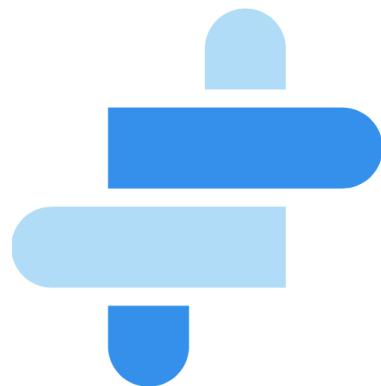


# Open Targets: Integrating genetics and genomics for drug discovery



12<sup>th</sup> March 2018

Denise Carvalho-Silva  
Open Targets  
EMBL-EBI -- UK

# Notes

This booklet is based on the February release (2018) of the Open Targets Platform.

Some useful links:

1) Science in Open Targets

<https://www.opentargets.org/science/>

2) About the Open Targets Platform

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

3) Platform FAQs

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

4) Videos and animations

<https://tinyurl.com/opentargets-youtube>

5) Cite us

<https://academic.oup.com/nar/article/45/D1/D985/2605745>

6) Open Targets REST API docs

<http://api.opentargets.io/v3/platform/docs>

7) GitHub

<https://github.com/opentargets>

Questions or suggestions?  
[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>7</b>
<b>HANDS-ON EXERCISES.....</b>	<b>29</b>
Exercise 1.....	29
Exercise 2.....	30
Exercise 3.....	31
<b>EXTRA HANDS-ON EXERCISES .....</b>	<b>33</b>
<i>Exercise 4.....</i>	<i>33</i>
<i>Exercise 5.....</i>	<i>33</i>
<b>QUICK GUIDE TO DATABASES .....</b>	<b>35</b>

## **OVERVIEW**

Open Targets is a partnership to transform drug discovery through the systematic identification and prioritisation of targets.

We work to create a research and development (R&D) framework that can be applied to a wide range of human diseases. We will share our results 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 Biogen (<https://www.biogen.com/>) joined the initiative. The consortium was rebranded to Open Targets in April 2016.

In drug discovery, the *validation* of a target refers to the creation of a specific entity that modulates the activity of a target 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 high failure rates.

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 science carried out in Open Targets
- The Open Targets Platform
- How to browse the web interface of the Platform

## **INTRODUCTION TO OPEN TARGETS**

Open Targets employs large-scale human genetics and genomics data to change the way drug targets are identified and prioritise. We have established a set of scientific projects to both integrate and generate data and analytical processes that implicate a target as valid.

Our experimental projects use CRISPR gene editing, induced pluripotent stem cells, single cell genomics, organoids to generate new data and provide insights in the validation of targets relevant to key therapeutic areas namely:

- Oncology
- Immunology
- Neurodegeneration

Our core bioinformatics and data pipelines team has developed the Open Targets Platform to provide easy access to biological data relevant to drug target identification and selection by a diverse audience of users.

More details on our projects can be found on our [Scientific Overview](#) page.

### **The Open Targets Platform**

The Open Targets Platform is a web application that integrates and displays publicly available biological data to facilitate the identification and selection of targets for new therapies.

We use biological data from different [data sources](#) to associate genes and diseases. Similar data sources are combined into the following data types: Genetic associations, Somatic mutations, Drugs, Affected pathways, RNA expression, Text mining, and Animal models.

This evidence is used for the ranking and [scoring of target-disease associations](#), which depends on the frequency of evidence, the confidence and severity. We then aggregate the evidence using the

sum of the [harmonic progression](#) to obtain the score at the data source and data type levels, as well as the overall score.

The latest release of the Platform (February 2018) contains:

- 20,974 targets
- 9,728 diseases
- 2,306,670 associations between targets and disease
- 5,905,247 evidence

The Open Targets Platform is for all types of users (biologists and otherwise), from both academia and industry. Our users can browse a target on a gene by gene (or disease by disease) basis, search for a batch of up to 200 genes ([Batch search](#)), carry out more complex queries using the REST-API, or download all evidence and association objects for downstream analyses.

## What can you do with the Open Targets Platform?

- Find which targets are associated with a disease,
- Find which diseases are associated with a target,
- Find the evidence supporting target-disease associations,
- Find the associations of a target with diseases from different therapeutic areas,
- Find target specific information in our target profile page,
- Find disease specific information in our disease profile page,
- Export data for target-disease association,
- Search for many targets at once with our batch search tool.

## Connect with us

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

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

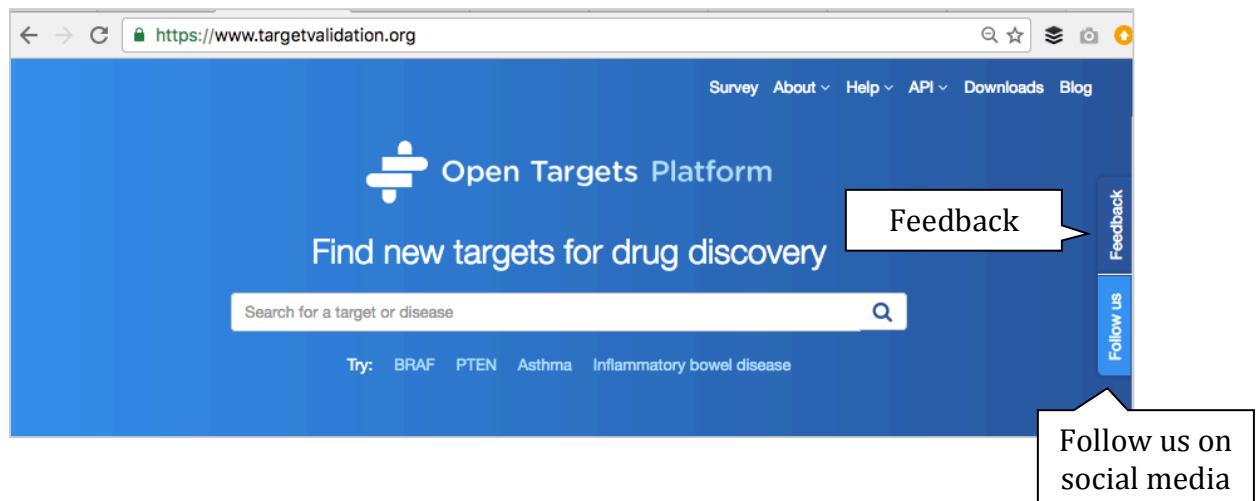
We will guide you through the website using multiple sclerosis (MS), as an example of a disease, then we will explore the evidence associationg *CD86* with MS.

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: Disease centric workflow

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



Select the first (best) hit:

multiple scl| 🔍

## multiple sclerosis

2080 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

experimental autoimmune encephalomyelitis  
[...] > central nervous system inf... > encephalomyelitis > experimental autoimmune e...

chronic progressive multiple sclerosis  
autoimmune disease > multiple sclerosis > chronic progressive multipl...

You will see a page like this:

Total number of targets associated with multiple sclerosis

2,735 targets associated with multiple sclerosis

💡 View disease profile

Filter the results

Filter by

Showing 1 to 50 of 2,735 targets

Search:

Data types (Genetic Associations Drugs, etc)

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Target name
IL2RA									interleukin 2 receptor sub...
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 subfamil...
IFNAR1									interferon alpha and beta ...
CD52									CD52 molecule

The current release of the Open Targets Platform (February 2018) lists 2640 targets associated with multiple sclerosis.

The data types supporting these results are based on Genetic association, Somatic mutations, Drugs, RNA expression, Text mining, and Animal models. There is no data on affected pathways available for this disease.

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', RNA tissue specificity 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 (i.e Reactome and SLAPenrich)
- 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) RNA tissue specificity

Select the organs (or anatomical system) where the target is significantly more expressed in the selected tissues than the mean of the other tissues

E) 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 and SLAPenrich:

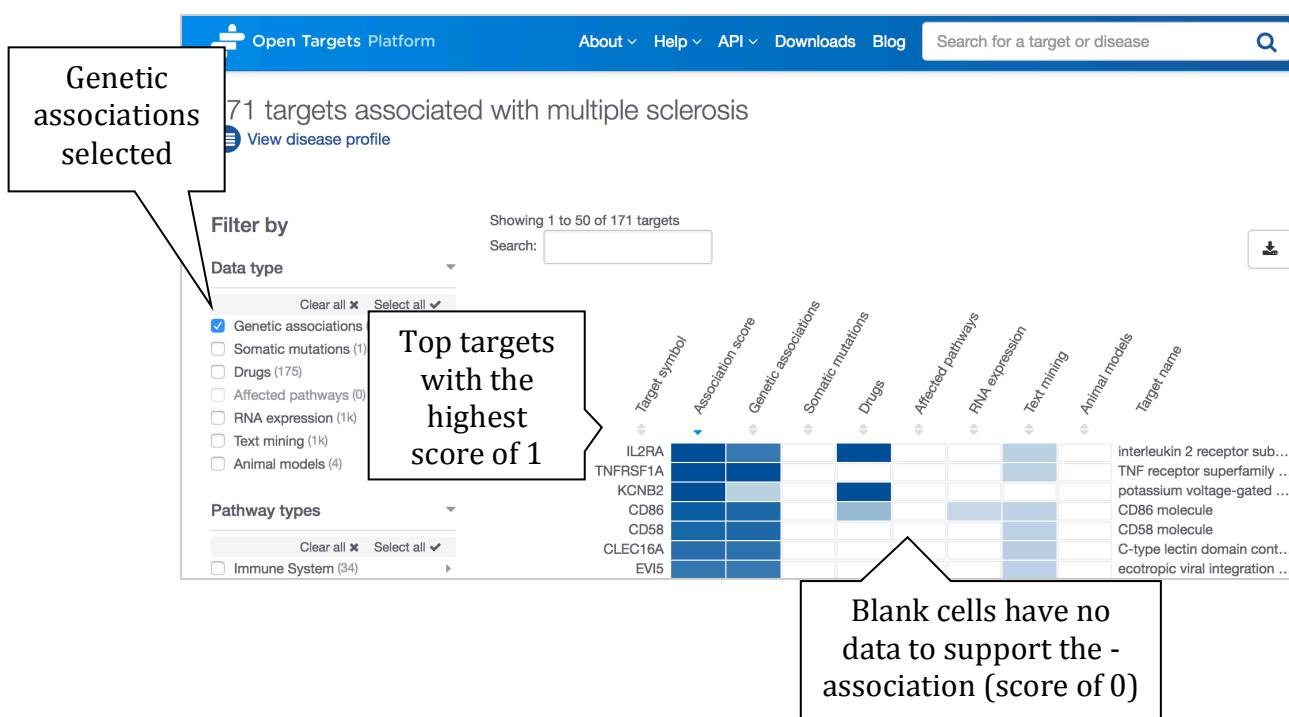
<http://www.reactome.org/>

<https://saezlab.github.io/SLAPenrich/>

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 371:



Those are targets associated with multiple sclerosis based on genetic variants only. Genetic variants are SNPs (single nucleotide polymorphisms or mutations, or short indels) from data sources such as GWAS catalog, UniProt, PheWAS, Genomics England PanelAPP, to name a few.

The table is sorted by default with the best hit on the top of the table i.e. *IL2RA*. The best hit is the target that contains the highest (overall) association score. The association score can vary from 0 to 1. We will sum up the scores from different data sources to obtain the overall score and due to this sum, the score can be higher than 1 but we will always cap it to 1.

Different weight is given to different data types when computing the score. From the highest to the lowest weight, this is the ranking of our data types:

Genetic association = somatic mutations = drugs = pathways >  
RNA expression > Animal models = Text mining

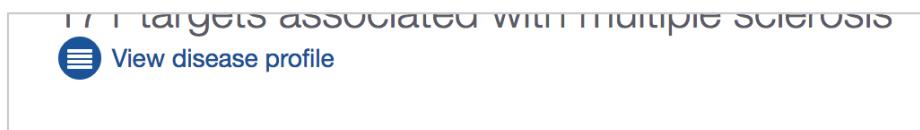
You can sort the table by alphabetical order of the list of targets, or by the association score (either overall or per data type e.g. Genetic associations, Drugs, Text mining, etc). The association score varies from 0 to 1, the closer to 1 the stronger the association. This score is computed for each 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):

Showing 1 to 50 of 168 targets								
Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models
IL2RA								
TNFRSF1A								
KCNB2								

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.

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':



Let's now expand the tab 'Drugs'. You will see a page like this:

A screenshot of the Open Targets Platform disease profile page for "multiple sclerosis". The page has a blue header bar with the "Open Targets Platform" logo on the left and a search bar on the right. The main content area shows the disease name "multiple sclerosis" and a "View associated targets" button. Below this, there is a section for "Synonyms" with terms like "MS (Multiple Sclerosis)", "Sclerosis, Disseminated", "Disseminated Sclerosis", "Sclerosis, Multiple", "Multiple Sclerosis, Acute Fulminating", "MS", and "MULTIPLE SCLEROSIS ACUTE FULMINATING". There are several tabs at the bottom: "Similar diseases (based on targets in common)" (with a "New" badge), "Phenotypes", "Drugs" (selected tab), "Bibliography" (with a "New" badge), and "Classification".

In the February 2018 release, we have 57 unique drugs in different phases of clinical trials with patients suffering from this condition. They will be targeting different proteins.

# multiple sclerosis

 View associated targets

Synonyms: MS (Multiple Sclerosis) Sclerosis, Disseminated Disseminated Sclerosis Sclerosis, Multiple

Multiple Sclerosis, Acute Fulminating MS MULTIPLE SCLEROSIS ACUTE FULMINATING

## Similar diseases (based on targets in common)

New

## Phenotypes

Feedback

## Drugs

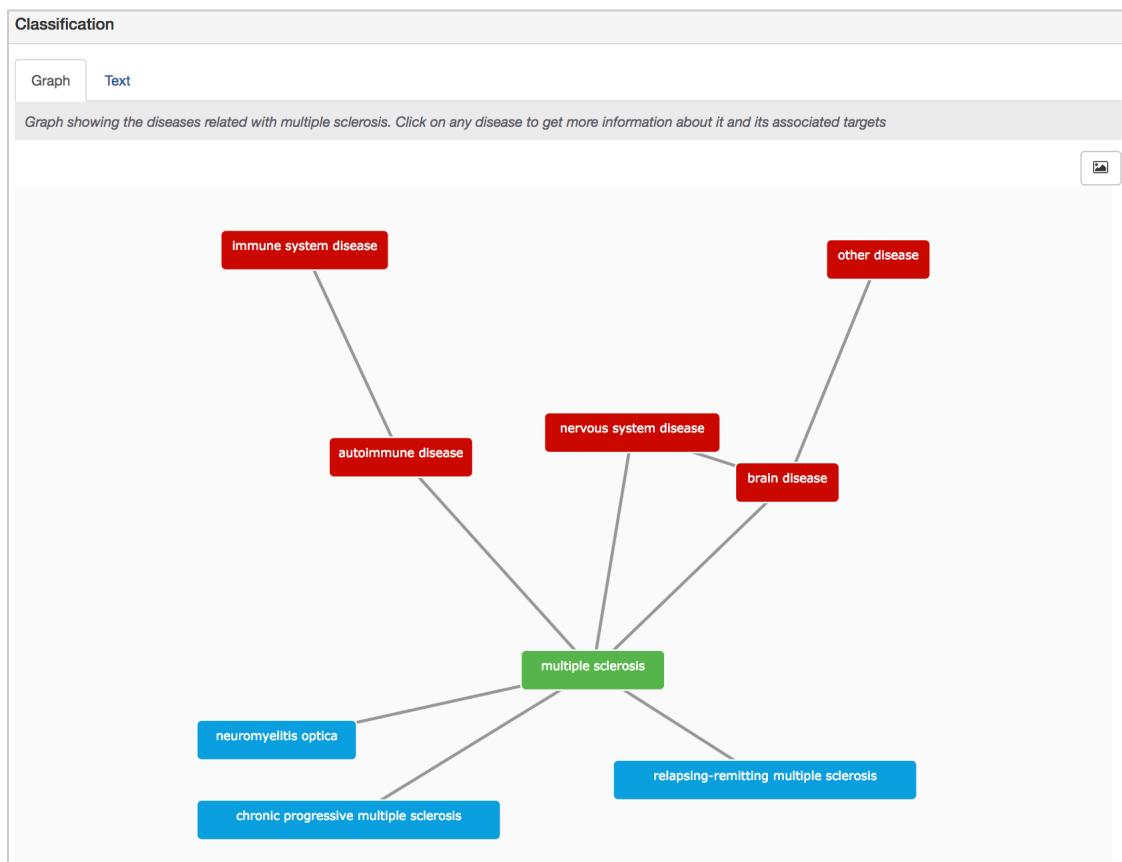
Drugs targeting multiple sclerosis.

Source: ChEMBL

Found 57 unique drugs: **ACETAMINOPHEN** **ALEMTUZUMAB** **AMANTADINE** **ARBACLOFEN PLACARBILO** **ASPIRIN** **BACLOFEN** **BRIMONIDINE** **CAFFEINE** **CETIRIZINE** **CHOLECALCIFEROL** **CORTICOTROPIN** **CORTICOTROPIN, REPOSITORY** **DACLIZUMAB** **DALFAMPRIDINE** **DEXTROMETHORPHAN** **DIMETHYL FUMARATE** **DONEPEZIL** **DOXYCYCLINE** **ERGOCALCIFEROL** **ESOMEPRAZOLE** **FINGOLIMOD**

You can filter (and sort) the drug table by disease, phase of clinical trial (e.g. III), class of the target (e.g. membrane receptor), and much more. You can also download this table in csv (comma separated value).

Next we can 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), and we developed the visualisation.



Multiple sclerosis is represented in green. Red nodes correspond to parental terms, whereas the children terms of multiple sclerosis are shown in blue (e.g. chronic progressive multiple sclerosis). 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
226 genes associated (Showing the first 10)	
CD52	<a href="#">See Evidence</a>
S1PR1	<a href="#">See Evidence</a>

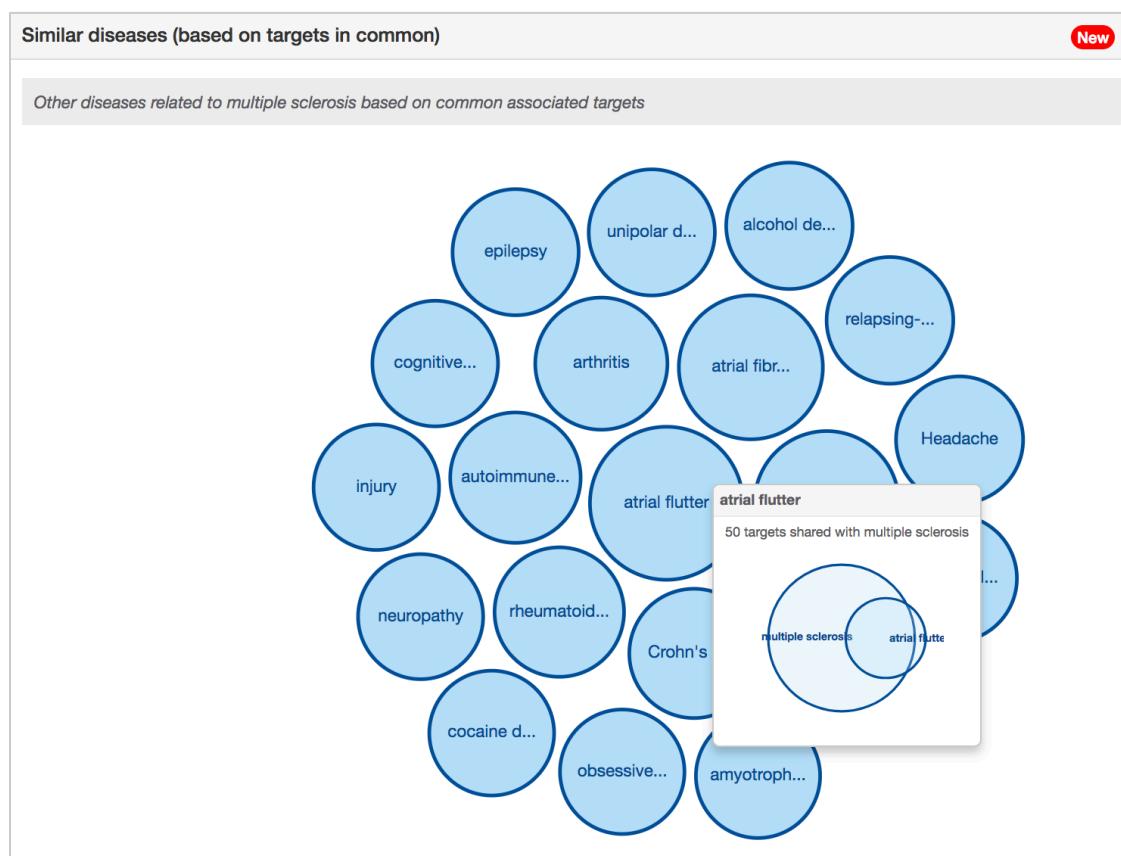
By using the EFO parent-child relationships, we can derive new associations that do not have direct evidence. For instance, IBD is an autoimmune disease and the direct evidence for targets associated with IBD are propagated to the higher autoimmune level to allow

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'.

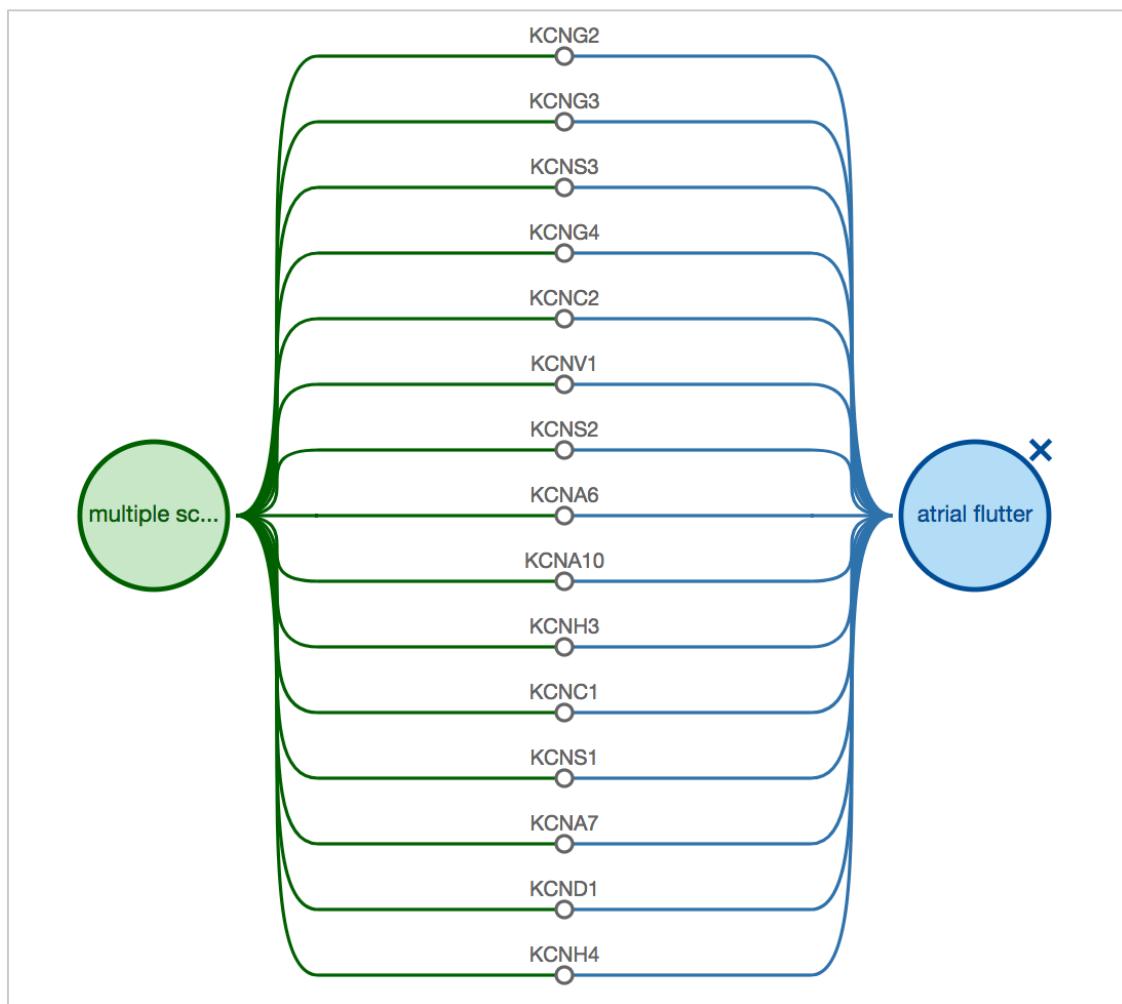
So you will get direct and indirect associations of diseases and genes. For more on this, check our blog post:

<https://blog.opentargets.org/direct-versus-indirect-evidence-should-you-care/>

There is more information on the disease on this page, including the new visualisation on similar (related) diseases:



Click on any of the bubbles to get details on the targets in common b and the evidence used for the associations:



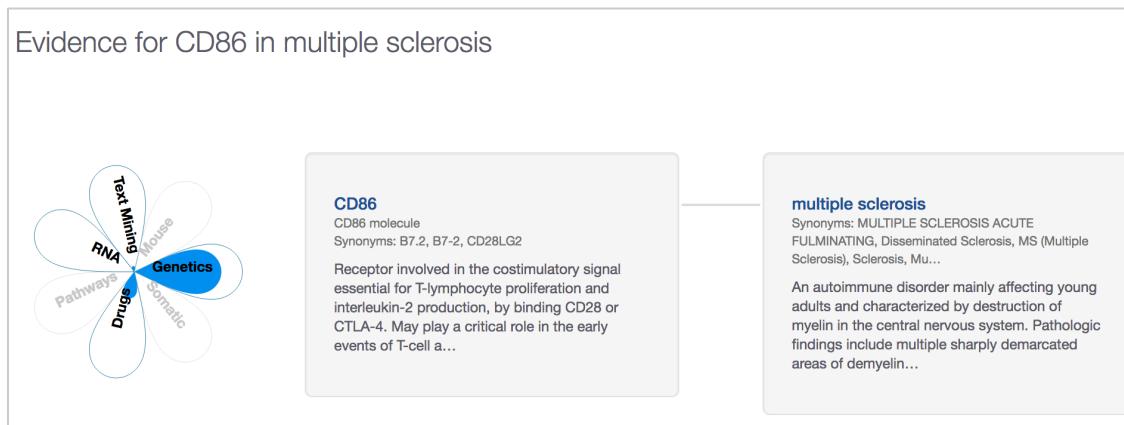
Let's now explore the evidence used to associate *CD86* and multiple sclerosis. We've chosen this target for follow up based on the availability of information from different data types.

If we click on any row corresponding to *CD86* in the table below:

Showing 1 to 50 of 295 targets  
Search:

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Target name
<a href="#">IL2RA</a>	1.00	0.55				0.07			interleukin 2 receptor subunit al...
<a href="#">TNFRSF1A</a>	1.00	1.00				0.06			TNF receptor superfamily mem...
<a href="#">KCNB2</a>	1.00	0.06	1.00						potassium voltage-gated chann...
<a href="#">SCN10A</a>	1.00	3.46e-5	1.00			0.04			sodium voltage-gated channel...
<a href="#">KCNQ1</a>	1.00	3.16e-5	1.00						potassium voltage-gated chann...
<a href="#">ESR1</a>	0.71	5.17e-5	0.70			0.05			estrogen receptor 1
<a href="#">CD86</a>	0.64	0.57	0.26	0.00	0.05				CD86 molecule

we will land in the evidence page for the association between a CD86 and multiple sclerosis:



The data types that support this association are (see 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)

Greyed out areas in the flower image above indicate there is no information for those data types, i.e. Affected Pathways, Animal models and Somatic mutations.

The genetic evidence that supports the association (*CD86* – multiple sclerosis) comes from the GWAS catalog and is processed by our pipeline.

These variants are known in public databases e.g. dbSNP (the hint is on the rsIDs, i.e. rs9282641, rs4308217, rs2255214, and rs2681424).

**Genetic associations**

Table Browser

**Common diseases**

Source: GWAS catalog, PheWAS catalog

Showing 1 to 5 of 5 entries

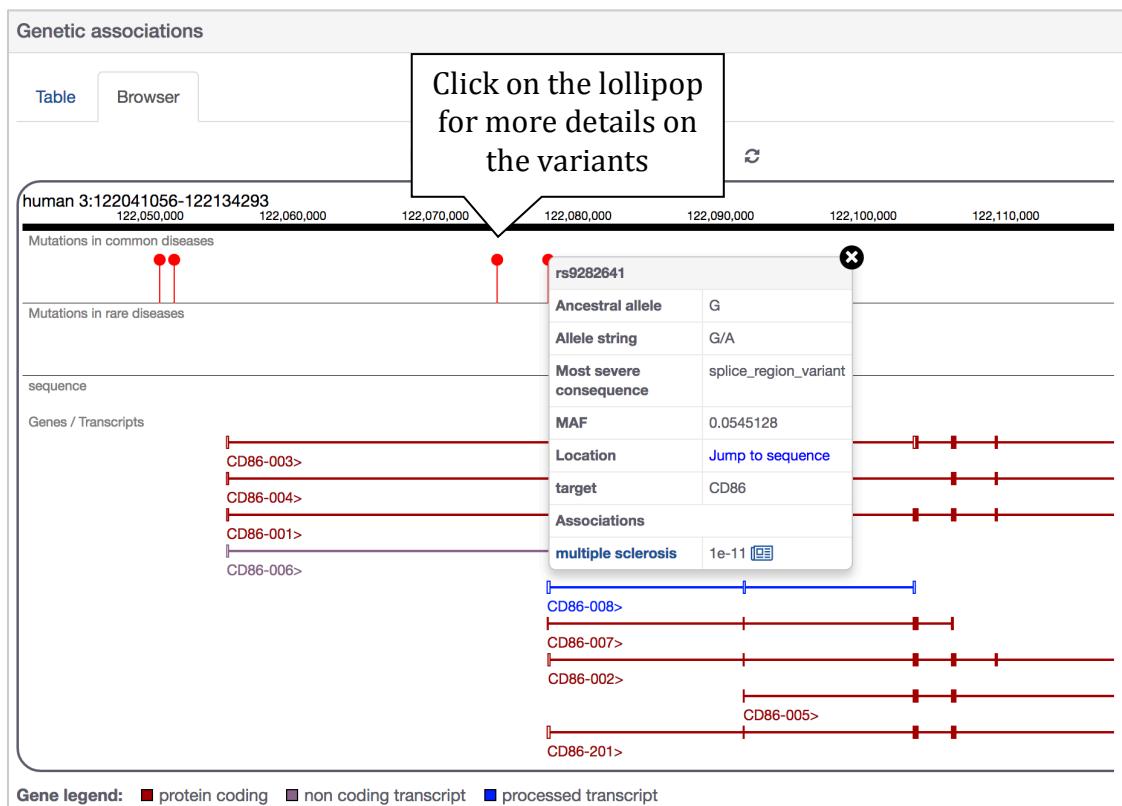
Search:  Download Feedback

Disease	Variant	Variant type	Evidence source	p-value	Publications
multiple sclerosis	rs9282641 ↗	splice region variant	gwas catalog ↗	1e-11	1 publication ↗
multiple sclerosis	rs2255214 ↗	upstream gene variant	gwas catalog ↗	1e-24	1 publication ↗
multiple sclerosis	rs4308217 ↗	intron variant	gwas catalog ↗	6e-8	1 publication ↗
multiple sclerosis	rs2255214 ↗	upstream gene variant	gwas catalog ↗	5e-8	1 publication ↗
multiple sclerosis	rs2681424 ↗	intron variant splice region variant upstream gene variant	gwas catalog ↗	2e-7	1 publication ↗

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 (depicted as lollipops).

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



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. There are three studies from clinicaltrials.gov, NCT01116427 and NCT00076934 (both for relapsing-remitting multiple sclerosis), and NCT00035529 (for multiple sclerosis).

You may also want to find out the research articles that have been mined for the co-occurrence of gene name and disease in the same sentence:

## Text mining

Source: Europe PMC

Shown are the 14 articles where target and disease are found in the same sentence.

Showing 1 to 10 of 14 entries



Year ▾

Disease ▾ Publication

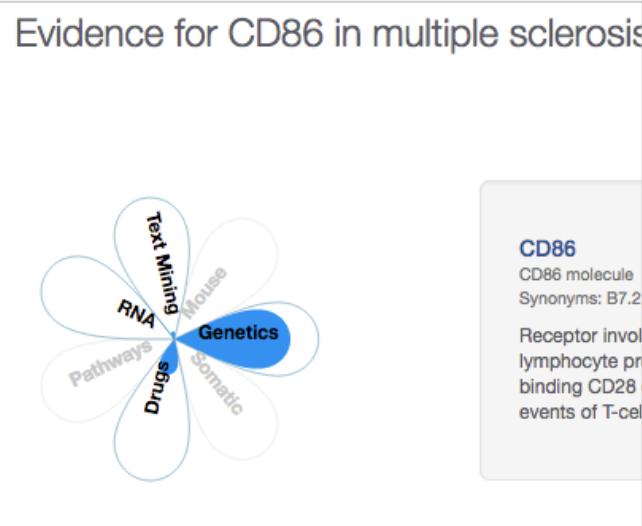
multiple sclerosis **Polymorphisms of RPS6KB1 and CD86 associates with susceptibility to multiple sclerosis in Iranian population.** 2017

Abdollah Zadeh R et al. *Neurol. Res.* 39(3):217-222  
PMID: 28079472

### Abstract

Multiple sclerosis (MS) is the most prevalent disorder of nervous system inflammation which involves demyelination of spinal cord; this process depends on both environmental and genetic susceptibility factors. In the present study, we examined the association between two SNPs in RPS6KB1 (rs180515) and CD86 (rs9282641) with MS in Iranian population. RPS6KB1 gene encodes p70S6K1 protein which plays a key role in mTOR signaling pathway, while CD86 gene codes a membrane protein type I which belongs to immunoglobulin super family act on co-stimulation signaling pathway. In this case-control study 130 patients with MS and 128 matched healthy controls were enrolled, genomic DNA was isolated and genotyping was performed using mismatched PCR-RFLP. The results were finally analyzed using SPSS. Our results showed significant difference in allelic frequency of SNP rs180515 among cases and controls ( $P = 0.004$ ). For this variation, AA genotype was shown to have protective effect ( $P = 0.016$  and OR = 0.6), while GG genotype was a susceptible genotype to MS ( $P = 0.04$  and OR = 2.2). Allelic frequency of SNP rs9282641 also showed significant difference between cases and controls ( $P = 0.006$ ). For this SNP, AG genotype had predisposing effect ( $P = 0.04$ , OR = 2.3), and GG genotype showed protective ( $P = 0.01$ , OR = 0.411). We successfully replicated the association of two novel SNPs introduced by a GWAS study, and MS in the Iranian population. This result can open ways for better understanding the mechanisms involved in MS.

Let's now scroll back to the top of the page and click on the 'CD86 link (next to the flower) to explore this gene in more detail.



The page you will land on is the Target profile page where you can find out the RNA expression levels of the gene across several tissues and whether or not it is associated with other diseases of the nervous system, among much more.

# 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.

(information provided by UniProt)

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](#)

## Drugs

## Protein Information

## Pathways

## Similar targets (based on diseases in common)

New

## Variants, isoforms and genomic context

## Protein interactions

## RNA and protein baseline expression

## Mouse phenotypes

New

## Protein Structure

## Gene Ontology

## Gene tree

## Bibliography

New

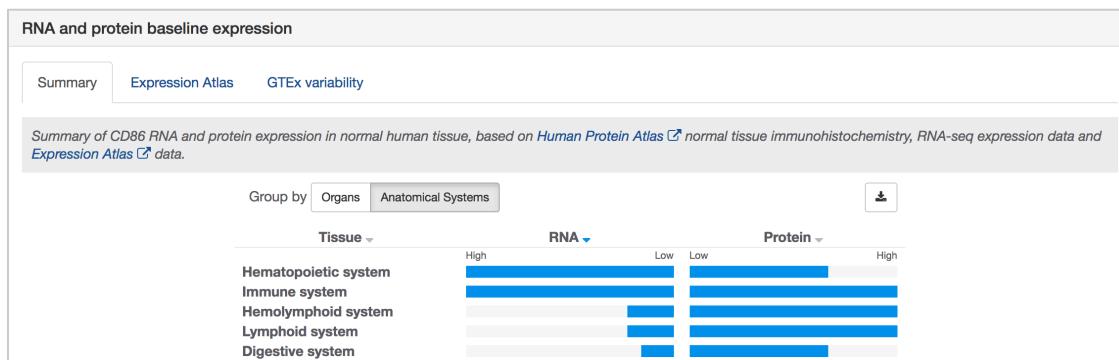
## Cancer hallmark

New

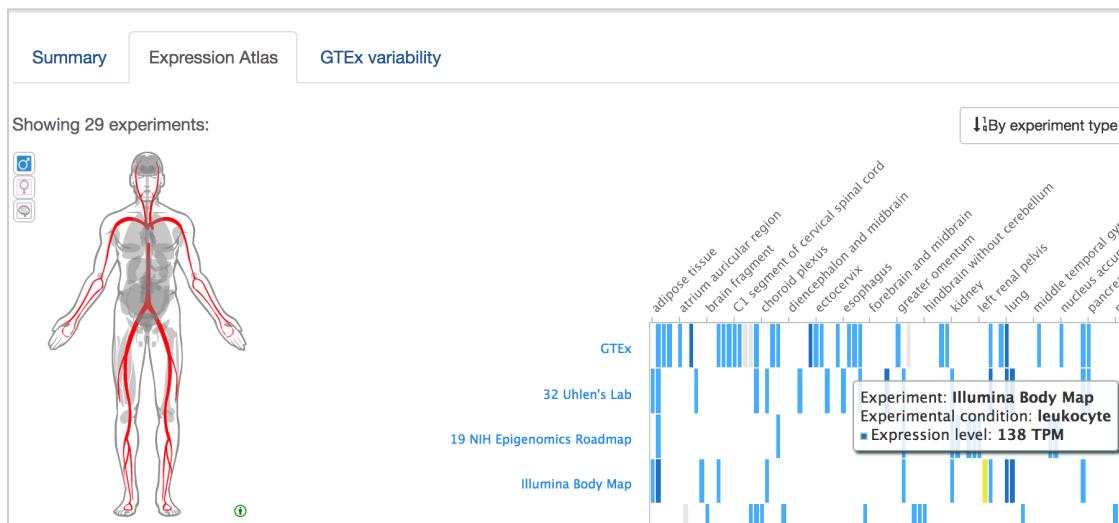
Expand the ‘RNA and protein baseline expression’ to find out in which tissues *CD86* is highly expressed.

You will find three tabs in there: “Summary”, “Expression Atlas” (data from several projects including the Illumina Body Map), “GTEX variability”.

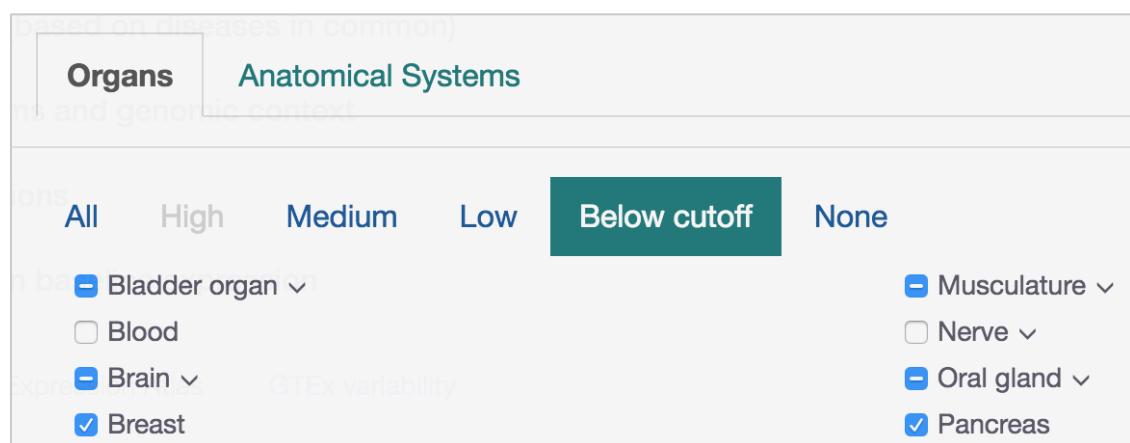
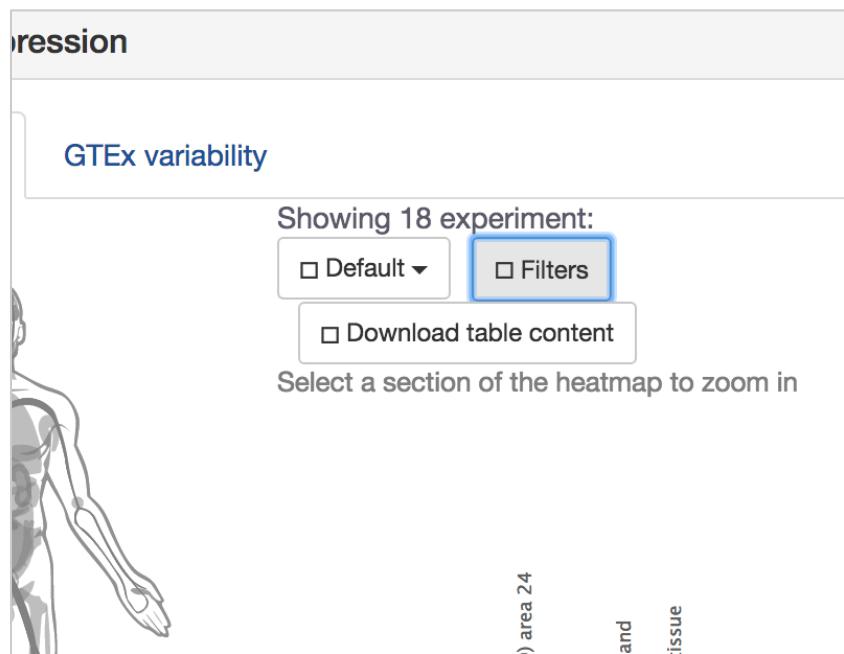
High expression at the mRNA level is in leukocytes, whereas at the protein level it's in the lymph node and spleen:



Check the 'Expression Atlas' tab for data coming from different projects e.g Illumina Body Map data and different techniques (RNASeq and microarray genotyping):



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



As per the Disease profile page, there is plenty of information on the Target profile page for you to explore. Go ahead and expand the different tabs there.

The latest news is in Bibliography with the inclusion of chips (as when searching for images in Google):

Bibliography New

dendritic cells t cell cd80 b7-2 maturation b7-1 cd28 cd40 t cells  
monocytes b cells antigen macrophages mice

✓ Concepts  
Genes Diseases Drugs Journal Authors

maturation and activation of human monocyte-derived dendritic

Sakhila Ghimire, Carina Matos, Massimiliano Caioni, Daniela Weber, Katrin Peter, Ernst Holler, Marina Kreutz, Kathrin Renner  
Immunobiology 2018 223(2):239-245

Indole is produced from L-tryptophan by commensal bacteria and further metabolized to indoxyl 3-sulfate (I3S) in the liver. Physiologic concentrations of I3S are related to a lower risk to develop graft versus host disease in allogeneic stem cell transplanted patients pointing towards an immunoregulatory function of I3S. Here we investigated the impact of I3S on the maturation of human monocyte-derived dendritic cells (DCs). Even pathophysiologic concentrations of I3S did not affect viability of mature DCs, but I3S decreased the expression of co-stimulatory molecules such as CD80 and CD86 on m ... [show more]

You can filter the results based on Concepts, Genes, Diseases, Drugs, Journal our Authors. You have the ability to mix and match different entities while filtering the bibliography.

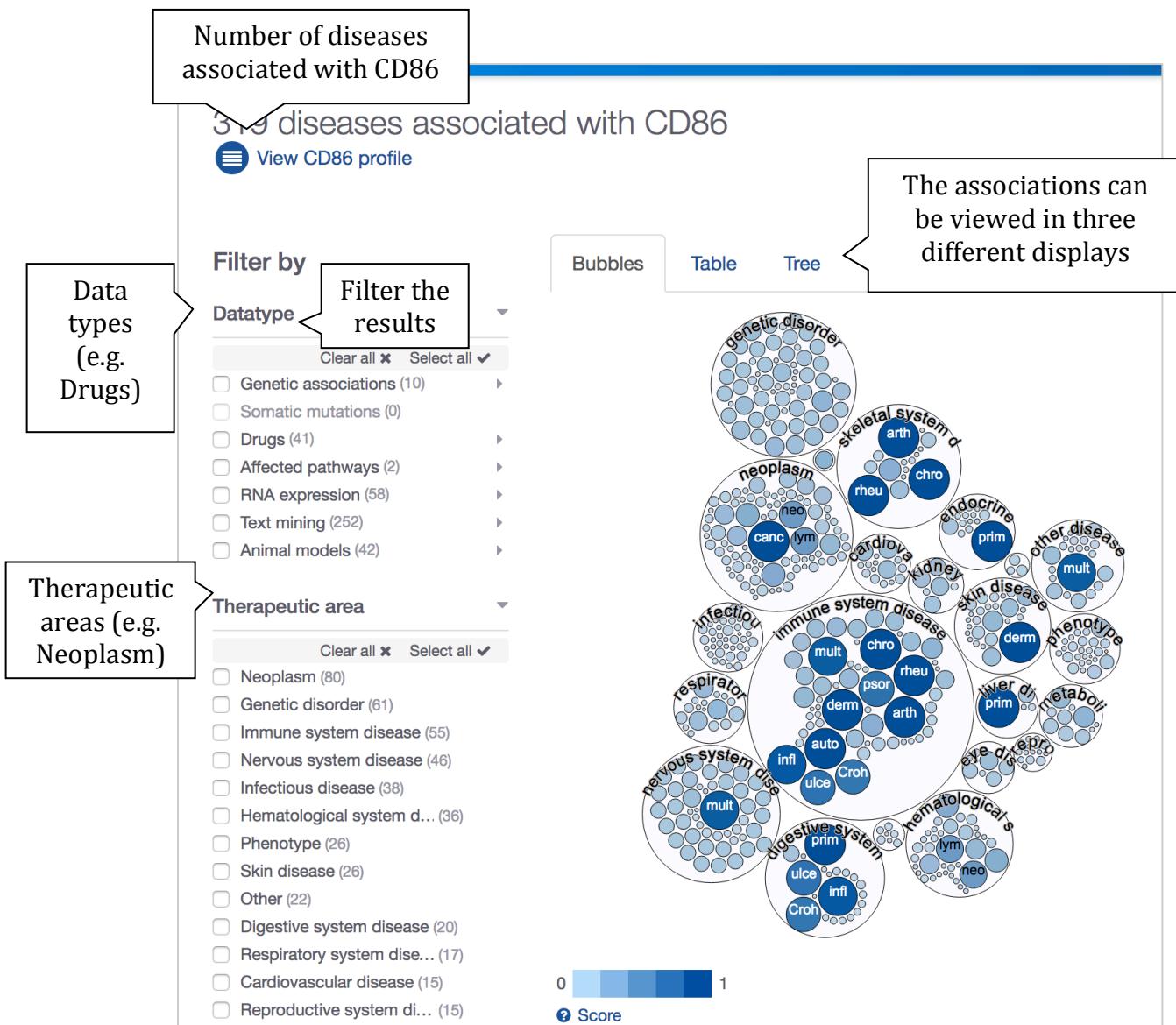
Let's now move from the Target profile page to the associations page for CD86. Click on 'View associated diseases' to find all diseases associated with the CD86 (apart from multiple sclerosis):

 Open Targets Platform About ▾

CD86

CD86 molecule |  View associated diseases

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:

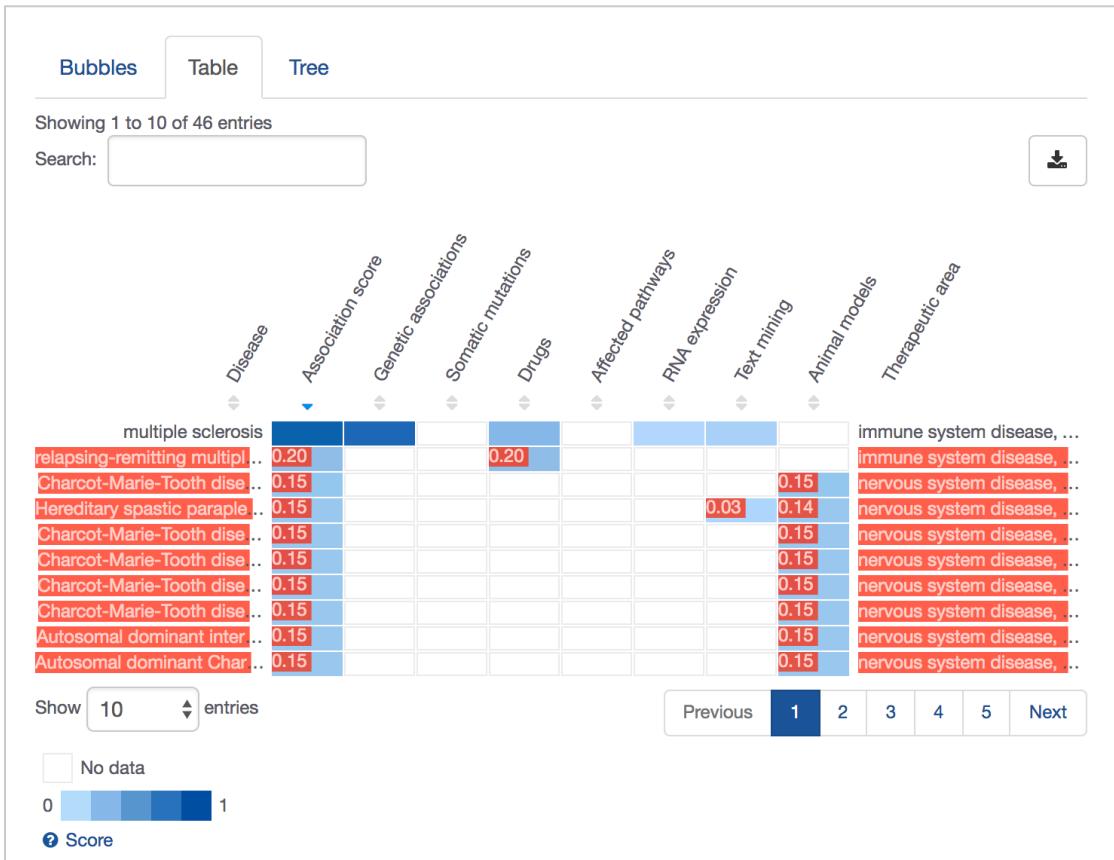
- Bubble view

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.



The screenshot shows a table interface for viewing associations between diseases and various biological and therapeutic areas. The columns are: Disease, Association score, Genetic associations, Somatic mutations, Drugs, Affected pathways, RNA expression, Text mining, Animal models, and Therapeutic area. The rows list diseases like multiple sclerosis, relapsing-remitting multiple sclerosis, Charcot-Marie-Tooth disease, etc. The 'Association score' column uses a color scale from white (0) to dark blue (1) to represent the strength of the association. A legend at the bottom left shows a white square for 'No data' and a color gradient from light blue to dark blue for 'Score'. A download icon is located in the top right corner of the table area. A red arrow points to this download icon.

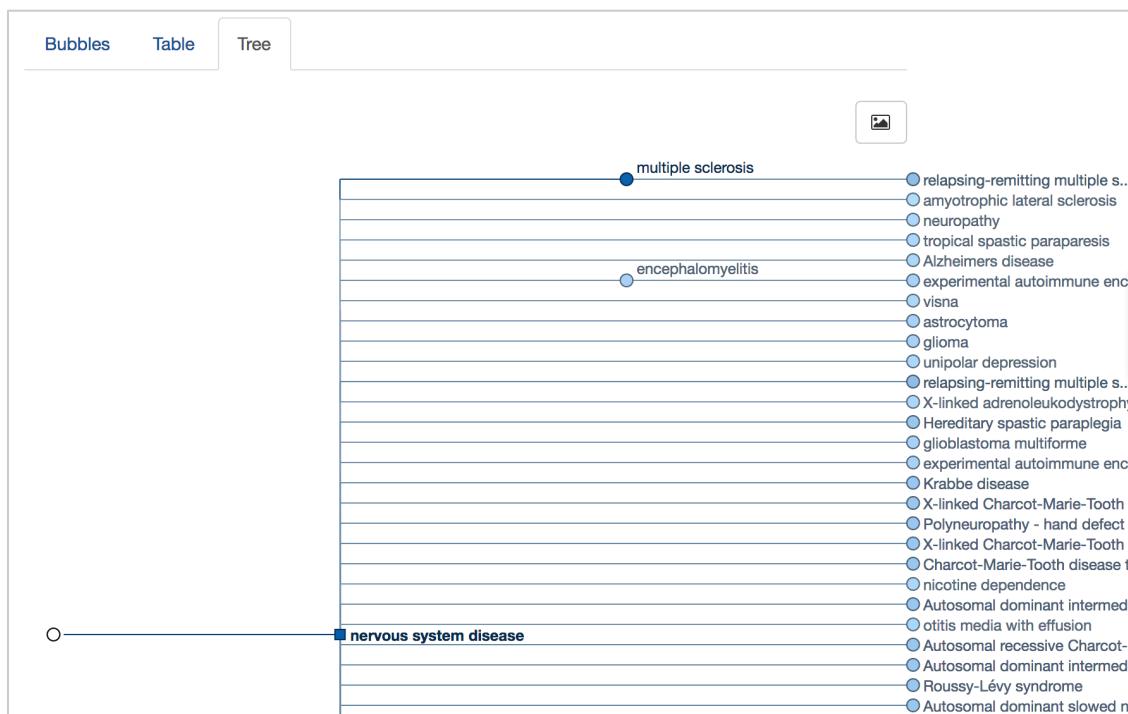
Disease	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Therapeutic area
multiple sclerosis									immune system disease, ...
relapsing-remitting multipl...	0.20		0.20						immune system disease, ...
Charcot-Marie-Tooth dise...	0.15					0.15			nervous system disease, ...
Hereditary spastic paraple...	0.15				0.03	0.14			nervous system disease, ...
Charcot-Marie-Tooth dise...	0.15					0.15			nervous system disease, ...
Charcot-Marie-Tooth dise...	0.15					0.15			nervous system disease, ...
Charcot-Marie-Tooth dise...	0.15					0.15			nervous system disease, ...
Charcot-Marie-Tooth dise...	0.15					0.15			nervous system disease, ...
Autosomal dominant inter...	0.15					0.15			nervous system disease, ...
Autosomal dominant Char...	0.15					0.15			nervous system disease, ...

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 47 of them (e.g. Hereditary spastic paraplegia, amyotrophic lateral sclerosis and neuropathy). The scores vary from just above 0.00 to 0.79 for glioma (and 0.98 for multiple sclerosis).

Click on any of the bubbles (or little squares/circles in the Tree view, or the row in the Table view) to explore more.

## 46 diseases associated with CD86

 View CD86 profile

## Filter by

## Data type

Genetic associations (4) ►

Somatic mutations (0) ►

Drugs (2) ►

Affected pathways (3) ►

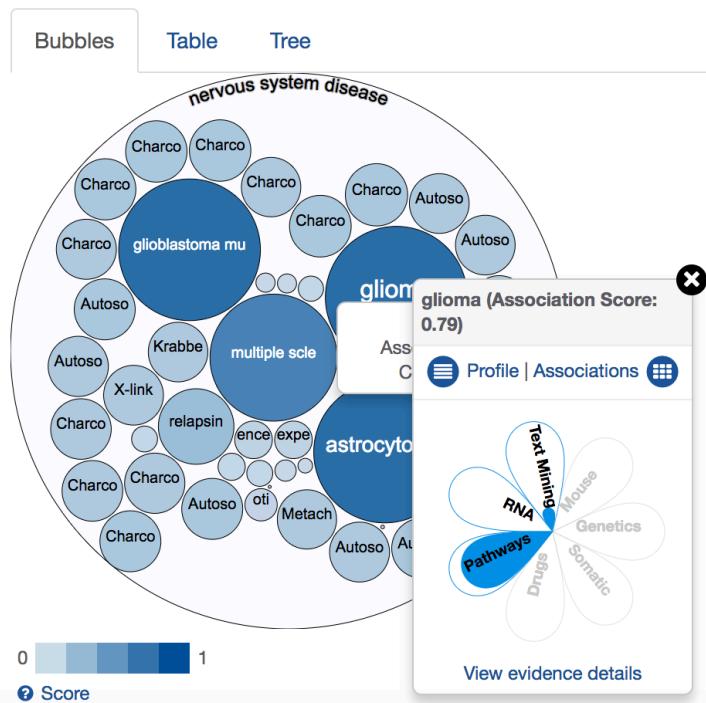
RNA expression (4) ►

Text mining (17) ►

Animal models (25) ►

## Therapeutic area

**selected**



---

## **HANDS-ON EXERCISES**

---

### **Exercise 1**

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

##### **BACKGROUND**

Biogen, one of our industry partners, 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) Which gene does this drug match to?
  - b) Which types of data support the association of this gene with Alzheimers? Which biochemical pathways seem to be affected by pathogenic mutations in this gene?
  - c) Still on the evidence page, expand the Genetic associations table to focus on the genetic variants linking this gene with Alzheimers. Are there any variants known as pathogenic by ClinVar database? Can you view these variants (or SNPs) in a genome browser display? Check where they map to on the human genome. Are they in the downstream or upstream region of the gene?
  - d) In addition to aducanumab, which other drugs target this gene (or protein)? Are there side effects to BAPINEUZUMAB?
  - e) In addition to Alzheimer's, can you name a few diseases from the nervous system where there is a strong confidence (association score >90%) in the association with the gene found in a)?
-

## **Exercise 2**

### **Searching for several targets at once**

#### **BACKGROUND**

The Drug Discovery Unit at the University of Dundee translates biology research into novel drug targets and candidate drugs to address unmet medical need across, such as infectious affecting mainly the developing world.

These are some of the targets that have been selected for further investigation: ODC1, GUCY1A2, GUCY1A1, CSF3R, CCR5, TLR7, and CD28.

#### **QUESTIONS**

- a) How can you search for all these targets at once using the web interface of the Open Targets Platform?
  - b) Which diseases do these targets map to?
  - c) Can you list a few of the pathways enriched in this list of targets? Find the Immune system pathway in this list and click on it get to a summary page, which includes an illustration of the pathway with the cells involved, receptors, etc.
  - d) Can you list a few of the drugs targeting some of these genes? Can you find the mechanism of action for APLAVIROC?
  - e) Do any of these targets seem to be interacting with each other?
-

## **Exercise 3**

### **Using the Open Targets Platform to assist space biology research at NASA's GeneLAB**

#### **BACKGROUND**

NASA's GeneLab has a diverse range of omic datasets available, which are related to space biology research. NASA's overall goal is to determine the biological risk of space travel and develop potential countermeasures to protect against the harm of long-term space missions. To do this, NASA aims at using existing therapeutic drugs for specific predicted diseases that can occur due to space travel. Through a systems biology approach and by using the existing NASA's GeneLab datasets, they have identified a list of key driving genes from multiple tissues from rodent experiments flown in space. Ultimately, NASA's GeneLab would like to determine possible clinical drugs that may be used as countermeasures to reduce associated health risk to long-term space missions.

Can the Open Targets Platform be used for this purpose? Let's find out.

#### **QUESTIONS**

a) Can you determine potential drugs for countermeasures against the following health risks to exposure to space:

- aI) cardiac arrhythmia
- aII) coronary artery disease
- aIII) contact dermatitis

b) We have a list of key driving genes from experiments with rodents and a subset of this is included below. Can you find the diseases these genes are associated with?

ACTN2

CYCS

DES

HRH1

IFIT2

IFNAR2

LDHA

MEF2C  
MYH8  
MYL1  
PFKM  
PML  
STAT1  
TNNI3  
CPT1A  
CPT2  
GPD2  
PPARA  
SLC2A1  
CDKN1A  
CEBPB  
DDIT4  
IFNGR1  
NFKBIA  
PTEN  
ACACA  
ACLY  
ALB  
CDKN1A  
CTNNB1  
FASN  
LCP2  
LPL  
STAT5A

- c) What pathways are over represented in the sublist of genes above?
- d) MECLIZINE is one of the drugs in clinical phase IV that targets HRH1. Can you list some of its known side effects?
- e) From the list of key genes provided can you determine potential drugs that can be given to astronauts as to reduce health risks associated with long-term space missions? Note you can use the batch search (maximum of 200 genes) or the REST API (for the entire list).
-

## **EXTRA HANDS-ON EXERCISES**

---

### ***Exercise 4***

#### ***LRRK2 in Parkinson's disease***

##### ***BACKGROUND***

*The LRRK2 gene encodes a protein with five putative functional domains: an N-terminal leucine-rich repeat (LRR) domain, a Roc (Ras of complex protein) domain that shares sequence homology to the Ras-related GTPase superfamily, a COR (C-terminal of Roc) domain, a mitogen-activated protein kinase kinase kinase (MAPKKK) domain, and a C-terminal WD40 repeat domain. A genetic variant in this gene is one of the most common causes of inherited Parkinson disease (Gandhi et al., 2008).*

##### ***QUESTIONS***

- a) *How long is the protein encoded by this gene/target? Can you find the protein domains listed above?*
  - b) *Are there similar targets to LRRK2 based on a common set of diseases? Take PDGFB, for example. Can you find a list of common diseases? What is the evidence available to associate each of these two targets (LRRK2 and PDGFB) with neurodegenerative disease?*
  - c) *Which tissue has the highest baseline expression at the protein level? Have a look at the visualisation showing the GTEx variability. Which tissue has the highest RNA expression?*
  - d) *Let's now have a look at the diseases associated with this target. Can you download a table containing all diseases from the nervous system where there is evidence for the associations from genetics and animal models?*
- 

### ***Exercise 5***

***Filter the target association table for Alzheimers based on a specific list of genes.***

## *BACKGROUND*

*This is a list of several genes associated with Alzheimer's from a list of various scientific papers:*

*HFE*

*PSEN1*

*TF*

*APOE*

*ADRB2*

*PSEN2*

*A2M*

*Can you search for the targets associated with Alzheimers but filter these results by using the list above to answer the following?*

## *QUESTIONS*

- a) *Which of these seven genes are significantly more expressed in the brain than other tissues?*
  - b) *Out of those expressed more in the brain, are there any classified as secreted proteins?*
  - c) *Where is this gene coding for a secreted protein located in the human genome (chromosome and coordinates)? Are there mutations in rare diseases (or phenotypes) mapping to this gene? Can you name a few of them?*
-

## **QUICK GUIDE TO DATABASES**

Here is a list of databases and projects that may be useful for you. Some of them are used as [data sources](#) for gene-disease associations available through our Platform.

### **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 (driver mutations), genes and pathways involved in cancer biology 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>

**Genomics England PanelApp** - The Genomics England PanelApp is a knowledgebase that combines crowdsourcing of expertise with curation to provide gene-disease relationships to aid the clinical interpretation of genomes within the 100,000 Genomes Project.

<https://panelapp.extge.co.uk/crowdsourcing/PanelApp/>

### **PheWAS Catalog**

The PheWAS (phenome-wide association studies) resources provide associations between a genetic variant and multiple phenotypes. It contains clinical phenotypes derived from the electronic medical record (EMR)-linked DNA biobank BioVU by the Center for Precision Medicine at the Vanderbilt University Medical Center.

<https://phewascatalog.org/>

## **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/>

**SLAPenrich** – It's a statistical framework for the identification of significantly mutated pathways, at the sample population level. We include in the Open Targets Platform the data obtained using SLAPenrich on somatic mutations from the The Cancer Genome Atlas across 25 different cancer types and a collection of pathway gene sets from Reactome.

<https://saezlab.github.io/SLAPenrich/>

## **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/>