

# Open Targets Platform

Analysis and visualisation for early drug discovery

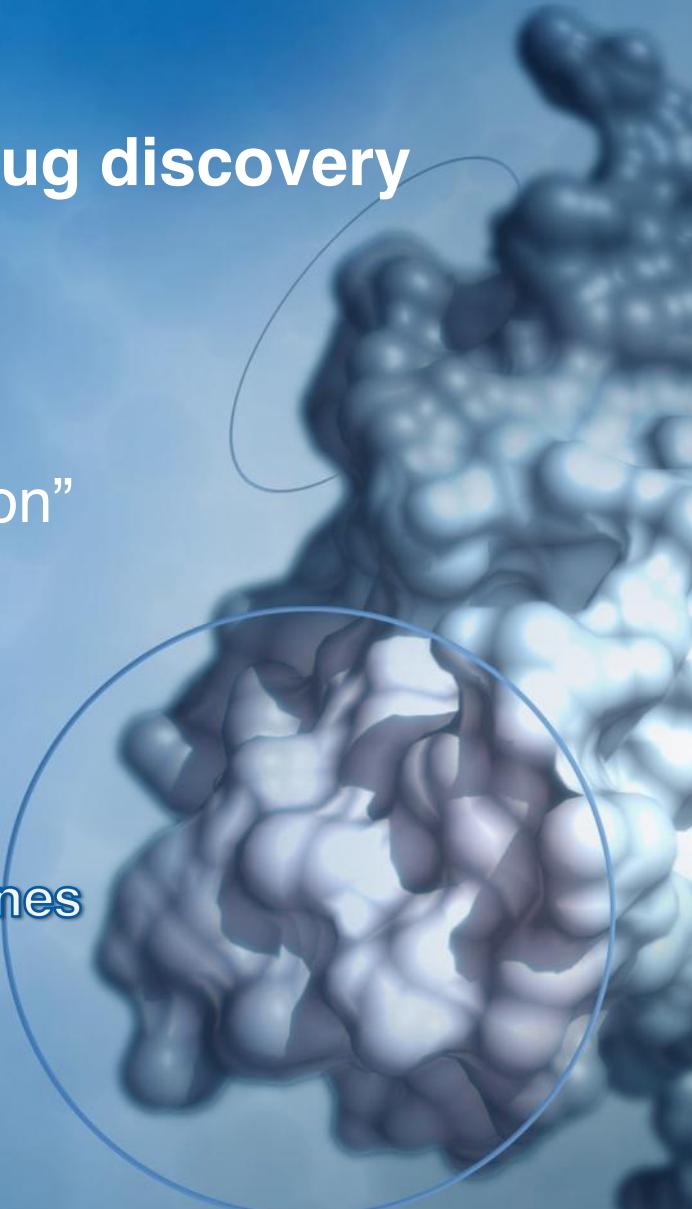
Part of “Introduction to Omics data integration”

**Denise Carvalho-Silva**

Core Bioinformatics and Computational Pipelines  
Open Targets, EMBL-EBI  
Wellcome Genome Campus



Open Targets



# Today 09:00-11:30

- Drug discovery: where to start
- The Open Targets Platform

*10:00-10:30 coffee break*

- Live demos and hands-on exercises
- Wrap up

# Objectives for the next 2.5 hours



**The Open Targets Platform:** analysis and visualisation

**The Open Targets Consortium**

**How to connect with us**



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

## 3. DELIVERY

# Where to start?

We are looking for better understanding of Hutchinson–Gilford progeria syndrome



Yes, and ultimately, interested in identifying drug targets for this extremely rare disease

Where could we start finding more about this disease and possible targets?

[All](#)[Images](#)[Videos](#)[News](#)[Shopping](#)[More](#)[Settings](#)[Tools](#)

About 120,000 results (0.63 seconds)



## [Hutchinson-Gilford progeria syndrome - Genetics Home Reference](#)

<https://ghr.nlm.nih.gov/condition/hutchinson-gilford-progeria-syndrome> ▾

Hutchinson-Gilford progeria syndrome is a genetic condition characterized by the dramatic, rapid appearance of aging beginning in childhood. Affected children ...

## [Progeria - Wikipedia](#)

<https://en.wikipedia.org/wiki/Progeria> ▾

Progeria is an extremely rare genetic disorder in which symptoms resembling aspects of aging ... It was also described independently in 1897 by Hastings Gilford. The condition was later named

**Hutchinson–Gilford progeria syndrome.**

[Sam Berns](#) · [Progeroid syndromes](#) · [Leon Botha](#) · [Failure to thrive](#)

## [Hutchinson-Gilford Progeria - NORD \(National Organization for Rare ...](#)

<https://rarediseases.org> › [For Patients and Families](#) › [Rare Disease Information](#) ▾

Progeria, or Hutchinson-Gilford progeria syndrome (HGPS), is a rare, fatal, genetic condition of childhood with striking features resembling premature aging.

**Format:** Summary ▾ **Sort by:** Most Recent ▾[Send to](#) ▾[Filters: Manage Filters](#)**Search Tip**

Sort by **Best Match** to display results from highest to lowest relevance to your search terms.

[Try it Now](#)**Search results****Items: 1 to 20 of 580**<< First < Prev Page  of 29 Next > Last >>

- [Nuclear lamins and progerin are dispensable for antioxidant Nrf2 response to arsenic and cadmium.](#)

1. Hashimoto K, Majumdar R, Tsuji Y.

Cell Signal. 2017 Feb 14;33:69-78. doi: 10.1016/j.cellsig.2017.02.012. [Epub ahead of print]

PMID: 28229933

[Similar articles](#)

- [Biomechanical Strain Exacerbates Inflammation on a Progeria-on-a-Chip Model.](#)

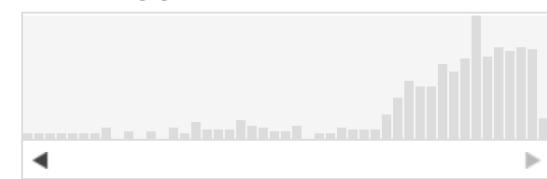
2. Ribas J, Zhang YS, Pitrez PR, Leijten J, Miscuglio M, Rouwkema J, Dokmeci MR, Nissan X, Ferreira L, Khademhosseini A.

Small. 2017 Feb 17. doi: 10.1002/smll.201603737. [Epub ahead of print]

PMID: 28211642

[Similar articles](#)

- [Metformin Alleviates Aging Cellular Phenotypes in Hutchinson-Gilford Progeria Syndrome Dermal Fibroblasts.](#)

**Results by year**Download CSV**Related searches**

[hutchinson-gilford progeria syndrome review](#)

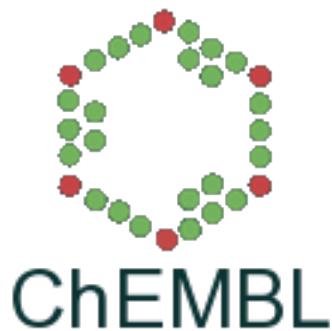
[lamin a truncation in hutchinson-gilford progeria](#)

[hutchinson-gilford progeria syndrome treatment](#)

**PMC Images search for Hutchinson-Gilford Progeria**

# Public databases for drug discovery

- European Bioinformatics Institute (EMBL-EBI)



- Fitting together like a jigsaw puzzle





I wish I could go to one place only  
and get as much information in an  
**easy** fashion



**Integration, Integration, Integration**  
X stop shop with  
extensive data, that I can  
trust and rank, with intuitive  
visualisation

That'd be fab! It'd be much quicker to carry  
out my experiments in the lab **identifying**  
**and prioritising** new targets.

# Open Targets Platform\*

- Developed by the Open Targets team at EMBL-EBI
- Intuitive GUI:
  - Identification of target–disease associations
  - Better understanding of disease biology
  - Access to comprehensive information on targets

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

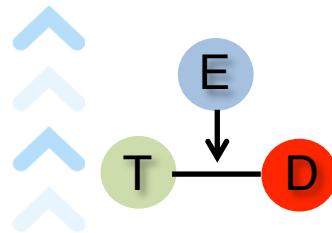
\* First release: December 2015

# Developing the Open Targets Platform

- 1 Interviewed ~100 people working with target identification at GSK
- 2 Made the selection of types and sources of data
- 3 Created a central place for storage (standard rules)
- 4 Integration: analysis, ranking, scoring
- 5 Designed a GUI with users for easy access of our analysis and visualisation

# Integration of existing data

## Public Databases and Pipelines



Open Targets experimental data: NEW  
Physiologically relevant and at scale

Oncology



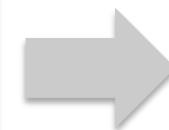
Immunity



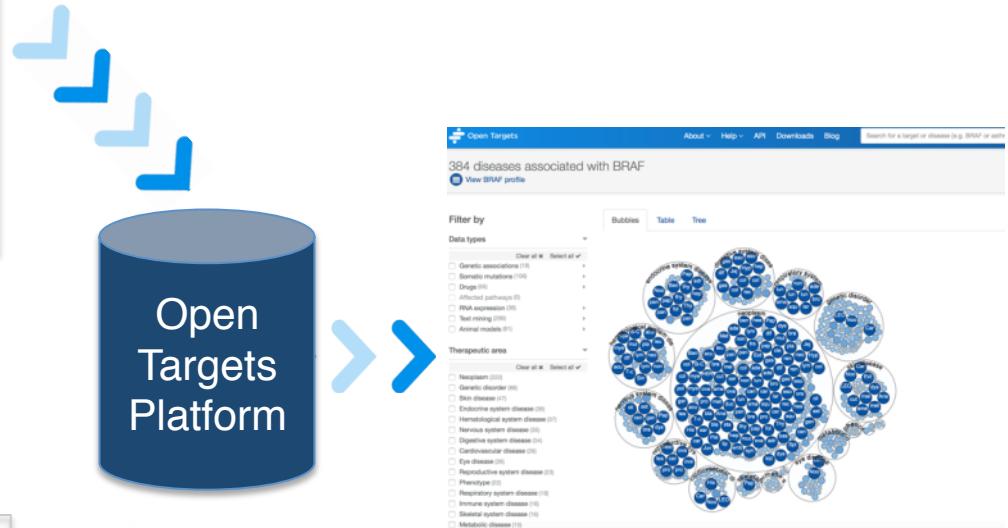
Cross-Disease



Neuro

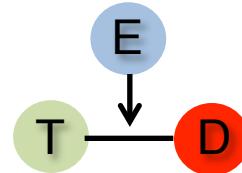


generated  
as we speak



Open Targets

# Evidence from public databases

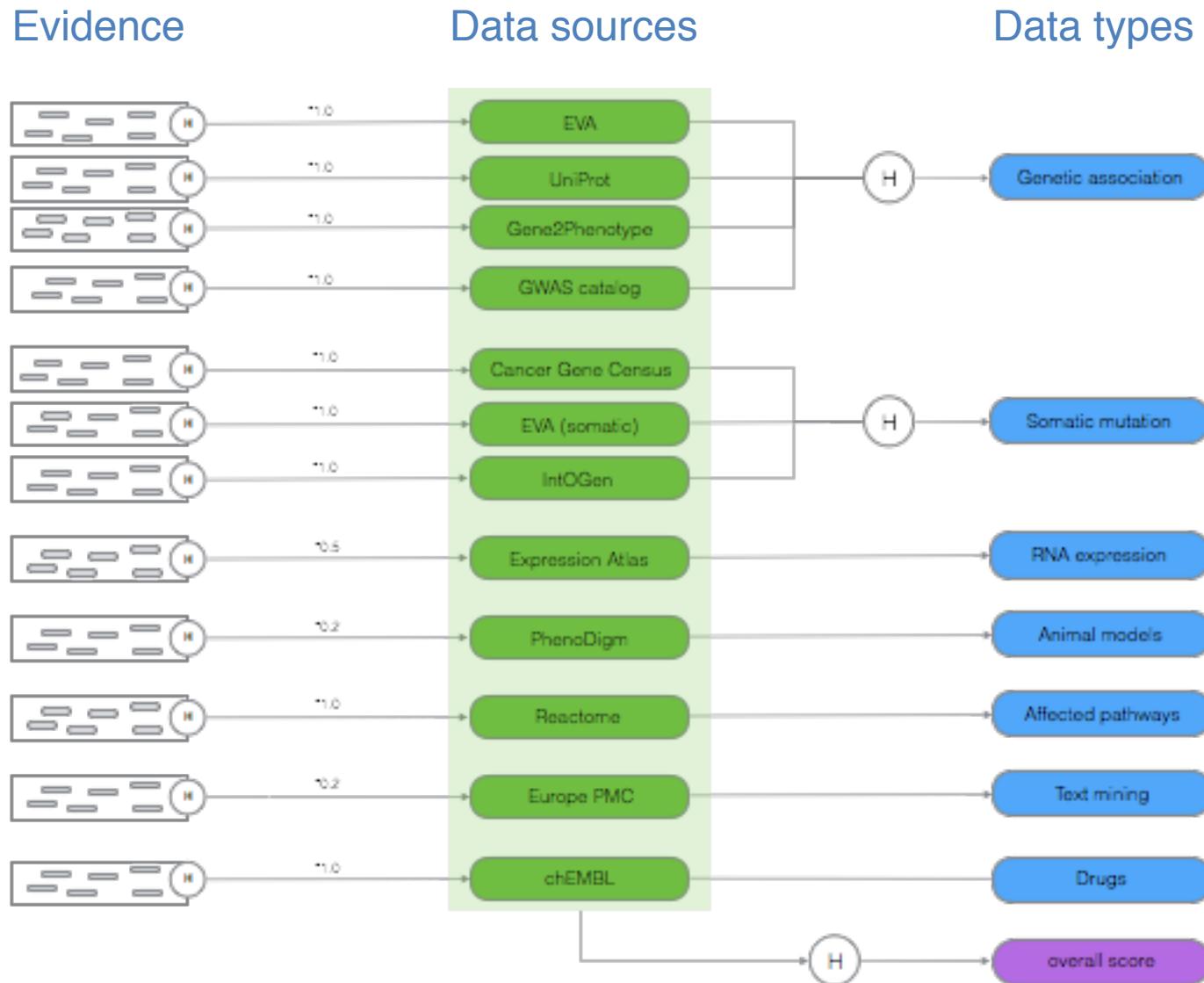


- Similar data sources are grouped into data types

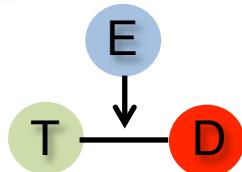
Data sources	Data types
GWAS catalog, UniProt, EVA, G2P	Genetic associations
Cancer Gene Census, EVA, IntOgen	Somatic mutations
Expression Atlas	RNA expression
ChEMBL	Drugs
Reactome	Affected pathways
Europe PMC	Text mining
PhenoDigm	Animal models
<b>Your favourite data?</b>	<b>Let us know!</b>

# Confidence score

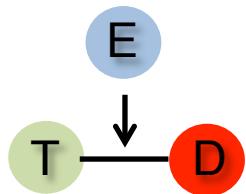
Oliver Stegle's team (EMBL-EBI)



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



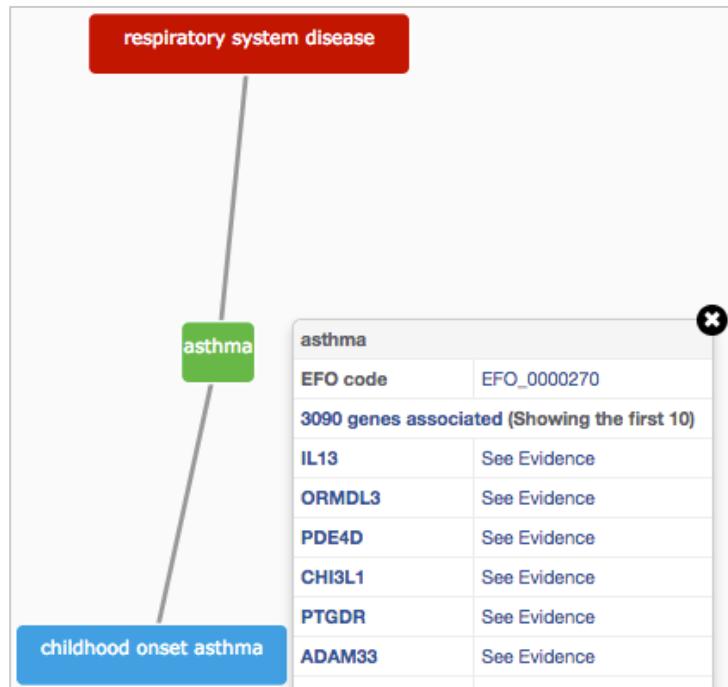
# How do we associate diseases and phenotypes w/ targets?



- 1 Map disease/phenotypes to an ontology using EFO and HPO terms
- 2 Use genes as proxies for our targets
- 3 Create target-disease evidence JSON documents
- 4 Calculate for each supporting evidence the likelihood of target T being associated with disease D
- 5 Compute integrated target-disease scores at the levels of data source, data type and overall score

# Experimental Factor Ontology\* (EFO)

- Smart dictionary → relationships between entities
- EFO: organise experimental variables (e.g. diseases)



hierarchy  
+  
controlled vocabulary

Easier and automatic integration  
Richer of annotation  
Consistency

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

# A use case

[Am J Hum Genet.](#) 2006 Jun; 78(6): 1011–1025.

Published online 2006 Apr 25. doi: [10.1086/504300](https://doi.org/10.1086/504300)

PMCID: PMC1474084

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

[Lude Franke](#),<sup>1</sup> [Harm van Bakel](#),<sup>1</sup> [Like Fokkens](#),<sup>1</sup> [Edwin D. de Jong](#),<sup>2</sup> [Michael Egmont-Petersen](#),<sup>3</sup> and [Cisca Wijmenga](#)<sup>1</sup>

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

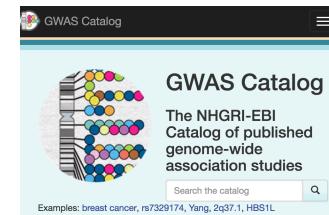
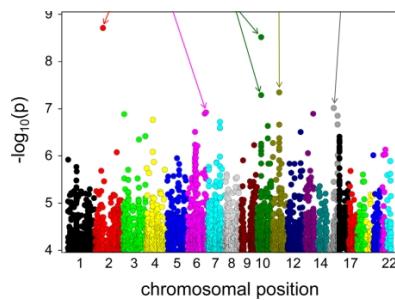
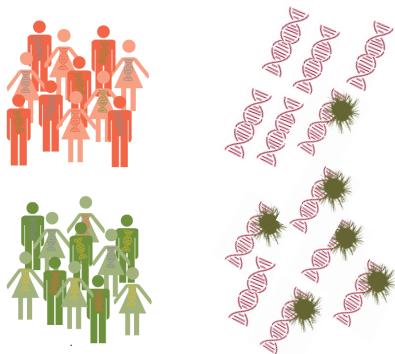
# Tutorial 1\*: Renal cell carcinoma

- How many targets are associated with this disease?
- Which data sources are used to support this association?
- Which targets are based on “Genetic association” evidence only?
- Which drugs are known to be in clinical trials for this disease?
- Which diseases are related to renal cell carcinoma?

\*coursebook: pages 9 -15

# Data sources: GWAS catalog

- Genome Wide Association Studies
- Array-based chips → genotyping 100,000 SNPs genomewide



<https://www.ebi.ac.uk/gwas/>

# Data sources: UniProt

- Protein: sequence, annotation, function



- Manual curation of variants: coding region, seen in patients

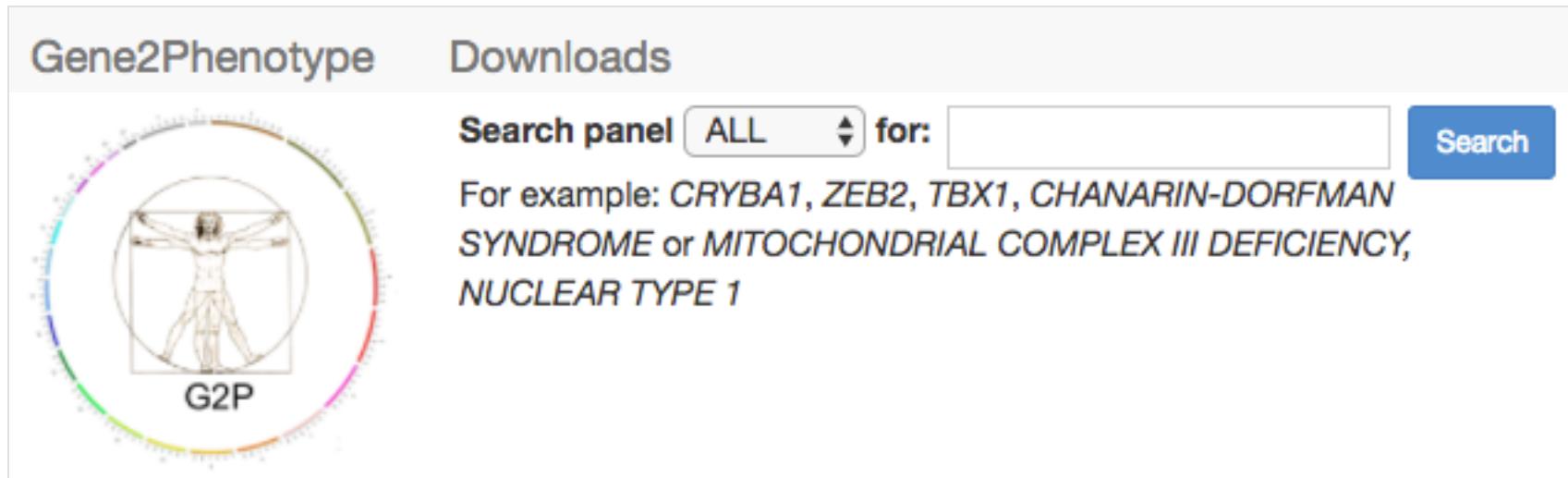
# Data sources: EVA

- Germline and somatic variants
- With ClinVar information for rare diseases

The screenshot shows the European Variation Archive (EVA) website. The top navigation bar includes links for Home, Submit Data, Study Browser, Variant Browser, Clinical Browser (which is highlighted in dark blue), GA4GH, API, FAQ, and Feedback. Below the navigation is a search bar with a placeholder 'Search' and a 'Go' button. To the left of the main content area is a sidebar with a 'Filter' section containing 'Reset' and 'Submit' buttons, a 'Position' dropdown set to 'GRCh37', a 'Filter By:' dropdown set to 'Chromosomal', and a text input field showing '2:48000000-49000000'. The main content area is titled 'ClinVar Browser' and displays a table of 960 records. The table has columns for Position, Affecte..., Most Severe Consequence..., Trait, Clinical Significance, and ClinVar ID. The first few rows show variants for the MSH6 gene on chromosome 2.

...	Posi...	Affecte... i	A...	Most Severe Consequence...	Trait	Clinical Significance	ClinVar ...
2	480...	MSH6	T/G	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	G/A	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	C/T	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	C/T	upstream_gen...	Lynch synd...	Benign	RCV000...
2	480...	MSH6	G/T	5_prime_UTR...	Lynch synd...	Uncertain s...	RCV000...
2	480...	MSH6	G/T	5_prime_UTR...	Hereditary ...	conflicting ...	RCV000...

# Data sources: Gene2Phenotype



Gene2Phenotype

Downloads

Search panel ALL for:  Search

For example: *CRYBA1, ZEB2, TBX1, CHANARIN-DORFMAN SYNDROME or MITOCHONDRIAL COMPLEX III DEFICIENCY, NUCLEAR TYPE 1*

G2P

- Variants, genes, phenotypes in developmental disorders
- Literature curation → consultant clinical geneticists in the UK

<https://www.ebi.ac.uk/gene2phenotype>

# Data sources: The Cancer Gene Census

Census

Breakdown

Abbreviations

*The cancer Gene Census is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer. The original census and analysis was published in [Nature Reviews Cancer](#) and supplemental analysis information related to the paper is also available.*

- Genes with mutations causally implicated in cancer
- Gene associated with a cancer plus other cancers associated with that gene

<http://cancer.sanger.ac.uk/cosmic/census>

# Data sources: IntOGen

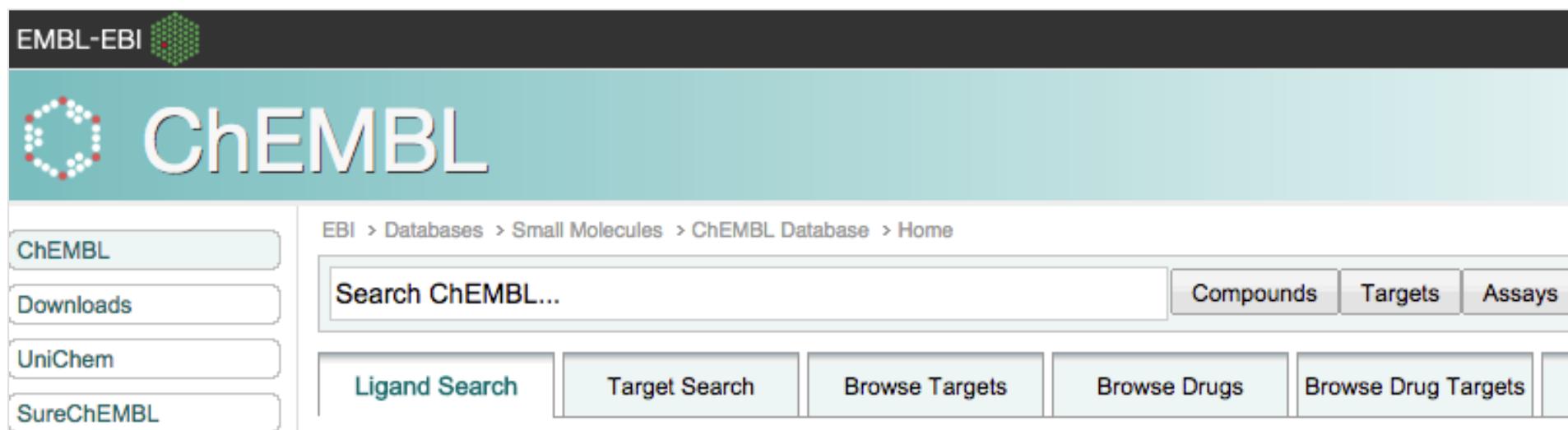
The screenshot shows the homepage of the intOGen website. At the top is a navigation bar with an orange gradient background. From left to right, it contains: the intOGen logo (a stylized orange 'i' icon followed by the word 'intOGen'), a 'Search' button with a magnifying glass icon, a 'Downloads' button with a download arrow icon, an 'Analysis' button with a gear icon, an 'About' button with a speech bubble icon, and a 'Sign In' button with a user profile icon.

The main content area features the intOGen logo again, this time larger and centered. To its right, the text 'Integrative Onco Genomics' is written in orange capital letters.

- Genes and somatic (driver) mutations
- Involvement in cancer biology

<https://www.intogen.org/search>

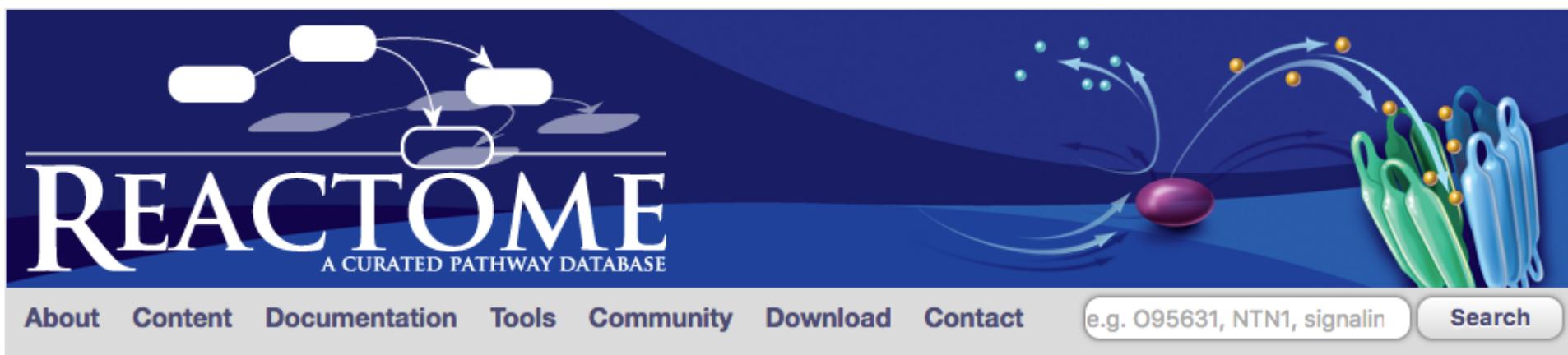
# Data sources: ChEMBL



The screenshot shows the ChEMBL database homepage. At the top left is the EMBL-EBI logo. The main header features the ChEMBL logo (a stylized circular pattern) and the word "ChEMBL". Below the header is a navigation bar with links to "ChEMBL", "Downloads", "UniChem", and "SureChEMBL". To the right of the navigation bar is a search bar containing the placeholder "Search ChEMBL...". Above the search bar is a breadcrumb trail: "EBI > Databases > Small Molecules > ChEMBL Database > Home". To the right of the search bar are three buttons: "Compounds", "Targets", and "Assays". Below the search bar are five buttons: "Ligand Search", "Target Search", "Browse Targets", "Browse Drugs", and "Browse Drug Targets".

- Known drugs linked to a disease and a known target
- FDA approved for marketing or clinical trials

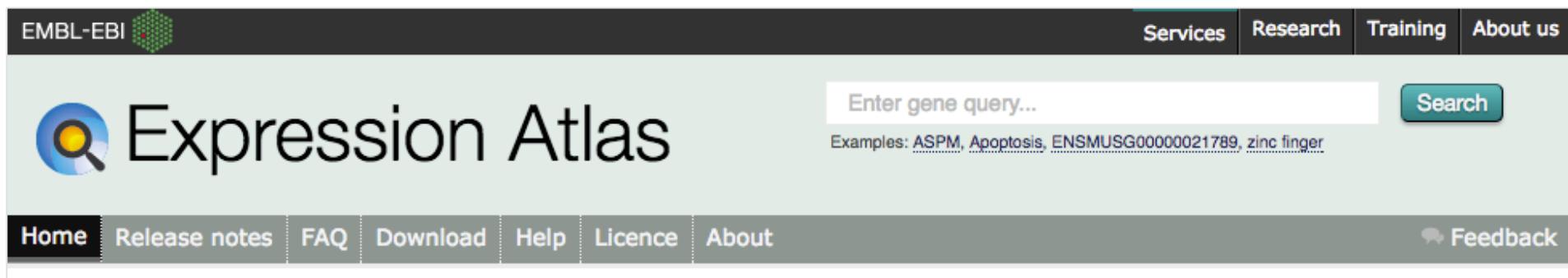
# Data sources: Reactome

The image shows the Reactome homepage. At the top, there is a dark blue header with the word "REACTOME" in large white letters, followed by "A CURATED PATHWAY DATABASE". Below the header, there is a navigation bar with links: "About", "Content", "Documentation", "Tools", "Community", "Download", and "Contact". To the right of the navigation bar is a search bar containing the placeholder text "e.g. O95631, NTN1, signalin" and a "Search" button. The main content area features a blue background with abstract biological illustrations, including a network of nodes and arrows at the top left and a green, branching structure with yellow spheres at the bottom right.

- Biochemical reactions and pathways
- Manual curation of pathways affected by mutations

<http://www.ebi.ac.uk/training/online/course/reactome-quick-tour>

# Data sources: Expression Atlas

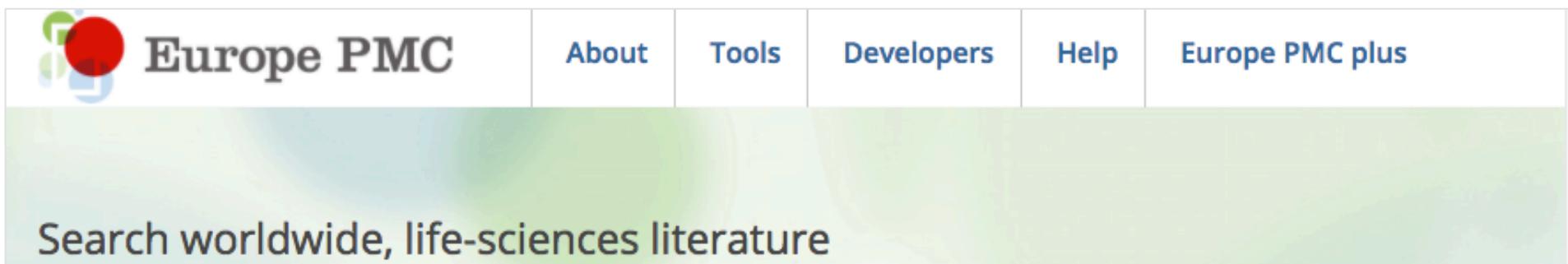


The screenshot shows the Expression Atlas website. At the top, there is a dark header bar with the EMBL-EBI logo on the left and navigation links for Services, Research, Training, and About us on the right. Below the header, the main title "Expression Atlas" is displayed, featuring a magnifying glass icon next to the word "Expression". To the right of the title is a search bar with the placeholder "Enter gene query..." and a "Search" button. Below the search bar, there is an example query: "ASPM, Apoptosis, ENSMUSG00000021789, zinc finger". A navigation menu at the bottom includes links for Home, Release notes, FAQ, Download, Help, Licence, and About, with the "Home" link being the active one. On the far right of the menu is a "Feedback" link.

- Baseline expression for human genes
- Differential expression (healthy versus diseases tissues)

<http://www.ebi.ac.uk/training/online/course/expression-atlas-quick-tour-0>

# Data sources: Europe PMC



A screenshot of the Europe PMC website. At the top left is the Europe PMC logo, which consists of a stylized 'E' made of overlapping colored circles (blue, green, red) followed by the text 'Europe PMC'. To the right of the logo is a horizontal menu bar with five items: 'About', 'Tools', 'Developers', 'Help', and 'Europe PMC plus'. Below the menu is a large search bar with the placeholder text 'Search worldwide, life-sciences literature'.

- Mining titles, abstracts, full text
- Target and disease: co-occurrence in the same sentence

<http://www.ebi.ac.uk/training/online/course/europe-pmc-quick-tour-0>

# Data sources: PhenoDigm

The screenshot shows the homepage of the PhenoDigm website. At the top is a dark header bar with the Wellcome Trust Sanger Institute logo on the left. To the right of the logo is a blue navigation bar with the following items: "ABOUT" with a dropdown arrow, "Who we are", "Careers", "Study", "Sex in Science", "Groups", "Campus", and a magnifying glass icon for search. Below the header is a large, bold title: "Welcome to PhenoDigm (PHENOtype comparisons for Disease and Gene Models)". Underneath the title is a horizontal menu bar with three items: "Diseases" (which is highlighted in blue), "Tissue phenotype associations", and "Secondary phenotypes".

- Semantic approach to associate mouse models with diseases

<http://www.sanger.ac.uk/resources/databases/phenodigm/>

# Coffee break

# Tutorial 2\*: MET-papillary renal cell carcinoma

- 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?

# Tutorial 3\*: your list of targets

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?

Exercises  
Pages 29

# Wrap up

Looking for omics data integration?

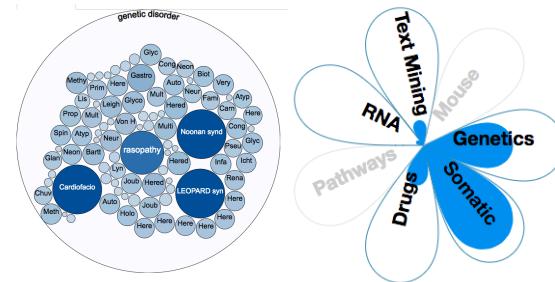
Open Targets Platform is the place:

- Target-disease associations
- Disease and target information
- Easy-to-use GUI
- Programmatic access

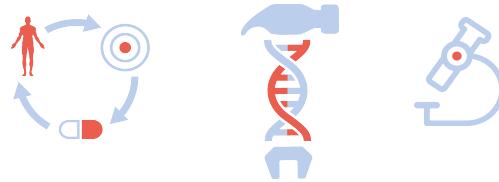
Oh Yes!  
And it is 100% free  
and open source

# Open Targets Consortium: two\* major areas of work

## Core bioinformatics pipelines



## Experimental projects



\* Concurrent

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

# Early stages in the 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.

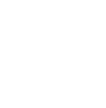
## 2. DEVELOPMENT



### 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



1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100



### 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

# 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?

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

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



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

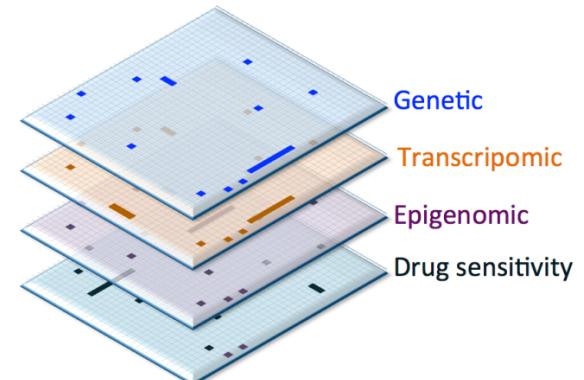
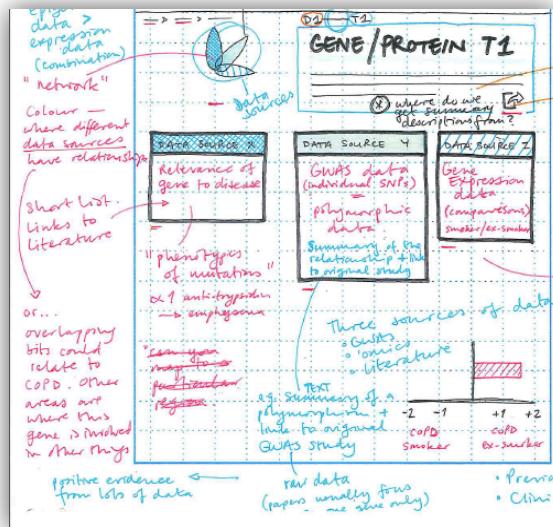
# Open Targets is unique

Addressing all areas of human disease

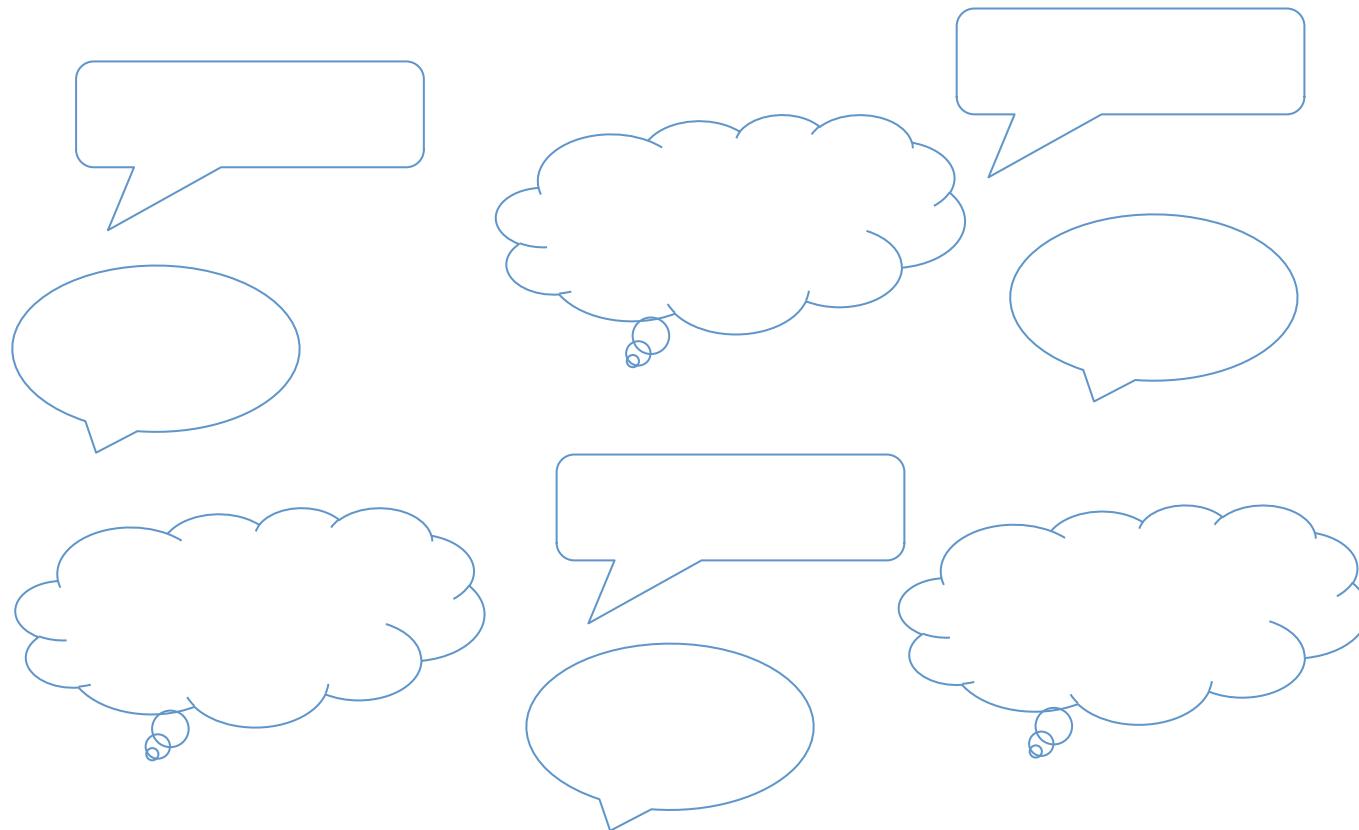
Putting our users first

Working genome wide

Bringing the partners together



# Your take home message



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

<sup>1</sup>Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, <sup>2</sup>GSK, Medicines Research Center, Gunnels Wood Road, Stevenage, SG1 2NY, UK, <sup>3</sup>Biogen, Cambridge, MA 02142, USA, <sup>4</sup>European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK, <sup>5</sup>Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK and <sup>6</sup>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](mailto:support@targetvalidation.org)

# Get in touch



@targetvalidate



support@targetvalidation.org



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



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



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