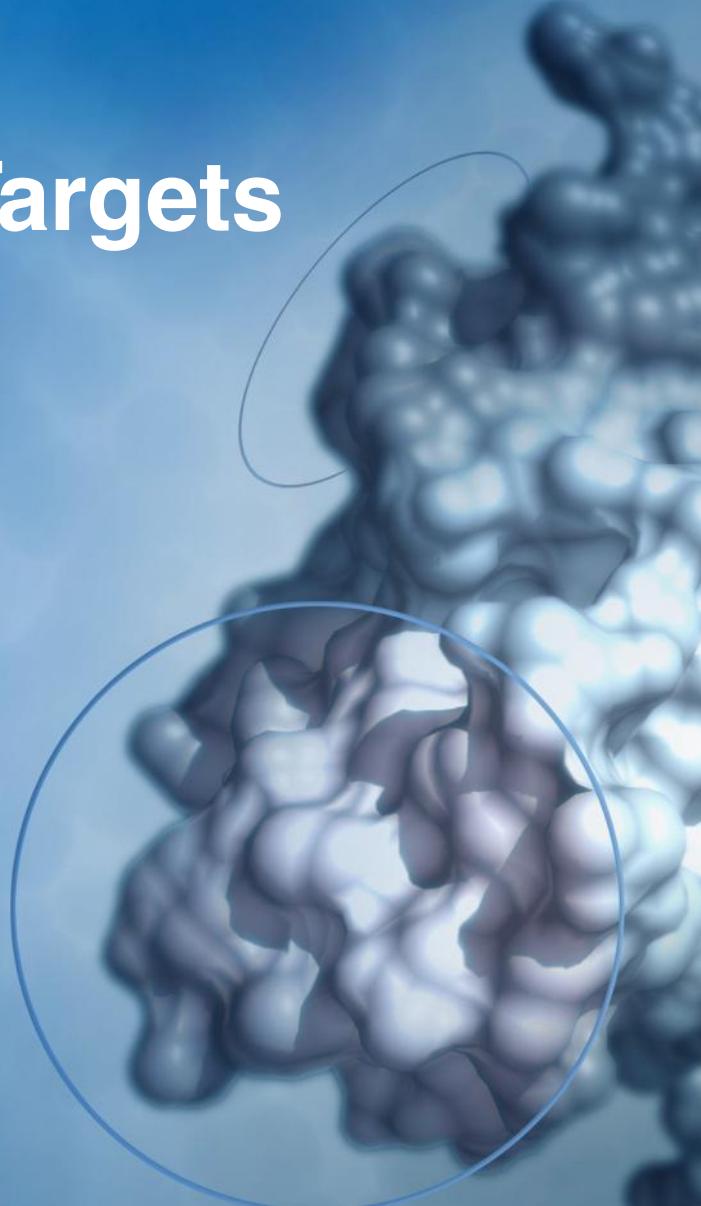


# Mining Gene and Disease Associations with Open Targets

2017 Lab Automation Virtual Conference  
May 24<sup>th</sup>-25<sup>th</sup> 2017

**Denise Carvalho-Silva, PhD**  
Wellcome Genome Campus, United Kingdom  
Open Targets Consortium  
Core Bioinformatics team



# What I will cover in this talk

- Open Targets Consortium
- Experimental projects and the Platform
- What makes Open Targets unique
- What is coming up

# Acknowledgments



# Drug discovery: 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

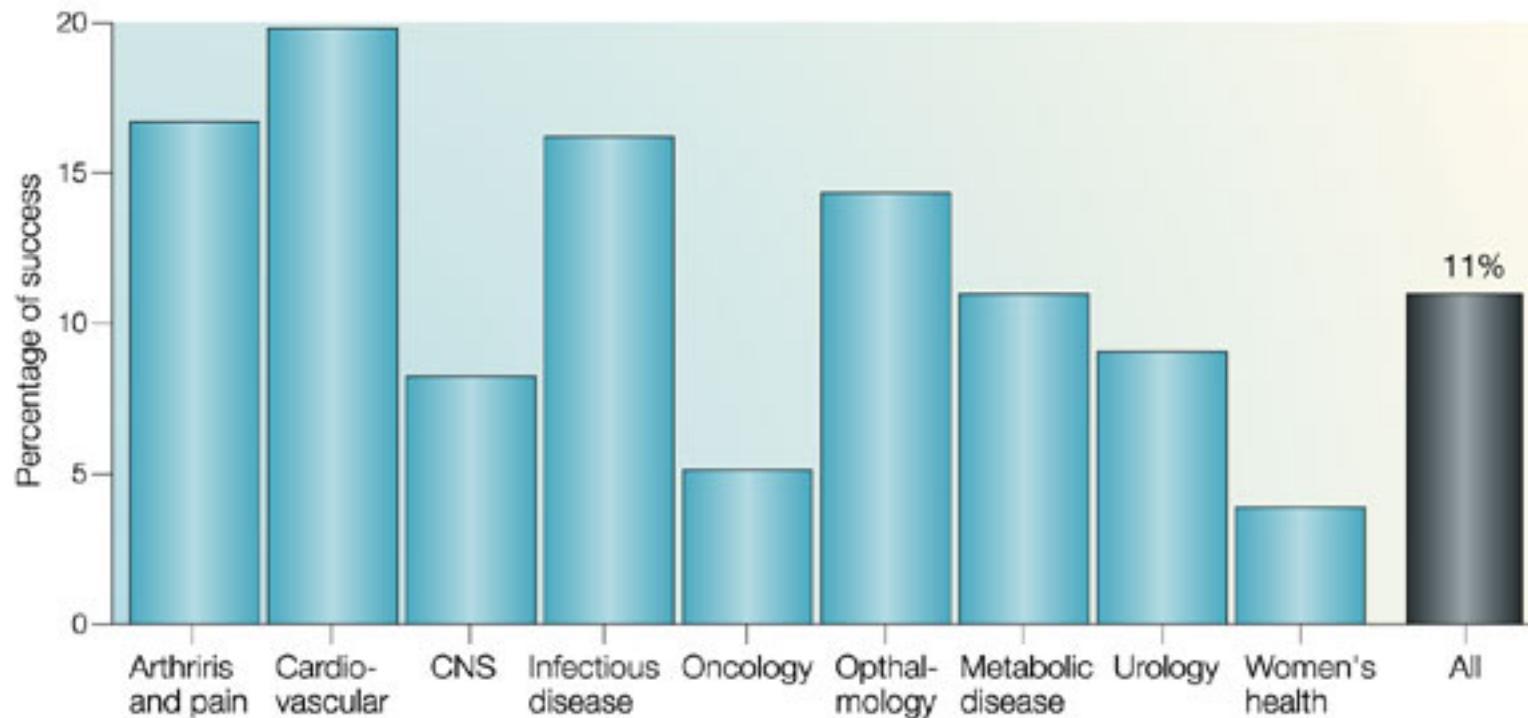
Human trials are completed. FDA approval. Industry is responsible for bringing a drug to market. Safety and evaluation continue after approvals.



PATIENT CARE

## 3. DELIVERY

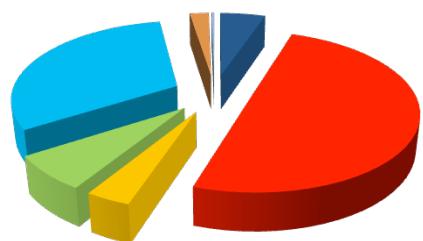
# Drug discovery: the challenges



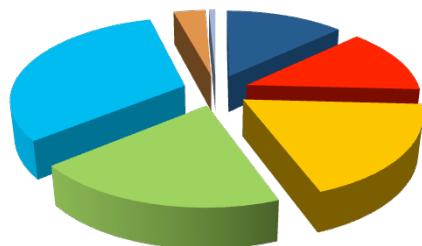
Lengthy, costly, low success rate, **high attrition rate**

# What are the causes for the attrition?

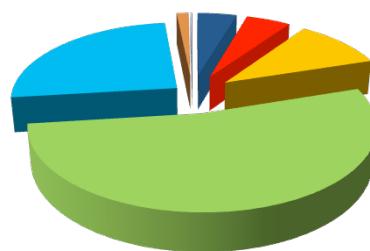
Pre-clinical



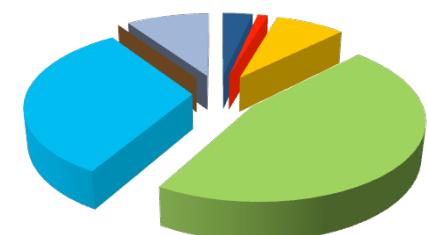
Phase I



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*



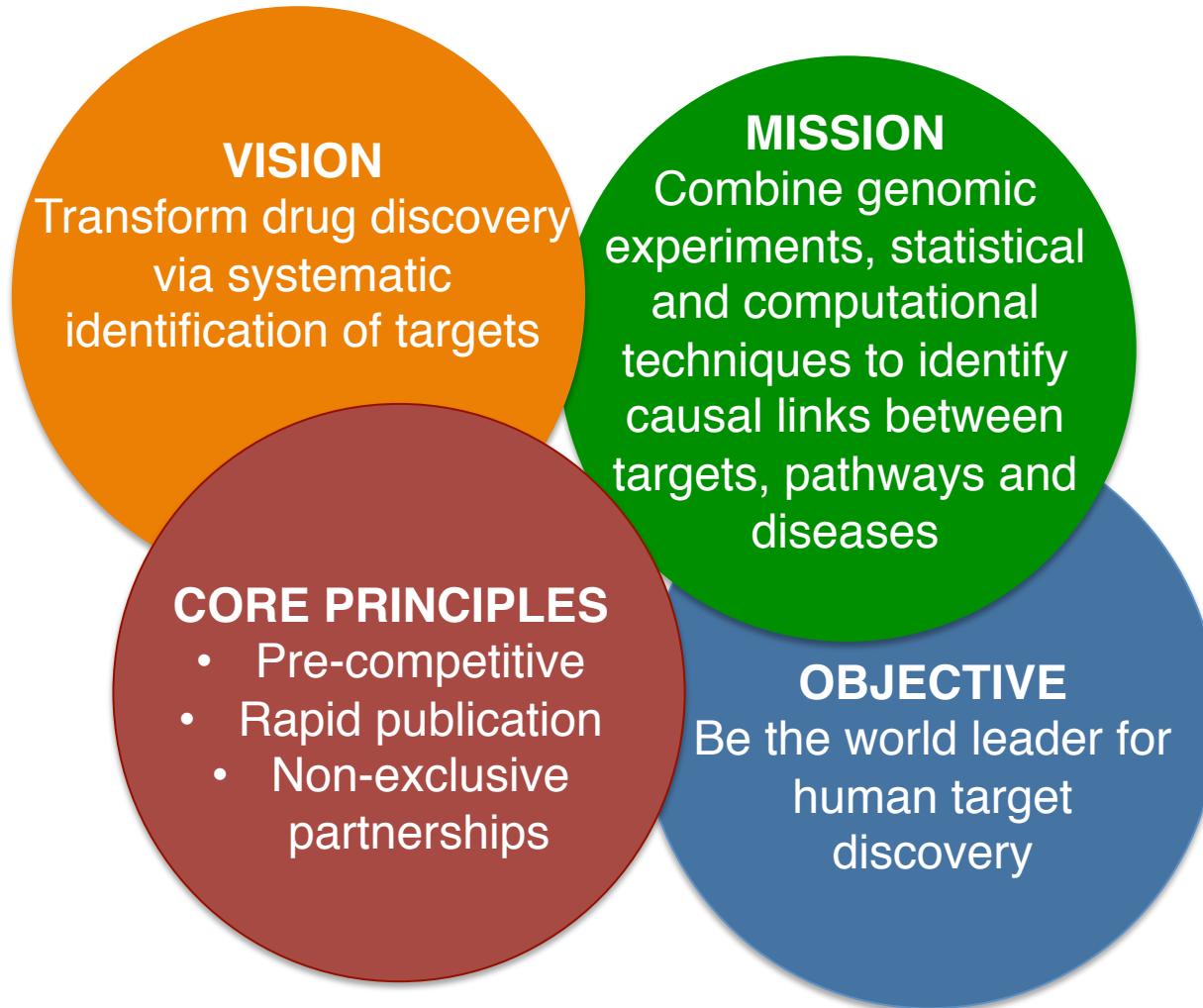
*Professor Dame  
Janet Thornton  
former Director, EMBL-EBI*

Yes, we can!  
And we should.

But one institution  
can not do it alone.

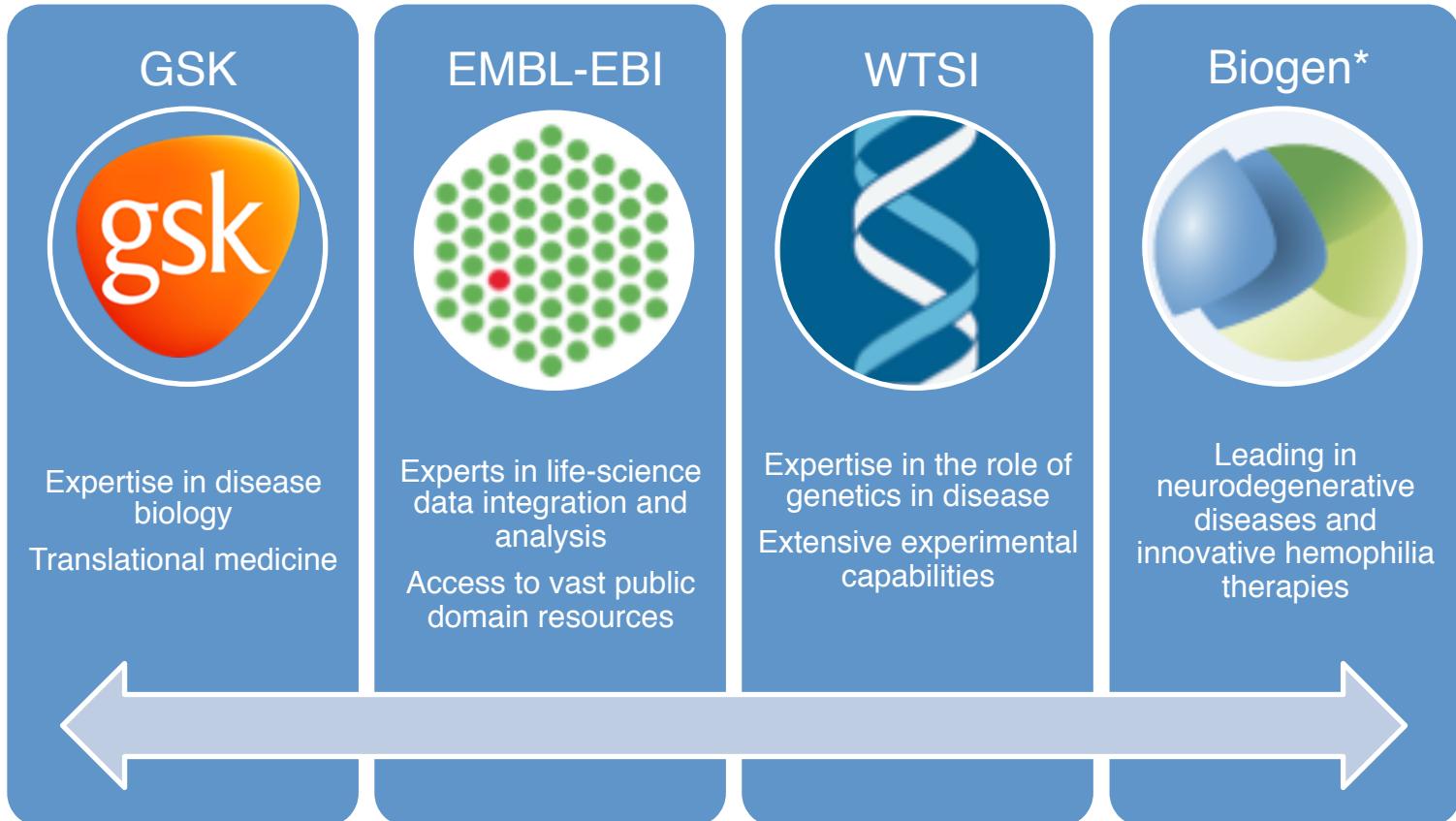


# Open Targets Consortium\*



\* Launched in March 2014  
Three founding partners

# Who is Open Targets now?



\* Biogen joined the consortium in February 2016

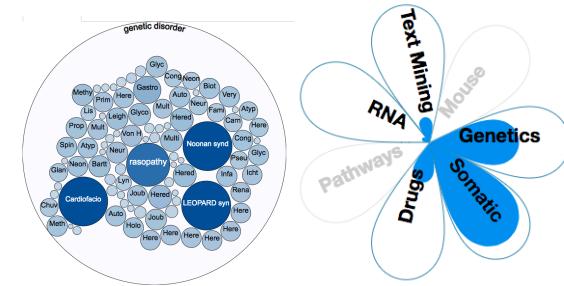
# Two major areas of work in Open Targets

## Experimental projects



Generate new evidence  
CRISPR/Cas9, Organoids

## Bioinformatics projects



Integration of available data  
Web interface, API and data dumps

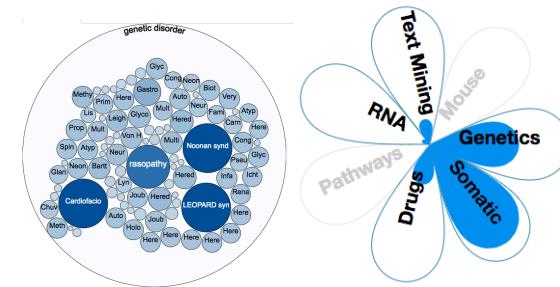
# Two major areas of work in Open Targets

## Experimental projects



Generate new evidence  
CRISPR/Cas9, Organoids

## Bioinformatics projects

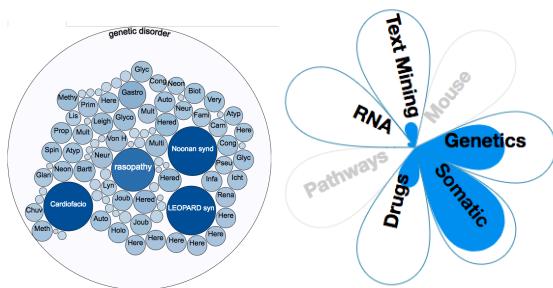


Integration of available data  
Web interface, API and data dumps

# Open Targets Platform\*

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

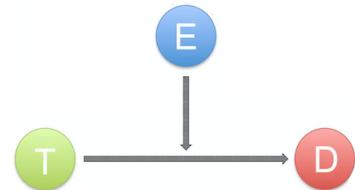
<https://www.targetvalidation.org/>



\* First release: December 2015

# Evidence from publicly available data

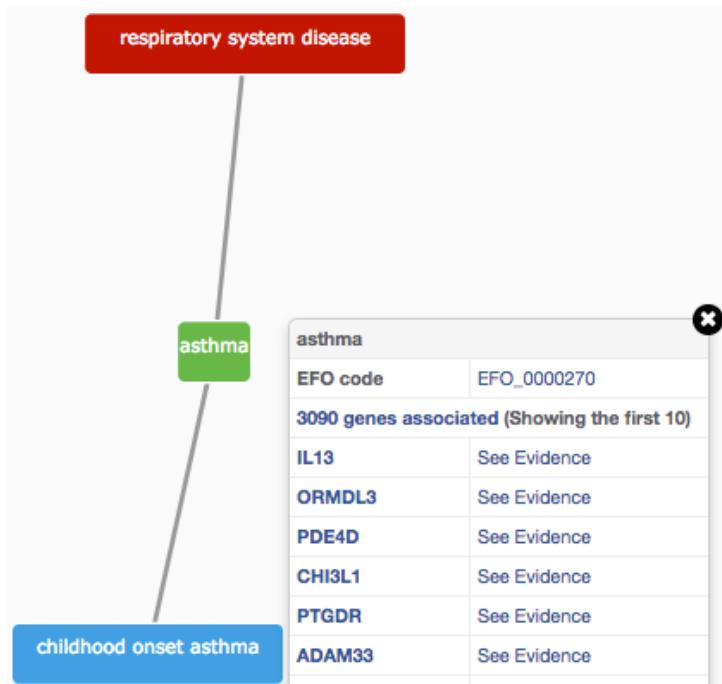
- Similar data sources are grouped into data types



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
<b>Your favourite data?</b>	<b>Let us know!</b>

# Experimental Factor Ontology\* (EFO)

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

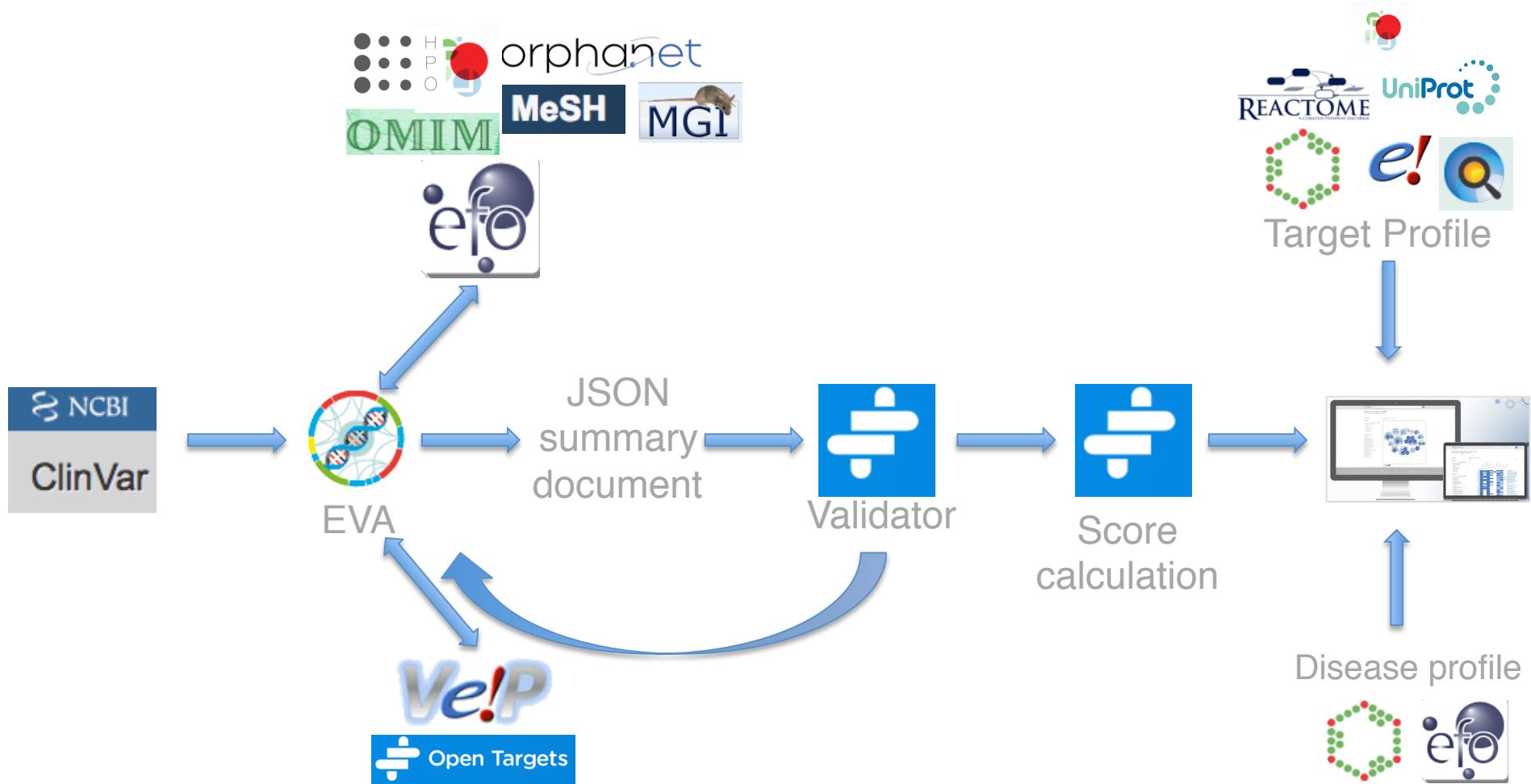


controlled vocabulary  
+  
hierarchy (relationship)

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

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

# Data flow pipeline



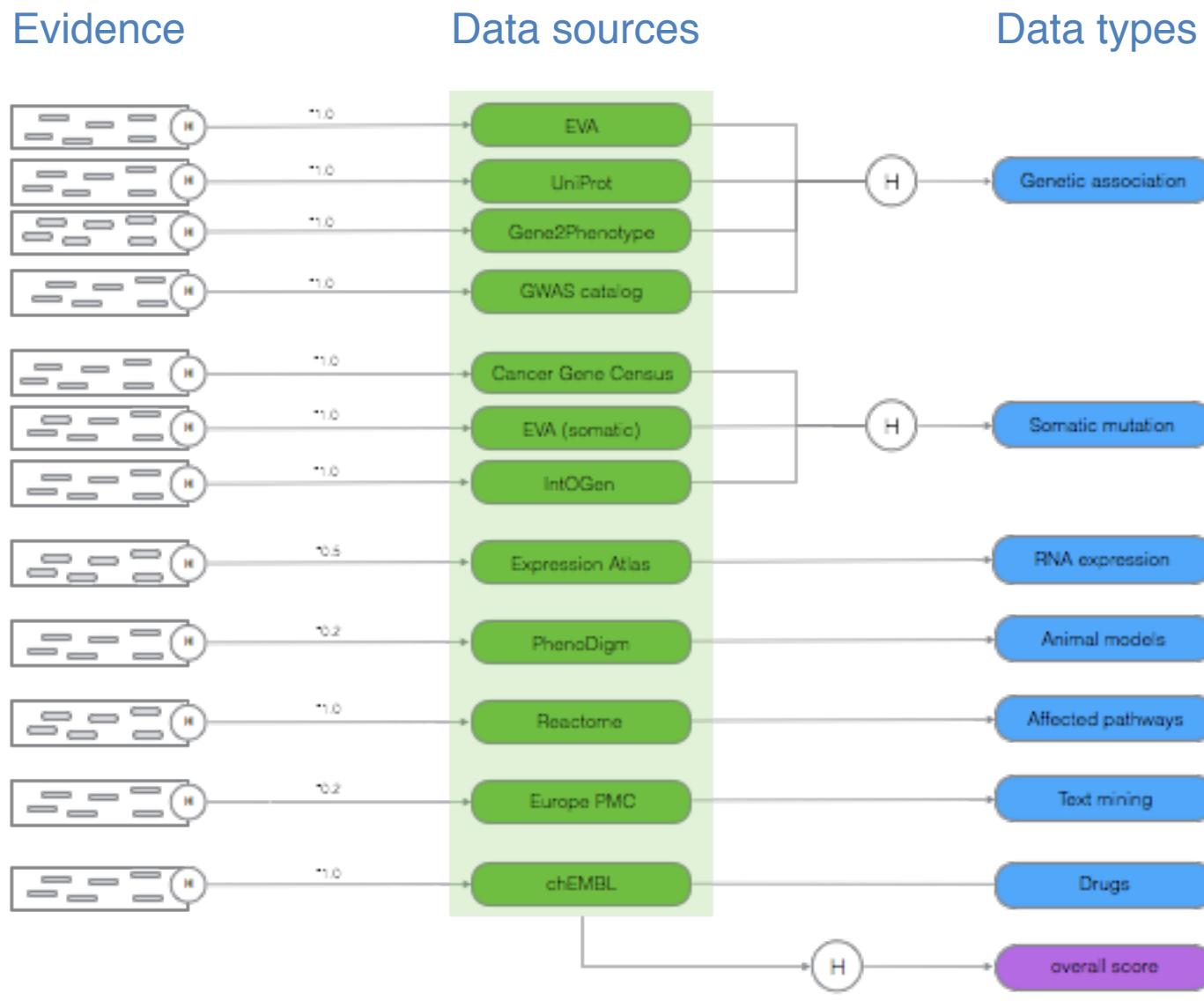
\* JSON summary document = IDs (gene, disease, papers) + curation (e.g. manual) + evidence + source + stats for the score

# JSON summary document

```
ArrayExpress Experiment overview"}], "experiment_overview": "Transcription profiling of human intestinal epithelium in patients suffering from inflammatory bowel disease.", "access_level": "public", "id": "866f4a506d5a01ebcea8bd84bb5485c8", "unique_association_fields": {"comparison_name": "'active ulcerative colitis' vs 'normal'", "geneID": "http://identifiers.org/ensembl/ENSG00000073605", "study_id": "http://identifiers.org/gxa.expt/E-MEXP-2083"}, "sourceID": "expression_atlas", "target": {"gene_info": {"symbol": "GSDMB", "geneid": "ENSG00000073605", "name": "gasdermin B"}, "id": "ENSG00000073605", "activity": "increased_transcript_level", "target_type": "transcript_evidence"}, "disease": {"id": "EFO_0000729", "efo_info": {"path": [[{"EFO_0000405", "EFO_0003767", "EFO_0000729"}, [{"EFO_0000540", "EFO_0005140", "EFO_0003767", "EFO_0000729}], "label": "ulcerative colitis", "therapeutic_area": {"labels": ["immune system disease", "digestive system disease"]}, "codes": ["EFO_0000405", "EFO_0000540"]]}, "biosample": {"name": "intestinal epithelial cell", "id": "http://purl.obolibrary.org/obo/CL_0002563"}, "data_release": "16.12", "type": "rna_expression", "scores": {"association score": 0.014397535094125722}}
```

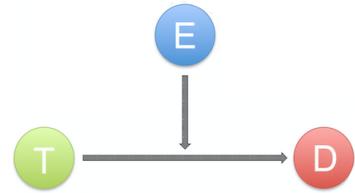
Validated against schema version: 1.2.3, evidence: {"confidence\_level": "low", "log2\_fold\_change": {"value": 1, "percentile\_rank": 97}, "unique\_experiment\_reference": "STUDYID\_E-GEOID-48258", "resource\_score": {"value": 0.0328, "type": "pvalue", "method": {"description": "Moderated t-statistics computed with limma version 3.16.8. Independent filtering performed using genefilter version 1.46.1, with gene variances as filter statistic. p-values adjusted using Benjamini & Hochberg (1995) FDR correction. By default, significant differential expression called if adjusted p <= 0.05."}}, "date\_asserted": "2015-07-20T22:01:23Z", "reference\_replicates\_n": 3, "evidence\_codes\_info": [{"eco\_id": "ECO\_0000356", "label": "differential gene expression evidence from microarray experiment"}]}, "test\_replicates\_n": 3, "is\_associated": true, "provenance\_type": {"database": {"version": "18-11-2016", "id": "Expression\_Atlas"}, "evidence\_codes": ["ECO\_0000356"], "reference\_sample": "control", "comparison\_name": "'fludarabine, 10 micromolar' vs 'control'", "test\_sample": "fludarabine, 10 micromolar", "urls": [{"url": "http://www.ebi.ac.uk/gxa/experiments/E-GEOID-48258?geneQuery=ENSG00000179889", "nice\_name": "Gene expression in Expression Atlas"}, {"url": "http://www.ebi.ac.uk/gxa/genes/ENSG00000179889", "nice\_name": "Basel"}}

# Score approach and aggregation



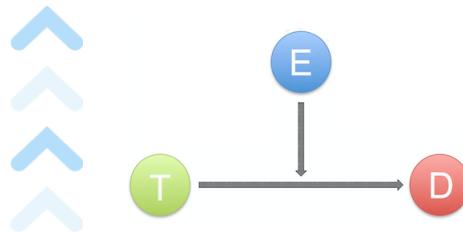
Evidence score and aggregation with the harmonic sum

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



# Currently: Integration of existing data

## Public Databases and Pipelines



Open  
Targets  
Platform

Target symbol	Association score	Genetic association	Somatic mutations	Drugs	Affects
TNFRSF1A					
KCNB2					
IL2RA					

Open Targets experimental data: NEW  
Physiologically relevant and at scale

Oncology



Immunology



Neurodegeneration



Generated  
as we speak

# 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?

What else can I find out about my drug target?



Open Targets

# Disease-centric workflow



Which targets are associated with a disease?

https://www.targetvalidation.org

Open Targets Platform

Find new targets for drug discovery

Search for a target or disease

Try: BRAF PTEN Asthma Inflammatory bowel disease

multiple scl

multiple sclerosis  
2735 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

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



Open Targets

168 targets associated with multiple sclerosis

[View disease profile](#)

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

Filter by

Filter the results

Datatype

[Clear all](#) [Select all](#)

- Genetic associations (168)
- Somatic mutations (1)
- Drugs (152)
- Affected pathways (0)
- RNA expression (1k)
- Text mining (1k)
- Animal models (4)

Data types  
(Genetic Associations,  
Drugs, etc)

Pathway types

[Clear all](#) [Select all](#)

- Immune System (32)
- Signal Transduction (28)
- Metabolism (11)
- Developmental Biology (10)
- Disease (10)
- Gene Expression (9)
- Hemostasis (8)
- Metabolism of proteins (8)
- Vesicle-mediated transport (7)
- Cell Cycle (6)
- Organelle biogenesis and main... (4)
- Neuronal System (2)
- Extracellular matrix organization (2)
- Cellular responses to stress (2)
- Transmembrane transport of s... (2)
- Muscle contraction (1)
- Programmed Cell Death (1)
- DNA Repair (1)

Pathway types  
(Metabolism,  
Cell cycle)

Target class

- Enzyme (18)
- Unclassified protein (8)
- Membrane receptor (6)
- Transcription factor (2)
- Secreted protein (1)
- Other membrane protein (1)
- Surface antigen (1)
- Transporter (1)

Target class  
(Receptor, Kinase)

Your target list

[Choose File](#) No file chosen

Available lists:  
[sampleList](#)

Upload your  
own list of  
targets

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected Pathways	RNA expression	Text mining	Animal models	Target name
IL2RA									interleukin 2 receptor subunit al...
TNFRSF1A									TNF receptor superfamily memb...
KCNB2									potassium voltage-gated chann...
CD86									CD86 molecule
CD58									CD58 molecule
CLEC16A									C-type lectin domain family 16 ...
EVI5									ecotropic viral integration site 5
TAGAP									T-cell activation RhoGTPase act...
TTC34									tetratricopeptide repeat domain...
PTGER4									prostaglandin E receptor 4
CMC1									C-X9-C motif containing 1
STAT3									signal transducer and activator ...
TNFSF14									tumor necrosis factor superfami...
AH11									Abelson helper integration site 1
A-DQA2									major histocompatibility comple...
CD24									CD24 molecule
PTPRC									protein tyrosine phosphatase, re...
NR1H3									nuclear receptor subfamily 1 gro...
CYP27B1									cytochrome P450 family 27 sub...
IL7R									interleukin 7 receptor
ZMZ1									zinc finger MIZ-type containing 1
SP140									SP140 nuclear body protein
CD6									CD6 molecule
ZFP36L1									ZFP36 ring finger protein like 1
CBLB									Cbl proto-oncogene B
CYP24A1									cytochrome P450 family 24 sub...
KIF1B									kinesin family member 1B
IL22RA2									interleukin 22 receptor subunit a...
NCOA5									nuclear receptor coactivator 5
FBXO48									F-box protein 48
A-DQB1									
DKKL1									
MALT1									
AGAP2									
C1orf106									
MPV17L2									
CD5									
ZFP36L2									
TIMMD1C									
IRF8									
MAPK1									
MERTK									
CLECL1									
SLC15A2									
BATF									
ODF3B									
IL12B									
MLANA									
BACH2									

[http://www.targetvalidation.org/  
disease/EFO\\_0003885/associations?  
fcts=datatype:genetic association](http://www.targetvalidation.org/disease/EFO_0003885/associations?fcts=datatype:genetic_association)



Open Targets Platform About ▾ Help ▾ API ▾ Downloads Blog Search for a target or disease

2735 targets associated with multiple sclerosis

[View disease profile](#)

Phenotypes  
Drugs  
Disease Classification

Classification

Graph Text

Graph showing the diseases related with multiple sclerosis. Click on any disease to get more information about it and its associated targets

multiple sclerosis

immune system disease

autoimmune disease

nervous system disease

brain disease

other disease

neuromyelitis optica

relapsing-remitting multiple sclerosis

Drugs

Source: ChEMBL

Found 32 unique drugs: ALEMTOZUMAB, BACLOFEN, BOTULINUM TOXIN TYPE A PURIFIED NEUROTOXIN, DICOLOFENAC, DIMETHYL FUMARATE, DULOXETINE, ECULIZUMAB, ERGOCALCIFEROL, FINASTERIDE, MIRABEGRON, MITOXANTRONE, NALTREXONE, NATALIZUMAB, OCRELIZUMAB, OFATUMUMAB, RILUZOLOMIDE, RITUXIMAB, SIMVASTATIN, Siponimod, TERIFLUONOMIDE

Showing 1 to 10 of 1,000 entries

Search:

Disease	Drug	Phase	Status	Type
multiple sclerosis	DALFAMPRIDINE	Phase IV	Completed	Small molecule
multiple sclerosis	DALFAMPRIDINE	Phase IV	Completed	Small molecule

Drug Information

Voltage-gated potassium channel blocker  
1 publication  
FDA

Open Targets

# Target-centric workflow



Which diseases are associated with a target?

What is the evidence supporting the association?

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



Open Targets

# Choose your favourite internet browser

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



# Searching for several targets at once



But we have more than one target we would like to pursue!

Yes, a list of 26, all possibly good candidates for IBD.

How can we get data for these in one go?



# Alternative ways to access the data

The screenshot shows a web browser window with the URL <https://www.targetvalidation.org/download> in the address bar. The page itself has a blue header with the Open Targets Platform logo and navigation icons. The main content area is titled "Data Download" and contains text explaining that all data from targetvalidation.org is available for download as compressed JSON files. It describes the availability of associations and evidence objects via API methods. Below this, a section titled "2017 Feb (Latest)" lists two download links: "Association objects (2016-12-09, 215MB, md5sum)" and "Evidence objects (2016-12-09, 4.35Gb, md5sum)".

All data from targetvalidation.org is available for download as compressed JSON files.

We provide downloads of all associations between target and disease calculated by the platform, as well as all the evidence used in calculating each associations. These are the same objects returned by the corresponding [/public/associations](#) and [/public/evidence](#) API methods. Head to the API documentation for further details.

**2017 Feb (Latest)**

- Association objects (2016-12-09, 215MB, md5sum)
- Evidence objects (2016-12-09, 4.35Gb, md5sum)

# Alternative ways to access the data



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

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

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



Open Targets

# What makes Open Targets unique?

Addressing many areas of human disease

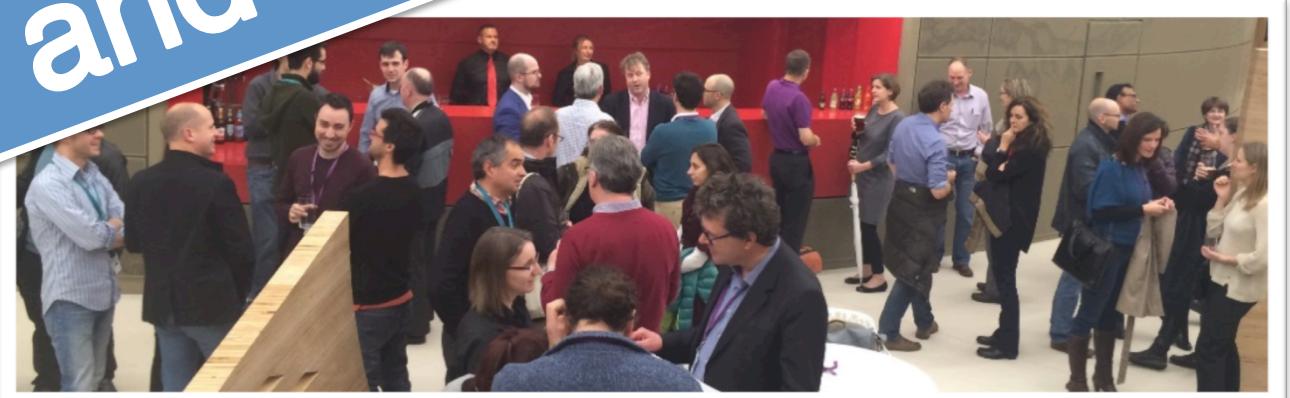
Putting our users first

Working genome wide

Bringing the partners together



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



# What is coming up next?

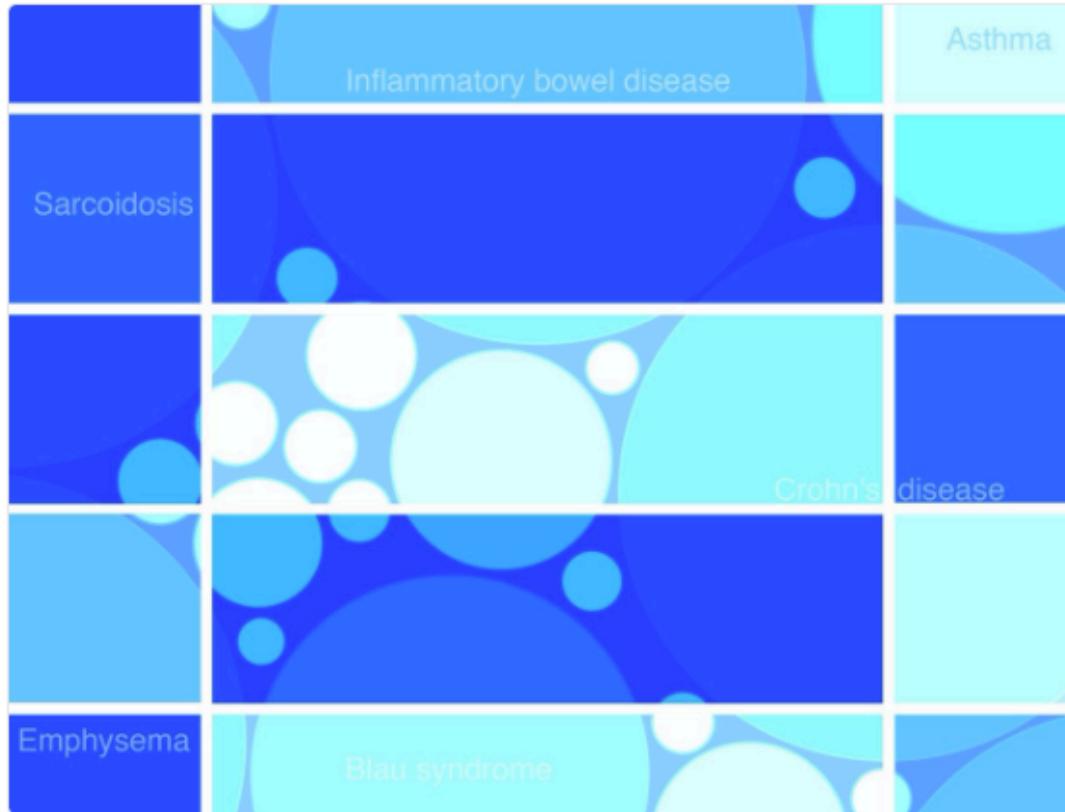




...And the 2017 Breakthrough Articles Award goes to...

#OpenTargets, as well as @MonarchInit & @denovodb! @NAR\_Open [buff.ly/2iGMXlc](http://buff.ly/2iGMXlc)

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



RETWEETS  
10

LIKES  
8



# Support, dissemination, GIFs



[support@targetvalidation.org](mailto:support@targetvalidation.org)



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



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



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



@targetvalidate



<http://imgur.com/a/JIDCP>

<http://imgur.com/a/LKDhp>



Open Targets