

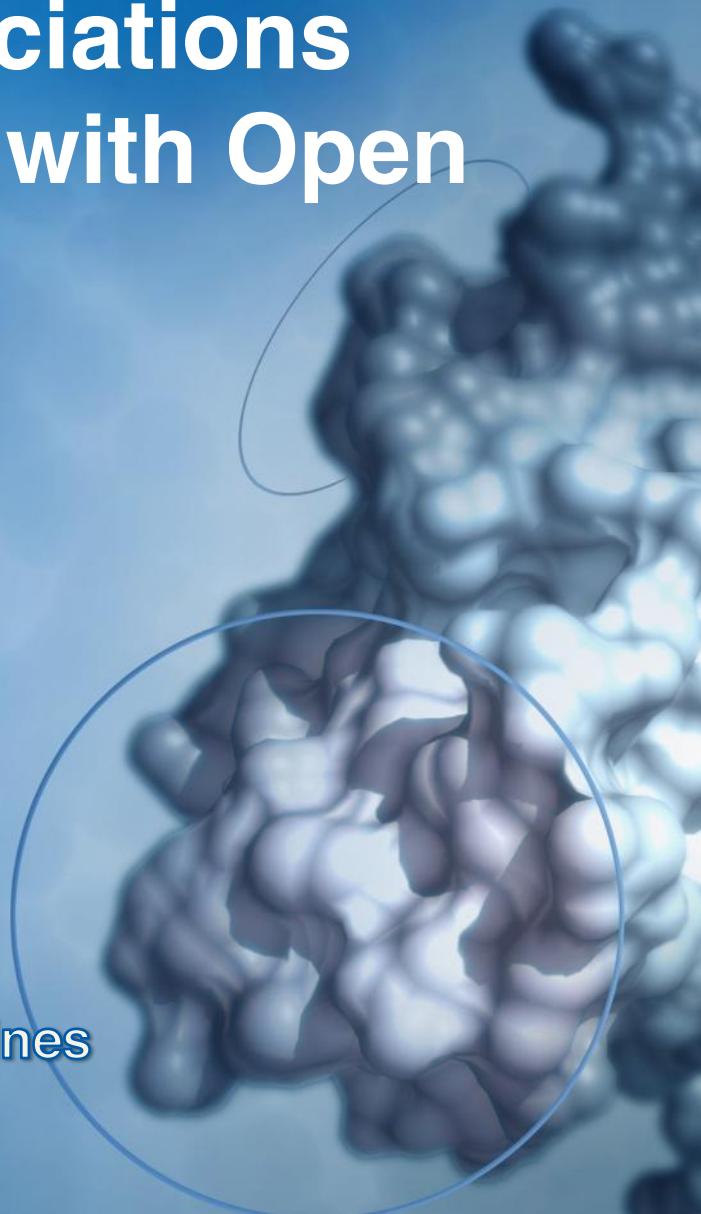
# Mining gene-disease associations and drug target validation with Open Targets

Denise Carvalho-Silva

Andrea Pierleoni

Miguel Pignatelli

Wellcome Genome Campus  
Open Targets  
Core Bioinformatics and Computational Pipelines



# Materials

<https://github.com/deniseOme/training>



slides



exercises



answers

# Today 13:00-17:00

- Drug discovery and validation
- Open Targets and the Target Validation Platform
- Walkthrough the Platform

*15:30-15:45 coffee/tea*

- Hands-on exercises
- Wrap up and feedback survey

# Course objectives

What is Open Targets?

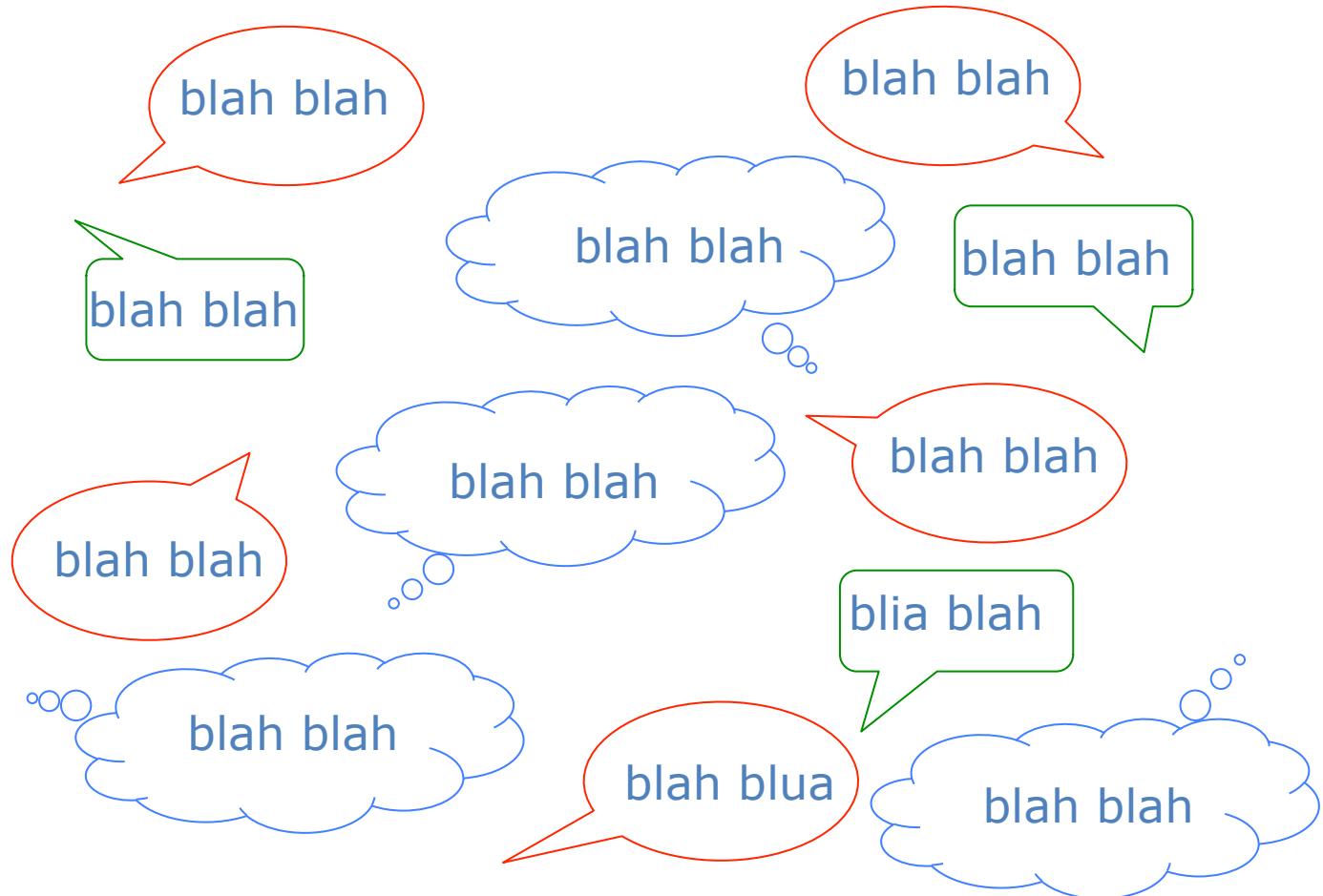
What type of data can you get from the **Target Validation Platform**?

How to **navigate** the platform?

How to **connect** with team



Open Targets



Open Targets

# Do we need programmes for drug development and discovery?

YES

NO

**Quora** Ask or Search Quora Ask Question

Infectious Diseases Medical Conditions and Diseases +1

**How many diseases today have no cure? Which ones?**



**Karen Tiede**, Hula hoop for your health.

641 Views · Most Viewed Writer in Infectious Diseases with 90+ answers

How many? **Most of them, probably.** We have treatments, and support, and care, and some of the time, you live through the disease, but it's not exactly that we "cured" it. You simply didn't have to die of it.

To add to Meghana's list,\* diabetes (both kinds), heart disease, arthritis, pretty much all of the auto-immune disorders, most of the mental illnesses, many of the genetic disorders (cystic fibrosis, for example).

The list of diseases that can actually be cured, rather than prevented or treated, is pretty short.

Written Aug 9, 2014 · View Upvotes · Answer requested by 1 person

Upvote | 4

Downvote Comment

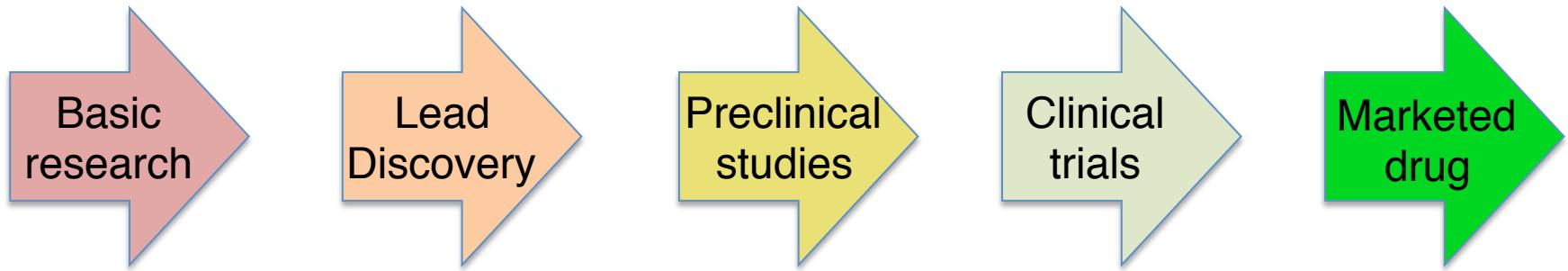


\* **Meghana Rastogi**, Research Scholar

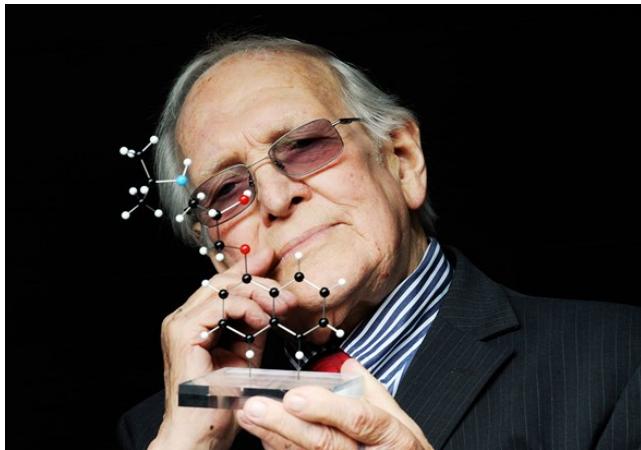
Ebola, HIV, Cancer, Dystonia are one of those disease which don't have any cure.

# Drug development and discovery

- Driven by this unmet clinical need



- 10 year journey → new drug from discovery to the market
- Cost for research and development: ~ 2,000,000,000 GBP
- 10% only → approved by the FDA (U.S Food and Drug Administration)



Sir James Black, winner of the  
1988 Nobel Prize in Physiology  
and Medicine

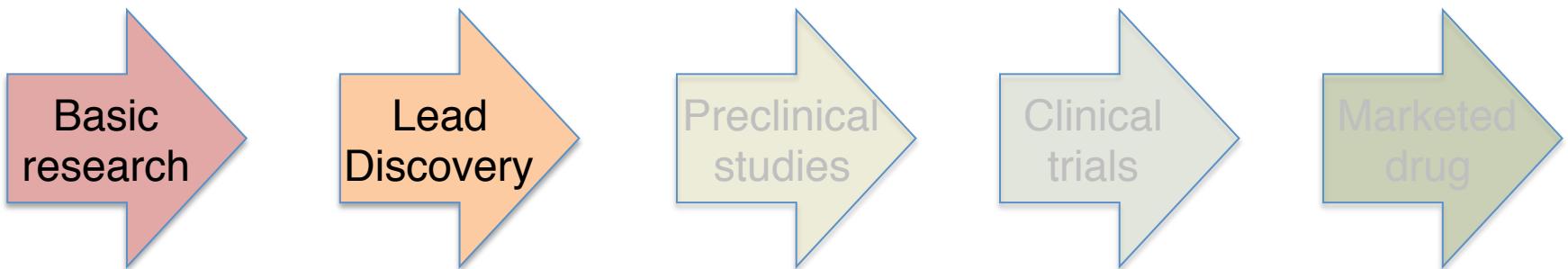
“The most fruitful basis for the discovery of a **new** drug is to start with an **old** drug”.

# Drugs do fail in the clinic

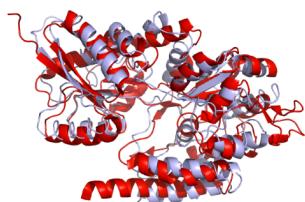
They don't work

They are not safe

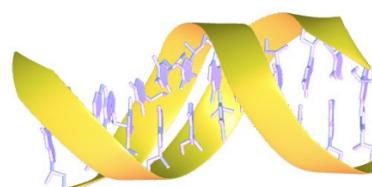
Important role of target identification and validation



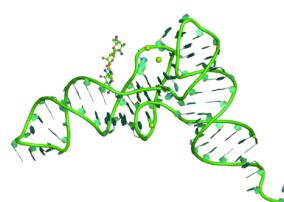
But what is a target?



protein



DNA



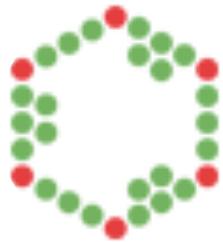
RNA



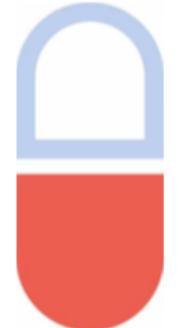
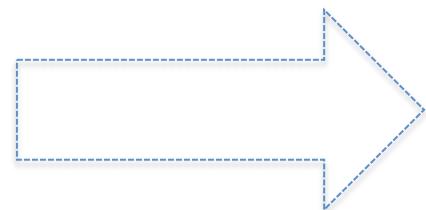
Open Targets

# Public databases for drug discovery

- European Bioinformatics Institute (part of EMBL)



- Supporting all stages of drug discovery
- ‘Fitting together like a jigsaw puzzle’ to build hypotheses



Open Targets

# Huge hurdle: data is everywhere



Wouldn't it be great to have a **one stop shop** with all these data **integrated**?



Yes, a database with **comprehensive, trustworthy** data that we all could access it for **free**?



That'd be fab! It'd be much quicker to carry out our experiments in the lab validating drugs and identifying new ones.



Open Targets



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



*Patrick Vallance, President  
Pharmaceuticals R&D  
GlaxoSmithKline*



*Professor Sir  
Mike Stratton  
Director, Sanger Institute*



- Target validation can (should) be improved
- One institution could not necessarily or easily do it alone
- Strong desire to collaborate based on highly complimentary skills set
- Existing strong relationships, real commitment to the mission

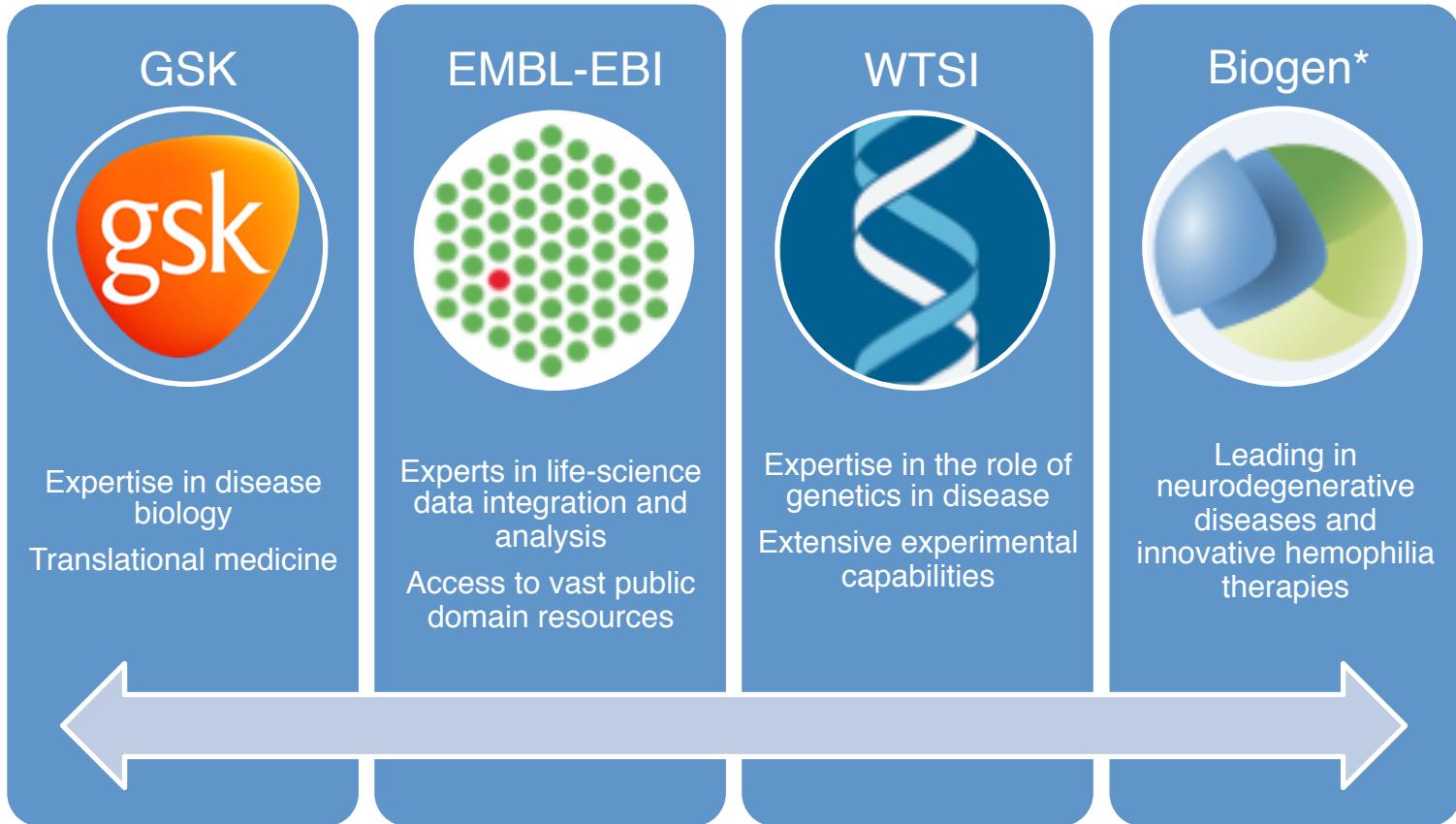
# The consortium

- Launched in March 2014
- Formerly known as **CTTV**
- Rebranded to Open Targets in April 2016
- Public-private initiative
- Aim: **transform** drug discovery
- How? Through the **identification** and **prioritisation** of targets

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

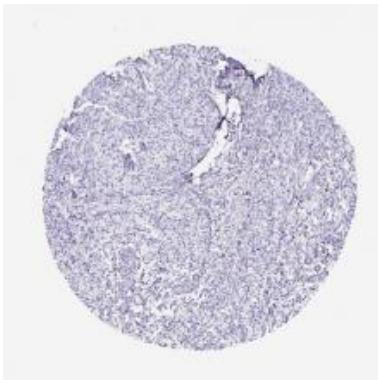
<http://www.ebi.ac.uk/about/news/press-releases/open-targets-new-name-new-data>

# Who is Open Targets?



\*Biogen joined the consortium in February 2016

# Open Targets Projects



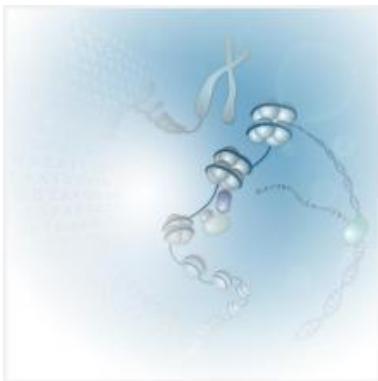
Oncology



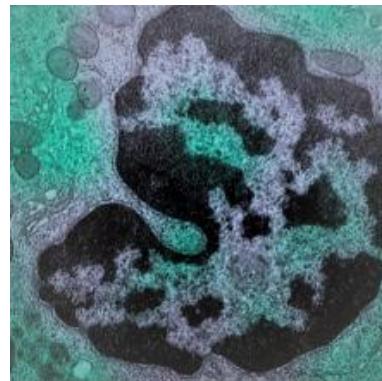
Respiratory



IBD



Cell line  
epigenomes



Inflammation  
and immunity



Core bioinformatics  
pipelines

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



Open Targets

# Scientific project portfolio

Oncology	Immunology	Neurodegeneration	Cross-Disease
 <b>Human Cellular experiments</b> <ul style="list-style-type: none"><li>• 1000 Cell-line Fusion Analysis</li><li>• CRISPR synthetic lethality screen</li><li>• NK cell receptors</li></ul>	 <b>Immunology</b> <ul style="list-style-type: none"><li>• IBD Organoids,</li><li>• MacroScreen</li><li>• Dendritic Cell Screen</li><li>• Asthma Single-cell</li></ul>	 <b>Neurodegeneration</b> <ul style="list-style-type: none"><li>• Ox. stress &amp; tau CRISPR</li><li>• Familial Alzheimer's single-cell RNAseq</li></ul>	 <b>Cross-Disease</b>
 <b>Genetics as A tool</b> <ul style="list-style-type: none"><li>• NGS Melanoma,</li><li>• Cancer Signaling Pathways</li></ul>	<ul style="list-style-type: none"><li>• IBD &amp; MS GWAS fine-mapping</li><li>• Bronchiectasis</li></ul>	<ul style="list-style-type: none"><li>• Parkinson's &amp; Alzheimer's GWAS resolution</li></ul>	<ul style="list-style-type: none"><li>• Influential Variants</li><li>• Metabolite GWAS</li></ul>
 <b>Enabling resource</b>	<ul style="list-style-type: none"><li>• Immune cell functional maps</li><li>• IBD BioResource</li></ul>	<ul style="list-style-type: none"><li>• Neuron functional maps</li></ul>	<ul style="list-style-type: none"><li>• CELLector</li><li>• Cell line Epigenomics</li></ul>
<b>Robust Data Integration</b>			

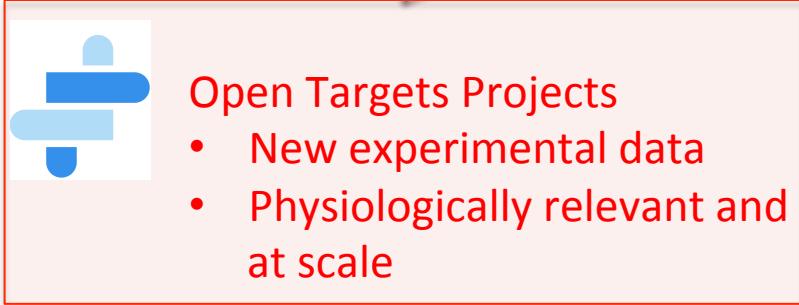
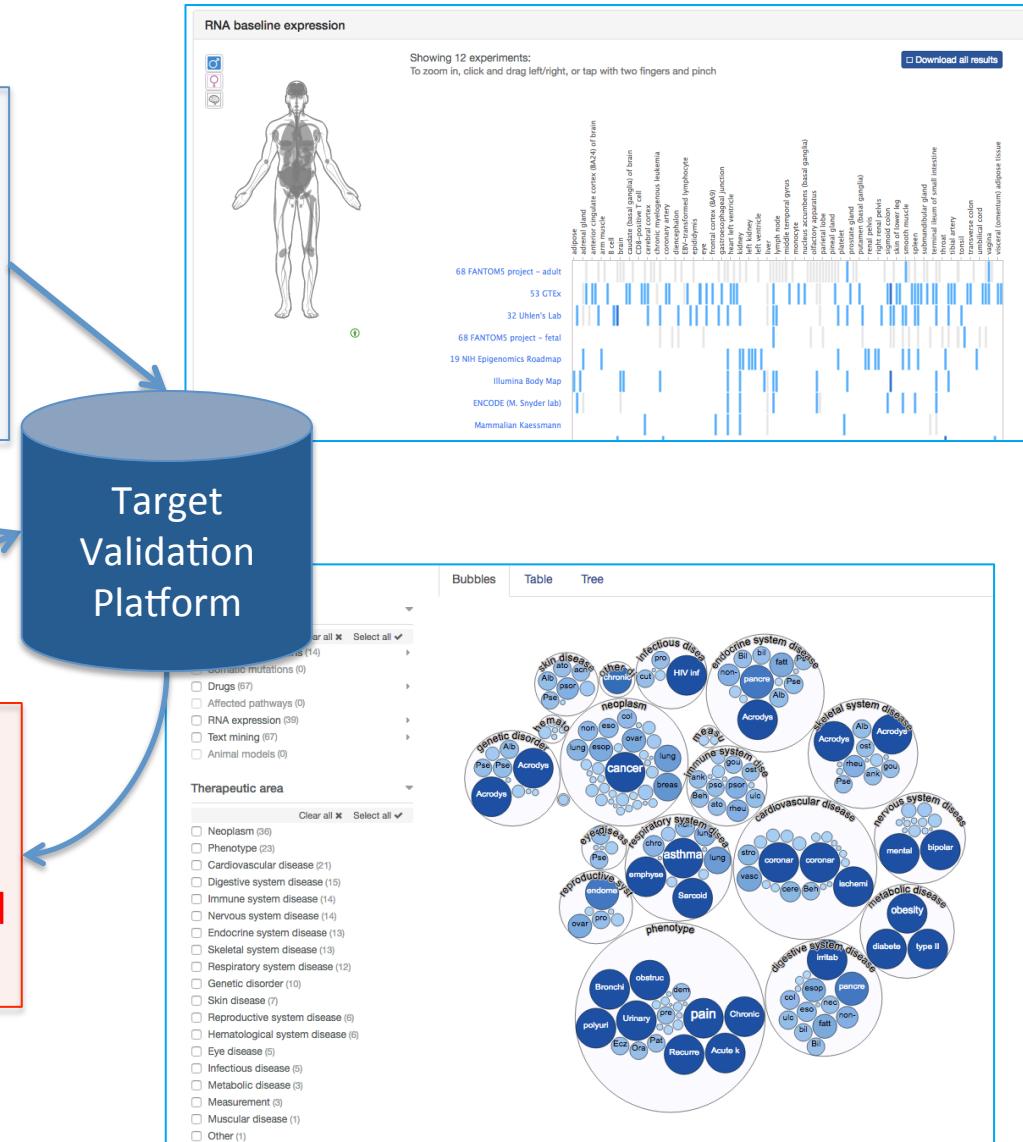
# Core bioinformatics pipelines

- Target Validation Platform: first release (Dec 2014)  
<https://www.targetvalidation.org/>
- Help scientists identify and prioritise relationships between targets and diseases
- Integration of new and existing data

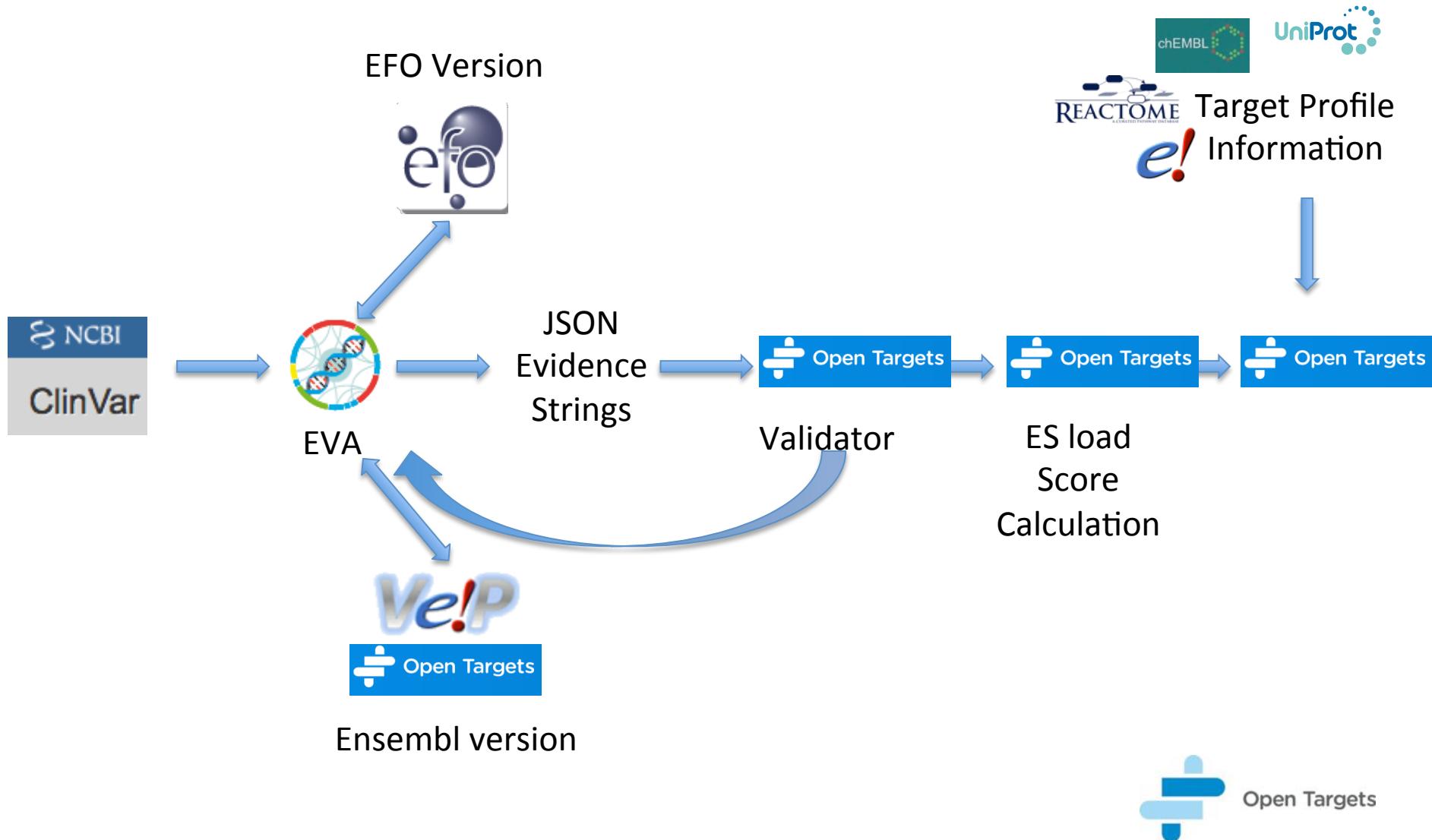
# Target Validation Platform



Target Validation Platform

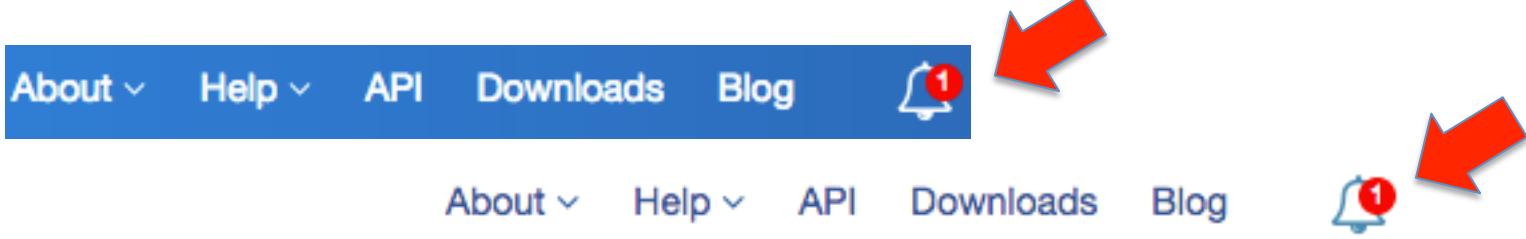


# Data flow pipeline



# Target Validation Platform

- Intuitive and easy-to-use web interface
- Constantly updated: new data, new web features\*

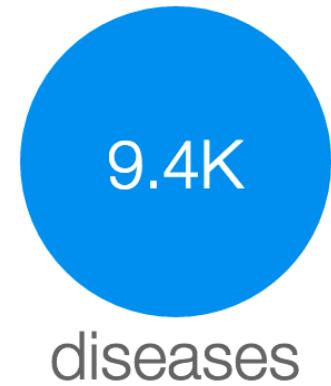
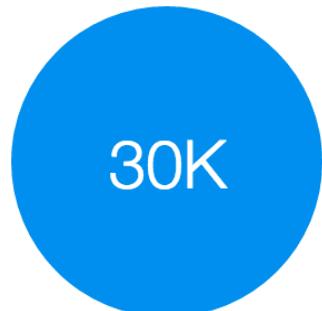


- Improvements driven by our user communities

\* <https://www.targetvalidation.org/release-notes>

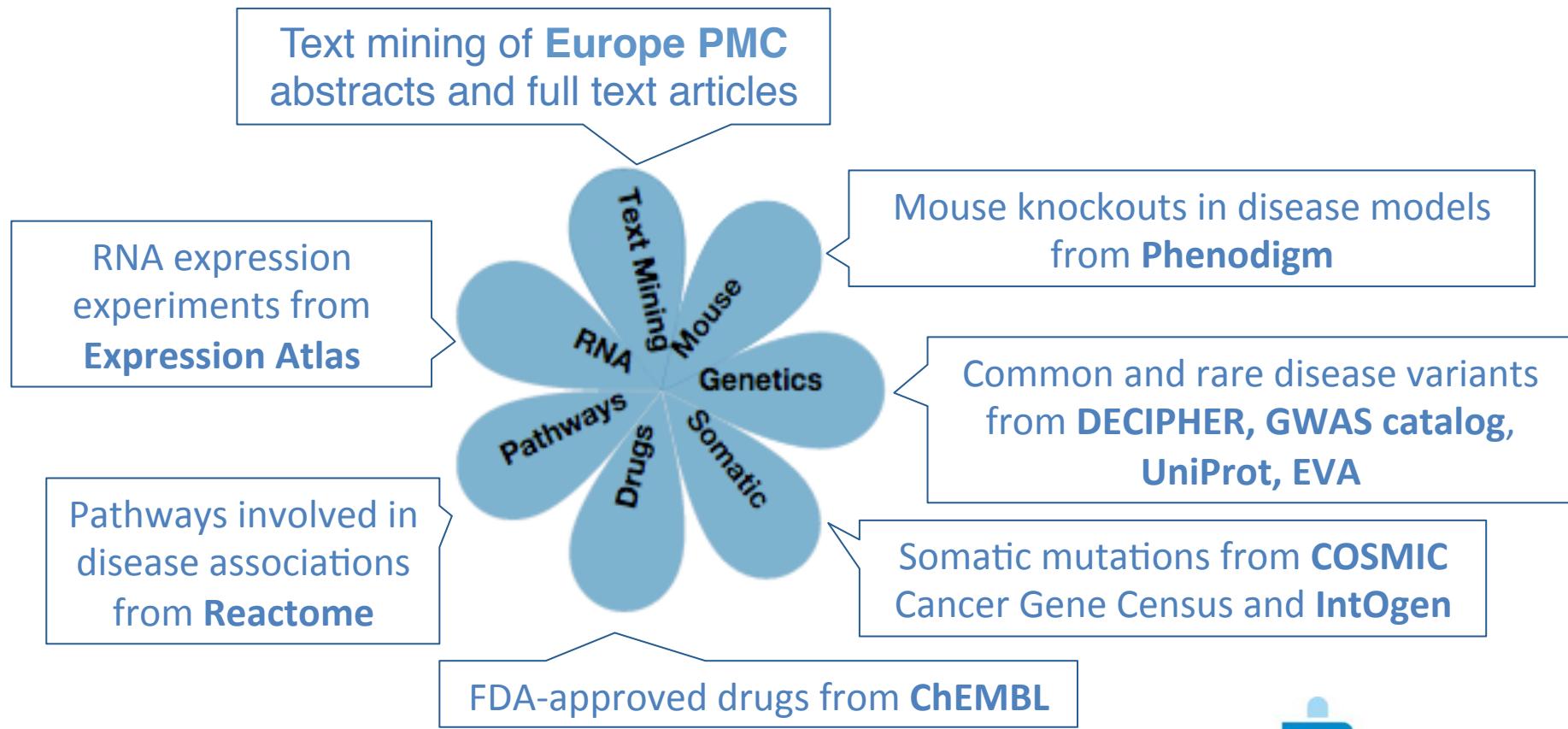
# Target Validation Platform

- Latest release: September 2016\*



# > 2 million associations

## Supported by publicly available data: the evidence



Open Targets

# How confident can we be of a target being associated with a disease?

- Our scoring scheme ( $s = F \times S \times C$ )
  - per evidence (e.g. drug A in phase IV,)
  - per data source       
  - per data type (e.g. Genetic association)         
  - Overall (capped at 1)
- Challenge: the joint computation of all evidence

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



Open Targets

# How strong are the associations?

- Look for our score and the shades of blue!



- No evidence to support the association: score is 0
- Sources providing their own score:

# We support decision-making

A) Which targets are associated with a disease?

B) What evidence supports this target-disease association?

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

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

C) Are there FDA drugs for this association?

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



# But how?

## Walkthrough the platform

# Hypothetical scenario for a use case

The screenshot shows the header of the *Nature Genetics* website. The logo "nature genetics" is on the left. To the right is a decorative background image of a DNA double helix. Below the logo is a horizontal navigation bar with links: Home, Current issue, Comment, Research, Archive (with a dropdown arrow), Authors & referees (with a dropdown arrow), and About the journal (with a dropdown arrow).

[home](#) ▶ [archive](#) ▶ [issue](#) ▶ [letter](#) ▶ [full text](#)

**NATURE GENETICS | LETTER**



## Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing

Alejandro Sifrim, Marc-Phillip Hitz, Anna Wilsdon, Jeroen Breckpot, Saeed H Al Turki, Bernard Thienpont, Jeremy McRae, Tomas W Fitzgerald, Tarjinder Singh, Ganesh Jawahar Swaminathan, Elena Prigmore, Diana Rajan, Hashim Abdul-Khalil, Siddharth Banka, Ulrike M M Bauer, Jamie Bentham, Felix Berger, Shoumo Bhattacharya, Frances Bu'Lock, Natalie Canham, Irina-Gabriela Colgiu, Catherine Cosgrove, Helen Cox, Ingo Daehnert, Allan Daly  
+ et al.

*Nature Genetics* 48, 1060–1065 (2016) | doi:10.1038/ng.3627

Received 23 December 2015 | Accepted 24 June 2016 | Published online 01 August 2016

# Congenital heart disease (CHD)

- How many targets are associated with CHD? Can you filter the search to find targets based on Genetic association only?
- Among the top targets with the highest association score i.e. 1, which target has the highest individual score based on Text mining only?
- Where does the association between target and disease based on animal models come from?
- Are there other cardiovascular diseases associated with this target?
- One of the targets in the paper that is associated with CHD is *PRKD1*. What are the drugs currently in clinical trials and targeting this gene?

# Exercises and Answers\*

\* Answers are provided in a separate booklet in digital form

# Alternative ways to access the data

Looking for our entire datasets?

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

- All target-disease associations: 179 MB
- All evidence: 1.7 GB

Looking for extracts of our datasets?

- API: REST calls, R and Python clients

# Some 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>



Open Targets

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 <b>TARGET_ID-DISEASE_ID</b>	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](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": "Reactome", "score": 80, "type": "Pathway-Disease"}]  
      }  
    }  
  ]  
}
```

- 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)

# Wrap up

Target Validation Platform is the place!

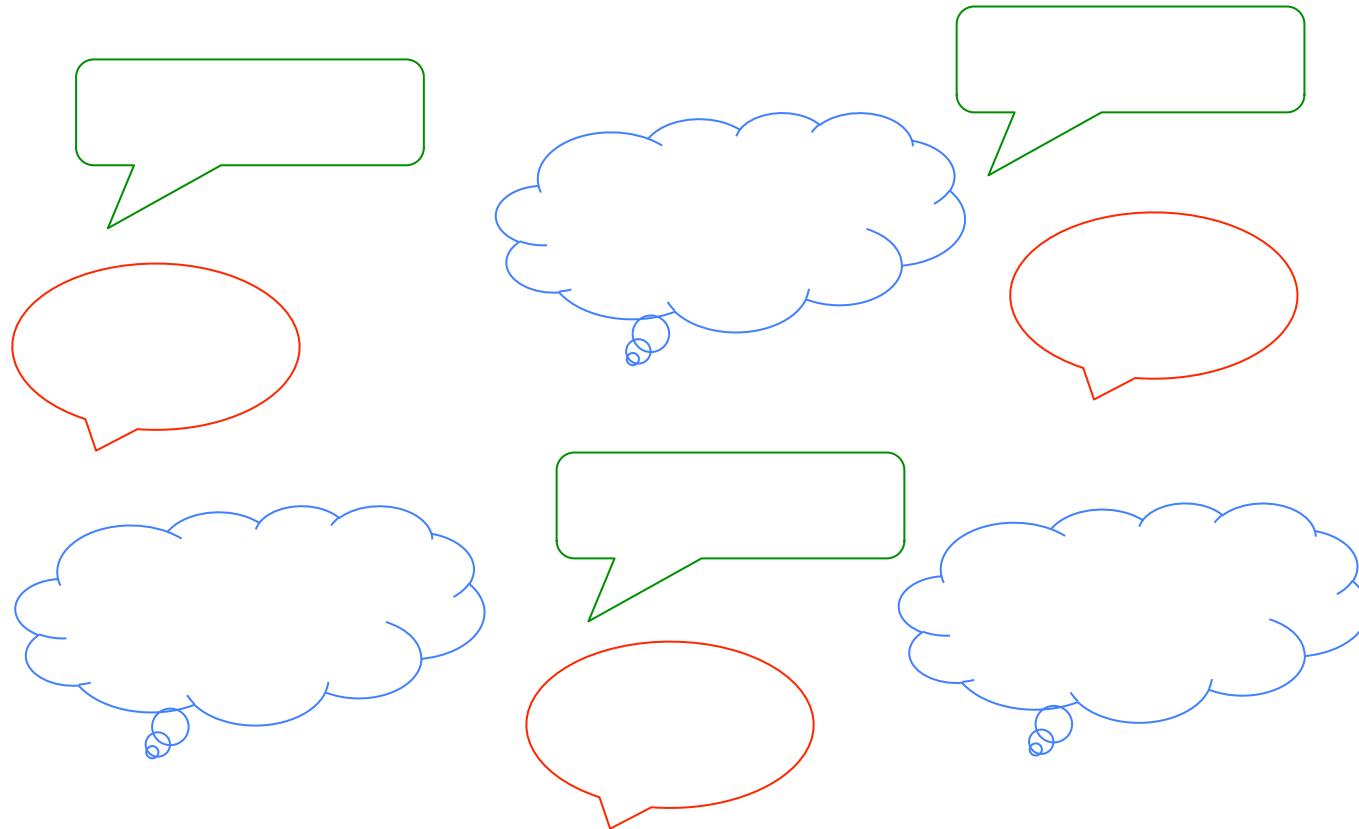
Drug discovery and development

Target-disease associations: different evidence

Information on target and diseases

Oh Yes!  
And all is 100% free

# Your take home message



Open Targets

# Short feedback survey

<https://goo.gl/forms/crSjmkTpEBMEL7oz2>

# Help and documentation

## Get in touch



@targetvalidate



support@targetvalidation.org



[www.facebook.com/OpenTargets/](https://www.facebook.com/OpenTargets/)

## The Target Validation Platform:

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

## Frequently Asked Questions:

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

## Open Targets Blog

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



Open Targets

# Acknowledgements



GWAS  
Catalog

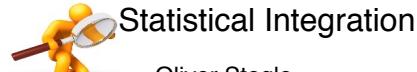
Atlas



Web production



EBI  
Database  
Admin



Oliver Stegle,  
Naruemon (Ploy)  
Pratanwanich,  
Verena Zuber



Jenny Cham,  
Francis Rowland  
Bren Vaughan

## Target Validation platform team



Ian Dunham



Jessica Vamathevan



Andrea  
Pierleoni



Denise  
Carvalho-Silva



Eliseo  
Papa



Gautier  
Koscielny



Luca  
Fumis



Mick  
Maguire



Miguel  
Pignatelli



Niki  
Karamanis



Samiul  
Hasan



Jeff Barrett  
Ewan Birney  
Holly Foster  
Philippe Sanseau



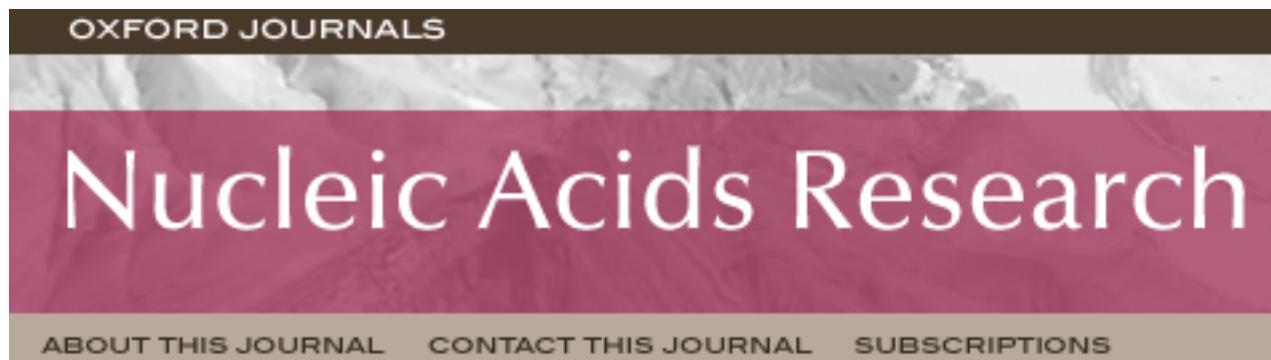
50 Biogen  
and 100 GSK  
beta testers



Open Targets

# How to cite us

Keep an eye out for our 2017 paper:



## **Database issue**

For the time being, include our URL:

[www.targetvalidation.org](http://www.targetvalidation.org)