Open Targets: integrating genetics, omics and chemical data for drug discovery

21st August 2019

Denise Carvalho-Silva
Maya Ghoussaini

Open Targets
Wellcome Genome Campus
United Kingdom

Notes

This booklet is based on the latest releases of Open Targets Platform and Open Targets Genetics, June 2019. Check the links below for more help:

Open Targets Platform:

Release blog posts
http://blog.opentargets.org/tag/release-notes/

Help documentation
https://docs.targetvalidation.org/

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

Open Targets Genetics:

Release blog posts
http://blog.opentargets.org/tag/genetics-releases/

Help documentation
https://genetics-docs.opentargets.org/

FAQs
https://genetics-docs.opentargets.org/faqs

Questions or suggestions?
support@targetvalidation.org
geneticsportal@opentargets.org

TABLE OF CONTENTS

HANDS-ON EXERCISES

Open Targets Platform

Exercise 1: Durvalumab and non-small cell lung carcinoma

BACKGROUND
AstraZeneca and MedImmune have announced the final overall survival results for the Phase III MYSTIC trial, a randomised, open-label, global trial of durvalumab. This a human monoclonal antibody that binds to PD-L1, blocking its interaction with PD-1 and CD80, to counter the tumour's immune-evading tactics and release the inhibition of immune responses.

QUESTIONS
a) How many targets have associations involving this drug in the Open Targets Platform? *Note: the extent of the match between the search term (durvalumab) and one of the targets returned on the search results page.*

b) Let's now focus on the target returned in the result of the previous question. Which data types and sources support the association of this target with non-small cell lung carcinoma?

c) Looking more closely at the evidence for the association between this target and non-small cell lung carcinoma, open the "Drugs" section and find "durvalumab" in the table provided. Are there any clinical trials that have been withdrawn by the FDA?

d) Now click on the hyperlinked drug name. You will be redirected to the "drug summary page". When was this drug first approved? In addition to non-small cell lung carcinoma, can you name other diseases where this drug is in clinical trials? Have any adverse effects been reported for this drug?

e) Is there any evidence showing a change in the **RNA expression** of this target? Does the change correspond to a decrease or increase in RNA levels when comparing 'primary tumor' vs 'adjacent normal tissue`? Note: an increase in RNA expression in the disease tissue versus normal would suggest the gene is up-regulated in the disease state, whereas a decrease would suggest the opposite i.e. the gene would be down-regulated in the diseased tissue.

f) Are there any papers from 2019 that support the link between this target and non-small cell lung cancer following **text mining**?

Exercise 2: Advancing research in the field of IBD

BACKGROUND
More than five million people worldwide live with inflammatory bowel disease (IBD). While the causes of IBD are unknown, several hypotheses have been suggested, such as genetic predisposition, environmental triggers, chronic and aberrant inflammation. Use the Open Targets Platform to answer the following:

QUESTIONS
a) How many targets are associated with IBD? How many of those are involved in the interleukin-4 and 13 signaling pathway (immune system) and are of target class 'Membrane receptor'. *Note: you will select these different options located in the left-hand side of the associations page for IBD.*

b) Continue with the selection from a). What is the only target from the list above that has evidence from the GWAS Catalog for associations with IBD?

c) Let's now see the genetic associations available for this target and IBD. How many SNPs are reported for this association? Is this gene associated with other diseases than IBD?

b) Let's find out some annotations for this target by exploring its profile page. Can you answer the following?

- Is this target tractable? Is there stronger evidence to support tractability for small molecules or antibody modalities?

- Which amino acids correspond to antigenic sequences in this protein? *Note that antigenic sequences are sequences where antibodies can bind to (Hint: Expand the Protein information tab in the target profile page to find this information*).

- In addition to Interleukin-4 and Interleukin-13 signaling, which other pathway (s) is this protein involved in?

- Which tissue has the highest level of RNA baseline expression? Does this correspond to the highest expression at the protein level as well?

- Can you list a few phenotype labels related to the immune system phenotype category in mice, when this gene is knocked-out? Download this table.

- In addition to mouse, can you use other model organisms (e.g. guinea pig, rat) to assess what the modulation of EGFR would entail? *Note: use the "Gene tree" tab in the profile page to get this information.*

Open Targets Genetics

Exercise 3: Demonstrating the role of genetics supporting an existing drug in retrospect

Search for LDL cholesterol and pick Willer CJ (2013) Nat Genet.

- How many loci are independently associated with LDL cholesterol at genome-wide significance (p-value < 5e-8) in this study?

In the Manhattan-like plot, zoom in on chromosome 5.

- How many independent associations can you find?

In the table, click on the lead variant, 5_75329662_C_A.

- Which genes are functionally implicated by 5_75329662_C_A in the Summary tab in the "Assigned genes" table?

- Can you rank these genes by the distance to canonical TSS (transcript start site), from closer to farther to the TSS? Note: by

default, the ranking is set for the overall V2G score, from the highest to the lowest.

- What functional evidence supports the link of this variant to gene *GCNT4*?

Click on the eQTL tab in the same table to view tissue and direction of effect.

- In which tissue is there eQTL evidence for HMGCR?
- What is the direction of effect?

Scroll down to the PheWAS plot.
You can see that 'LDL cholesterol' (Study ID GCST002222) is the most significantly-associated trait.

- What other traits are associated with this variant at phenome-wide significance? *Note the direction of effect by the triangles pointing upward or downward.*

There are several related studies from GWAS Catalog and UK Biobank associated with this variant e.g. 'Disorders of lipoid metabolism' (SAIGE_272), 'Hyperlipidemia' (SAIGE_272_1) as well as simvastatin and atorvastatin (Treatment/medication codes).

Now, let's look closer at the HMGCR locus. In the table below the PheWAS plot, click on the 'Locus Plot' button for the UK Biobank study ' high cholesterol |Non-cancer illness code, self-reported'.

Use the drop-down to toggle between LD and fine mapping at this locus. The table below the figure displays the variants tagging this lead variant and the genes functionally implicated by these tag variants.

To learn more about HMGCR and to get a list of other studies associated with this gene, click on HMGCR in the table.
The link, ' Is there known drug data? ', directs you to the Open Targets Platform where you can view additional information about the gene including drugs approved or in clinical trials.

Exercise 4: Identifying diseases and molecular QTLs that colocalise with a disease-associated signal

BACKGROUND

VEDOLIZUMAB is an approved (Phase IV) drug targeting ITGA4 to treat Crohn's disease. We will use Open Targets Genetics to find if independent association signals at a ITGA4 are consistent with having a shared casual variant. This is known as colocalization.

QUESTIONS

a) Which studies have evidence of colocalisation with molecular QTLs for ITGA4?

b) One of the lead variants in this colocalisation study is 2_181443625_A_G. Click on any of the **Colocalisation** buttons in the table. *Note all buttons will lead to the same colocalisation page between study and the variant (i.e. 2_181443625_A_G)*.

c) Let's first look at the 'QTL colocalisation' section. What molecular traits (e.g. eQTL, pQTL) colocalise with this Crohn's disease (de Lange KM, 2017) signal? What tissues is the colocalisation reported?

d) Let's now look at the 'GWAS study colocalisation'. What GWAS traits colocalise with Crohn's disease (de Lange KM, 2017) at this locus? What is the evidence for this?

Scroll down to the 'Credible Set Overlap' section. This visualisation shows the overlap between fine mapping credible sets that colocalise with Crohn's disease (de Lange KM, 2017) at this locus and can help us to identify which variants at this locus are most likely causal.

e) Do the 'Crohn's disease' and 'CEDAR: ITG4A in Monocyte CD14' signals look the same in the regional plots?

f) Can you identify the set of likely causal variants at this locus based on the Crohn's disease and ITG4A Monocyte signals? Download this information in "TSV" for further prioritization analysis.

EXTRA HANDS-ON EXERCISES

Exercise E1: Assessing the specificity of a list of targets for Barrett's esophagus

BACKGROUND
A biologist working on translation medicine at the Imperial College of London has a list of genes linked to Barrett's esophagus but he/she would like to know how specific this list to that disease or whether this set of genes could be therapeutic targets for other diseases of the digestive system.*

QUESTIONS
a) Are there other diseases affecting the oesophagus where this list of targets is also specific to?

b) What is the most enriched pathway for this list of genes?

c) Which of those genes seem to be the most tractable ones for either small molecule or antibody.

d) Are there any drugs currently in clinical trials phase I or II targeting any of these genes?

e) Is any protein-protein interaction predicted to exist in this set of genes?

**The list can be download from https://tinyurl.com/batch-kogo-0219.*

Exercise E2: Filtering Alzheimer's disease associations based on a list of targets

BACKGROUND
A drug discovery scientist at Alzheimer's Research UK has a list of eight (n=8) genes as possible drug targets in Alzheimer's disease (AD).

Can you upload this list to the Open Targets Platform associations page for Alzheimer's and answer the questions below? *Note: you will need to save this list as a .txt file, one gene name per row.*

HFE
PSEN1
PRO1557
APOE
ADRB2
PSEN2
CPAMD5
BACE1

QUESTIONS

a) How many of these targets have data on animal models used as evidence for the association with Alzheimer's? Is the association supported by direct or indirect evidence?

b) Let's go back to the associations page listing all eight targets from your list. Which of those eight targets have higher levels of mRNA expression in the cerebral cortex than in any other tissue (this is known in Open Targets as RNA tissue specificity)?

c) Now go to the target profile page for one of your targets in the list, BACE1. Are there any known safety concerns for this target?
and what are the phenotypes in animal models, such as mouse?

Exercise E3: 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) No drug is currently available to target LRRK2. There may be other compounds, such as chemical probes (small molecules that will alter the function of biological target)? Can you use the Open Targets Platform to find which chemical probes, if any, are available that could be used to modulate the function of this protein?

c) Can this protein be targeted by either a small molecule or antibody (tip: have a look at the Target tractability information)?

d) Can you list some of the proteins that interact with LRRK2? Can you download this image?

e) Let's now have a look at the diseases associated with this target. Can you name a few diseases in the "Nervous system disease" and "digestive system disease" therapeutic areas for which there is evidence for Genetic associations in this gene?
