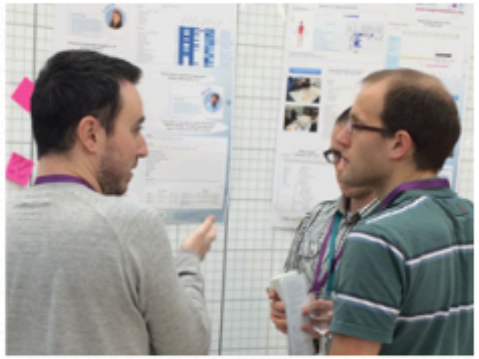
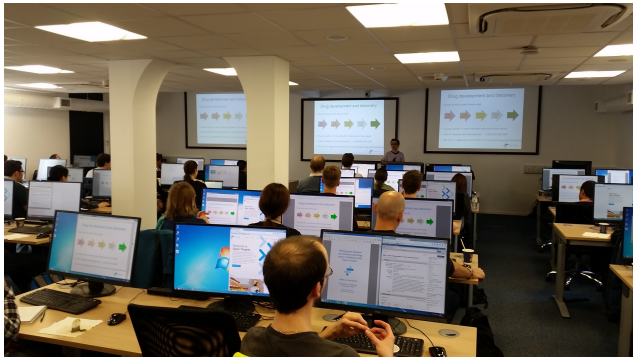
[1]Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, [2]GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK, [3]Biogen, Cambridge, MA 02142, USA, [4]European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, [5]Wellcome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

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

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

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

Open Targets


Open Targets

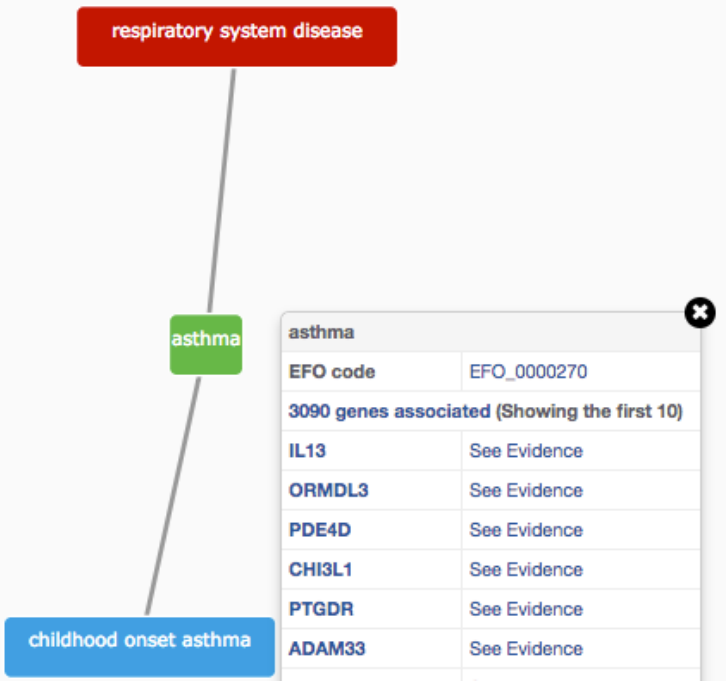
Extra slides

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

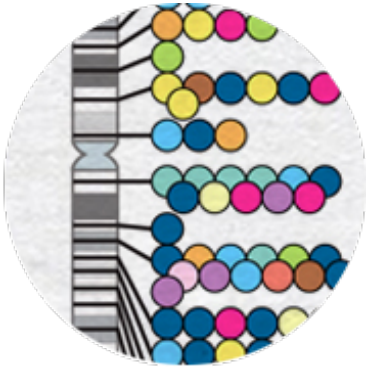
e.g. GWAS Catalog

$$S = f * s * c$$

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*

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

where S1,S2,...,Si 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

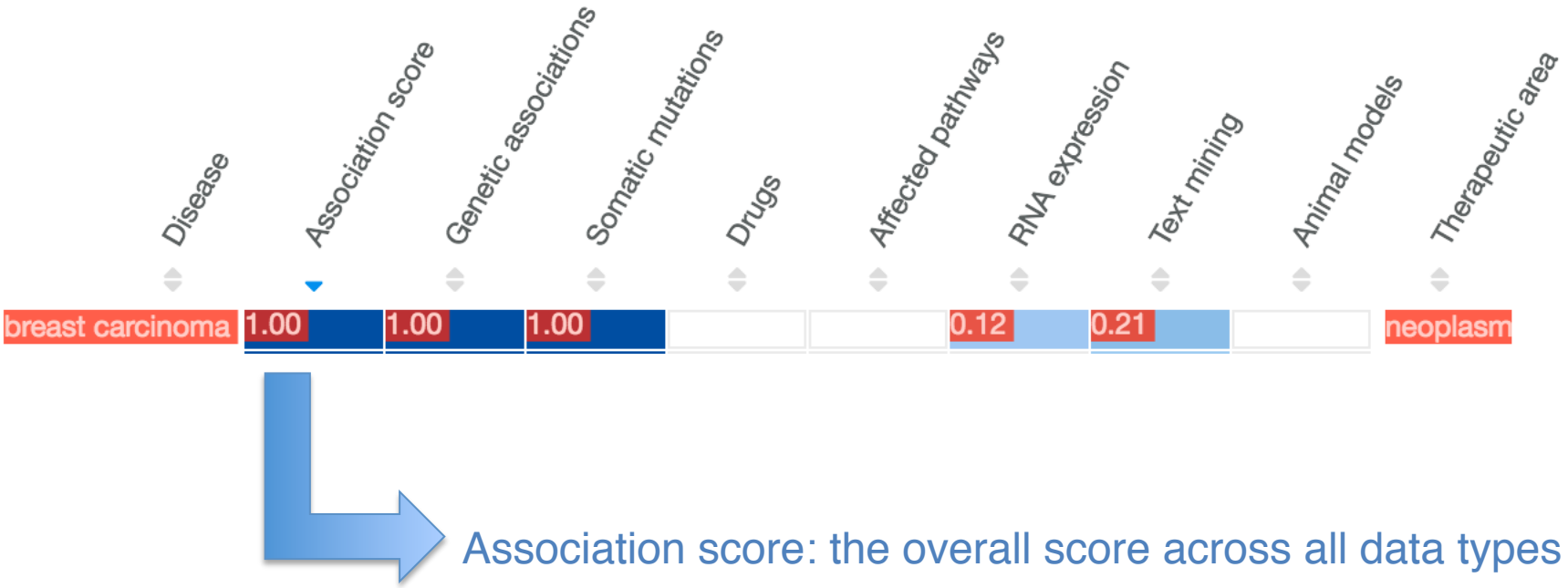
* PMID: 19107201, PMID: 20118918

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",
```


- 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

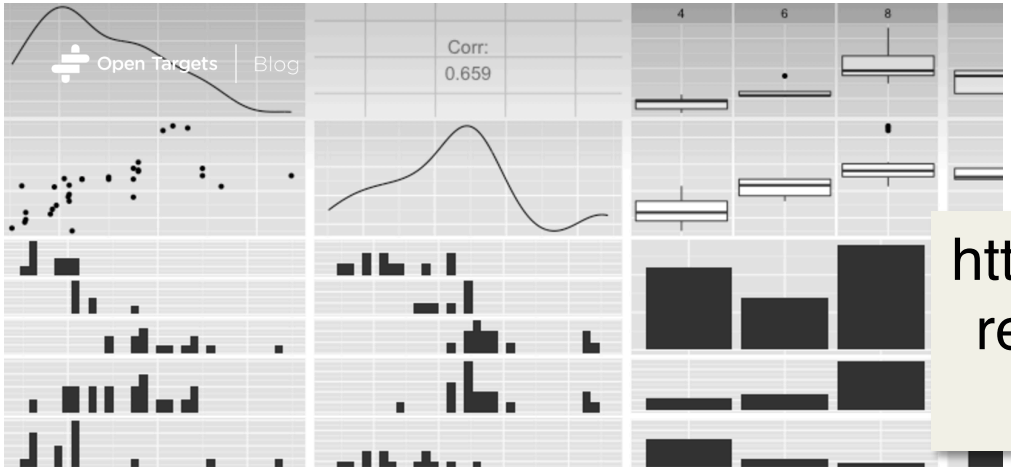
Docs » opentargets - Python client for targetvalidation.org

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/

How to access Open Targets with R