

# **Mining gene-disease associations with Open Targets**



**European Molecular Biology Laboratory  
European Bioinformatics Institute  
17<sup>th</sup> March 2017**

**Denise Carvalho-Silva**

**Open Targets  
Wellcome Genome Campus  
Hinxton, United Kingdom**

# Notes

This webinar is based on the December 2016 release of our Open Targets Platform.

Some useful links:

1) About the Open Targets Consortium

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

2) About the Open Targets Platform

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

3) Our publication

[www.bit.ly/OpenTargets](http://www.bit.ly/OpenTargets)

4) Details on the latest Platform release

<https://blog.opentargets.org/open-targets-platform-our-new-release-is-out-2/>

Feel free to tackle questions relative to your own research instead of following the ones provided in this coursebook.

Questions or Feedback?

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

## **TABLE OF CONTENTS**

<b>OVERVIEW.....</b>	<b>4</b>
<b>INTRODUCTION TO OPEN TARGETS.....</b>	<b>5</b>
<b>OPEN TARGETS PLATFORM: LIVE DEMOS.....</b>	<b>8</b>
Demo 1.....	9
Demo 2.....	16
Demo 3.....	28
<b>HANDS-ON EXERCISES.....</b>	<b>25</b>
Exercise 1.....	25
Exercise 2.....	26
Exercise 3.....	27
<b>EXTRA HANDS-ON EXERCISES .....</b>	<b>33</b>
Exercise 4.....	33
Exercise 5.....	34
<b>QUICK GUIDE TO DATABASES .....</b>	<b>35</b>

## **OVERVIEW**

Open Targets is a public-private initiative to generate evidence on the validity of therapeutic targets based on genome-scale experiments and analysis. We are working to create an R&D framework that applies to a wide range of human diseases, and we want to share this data openly with the scientific community.

The consortium was launched in March 2014 under the name of Centre for Therapeutic Open Targets (CTTV) and started with GlaxoSmithKline (<http://www.gsk.com/>), the Wellcome Trust Sanger Institute (<http://www.sanger.ac.uk/>) and the European Bioinformatics Institute (<http://www.ebi.ac.uk/>). In February 2016, a fourth institution namely Biogen (<https://www.biogen.com/>) joined the initiative and the consortium was rebranded to Open Targets in April 2016.

In the process of drug discovery, the *validation* of a target refers to the creation of a specific entity that modulates that target's activity to provide therapeutic benefit to individuals with a disease. The ultimate validation of a target is the creation of an effective therapeutic molecule. This is a long and costly endeavour with more failures than successes. The goal of Open Targets is to transform this process by predicting if the modulation of a target is likely to provide therapeutic benefit. This would be done much earlier in the drug discovery process than is currently possible and far in advance of having a final, approved medicine.

Points covered in this workshop:

- The projects of Open Targets consortium
- An introduction to the Open Targets Platform
- Browsing the Platform
- Pointing to alternative ways to access the data

## **INTRODUCTION TO OPEN TARGETS**

Open Targets employs large-scale human genetics and genomics data to change the way drug targets are identified and validated. We have established a set of projects to develop both the data and analytical processes that implicate targets as valid, and the core platform to provide the information to a diverse audience of users.

The core bioinformatics team develops pipelines and a database to integrate existing target data. The core also designed, created and maintains the Open Targets Platform, a public web portal to serve the integrated data and views.

Our experimental projects focus on providing insights in the identification of targets relevant to key therapeutic areas namely:

- Oncology
- Inflammatory bowel diseases (IBD)
- Respiratory disease
- Inflammation and immunity
- Neurodegenerative diseases

We also aim to develop standard epigenome profiles of cell models in use within the pharmaceutical industry and academia and establish a systematic approach for the determination of human biological and disease relevance.

More details can be found in our [Projects](#) page.

### **Retrieving data from Open Targets with our Platform**

The Open Targets Platform is a web application that integrates and displays publicly available biological data to foster the discovery and prioritisation of targets for new therapies. We use data sources as diverse as Gene2Phenotype, IntOGen, GWAS, UniProt, ChEMBL, Expression Atlas, Cancer Census, Reactome and EuropePMC as pieces of evidence to support target-disease associations. The associations are scored using objective statistical and computational techniques.

In our release (December 2016), the Platform provides information on 31,071 targets; 8,659 diseases; 4.9 million evidence; and 2.5 million target-disease associations.




In addition to the web application, we include the data dumps, a REST API and a Python client.

The Open Targets Platform is aimed at users from both academia and industry, whether they want to browse a target on a gene by gene (or disease by disease) basis, carry out more complex queries using the API, or download all evidence and association objects for downstream analyses.

### **Synopsis: what can I do with the Open Targets Platform?**

- Find out which targets are associated with a disease
- Explore the evidence supporting this target-disease association
- Export a table with the FDA drugs currently in clinical trials
- Discover if there other diseases associated with a given target
- Get the association of a target with diseases from different therapeutic areas
- Find target specific information, such as baseline expression, protein structure, alternatively spliced transcripts, gene trees
- Get disease target specific information, such as a classification based on the ontology of the disease and the drugs mapped to it

## Help documentation and support

-  [Data sources](#) in the Open Targets Platform
-  View our [FAQs](#)
-  [Email us](#)

## Connect with us

- ❖ [Open Targets Blog](#)
- ❖ Follow us on [Twitter](#)
- ❖ Check our page on [Facebook](#) and [LinkedIn](#)

## Further reading

Koscielny, G. *et al.* Nucleic Acids Res (2017 Database Issue):  
<http://nar.oxfordjournals.org/content/early/2016/11/29/nar.gkw1055>

A breakthrough article from Nucleic Acids Research:  
<http://www.narbreakthrough.com/>

## **OPEN TARGETS PLATFORM: LIVE DEMOS**

We will guide you through the website using multiple sclerosis, as an example and exploring the *CD86* gene.

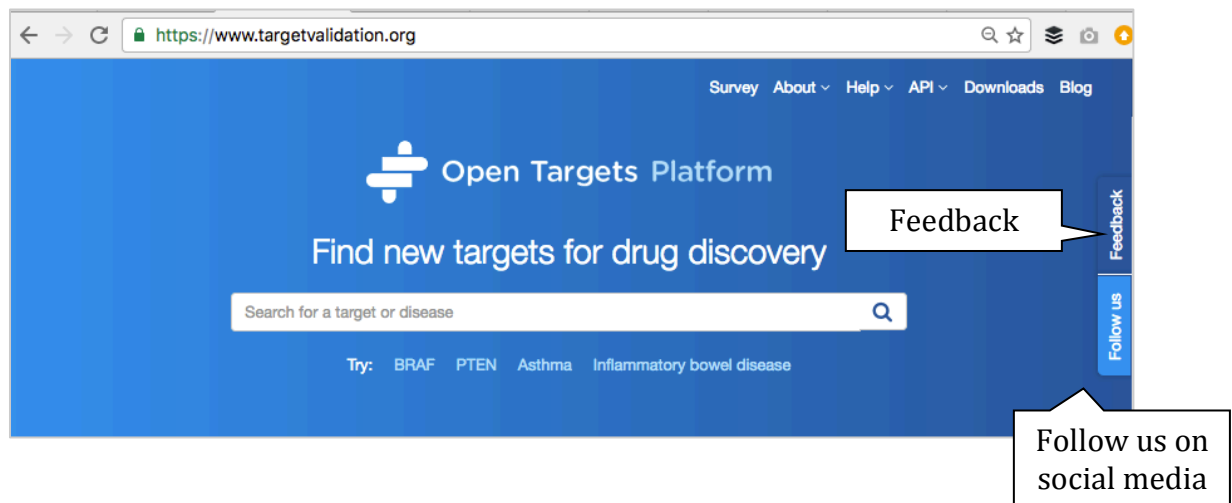
The following points will be addressed during the walkthrough:

- How to find targets associated with multiple sclerosis
- How to filter down the number of targets based on specific type of evidence
- How to find the data sources used to support the target-disease association
- How to look for other diseases (than multiple sclerosis) associated with a given target
- How to visualise a target in a browser-like view
- How to find out how strong the association between a target and a disease is
- How to find drugs currently in clinical trials
- How to filter **our** associations by **your** list of target genes

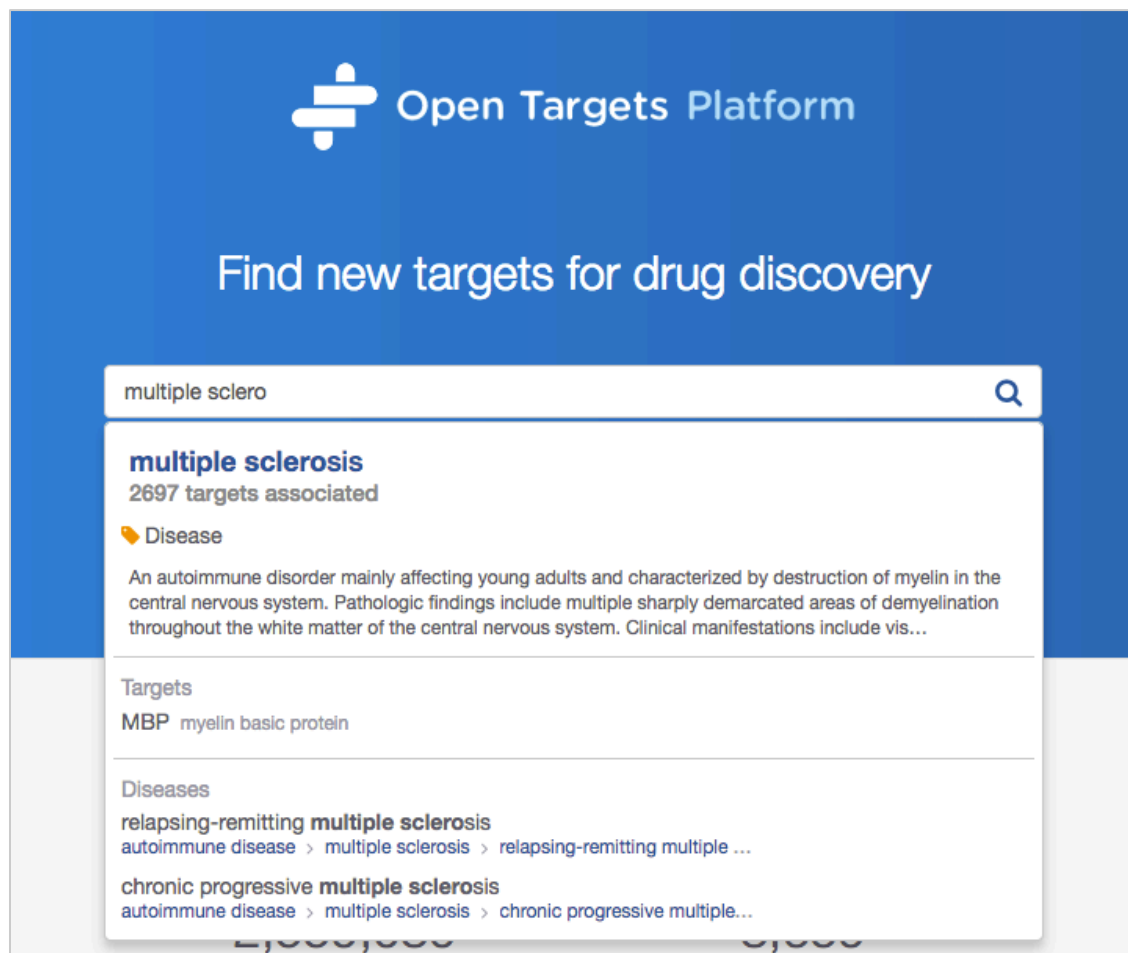


## Demo 1

Go to [www.targetvalidation.org](https://www.targetvalidation.org) and search for multiple sclerosis.



Select the first (best) hit:



You will see a page like this:

Total number of targets associated with multiple sclerosis

2697 targets associated with multiple sclerosis

View disease profile

Filter by

Data types

- ☒ Genetic associations (181)
- ☐ Somatic mutations (0)
- ☐ Drugs (129)
- ☐ Affected pathways (0)
- ☐ RNA expression (1k)
- ☐ Text mining (1k)
- ☐ Animal models (4)

Pathway types

- ☐ Immune System (602)
- ☐ Signal Transduction (473)
- ☐ Metabolism (329)

Filter the results

Showing 1 to 50 of 2,697 targets

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Target name
TNFRSF1A									TNF receptor superfamily ...
MS4A1									membrane spanning 4-do...
KCNB2									potassium voltage-gated ...
VDR									vitamin D (1,25- dihydroxy...
PTGS2									prostaglandin-endoperoxi...
S1PR1									sphingosine-1-phosphate...
NR3C1									nuclear receptor subfamili...
IFNAR1									interferon alpha and beta ...
CD52									CD52 molecule

Data types (Genetic Associations, Drugs, etc)

The current release of the Open Targets Platform (December 2016) lists 2697 targets associated with multiple sclerosis.

The data types supporting these results are based on Genetic association, Drugs, RNA expression, Text mining, and Animal models.

Check our help page to find out more about our data sources: [https://targetvalidation.org/data\\_sources](https://targetvalidation.org/data_sources).

You can filter the number of associations by 'Data types', 'Pathway types', 'Target class' and by uploading 'Your target list':

#### A) Data types

- Genetic associations (e.g. GWAS catalog)
- Somatic mutations (e.g. Cancer Gene Census, EVA)
- Drugs (from ChEMBL)
- Affected Pathways (from Reactome)
- RNA expression (from Expression Atlas)
- Text mining (from EuropePMC)
- Animal models (from PhenoDigm)

B) Pathway types

Signal Transduction

Metabolism

...

C) Target class

Enzyme

Membrane receptor

...

D) Your target list

Upload your own list of genes (in official gene symbols from HGNC or Ensembl Gene IDs)

What are **Data types**, **Pathway types** and **Target class**?

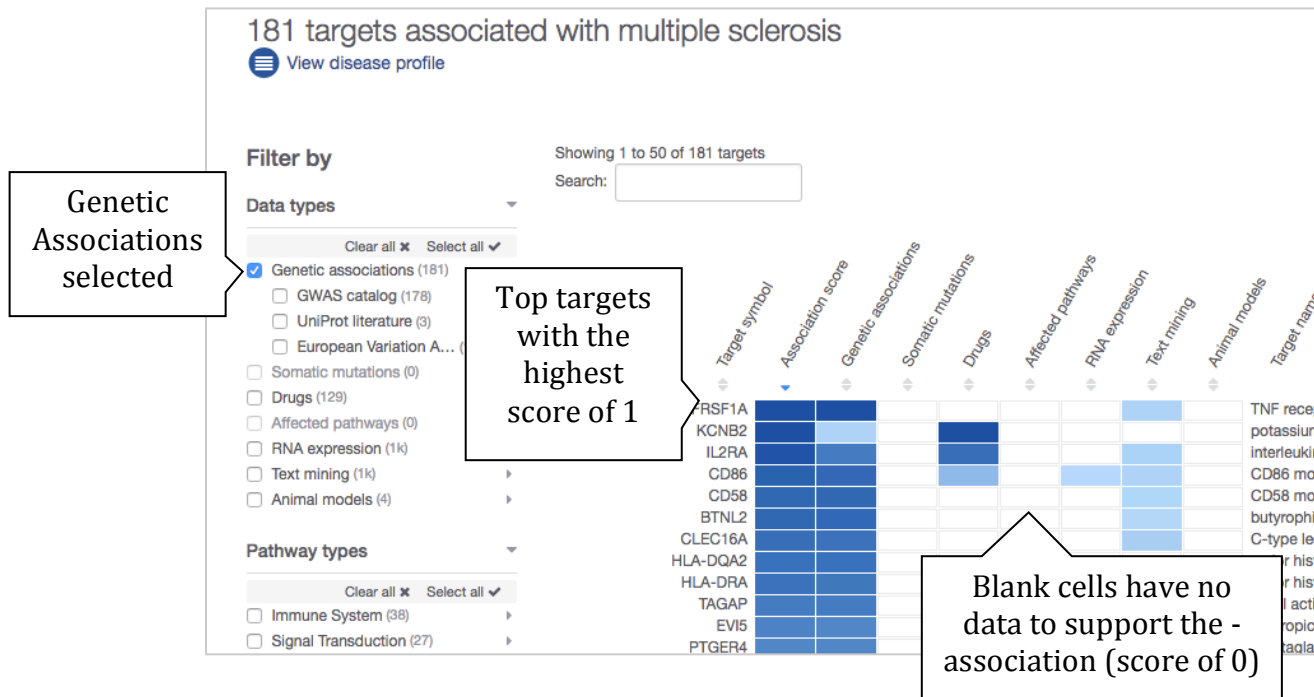
We collect data from various sources and combine them into categories called Data types. Example of data sources are GWAS catalog and UniProt, both combined into Data types. Note that data from an individual source can contribute to different Data types, e.g. data from EVA is observed in two data types, Genetic associations and Somatic mutations.

'Pathway types' are defined by Reactome:

(<http://www.reactome.org/>),

whereas the categories within 'Target class' are defined by ChEMBL (<https://www.ebi.ac.uk/chembl/>).

Let's now filter the data to focus on 'Genetic associations' (Data type) only. The number of targets goes down to 181:



These are targets associated with multiple sclerosis based on genetic variants only. Genetic variants are SNPs (single nucleotide polymorphisms or mutations, or short indels) from the GWAS catalog, UniProt and EVA (European Variation Archive) sources only.

The table is sorted by default with the best hit on the top of the table i.e. *TNFRSF1A*. The best hit is the target that contains the highest (overall) association score. Different weight is given to different data types when computing the score:

From highest to lowest weight of all data types:

Genetic association = somatic mutations = drugs = pathways

RNA expression

Animal models = Text mining

You can sort the table by alphabetical order of the list of targets, or by the association score (either overall or per data type e.g. Genetic associations, Drugs, Text mining, etc). The association score varies from 0 to 1, the closer to 1 the stronger the association. This score is computed for each piece of evidence that is used to support the association. Individual scores within data sources and data types are combined to give the overall score ('Association score' column in the table below):

Click here to sort the results by alphabetical order of the gene symbols

Click on the arrows to sort the results by score values of individual data types e.g. Text mining.

Target symbol	Association score	Genetic association	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Target name
TNFRSF1A									TNF receptor superfamily ...
KCNB2									potassium voltage-gated ...
IL2RA									

To see which drugs are used in the clinical trials with patients with multiple sclerosis (or children terms of this disease), click on the 'View disease profile':

Open Targets Platform

About ▾ Help ▾ API ▾ Download

2697 targets associated with multiple sclerosis

[View disease profile](#)

Let's now expand the tab 'Drugs' to get a list of all drugs (n = 40 unique drugs in the December release) in different phases of clinical trials. You can filter (and sort) the table by disease, phase of clinical trial (e.g. IV, the most advanced phase), class of the target (e.g. membrane receptor), etc. You can also download this table in csv (comma separated value):

**Drugs**

Source: ChEMBL

Found 40 unique drugs: [ABATACEPT](#) [ALEMTUZUMAB](#) [BACLOFEN](#) [BOTULINUM TOXIN TYPE A PURIFIED NEUROTOXIN COMPLEX](#) [CHOLECALCIFEROL](#) [CORTICOTROPIN](#) [DACLIZUMAB](#) [DALFAMPRIDINE](#) [DEXTROAMPHETAMINE](#) [DICLOFENAC](#) [DIMETHYL FUMARATE](#) [DULOXETINE](#) [ECULIZUMAB](#) [FINGOLIMOD](#) [GSK239512](#) [HYDROXYUREA](#) [INTERFERON BETA-1A](#) [INTERFERON BETA-1B](#) [LAMOTRIGINE](#) [LISDEXAMFETAMINE](#) [MEMANTINE](#) [METHYLPHENIDATE](#) [METHYLPREDNISOLONE](#) [MIRABEGRON](#) [MITOXANTRONE](#) [MYCOPHENOLATE MOFETIL](#) [NALTREXONE](#) [NATALIZUMAB](#) [NERISPIRIDINE](#) [OCRELIZUMAB](#) [OXCARBAZEPINE](#) [PEGINTERFERON BETA-1A](#) [PREDNISOLONE](#) [PREDNISOLONE PHOSPHORIC ACID](#) [PREDNISONE](#) [RILUZOLE](#) [RITUXIMAB](#) [SIMVASTATIN](#) [Siponimod](#) [TERIFLUNOMIDE](#)

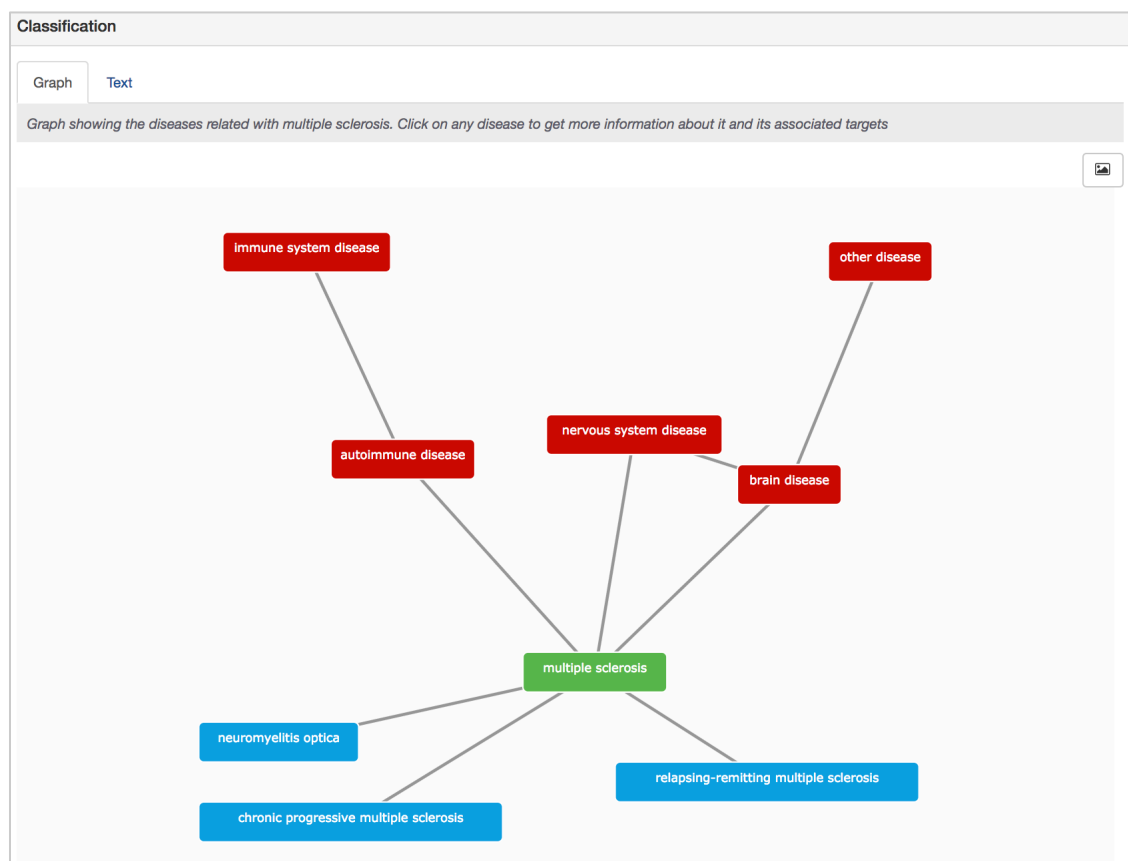
Showing 1 to 10 of 1,000 entries

Search:

[Download](#)

Drug Information							Gene-Drug Evidence	
Disease	Drug	Phase	Status	Type	Mechanism of action	Activity	Target class	Evidence source
relapsing-remitting multiple sclerosis	<a href="#">INTERFERON BETA-1A</a>	Phase IV	Completed	Protein	Interferon alpha/beta receptor agonist <a href="#">1 publication</a> <a href="#">DailyMed</a>	agonist	Membrane receptor	Curated from Clinical Trials Information

Scroll down to view the disease ontology (disease relationship) in the 'Classification' tab. The data on the relationship of diseases comes from the EFO (Experimental Factor Ontology).



Multiple sclerosis is the disease of interest in this tutorial, and is represented in green. Red nodes correspond to parental terms in relation to multiple sclerosis, whereas its children terms (e.g. chronic

progressive multiple sclerosis) are shown in blue. Click on any of disease names to get the targets associated with them:

relapsing-remitting multiple sclerosis

relapsing-remitting multiple sclerosis

EFO code

EFO\_0003929

215 genes associated (Showing the first 10)

CD52

See Evidence

S1PR1

See Evidence

MS4A1

See Evidence

S1PR4

See Evidence

ITGB7

See Evidence

S1PR5

See Evidence

KEAP1

See Evidence

S1PR2

See Evidence

ITGB1

See Evidence

IFNAR2


See Evidence

By using the EFO parent-child (subclass of) relationships, we derive new associations that may not have direct evidence. For instance, IBD is an autoimmune disease and the direct evidence of targets associated to IBD are propagated to the higher autoimmune level to allow users to find common targets across groups of related diseases (e.g. Ulcerative Colitis, Crohn's disease and IBD). In EFO, 'asthma' is a 'respiratory system disease' and 'childhood onset asthma' is a subclass of 'asthma'. Both evidence from 'asthma' and 'childhood onset asthma' are propagated to 'respiratory system disease'

## Demo 2

Let's now explore the evidence used to associate *CD86* and multiple sclerosis. We've chosen this target for follow up based on the availability of information for different data types.

Click on any row corresponding to CD86 in the table below:



2,697 targets

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Target name
TNFRSF1A	1.00	1.00				0.07			TNF receptor superfamily memb.
CD58	0.83	0.82				0.05			CD58 molecule
BTNL2	0.82	0.81				0.03			butyrophilin like 2
CD86	0.87	0.80		0.25	0.01	0.05			CD86 molecule

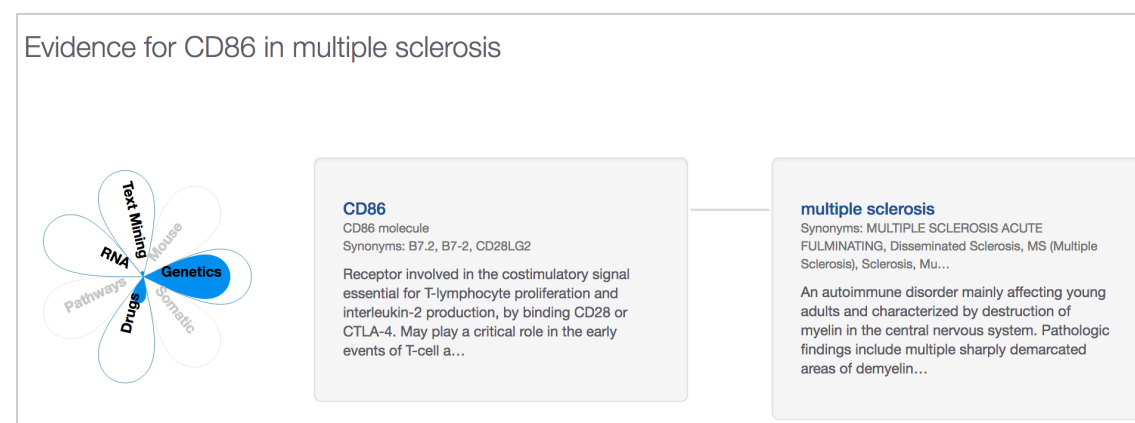
The data types that support this association are (check the flower and the coloured petals):

Genetic association

Drugs

RNA expression

Text mining



**Note:** If you wish to suggest data types or resources we could/should incorporate into our Platform, please get in touch: [support@targetvalidation.org](mailto:support@targetvalidation.org)



The genetic evidence that supports the association (*CD86* – multiple sclerosis) comes from the Gwas catalog. They are all known in public databases i.e. dbSNP (hint: the rsID such as rs9282641, rs4308217 and rs2681424).

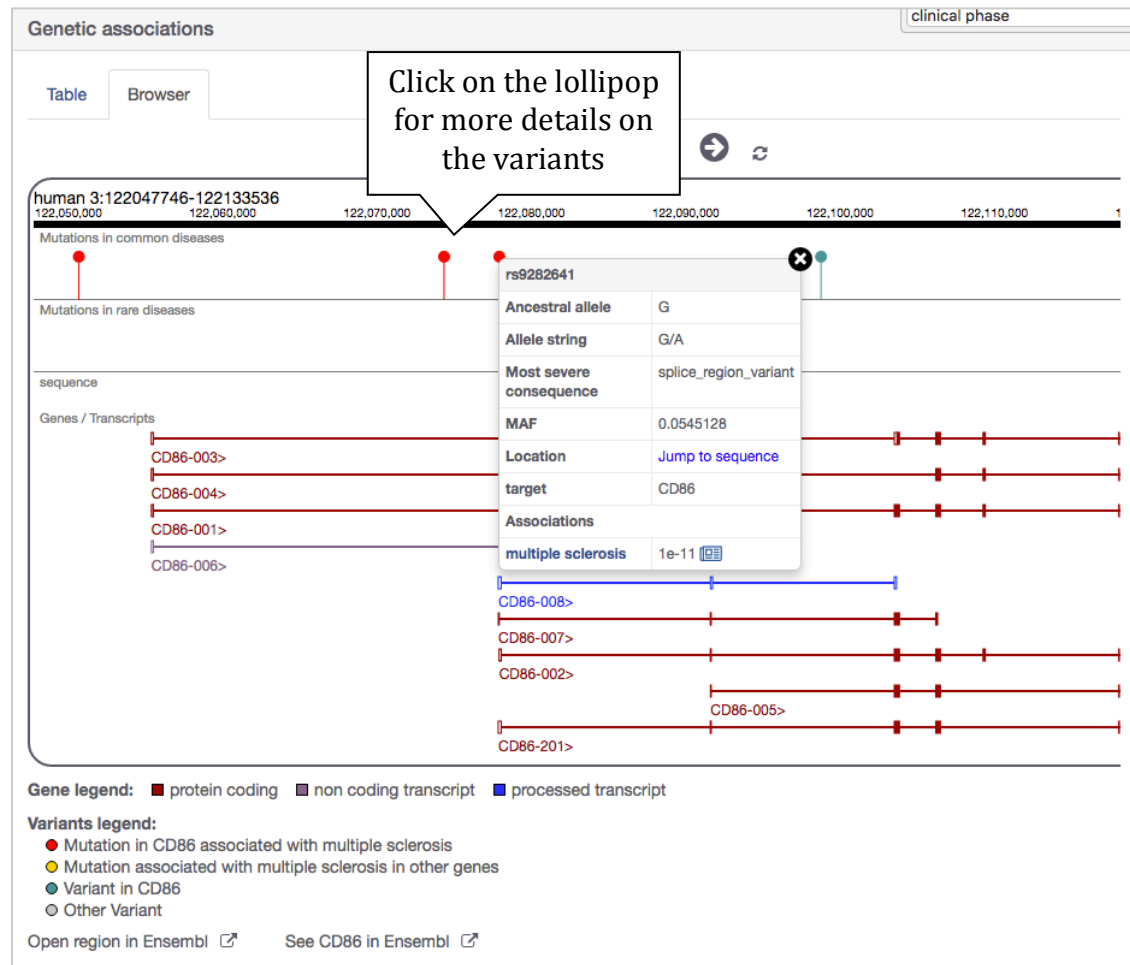
Genetic associations						
<div> <div>Table</div> <div>Browser</div> </div>						
<b>Common diseases</b> Source: GWAS catalog Showing 1 to 3 of 3 entries Search: <input type="text"/> <div>Download</div>						
Disease	Variant	Gene-Variant Evidence		Variant-Disease Evidence		Publications
		Variant type	Evidence source	Evidence source	P-Value	
multiple sclerosis	<a href="#">rs9282641</a>	splice region variant	Open Targets pipeline	<a href="#">gwas catalog</a>	1e-11	1 publication
multiple sclerosis	<a href="#">rs4308217</a>	intron variant	Open Targets pipeline	<a href="#">gwas catalog</a>	6e-8	1 publication
multiple sclerosis	<a href="#">rs2681424</a>	upstream gene variant	Open Targets pipeline	<a href="#">gwas catalog</a>	2e-7	1 publication
<div> Show 10 entries <div>Previous 1 Next</div> </div>						

In addition to the table above, you can also explore the ‘Genetic associations’ data in a Browser view.

In the Browser view, you can zoom in and out, scroll along the genome and find out more about the gene (s), transcript (s), and the genetic variants (represented as lollipops) in the genomic region depicted.

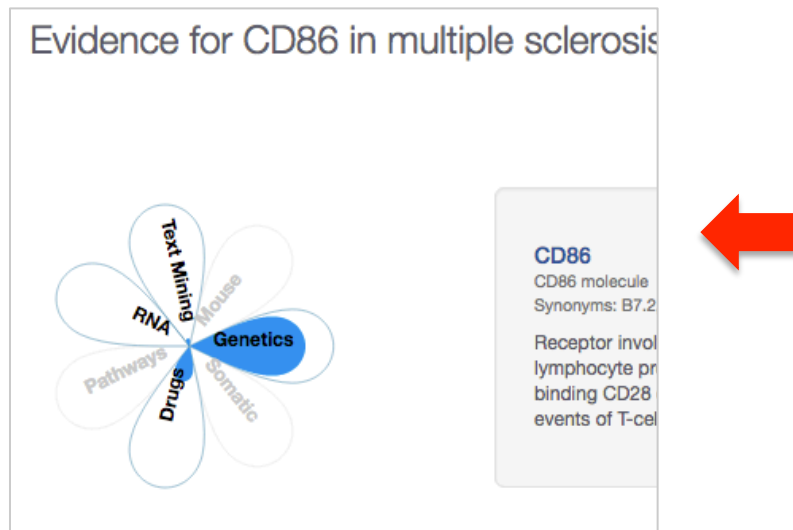
**Note:** The assembly we use is human assembly GRCh38, also known as hg38.

We also provide links to Ensembl.



To find out if there are drugs currently in clinical trials targeting *CD86* in patients with multiple sclerosis, let's expand the 'Drugs' tab. There is just one drug (i.e. ABATACEPT) on phase II of clinical trials.

We can now scroll back to the top of the page and click on the 'CD86' link (next to the flower) to explore this target in more details, such as to find out its RNA expression levels across several tissues and whether or not it is associated with other diseases of the nervous system.



You will land on a page like this:

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

**CD86**  
 CD86 molecule [View associated diseases](#)

Receptor involved in the costimulatory signal essential for T-lymphocyte proliferation and interleukin-2 production, by binding CD28 or CTLA-4. May play a critical role in the early events of T-cell activation and costimulation of naive T-cells, such as deciding between immunity and anergy that is made by T-cells within 24 hours after activation. Isoform 2 interferes with the formation of CD86 clusters, and thus acts as a negative regulator of T-cell activation.

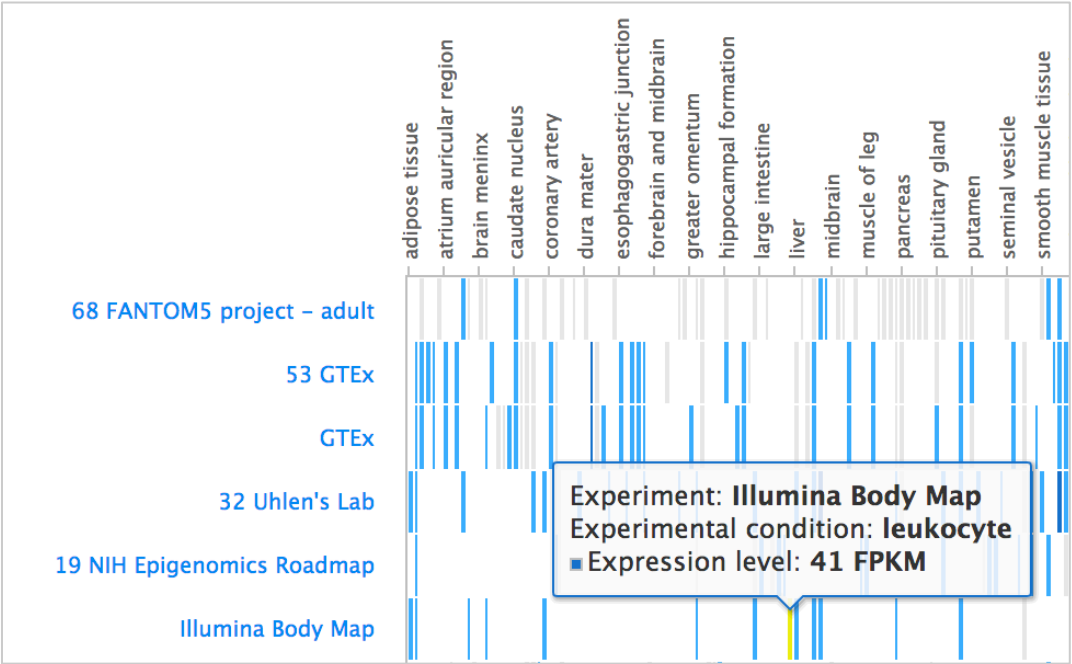
Synonyms: B7.2 B7-2 CD28LG2 B70 FUN-1 Activation B7-2 antigen T-lymphocyte activation antigen CD86 CTLA-4 counter-receptor B7.2 B-lymphocyte antigen B7-2 BU63

- Protein Information (from UniProt)
- Variants, isoforms and genomic context
- Protein baseline expression
- RNA baseline expression
- Gene Ontology
- Protein Structure
- Pathways
- Drugs
- Gene tree
- Bibliography

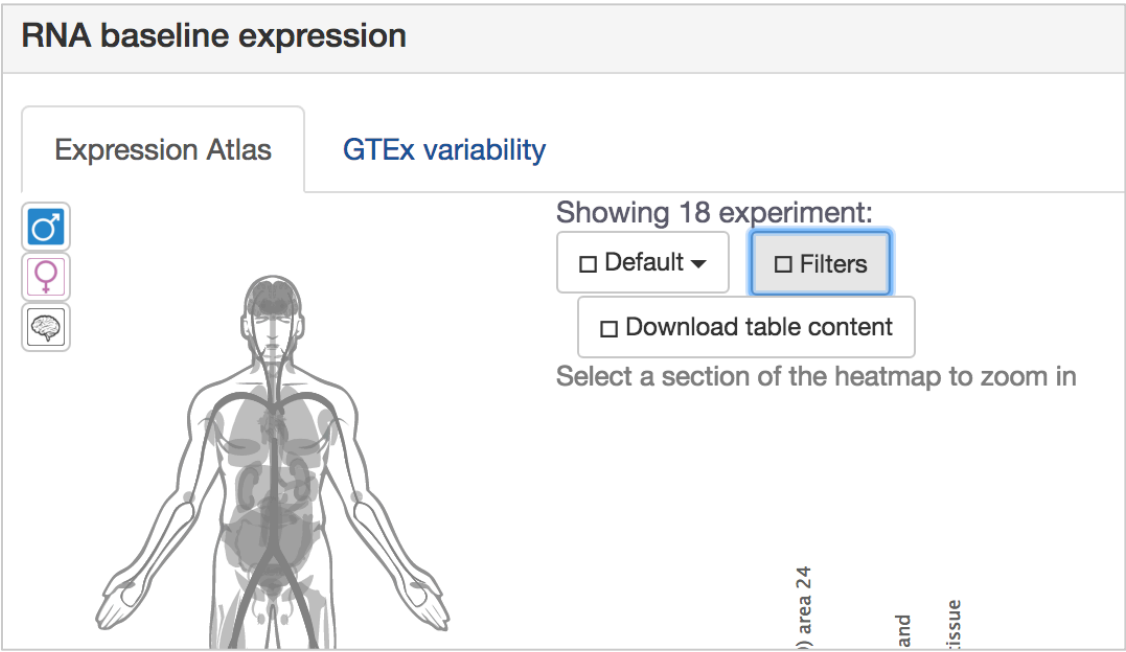
Expand the 'RNA baseline expression' to find out in which tissues *CD86* is highly expressed.

You will find two tabs in there: one with Expression Atlas data (with data from several projects including the Illumina Body Map), and one with GTEx data only:

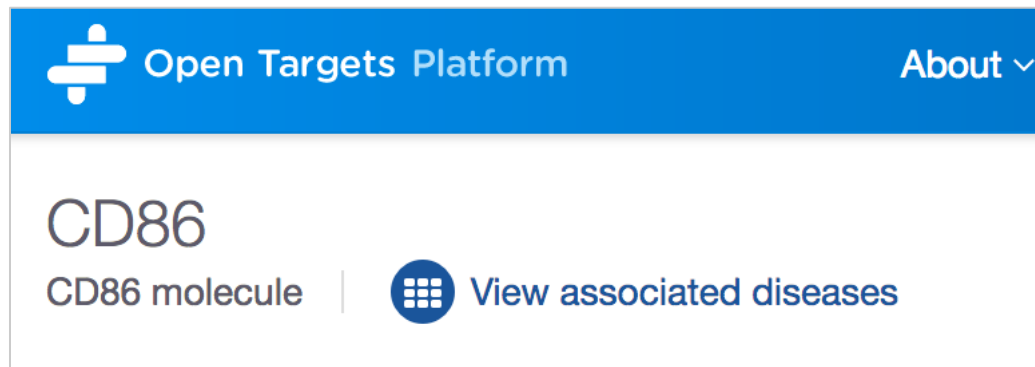
According to the Illumina Body Map data, leukocytes are the cells with the highest expression level for the *CD86* gene:



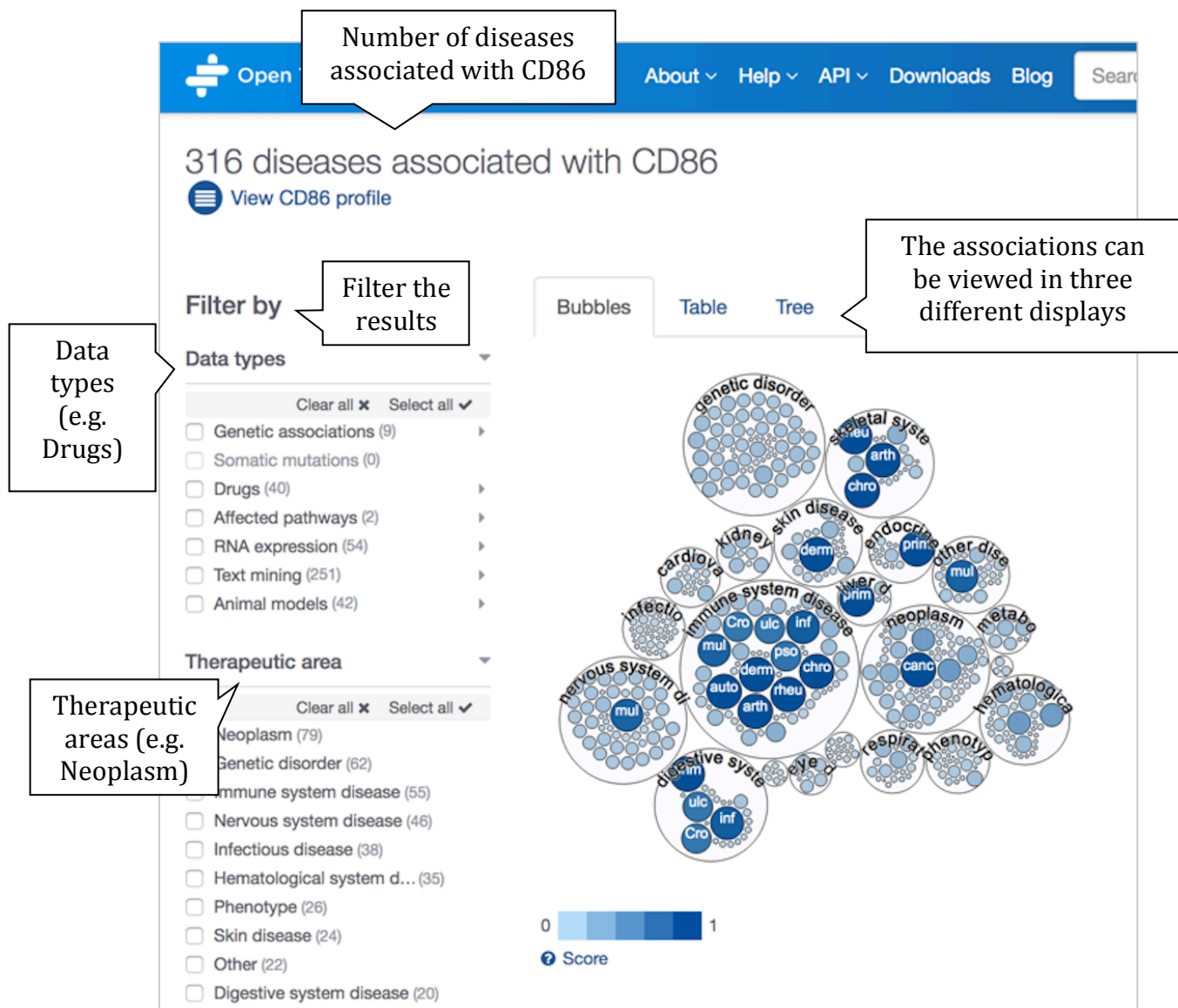
You can zoom this view in and out and apply filters such as 'Expression Value' (below cutoff, low, medium or high) and Anatomical Systems.



Let's now click on 'View associated diseases' to find out all diseases associated with the *CD86* (not only multiple sclerosis):



You will land on a page like this:



There are three different displays that can be used to view the diseases associated with any given target:

- Bubble view

In this view, we group diseases into 'bubbles' based on the disease ontology. Large bubbles correspond to a therapeutic area and consist of smaller bubbles representing diseases within this area. A disease can belong to several therapeutic areas and therefore can appear within more than one large bubble. The strength of the association between the target and a disease is represented by the size of the bubble and the shade of its blue colour; the larger the bubble and the darker the blue, the stronger the association.

- Table view

In this view, we list all diseases associated with a target, ordered by the association score, which is colour coded. When there is no evidence to support the association, the cells in this table are coloured in white (score of zero). You can show the 10 first entries and get the pagination for the remaining entries.

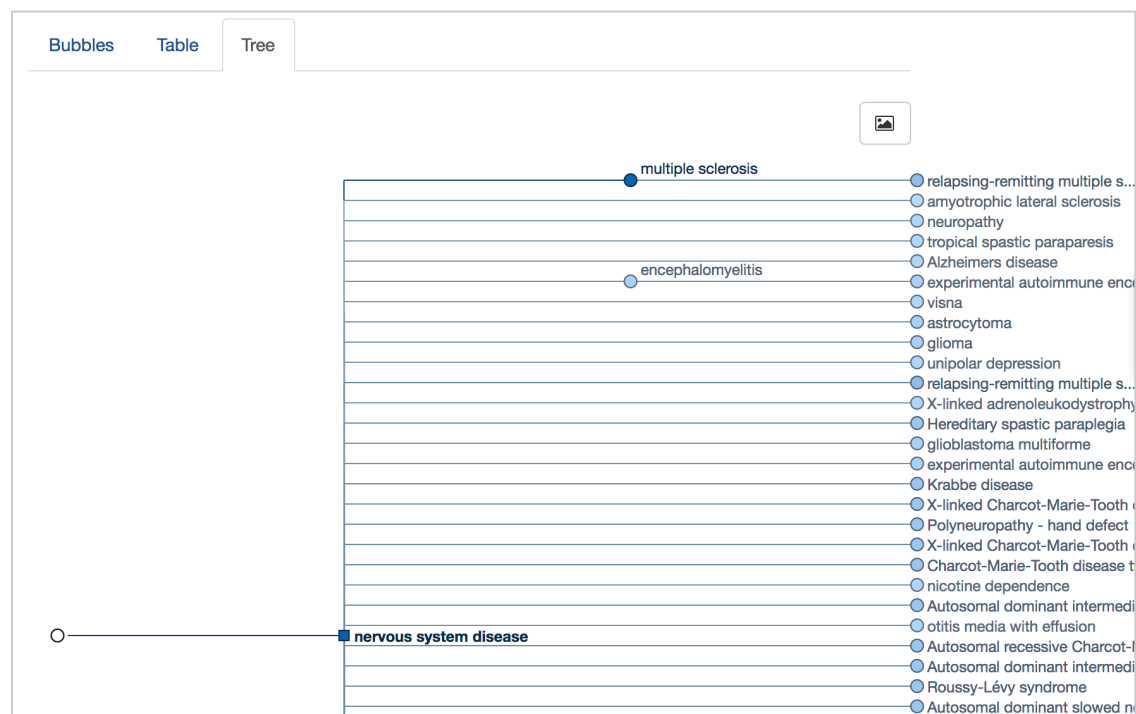


**This table can be exported in csv format** (look for the download button).

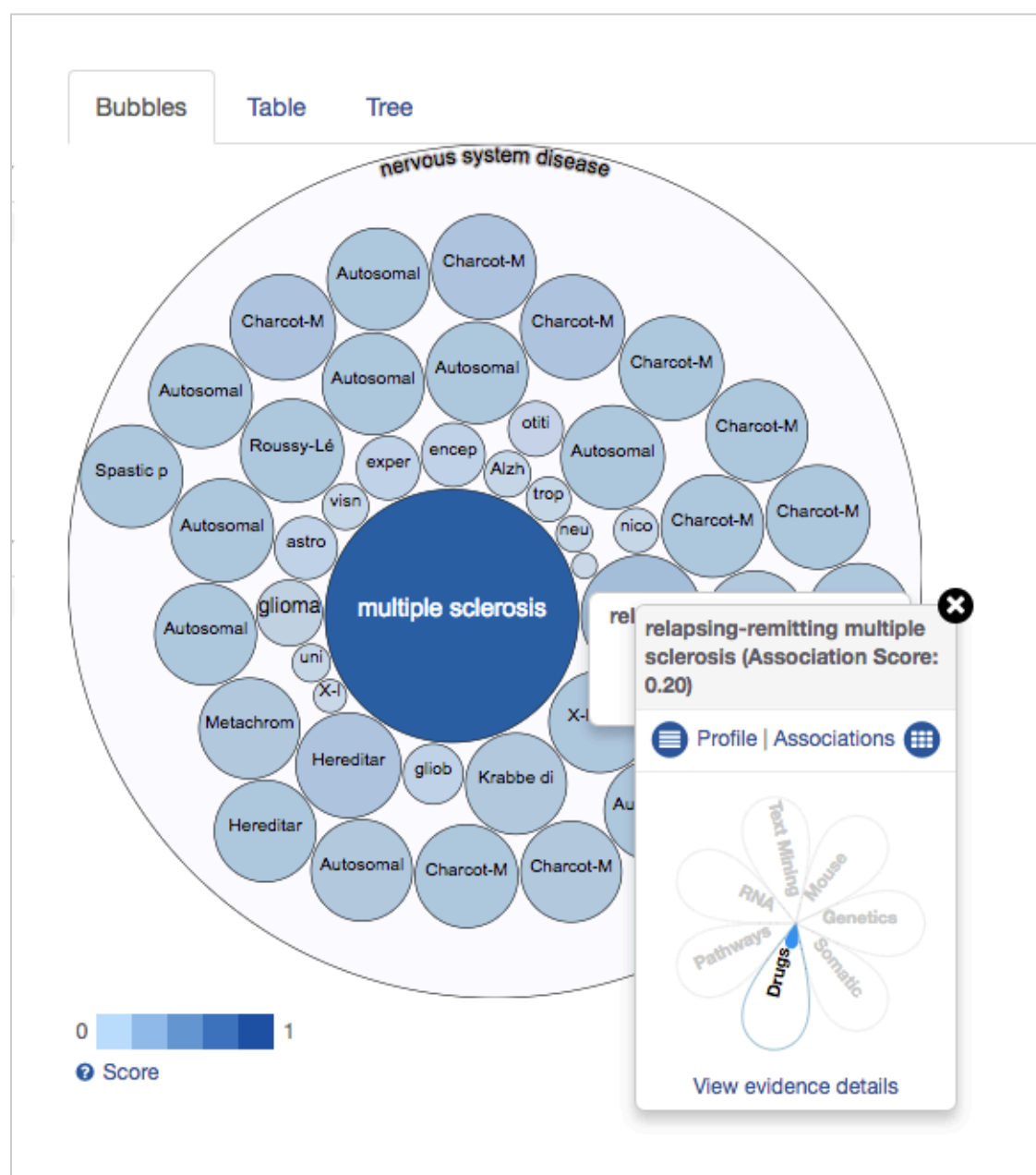
*Tip: We colour code the cells in the table in different shade of blue as a visual way to convey the strength of the association (strongest association is coloured in dark blue). The score varies from 0 to 1. Hover over the cells in the table to view the numbers. Alternatively, you can select the cells in the table so that you can view the numerical values.*

- Tree view

In the Tree view, you can visualise the evidence across the therapeutic areas in a tree format that represents the relationships of diseases. Therapeutic areas have a square symbol (e.g. Genetic disorders), while the diseases (e.g. ovarian carcinoma) are represented as circles. The squares and circles are colour coded in blue, and the darker the blue, the stronger the association:



Regardless the view you choose to explore, you can filter the data to find out if there are other “Nervous system” diseases associated with this target. There are 46 of them (e.g. Hereditary spastic paraplegia, amyotrophic lateral sclerosis and neuropathy). The scores vary from 0.01 to 0.20.





---

## **HANDS-ON EXERCISES**

---

### **Exercise 1**

#### **Aducanumab, antibody that may slow the progress of Alzheimer's disease**

##### **BACKGROUND**

The most common cause of dementia, Alzheimer's disease is a brain disorder that leads to impairment of intellectual functions and social skills, resulting in significant disturbance of daily living activities. Dementia is caused by the destruction of cells in certain parts of the brain primarily involved in memory and mental function.

##### **SIGNIFICANCE**

Alzheimer's disease is the 6<sup>th</sup> leading cause of death in the United States and it's expected 16 million Americans will be living with Alzheimer's by 2050. (source: alz.org).

Biogen in collaboration with Neurimmune has been conducting clinical studies investigating a potential new treatment (Aducanumab) for early Alzheimer's disease. Aducanumab is an antibody that binds to and may reduce amyloid plaques from the brain, potentially slowing the progress of the disease.

##### **QUESTIONS**

a) Can you use the platform to find out if this drug is currently used in trials with other dementia related diseases?

b) Which genes does this drug match to?

c) Can you list the diseases from the nervous system where there is a strong confidence (>90%) in the association with this gene?

Let's focus now on the different evidence used to associate *APP* with Alzheimer's.

d) Are there any known genetic variants (i.e. with a reference ID such as rs123456) listed in the Genetic associations table? Can you find some of the papers that support this association?

e) Click on the 'Browser' link to view the mutations is a graphical display. Are there variants associated with other traits (or diseases) in the region of the *APP* gene?

f) Which biochemical pathways seem to be affected by pathogenic mutations in this target?

Let's now have a look at the target itself by clicking on the target name. This page provides a profile for a target and it's where you can explore more information on a target such as data on RNA baseline expression, gene tree with orthologous genes in other species, etc.

g) Which tissues according to the GTEx project does this target seem to be highly expressed?

---

## Exercise 2

### ***MS4A1* as a possible drug target in the treatment of non-Hodgkin's lymphoma**

#### BACKGROUND

The B-lymphocyte antigen CD20 is an activated-glycosylated phosphoprotein expressed on the surface of all B-cells beginning at the pro-B phase and progressively increasing in concentration until maturity. In humans, the *MS4A1* gene encodes antigen CD20.

#### SIGNIFICANCE

CD20 is the target of monoclonal antibodies (mAb) in the treatment of all B cell lymphomas, leukemias and autoimmune diseases. Some of these active agents (mAb) are on clinical trials for non-Hodgkin's lymphoma. Others anti-CD20 mAb have been approved by the FDA for B-cell chronic lymphocytic leukemia.

#### QUESTIONS

a) How many diseases within the broader Therapeutic area 'Hematological system' are associated with this target? How many of these are based on 'Drugs' only?

b) In addition to the data evidence 'Drugs', are there other types of evidence supporting the association between *MS4A1* and non-Hodgkin's lymphoma?

c) Let's now explore some disease information available for non-Hodgkin's lymphoma. Can you list all drugs in phase IV of clinical trials for the treatment of this disease which status is completed (i.e. no longer recruiting patients for the clinical trials)?

d) Can you find the different subtypes i.e. the children terms of non-Hodgkin's lymphoma in its ontology? Can you download this image?

*Note: you may want to click on the children diseases and see which targets have been associated with them.*

---

### **Exercise 3**

#### ***RS1*, a gene playing a role in the retina layer formation**

##### **BACKGROUND**

##### **QUESTIONS**

a) What are the eye diseases that seem to be associated with this target based on a combination of genetic evidence (e.g. SNPs from GWAS), text mining of the literature and mouse models?

b) How long is the protein encoded by this gene/target?

c) Which chromosome does this gene map to on the human genome? Is it a forward or reverse stranded gene?

d) Which tissue has the highest RNA baseline expression from the FANTOM5 project?

---

### Demo 3

Filter the target association table for Alzheimer's based on a list of known targets.

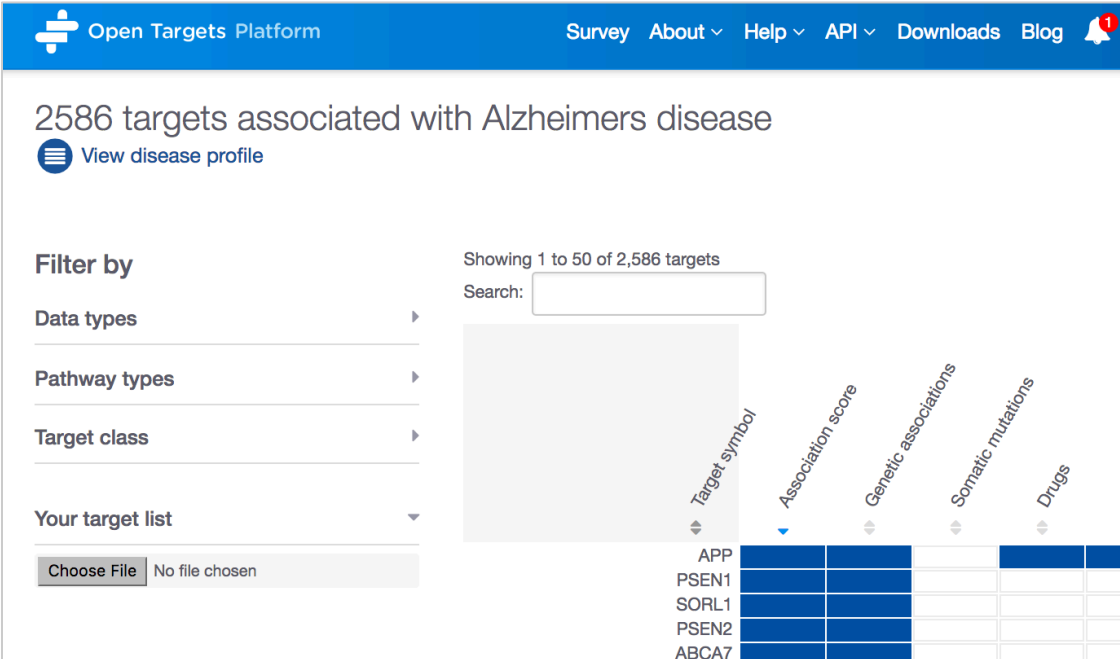
This is a list of several genes that seem to be associated with Alzheimer's based literature searches:

*HFE*  
*PSEN1*  
*TF*  
*APOE*  
*ADRB2*  
*PSEN2*  
*A2M*

Which information is available in the Open Targets Platform that could help you to choose and prioritise a target out of that list for follow up? We can input this data as a list (in .txt) to filter the associations for Alzheimer's in Open Targets.

a) Which of those seven genes have the strongest association w/ Alzheimer's?

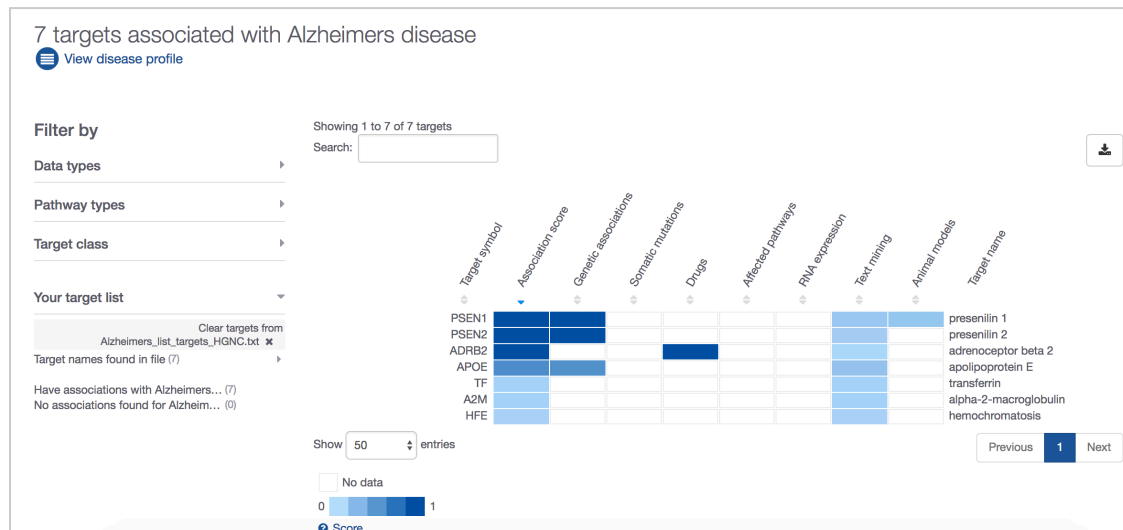
Firstly, let's search for Alzheimer's and then upload our target list onto the Platform:



The screenshot shows the Open Targets Platform interface. The top navigation bar includes links for Survey, About, Help, API, Downloads, Blog, and a notification icon. The main content area displays '2586 targets associated with Alzheimers disease' and a 'View disease profile' button. On the left, there is a 'Filter by' section with expandable categories: Data types, Pathway types, Target class, and Your target list. The 'Your target list' section includes a 'Choose File' button and the text 'No file chosen'. A red arrow points to this section. The main table shows 'Showing 1 to 50 of 2,586 targets' and a search bar. The table headers are Target symbol, Association score, Genetic associations, Somatic mutations, and Drugs. The first five rows of the table are:

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs
APP				
PSEN1				
SORL1				
PSEN2				
ABCA7				

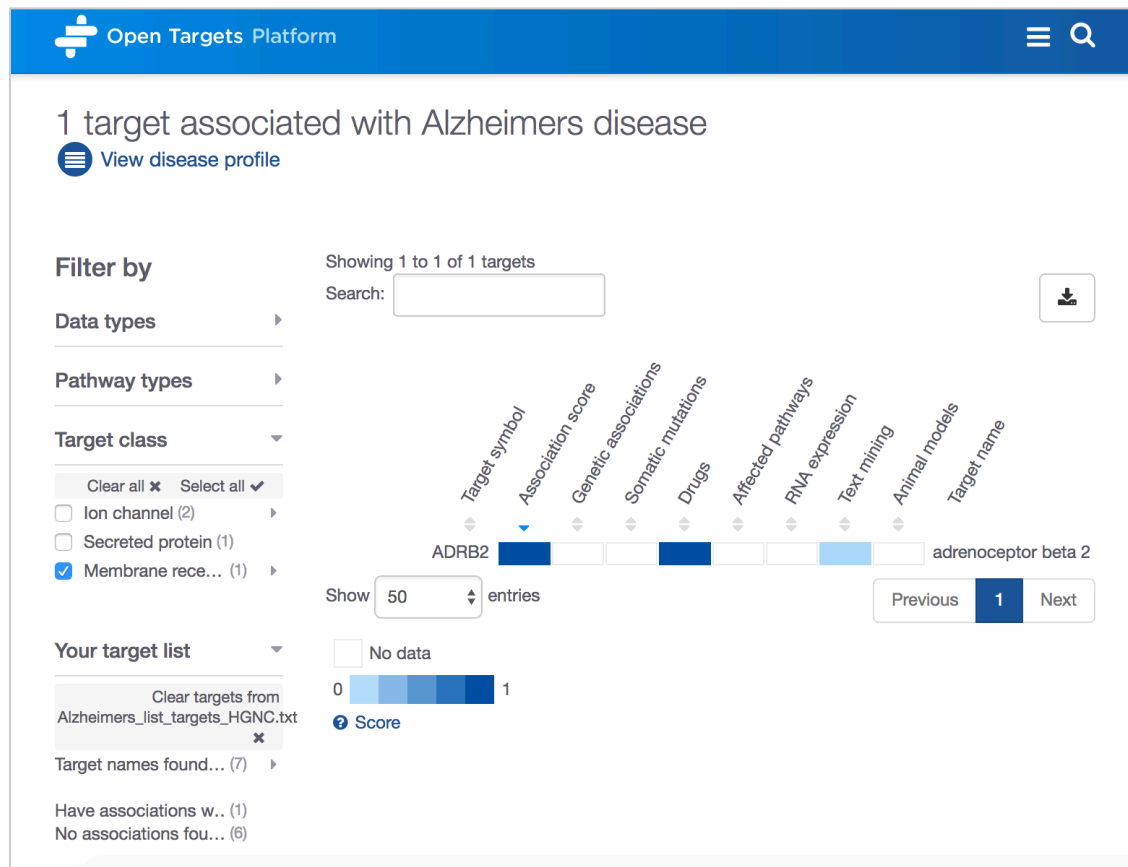
Now, we can upload a text file containing our list of genes (either as official gene symbols from HGNC e.g. *SOX3*, or Ensembl gene IDs e.g. ENSG00000134595). The filter 'Your target list' is at the left hand side of the association page (you will need to scroll down to see it):



The list should be uploaded and the resulting table will show the results of our analyses for your list of seven genes (i.e. targets). This will help you prioritise which targets to follow up.

b) Are there any targets, which are membrane receptors?

We can now use the filter 'Target class' to focus on membrane receptors only. There is only one target that is a membrane receptor, *ADRB2*.



**Open Targets Platform**

1 target associated with Alzheimer's disease

[View disease profile](#)

**Filter by**

**Data types**

**Pathway types**

**Target class**

Clear all ✕ Select all ✓

☐ Ion channel (2)

☐ Secreted protein (1)

☒ Membrane receptor (1)

**Your target list**

Clear targets from  
Alzheimers\_list\_targets\_HGNC.txt ✕

Target names found... (7)

Have associations w... (1)

No associations found... (6)

Showing 1 to 1 of 1 targets

Search:

Download

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Target name
ADRB2									adrenoceptor beta 2

Show 50 entries

Previous 1 Next

No data

0 1

Score

c) Which amino acids of this membrane receptor correspond to the extracellular domain?

To explore more about the target itself, we can click on any cell of the resulting table then click on the target name, next to the flower:

Open Targets Platform

## Evidence for ADRB2 in Alzheimers disease

**ADRB2**  
adrenoceptor beta 2  
Synonyms: ADRBR, BAR, B2AR, ADRB2R

Beta-adrenergic receptors mediate the catecholamine-induced activation of adenylate cyclase through the action of G proteins. The beta-2-adrenergic receptor binds epinephrine with an approximately 30-...



We will land on a page like this:

Open Targets Platform

## ADRB2

adrenoceptor beta 2 | View associated diseases

Beta-adrenergic receptors mediate the catecholamine-induced activation of adenylate cyclase through the action of G proteins. The beta-2-adrenergic receptor binds epinephrine with an approximately 30-fold greater affinity than it does norepinephrine.

Synonyms: ADRB2 BAR B2AR ADRB2R Beta-2 adrenoceptor Beta-2 adrenoreceptor Beta-2 adrenergic receptor

Protein Information (from UniProt)

Variants, isoforms and genomic context

Protein baseline expression

RNA baseline expression

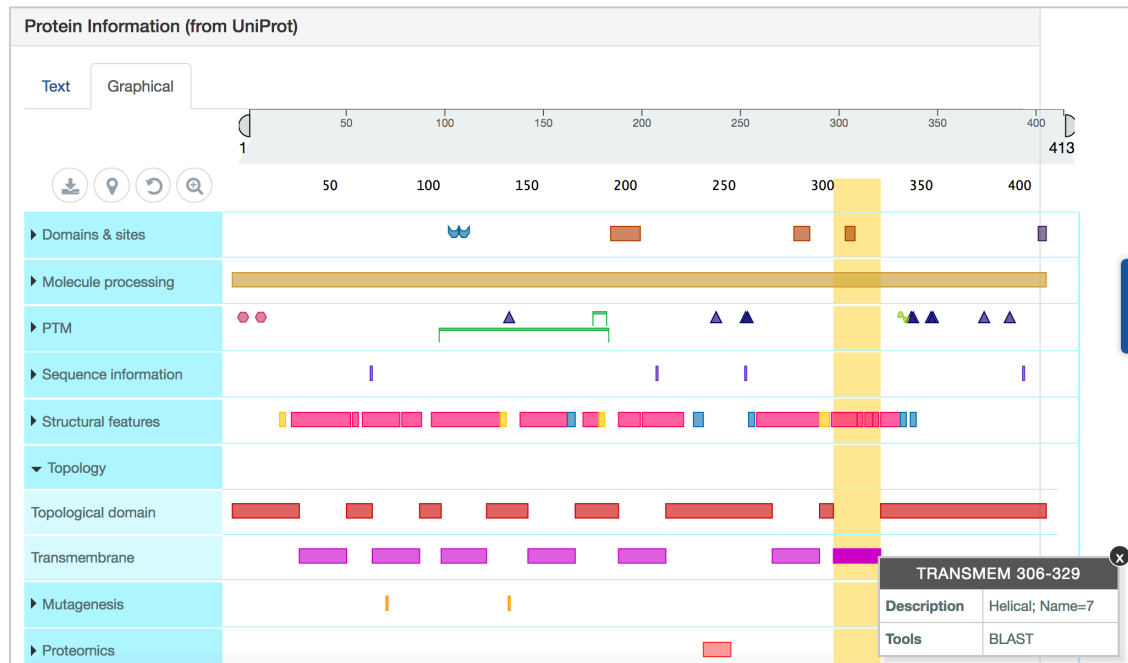
Gene Ontology

Protein Structure

Pathways

Drugs

To find what which amino acids of this membrane receptor correspond to the extracellular domain, we need to expand the tab 'Protein Information (from UniProt)', click on 'Graphical', then expand the 'Topology' section in the image:



This membrane receptor is a G-coupled transmembrane receptor, hence there are seven transmembrane domains. Click on the 'purple-ish' boxes in 'Topology' to get a pop up window with more information on which amino acids of the protein span the entirety of the cellular membrane.



---

## **EXTRA HANDS-ON EXERCISES**

---

***If you have finished exercises 1, 2 and 3 above, you may want to try these too:***

### **Exercise 4**

***Using the Open Targets Platform to find out if the modulation of a target by a drug poses any possible unsafe interactions or effects.***

#### **BACKGROUND**

*The main goals of drug development are effectiveness and safety. Although no drug is 100% safe (they all have side effects), the benefits of the drugs should outweigh the known risks.*

#### **SIGNIFICANCE**

*Many drugs used on the treatments of diseases can interfere with other physiological processes and even cause death when taken in excess. One of the ways to start assessing the safety of a new compound is to look at which target it modulates, whether or not this target is involved in other therapeutic areas such as cardiovascular and reproductive system, and the expression of the gene (or protein) in normal tissues.*

#### **USE CASE**

*Abemaciclib has been shown in vitro to be a selective ATP-competitive inhibitor of CDK6 kinase activity. CDK6 has been shown to phosphorylate and thus regulate the activity of tumor suppressor protein Rb. Expression of this gene is increased in some types of cancer.*

*Abemaciclib is under investigation in patients with breast carcinoma among other types of cancer.*

#### **QUESTIONS**

*a) Which data supports the association between CDK6 and breast carcinoma?*

*b) Are there other drugs in addition to abemaciclib used in clinical trials modulating the same target for breast carcinoma? Can you get to the original data?*

*c) Are there studies showing a decreased level of RNA expression of this gene in breast carcinoma?*

*d) Is this target associated with cardiovascular diseases with a strong confidence (i.e. score of 0.80 or above)?*

---

## **Exercise 5**

***How can I retrieve all disease associations for three genes of interest, all at once?***

### **BACKGROUND**

So far you have used the website [www.targetvalidation.org](http://www.targetvalidation.org) to search for target-disease associations on a gene by gene (or disease by disease) basis. You may want to access and retrieve data on several genes or several diseases. For this, you can access our data in programmatic way using our REST API (or Python, R clients)

### **USE CASE**

The following three genes have been associated with gastric carcinoma:

ENSG00000141736

ENSG00000141510

ENSG00000132356

### **QUESTIONS**

*a) "How can I find out all diseases (besides gastric carcinoma) associated with those three Ensembl gene IDs?"*

*b) "Which diseases have got the highest overall association score for each of those three genes?"*

*c) Can I download the above list in TAB format?"*

*Interested in other use cases using our REST API? Check our [blog posts](#).*

---

---

## **QUICK GUIDE TO DATABASES**

---

Here is a list of databases and projects that may be useful for you to explore:

### **PROTEINS**

**UniProtKB** – The “Protein knowledgebase” is a comprehensive set of protein sequences. It is divided into two parts: TrEMBL and Swiss-Prot. The later is manually annotated and reviewed, therefore provides a set of protein sequences of high quality.

<http://www.uniprot.org/>

### **GENE NOMENCLATURE COMMITTEES**

**HGNC** – The HUGO Gene Nomenclature Committee assigns unique names and symbols to every single human gene, whether they are coding or not. These gene names and symbols are the official ones for human genes.

<http://www.genenames.org/>

**MGI** – The HGNC counterpart for naming mouse genes and symbols.

<http://www.informatics.jax.org/>

### **GENETIC VARIANTS and SOMATIC MUTATIONS**

**GWAS catalog**– The catalog of Genome Wide Association Studies (GWAS) provides genetic variants (e.g. SNPs) that are associated with a disease.

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

**EVA** – The European Variation Archive (EVA) provides genetic variants and somatic mutations (associated with cancer).

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

**Cancer Gene Census** – A catalogue of genes for which mutations have been causally implicated in cancer. The Catalogue of Somatic Mutations in Cancer (COSMIC) at the Wellcome Trust Sanger Institute provides us with the set of genes associated with specific cancers in the Cancer Gene Census, in addition to other cancers associated with that gene in the COSMIC database.

[www.cancer.sanger.ac.uk/census/](http://www.cancer.sanger.ac.uk/census/)

**IntOGen** - It provides evidence of somatic mutations, genes and pathways involved in tumorigenesis from 6,792 samples across 28 cancer types.

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

**Gene2Phenotype** - The data in Gene2Phenotype (G2P) provides evidence of genetic variants that are manually curated from the literature by consultant clinical geneticists in the UK. This is provided by DECIPHER, a database of genomic variants and phenotypes in patients with developmental disorders.

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

## **DRUGS**

**ChEMBL** - The ChEMBL database at the EMBL-EBI provides evidence from known drugs that can be linked to a disease and a known target.

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

## **RNA EXPRESSION**

**Expression Atlas** – The Expression Atlas at EMBL-EBI provides information on genes that are differentially expressed between normal and disease samples, or among disease samples from different studies. In addition to differential expression, they provide baseline expression information for each gene.

<https://www.ebi.ac.uk/gxa/home>

## **AFFECTED PATHWAYS**

**Reactome** – The Reactome database at the EMBL-EBI contains pathway information on biochemical reactions sourced from manual

curation. It identifies reaction pathways that are affected by pathogenic mutations.

<http://www.reactome.org/>

### **ANIMAL MODELS**

**Phenodigm** - The Phenodigm resource at the Wellcome Trust Sanger Institute provides evidence on associations of targets and disease. It uses a semantic approach to map between clinical features observed in humans and mouse phenotype annotations.

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

### **TEXT MINING**

**Europe PMC** - The Europe PubMed Central at the EMBL-EBI mines the titles, abstracts and full text research articles from both PubMed and PubMed Central to provide evidence of links between targets and diseases.

<http://europepmc.org/>