**­­­­** Open Targets: integrating genetics, genomics and drug information for translational research and drug discovery

University of Cambridge



July 2020

Dr Denise Carvalho-Silva

Open Targets | EMBL-EBI

United Kingdom

(Coursebook created, written and edited by Denise Carvalho-Silva)

**Welcome to our training session!**

This booklet is based on the [release 20.06 of the Open Targets Platform](https://blog.opentargets.org/2020/06/17/open-targets-platform-20-06-has-been-released/).

Check the links below for more information:

**Open Targets Platform:**

Help documentation

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

FAQs

<https://docs.targetvalidation.org/faq/frequently-asked-questions>

Release blog posts

<http://blog.opentargets.org/tag/release-notes/>

Questions or suggestions?

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

How to cite the Open Targets Platform?

http://bit.ly/cite-us

**TABLE OF CONTENTS**

OVERVIEW 4

INTRODUCTION TO OPEN TARGETS 5

OPEN TARGETS PLATFORM: WALKTHROUGH 8

Demo 1: Disease centric workflow 8

Demo 2: Target and disease annotations 15

Demo 3: Target centric workflow 22

HANDS-ON EXERCISES 25

Exercise 1: Adalimumab and Crohn's disease 25

Exercise 2: Advancing research in the field of chronic obstructive pulmonary disease 25

Exercise 3: Filtering prostate cancer associations based on a list of eight targets 26

Video tutorials 27

# 

# 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 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, the Wellcome Sanger Institute and the EMBL-EBI. In February 2016, Biogen joined the initiative and the partnership was rebranded to Open Targets in April 2016. Since then we have welcomed three new partners: Takeda in 2017, and both Celgene (now Bristol-Myers Squibb company) and Sanofi in 2018.

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. However, 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 webinar:

* Open Targets partnership
* Open Targets Platform and how to browse its website
* Therapeutic areas for experimental research at Open Targets

# 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, data pipelines and core Genetics team have developed a suite informatics resources, namely Open Targets Platform, Open Targets Genetics, and LINK (the Open Targets LIterature coNcept Knowledgebase).

More details on our projects can be found on the [Scientific Overview](https://www.opentargets.org/science/) page and the complete list of publications and informatics resources is available on the [Resources](https://www.opentargets.org/resources/) page. Two publications worth noting specially researchers and drug discovery scientists interested in immunology and inflammation are:

* Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines
* Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases

**Open Targets Platform**

The first Open Targets informatics resource was launched in December 2015 and originally called Target Validation Platform, a database allowing for easy access to data relevant to drug target identification and selection. The tool was then renamed to Open Targets Platform.

In the Open Targets Platform, we use genetics, omics and chemical data from different [data sources](https://docs.targetvalidation.org/data-sources/data-sources) to associate genes and diseases. Similar data **sources** are combined into the following data **types** in the Open Targets Platform:

Genetic associations

Somatic mutations

Drugs

Pathways & systems biology

RNA expression

Text mining

Animal models

Different types of evidence (e.g. SNPs, scientific literature) from the data sources are used to compute the [association score](https://docs.targetvalidation.org/getting-started/scoring), which depends on the frequency of evidence, the confidence on the evidence and on its severity (for example, if a genetic variant changes the amino acid sequence of a protein, its severity is higher than variants that lie in the intronic region of a gene). We then aggregate the evidence score using the sum of the [harmonic progression](https://en.wikipedia.org/wiki/Harmonic_progression_(mathematics)) to obtain the score at the data source and data type levels, as well as the overall score. The association score can be used to rank target and disease associations in the Platform.

In the latest release of the Platform (June 2020, release 20.06) we have:

* 27,596 targets
* 13,914 diseases
* 7,282,832 target-disease associations

The Open Targets Platform is an open source and open access easy-to-use tool that can be applied in a variety of use cases in academia and pharma industry.

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

* Find targets associated with any disease
* Find diseases associated with any given target
* Find the evidence supporting target-disease associations
* Find annotations for targets e.g. tractability data, safety data
* Find annotations for diseases e.g. all drugs in trials or marketed
* Search for a list of up to 200 targets with the batch search tool
* Carry out more complex queries using the [REST API](https://docs.targetvalidation.org/programmatic-access/rest-api)
* [Download](https://www.targetvalidation.org/downloads/data) all evidence and association objects for all human genes and diseases to carry out your own downstream analysis

**Connect with us**

* [Open Targets Blog](file:///C:\Users\gk680303\Downloads\blog.opentargets.org)
* Follow us on [Twitter](http://www.twitter.com/targetvalidate), [Facebook](https://www.facebook.com/OpenTargets/), [LinkedIn](https://www.linkedin.com/company/centre-for-therapeutic-target-validation)
* Watch our video tutorials and short animations on [YouTube](https://www.youtube.com/channel/UCLMrondxbT0DIGx5nGOSYOQ)

# OPEN TARGETS PLATFORM: WALKTHROUGH

We will guide you through the website exploring psoriasis and the evidence used for the association of this disease with > 7000 targets.

The following points will be addressed during the walkthrough:

* + How to find targets associated with psoriasis
  + How to filter down the number of targets based on specific types of evidence, pathway types, target class, etc
  + How explore the evidence behind the NFKBIA-psoriasis association
  + How to find other diseases associated with NFKBIA
  + How to visualise the NFKBIA gene and its genetic variants on the genome
  + How to find drugs currently in clinical trials for psoriasis

## ****Demo 1: Disease centric workflow****

Go to [www.targetvalidation.org](http://www.targetvalidation.org) and search for psoriasis.

A screenshot of a cell phone

Description automatically generated

Select the first hit. You will be directed to a page like this:

A screenshot of a cell phone

Description automatically generated

Data types

(Somatic mutations,

Drugs, etc)

Total number of targets associated with psoriasis

Click on the arrows to sort the rows by (increased or decreased) score values of individual data types.

Filter the results

The current release of the Open Targets Platform (June 2020) lists 5700 targets associated with psoriasis.

The table in the associations page (see the image above) shows by default the best targets at the top of the table. The first target i.e. CARD14 is the target that contains the most compelling set of supporting evidence behind the association with psoriasis. This is summarised by the overall association score. The association score varies from 0 to 1; the closer to 1 the more evidence we have for an association.

This score is computed in four steps at the following levels:

* Evidence
* Data source
* Data type
* Overall

The top 100 individual scores for **evidence** within a data source are aggregated into one **data source score** followed by the aggregation of data source scores into a one **data type score**. The Overall association score column in the above table is the aggregation of all data source scores.

*Note 1: whenever we aggregate scores, we apply a harmonic sum.*

*Note 2: The ranking in the image above (e.g. CARD14 > IL36RN> IL12B) relies on the number of evidence available for the individual associations. Although the overall association score is capped at 1 for all those three targets, we have more evidence available for* CARD14*; hence it comes before IL36RN in the ranked table.*

We also apply different weight to different data types whilst computing the association score. RNA expression, animal models and text mining are all down weighted by a factor of 0.2, whereas Sysbio, PROGENy and SLAPenrich are down weighted by a factor of 0.5. Check the help documentation for more details on the [association score](https://docs.targetvalidation.org/getting-started/scoring) and [data sources](https://docs.targetvalidation.org/data-sources/data-sources) available in the Open Targets Platform.

*Note 1: that you can sort the table by alphabetical order of target names, or by the association score values, either overall score or per data type e.g. Genetic associations, Drugs, Text mining, etc.*

The association table listing all > 5,700 targets associated with psoriasis can be filtered by six categories:

1. Data types
2. Pathway types
3. Target class
4. RNA tissue specificity
5. Target tractability
6. Your target list

Let’s have a look at the individual options for each of these filters below:

**1) Data types:** we collect evidence from various data sources and combine them into categories called Data types. Examples of data sources are Open Targets Genetics Portal and UniProt, which are combined into the Genetic associations 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.

These are the data types where evidence was used for the associations with ovarian carcinoma:

Genetic associations

Somatic mutations

Drugs

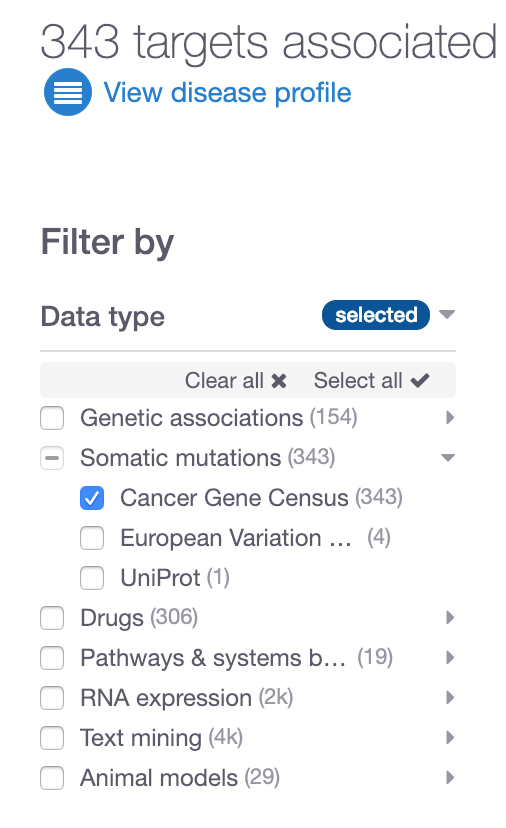
Pathways & systems biology

RNA expression

Text mining

Animal models

*Note: as our data types above can be made up of different data sources, click on the grey arrows to expand the options and filter the table to focus on specific data sources, such as Cancer Gene Census (data source) within Somatic mutations (data type):*

**

**2) Pathway types**: Reactome is the resource that provides us with pathway classification e.g. immune system (and its subtypes e.g. neutrophil degranulation), hemostasis (platelet degranulation), etc.

These are some of the pathway types currently available:

Immune System

Signal Transduction

Metabolism

…

Note that some of these pathways can be further subdivided into different subtypes, e.g. immune system can be broken down into neutrophil degranulation, Antigen processing and many others:

A screenshot of a cell phone

Description automatically generated

**3) Target class:** ChEMBL provides us targets into grouped into different classes such as Enzyme, Ion channel, Membrane receptor, etc.

**4) RNA tissue specificity:** RNA tissue specificity: the tissue specificity of a target is computed as the number of standard deviations from the mean of the log RNA expression of the target across the available tissues. This is a standard z-score calculation. A target is considered to be tissue specific if the z-score is greater than 0.674 (or the 75th percentile of a perfect normal distribution). We remove data for under-expressed targets before the z-score calculation. This RNA expression data comes from Expression Atlas.

This filter allows users to select the organs (or anatomical system) where the target is significantly more expressed in the selected tissues than the mean of the other tissues.

**5) Target tractability:** tractability of a target is the confidence that we can identify a modulator that interacts with the target to elicit a desired biological effect. Currently in the associations page you can filter whether targets can be modulated by a Small molecule or Antibody.

**6) Your target list**: in the associations page for a given disease, you can also filter the targets based on your own targets of interest. You can upload a list of targets (as .csv or .txt) and restrict the table to show only the targets in your list. This can help you to see the evidence Open Targets has integrated for your targets. Your own list of genes should be noted in official gene symbols from HGNC or Ensembl Gene IDs.

Note that in this associations page, there are two views, the Associations view and the Prioritisation view. The latter shows whether the available targets can be modulated by a small molecule, antibody or neither:

A screenshot of a cell phone

Description automatically generated

Now that we have looked at the different filters and explored the Prioritisation view, let’s now restrict the list of targets that are predicted to be tractable (druggable) by small molecule. Let’s now focus on NFKBIA.

Click on any of the cells to go to the evidence page for IL36RN in psoriasis.

Predicted tractable

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

Search for a target e.g NFKBIA

Click on any of the blue cells in the table to go to the evidence page and explore the underlying data behind the NFKBIA-psoriasis association:

A screenshot of a cell phone

Description automatically generated

Disease profile page

Target profile page

The coloured petals in the flower plot represent the data types that support this association. They are:

* Genetic associations
* RNA expression
* Text mining
* Animal models

**Note**: If you wish to suggest data types or resources we could/should incorporate into our Platform, please get in touch:

support@targetvalidation.org

Grey areas in the flower plot above indicate there is no evidence for those data types.

Let’s now scroll down on the page and expand the tabs available, for example, **Genetic associations** (with links to Open Targets Genetics, which provides us with common variant data), **RNA expression**, to find which studies show differential expression of this gene in health x disease comparisons, and **Text mining** to find the latest papers where the target and disease names occur in the same sentence and provide text evidence for the association.

Once you have explored the evidence, you can move on to look at target and disease annotations in the [target profile](https://docs.targetvalidation.org/getting-started/getting-started/target-profile) and [disease profile](https://docs.targetvalidation.org/getting-started/getting-started/disease-profile) pages, respectively.

## ****Demo 2: Target and disease annotations****

Let’s now scroll back up to the top of the evidence page and click on the “NFKBIA” link:

A screenshot of a cell phone

Description automatically generated

Click on the target link to get to the target profile page

You are now in the target profile page of this gene:

<https://www.targetvalidation.org/target/ENSG00000100906>

A screenshot of a social media post

Description automatically generated

This is the target profile page where you can find gene specific information for NFKBIA, e.g. target tractability, Protein information, RNA and protein baseline expression levels, Mouse phenotypes and more.

In Target tractability, for example, you will find that NFKBIA is predicted to be tractable by small molecules and antibodies, but not other modalities (such as oligonucleotides).

A screenshot of a social media post

Description automatically generated

You can download results of the target tractability pipeline from the [Data download](https://www.targetvalidation.org/downloads/data) page. Look for the “target tractability” file name.

Let’s now expand the RNA and protein baseline expression to find out in which organs or anatomical systems NFKBIA is highly expressed.

A screenshot of a social media post

Description automatically generated

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

In the Summary tab, you can compare the mRNA and protein expression side by side and learn that the expression in the reproductive system is higher at the protein than RNA level. You can click on the tissue names to get further granularity such as Esophagus, small intestine, etc within the digestive system.

Note that from the target profile page you can also explore other diseases associated with NFKBIA, apart from ovarian psoriasis, if you click on View associated diseases.

A screenshot of a cell phone

Description automatically generated

Let’s now go back to the previous evidence page (the flower page):

A screenshot of a cell phone

Description automatically generated

Click on the disease link to get to the disease profile page

Click on the disease name and explore the annotations for psoriasis:

<https://www.targetvalidation.org/disease/EFO_0000676>

Let’s check the Similar diseases (based on targets in common) tab in the disease profile page:

A screenshot of a cell phone

Description automatically generated

For each pair of diseases, we compute the overlap of shared targets against the total number of connections to both targets, correcting each pair by the significance and the specificity of these connections.

This procedure will consider targets that are specifically linked to fewer diseases more relevant than targets that are commonly linked to many types of diseases.

You will see that the target profile page has a similar visualisation under the tab Similar targets (based on diseases in common). There we will compute a closer distance between two targets sharing a rare disease than two targets sharing diseases that are highly connected to many genes, such as cancer.

You can click on any of the bubbles to get details on the targets in common between any two diseases (e.g. psoriasis and dermatitis) and the evidence used for the associations:

A screenshot of a computer

Description automatically generated

Scroll down to view the disease ontology (disease concepts and relationships) under the Classification tab.

A screenshot of a social media post

Description automatically generated

Psoriasis is represented in green. Red nodes correspond to parental terms, whereas the children terms are shown in blue (e.g. psoriasis vulgaris). Click on any of the disease names to get a pop-up box with the (first) 10 targets associated with any disease:

A screenshot of a social media post

Description automatically generated

The disease ontology has an important role in the Open Targets Platform to reveal target-disease associations that are supported by what we refer to as indirect evidence. For more on this concept, check our [direct versus indirect](https://blog.opentargets.org/direct-versus-indirect-evidence-should-you-care/) blog post. For an example of the role of indirect evidence, lets use IBD as an example. IBD is an autoimmune disease that will have direct evidence for its association with its targets. We can propagate this direct evidence up to higher terms in the ontology of IBD and use this evidence to associate target X with autoimmune disease (a parent term of IBD). This procedure can allow us to find common targets across groups of related diseases (e.g. Ulcerative Colitis, Crohn's disease and IBD) even when direct evidence is not available.

## ****Demo 3: Target centric workflow****

You can also use the Open Targets Platform to search for targets (genes and proteins). Let’s search for ADORA1 and explore some the data available in the Open Targets Platform.

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

* 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 .csv button).

*Tip: The different shades of blue in the table convey the strength of the association based on the available evidence (strongest association is represented 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.*

* 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.

* Tree view

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

For all these three different views, you have the option to filter the data according to Data type or Therapeutic area. You can, for example, explore all diseases in the therapeutic area “Immune System Disease”.

This is the bubbles view for all diseases in the immune system area associated with ADORA1.

A screenshot of a cell phone

Description automatically generated

Click on any bubble such as Crohn’s disease to get a pop up with the following options:

* “Profile” to get to the profile page of Crohn’s disease
* “Associations” to get all the target associations for Crohn’s
* “View evidence details” to see the underlying evidence for the associations

---------------------------------------------------------------------------------------------------------

End of the walkthrough

---------------------------------------------------------------------------------------------------------

# HANDS-ON EXERCISES

## Exercise 1: Adalimumab and Crohn's disease

BACKGROUND

Adalimumab (brand name Humira) is a medicine used to reduce signs and symptoms, and to achieve and maintain clinical remission in adults with moderate to severe Crohn’s disease who have not responded well to certain other medications. Adalimumab is an immunosuppressive medication and is used to treat other autoimmune diseases. Let’s use the **Open Targets Platform** to find out which target is modulated by Adalimumab.

QUESTIONS

a) Search for Adalimumab. How many diseases are returned? Which targets are modulated by this drug?

b) Let’s now focus on the target returned above. How many diseases from the therapeutic area “Immune system disease” and with evidence from the Open Targets Genetics Portal (under the data type called “Genetic associations) are listed in the Open Targets Platform?

c) Let’s look at one association between this targets and TNF in rheumatoid arthritis. Which data supports this association? In addition to adalimumab, are there any drugs currently in **Phase III** of clinical trials, whose **status** is “recruiting”, and type is “protein”? If you click on this drug, you will be redirected to the “drug summary” page where you can see a summary of some of the post-marketing adverse events reported for this drug (under “Pharmacovigilance”).

## Exercise 2: Advancing research in the field of chronic obstructive pulmonary disease

BACKGROUND

Chronic obstructive pulmonary disease (COPD) is a type of obstructive lung disease characterized by long-term breathing problems and poor airflow. Let’s use the **Open Targets Platform** to explore some of the targets associated with this disease.

QUESTIONS

a) How many targets that are membrane receptors (therefore target class amenable to antibody drug modality), which are expressed more in the lung than any other tissues (under “RNA tissue specificity”) are associated with chronic obstructive pulmonary disease (COPD)? Let’s look at the “Target tractability” filter at the left-hand side of the page. How many of the targets that meet the conditions above have clinical precedence?

b) Let’s now have a look at another target associated with COPD, one with strong evidence from the data type “Drugs”; the target is CHRM3. Look at the evidence for genetic associations. What is the variant ID (e.g. rs123) for the genetic variant from the Open Targets Genetics Portal used to associate COPD with CHRM3.

c) Is there any target safety related information for CHRM3? Explore the list and the source of this information. For more details, check the [Target safety](https://docs.targetvalidation.org/getting-started/target-safety) help page.

d) Can you find the mouse phenotypes that mimic some of the human phenotypes in the respiratory system in knockout mice?

## Exercise 3: Filtering prostate cancer associations based on a list of eight targets

BACKGROUND

A drug discovery scientist at CRUK Therapeutic Discovery Laboratories in London has a list of eight targets that seem to be associated with prostate carcinoma based on the literature. These are CHEK2, AR, CDKN1A, CDK12, PTEN, TSC2, KLF6, and MXI1. This list is saved as exercise3.txt file.

Go to the **Open Targets Platform** and search for prostate carcinoma. In the association page listing all targets for this disease, look for **Your target list** section at the left-hand side menu. Click on the “Choose file” button, search for your file (exercise3.txt) and upload this to the Open Targets Platform.

Once you have done this, you will be able to answer the following:

QUESTIONS

a) Which of those eight targets have higher levels of mRNA expression in the prostate gland than in any other tissue? Note: this is known in the Open Targets Platform as RNA tissue specificity. Prostate gland is under the “Reproductive structure”.

Let’s now explore the evidence for the association between one the targets and prostate carcinoma. The target we will explore is the androgen receptor (AR).

b) Are there any variants or mutations that cause a missense change in the AR protein and that has clinical significance of being pathogenic? Can you find the publications that support this?

c) Are there any drugs in Phase III of clinical trials that have been completed where AR is modulated in patients suffering from prostate carcinoma? Can you download this data in .csv?

d) How many unique drugs in total are in clinical trials (or already marketed) for prostate carcinoma? Note the different types of drugs available, not only small molecule and antibodies. Can you name the only drug that is of type “Oligonucleotide” in trials for this disease?

## Video tutorials

Check the [Open Targets YouTube channel](https://www.youtube.com/channel/UCLMrondxbT0DIGx5nGOSYOQ) for recorded webinars, quick animations and video tutorials on the Open Targets Platform and Open Targets Genetics.