

Mining gene-disease associations for drug identification and discovery with Open Targets



**Hands-on Workshop
Answer booklet**

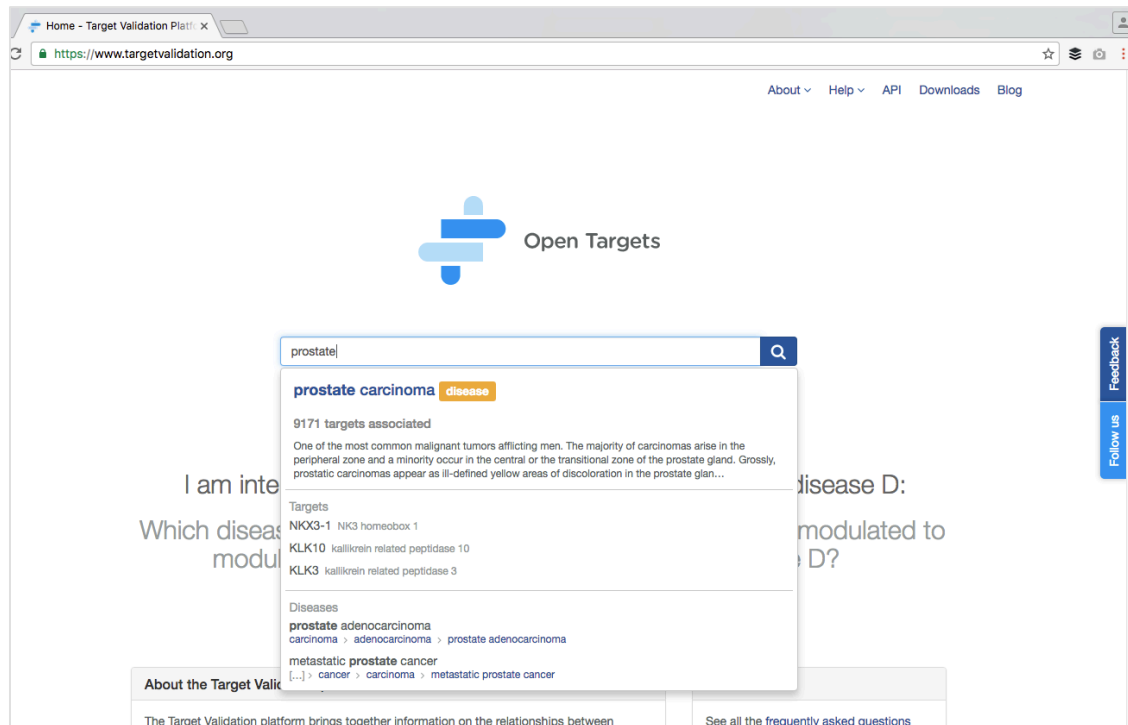
**CRUK Manchester Institute
22nd November 2016**

**Denise Carvalho-Silva
Open Targets Outreach**

Answers to exercises 1 and 2, pages 23-25 of coursebook

Exercise 1 – Prioritising targets for drug discovery in prostate carcinoma

a) Go to www.targetvalidation.org and search for *prostate carcinoma*:

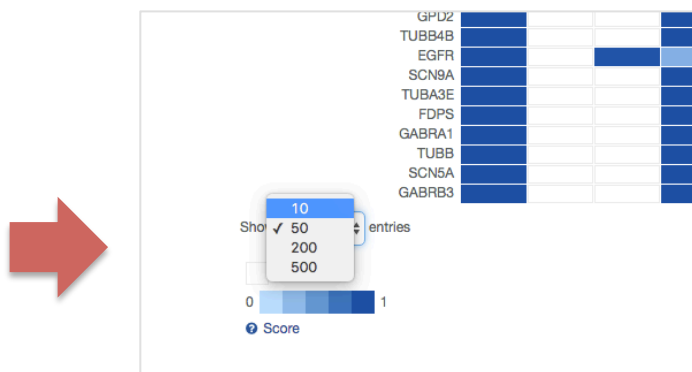


Select the first (best) hit. You will then see a page like this, which lists 9171 targets associated with prostate carcinoma:

The screenshot shows the results page for '9171 targets associated with prostate carcinoma'. The page includes a search bar, a filter section, and a table of results. The filter section allows users to filter by 'Data types' (Genetic associations, Somatic mutations, Drugs, Affected pathways, RNA expression, Text mining, Animal models) and 'Pathway types'. The table displays 10 entries (rows) with columns for Target symbol, Association score, Genetic associations, Somatic mutations, Drugs, Affected pathways, RNA expression, Text mining, Animal models, and Target name. The first entry is PTEN, which is associated with all the listed data types and pathways.

Target symbol	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Target name
PTEN									phosphatase and tensin homolog
CHEK2									checkpoint kinase 2
KLK6									Kruppel like factor 6
FGFR4									fibroblast growth factor receptor 4
AR									androgen receptor
HOXB13									homeobox B13

Scroll down and select to see the results with 10 entries (rows) only:

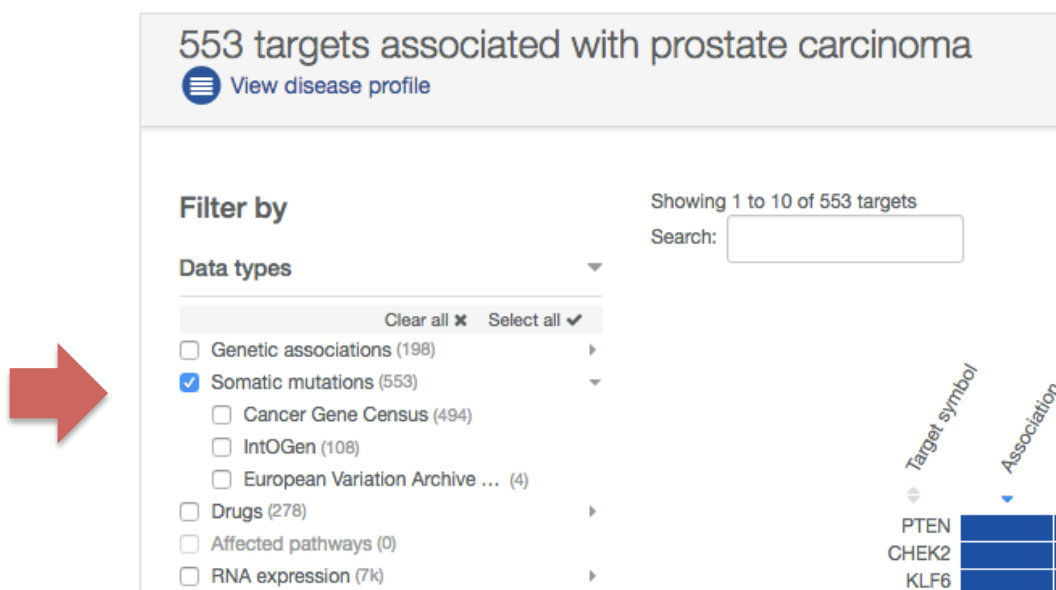


The first 10 rows will show the top 10 targets associated with prostate cancer. These will have the highest score (score of 1): *PTEN*, *CHEK2*, *KLF6*, *FGFR4*, *AR*, *HOXB13*, *ABL1*, *PDGFRB*, *KIT*, and *CACNA1D*.

The confidence on the target-disease association is indicated by the association score, which ranges from 0 to 1 (from no association to the strongest association). The score is computed individually for each piece of evidence (e.g. a drug on phase I), followed by the score computed for the data sources (e.g. ChEMBL), then a score for the data type (e.g. Genetic associations) and the overall score (a harmonic sum of the individual scores). The overall score is shown in the first column in the above table. More details on the scoring can be found below:

https://github.com/CTTV/association_score_methods

b) Restrict the results by filtering the table to show the targets associated with prostate cancer based on Somatic mutations only:



This refined list restricted to somatic mutations only does not match 100% the list resulting from step (a) above. There are no somatic mutations described in the *HOXB13* gene, so *HOXB13* is no longer in the top 10 genes when restricting the data based on somatic mutations.

For more details on the data we currently use to associate a gene to a disease can be found below:

https://www.targetvalidation.org/data_sources

c) Let's now focus on one of these targets namely *FGFR4* to find out more about some of the evidence that seems to support the association between *FGFR4* and prostate carcinoma.

Click on the gene name itself or on any cell in the gene table that corresponds to the *FGFR4* row:



This will take you to a page similar to this:

Evidence for FGFR4 in prostate carcinoma

Text Mining
Pathways
RNA
Drugs
Genetics
Somatic

FGFR4
fibroblast growth factor receptor 4
Synonyms: JTK2, CD334, TKF
Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factors and plays a role in the regulation of cell proliferation, differentiation and migration, and in regulation of I...

prostate carcinoma
Synonyms: Cancer of Prostate, Cancer of the Prostate, Carcinoma of Prostate, Carcinoma of the Prostate, Prosta...
One of the most common malignant tumors afflicting men. The majority of carcinomas arise in the peripheral zone and a minority occur in the central or the transitional zone of the prostate gland. Gr...

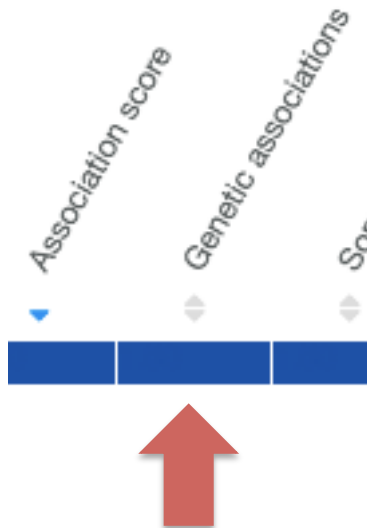
Genetic associations
Somatic mutations
Drugs
Affected pathways
RNA expression
Text mining
Animal models

The evidence used to support the association is shown in different tables (the tabs that are greyed out have no data: there is no data on


pathways, RNA expression, animal models to support this association).


Expand the 'Genetic associations' tab.

Note: if you click on the cell containing the data relative to Genetic associations (see below):



you will automatically land on a page where the tab containing the Genetic association will be already opened:





FGFR4
 fibroblast growth factor receptor 4
 Synonyms: JTK2, CD334, TKF

Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factors and plays a role in the regulation of cell proliferation, differentiation and migration, and in regulation of l...

prostate carcinoma
 Synonyms: Cancer of Prostate, Cancer of the Prostate, Carcinoma of Prostate, Carcinoma of the Prostate, Prosta...

One of the most common malignant tumors afflicting men. The majority of carcinomas arise in the peripheral zone and a minority occur in the central or the transitional zone of the prostate gland. Gr...

Genetic associations

Table

Browser

Rare diseases
 Source: UniProt, European Variation Archive (EVA), UniProt literature, Gene2Phenotype

Showing 1 to 4 of 4 entries

Search:

Disease	Mutation	Gene-Disease Evidence	Evidence source	Publications
		Mutation consequence		
prostate adenocarcinoma	N/A	Curated evidence	Further details in UniProt database	10 publications
prostate adenocarcinoma	rs351855	missense variant	Further details in UniProt database	7 publications
prostate carcinoma	N/A	Curated evidence	Further details in UniProt database	10 publications
prostate carcinoma	rs351855	missense variant	Further details in UniProt database	7 publications

Yes, there is one known genetic variant (rs351855) in this gene that is associated with prostate carcinoma.

Click on the 7 'publications' link to see the papers supporting the association:

Europe PMC

About Tools Developers Help Europe PMC plus

Search worldwide, life-sciences literature

EXT_ID:18756523 OR EXT_ID:11781352 OR EXT_ID:18670643 OR EXT_ID:20876804 OR EXT_ID:218822 **Search** Advanced Search

E.g. "breast cancer" HER2 Smith J

Results RSS Save Search Recent Activity Export

1 - 7 of 7 results Sort by: Relevance | Date | Times Cited

☐ Select results 1 - 7

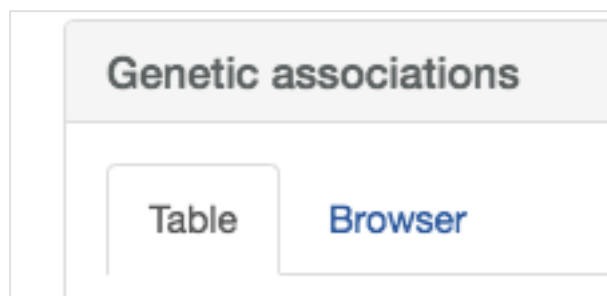
☐ [Germline variant FGFR4 p.G388R exposes a membrane-proximal STAT3 binding site.](#)
(PMID:26675719)
Ulaganathan VK, Sperl B, Rapp UR, Ullrich A
Nature [2015, 528(7583):570-574]
Cited: 0 times

☐ [PAX3-FOXO1 and FGFR4 in alveolar rhabdomyosarcoma.](#)
(PMID:21882254)
Marshall AD, van der Ent MA, Grosveld GC
Mol Carcinog [2012, 51(10):807-815]
Cited: 4 times

☐ [FGFR4 Gly388Arg polymorphism contributes to prostate cancer development and progression: a meta-analysis of 2618 cases and 2305 controls.](#)
(PMID:21349172 PMCID:PMC3049742) Free full text article
Xu B, Tong N, Chen SQ, Hua LX, Wang ZJ, Zhang ZD, Chen M
BMC Cancer [2011, 11:84]
Cited: 14 times

Popular content sets
[Full Text articles only \(3\)](#)
[Open Access articles only \(1\)](#)

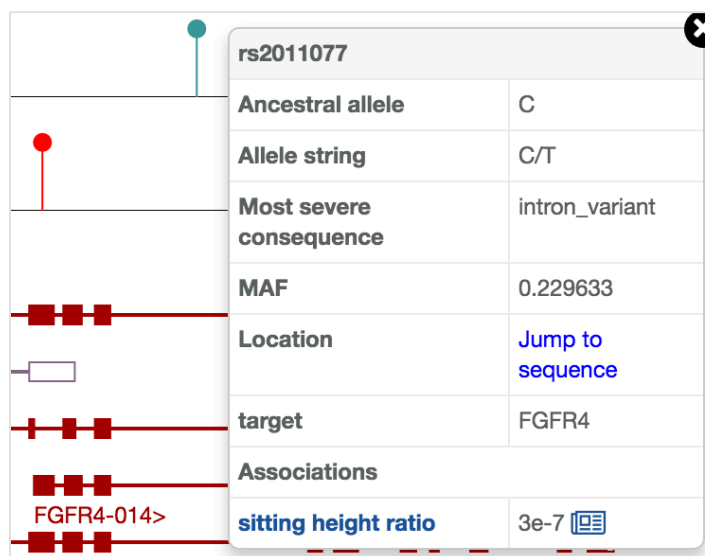
d) The Genetic associations data can also be visualised in a graphical display. Click on the 'Browser' link:



You will see the transcripts annotated in that gene and the variants (SNPs or mutations) that map to the region. Look at the legend to find out what the colours mean. This browser view is interactive and dynamic: you can zoom in and out and scroll along the genomic region.



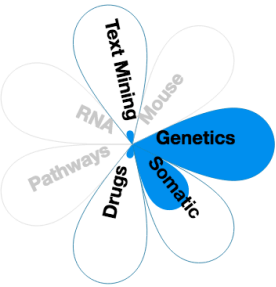
There are few variants associated with other traits such as body mass index, waist-hip ratio and body height. You may want to zoom out to view more variants up or downstream of the gene, and then click on the lollipop (the variant) for more details:



e) Let's now have a look at the target itself outside the specific context of any disease.

Still on the same page as above, click on the hyperlink FGFR4:

Evidence for FGFR4 in prostate carcinoma




FGFR4
fibroblast growth factor receptor 4
Synonyms: JTK2, CD334, TKF

Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factors and plays a role in the regulation of cell proliferation, differentiation and migration, and in regulation of l...

You will end up in a page like this:

<https://www.targetvalidation.org/target/ENSG00000160867>

Open Targets About Help API Downloads Blog

FGFR4
fibroblast growth factor receptor 4  View associated diseases

Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factors and plays a role in the regulation of regulation of lipid metabolism, bile acid biosynthesis, glucose uptake, vitamin D metabolism and phosphate homeostasis. CYP7A1, the rate-limiting enzyme in bile acid synthesis, in response to FGF19. Phosphorylates PLCG1 and FRS2. Ligand cascades. Activation of PLCG1 leads to the production of the cellular signaling molecules diacylglycerol and inositol 1,4,5 recruitment of GRB2, GAB1, PIK3R1 and SOS1, and mediates activation of RAS, MAPK1/ERK2, MAPK3/ERK1 and the M signaling pathway. Promotes SRC-dependent phosphorylation of the matrix protease MMP14 and its lysosomal degradat

Synonyms: JTK2 CD334 TKF FGFR-4 2.7.10.1 Fibroblast growth factor receptor 4

Protein Information (from UniProt)

Variants, isoforms and genomic context

Protein baseline expression

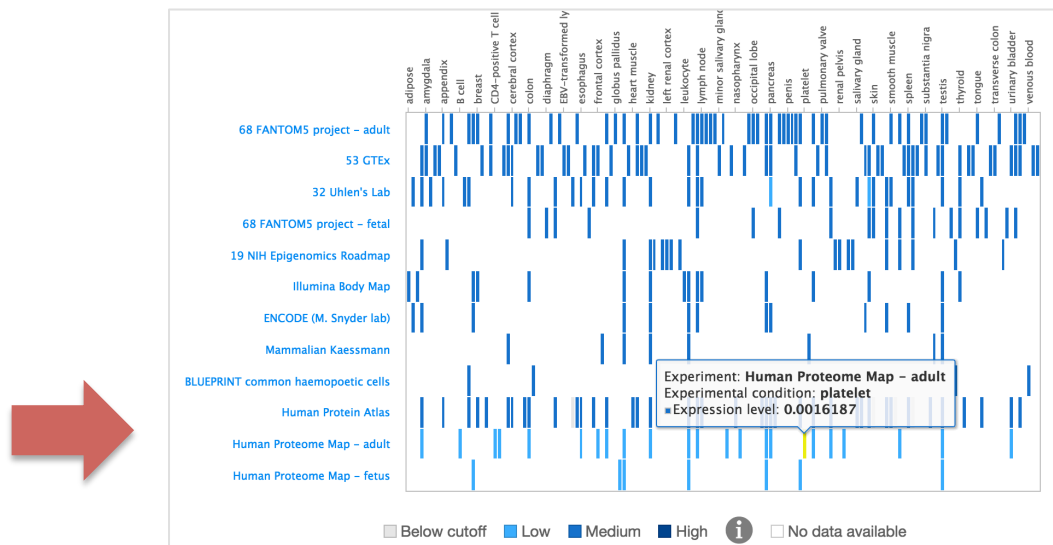
RNA baseline expression

Gene Ontology

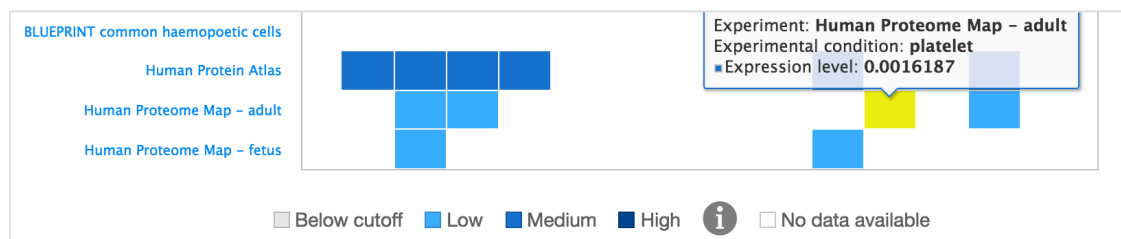
Protein Structure

9

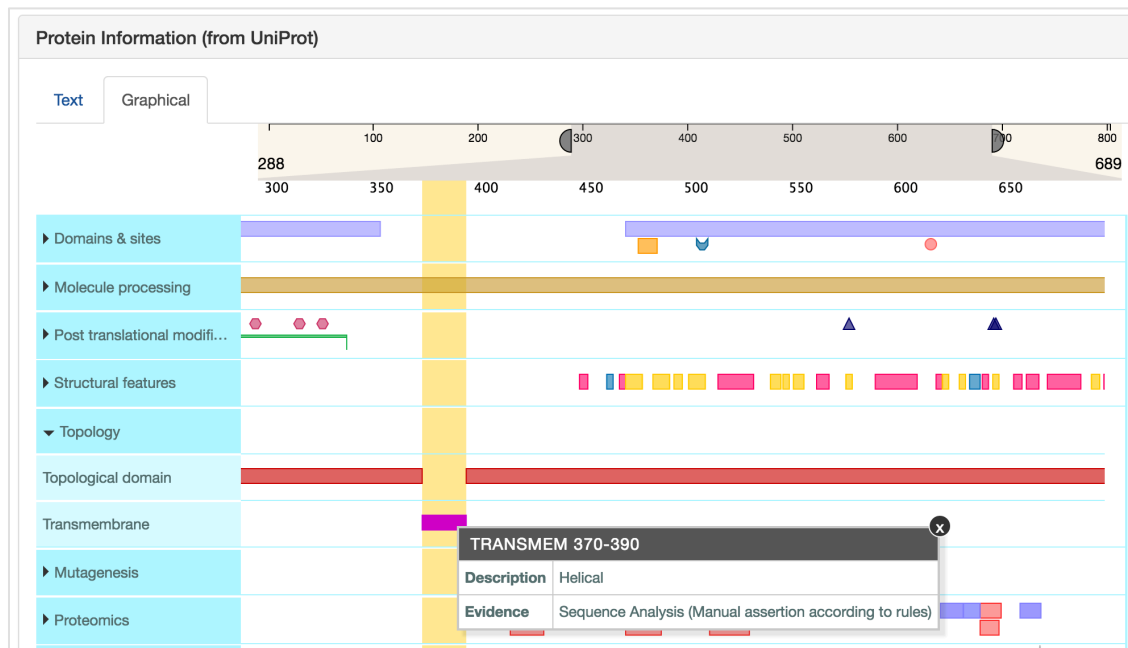
Click on RNA baseline expression to find out the tissue with the highest expression level according to Human Proteome Map (in adult tissues) is platelet:



You can also zoom in to get an image like this:



f) Click on the Protein information (from UniProt) tab. Now, click on the Graphical view option, then click on the Topology menu to see the annotated domains: extracellular, transmembrane and intracellular. The transmembrane (TM) domain goes from amino acid 370 to 390. Gene *FGFR4* codes for a receptor, so one should expect a transmembrane domain annotated in the protein.



Exercise 2 – *MS4A1* as a possible drug target in the treatment of non-Hodgkin's lymphoma

You can also use the search box at the top right corner of the pages in the Open Targets Platform (you do not need to go back to the homepage):

Open Targets About Help API Downloads Blog

MS4A1

MS4A1 target

membrane spanning 4-domains A1

411 diseases associated

This protein may be involved in the regulation of B-cell activation and proliferation.

Targets

MS4A12 membrane spanning...

MS4A7 membrane spanning 4...

Feedback

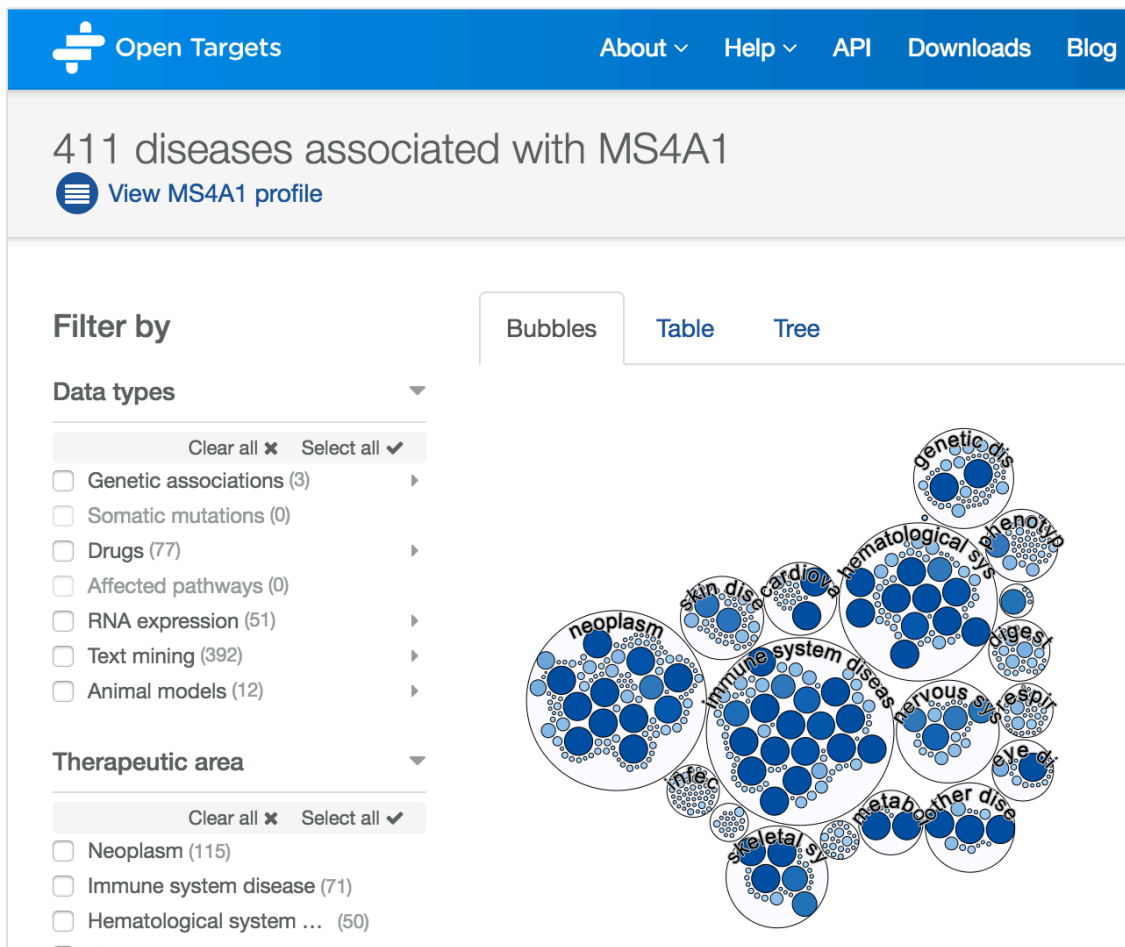
FGFR4

fibroblast growth factor receptor 4

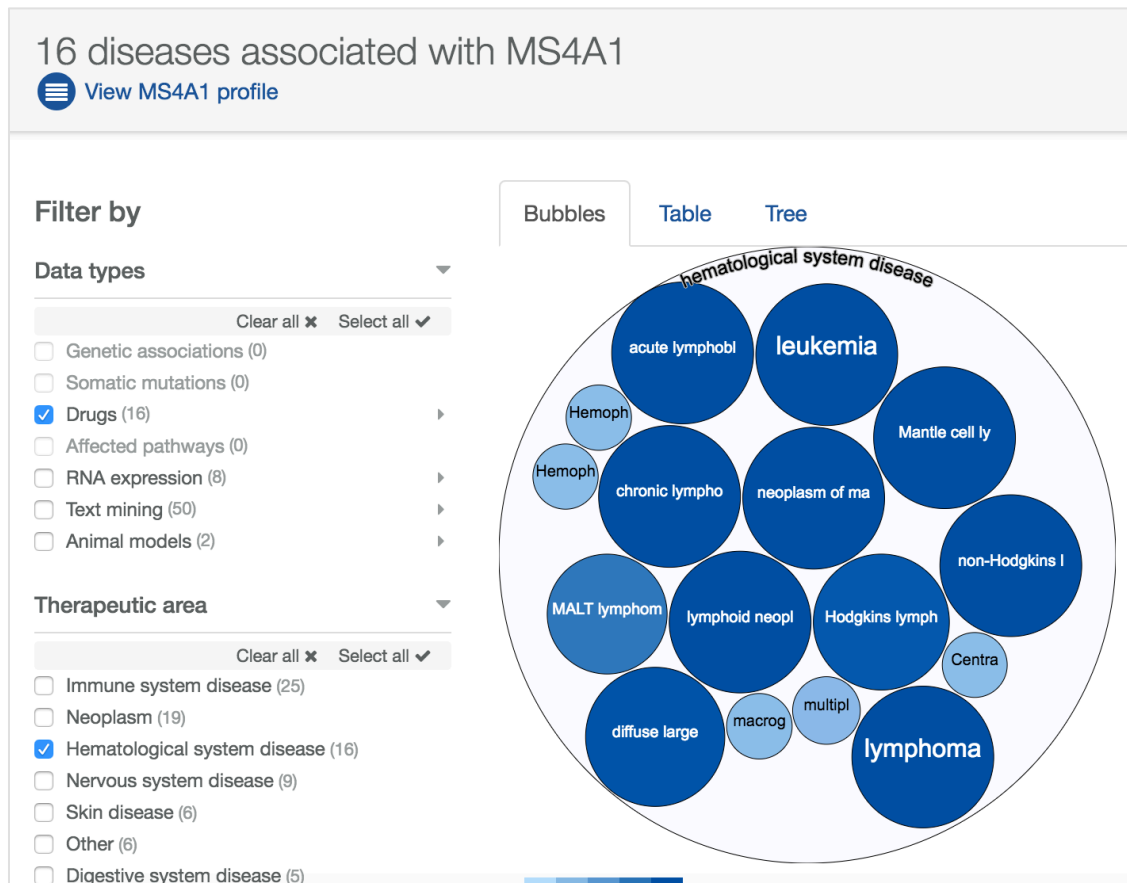
View associated diseases

Tyrosine-protein kinase that acts as cell-surface receptor for fibroblast growth factor. It is involved in the regulation of cell proliferation, differentiation and migration, and in regulation of biosynthesis, glucose uptake, vitamin D metabolism and phosphate homeostasis. It is also involved in the regulation of the expression of CYP7A1, the rate-limiting enzyme in bile acid biosynthesis. Phosphorylates PLCG1 and FRS2. Ligand binding leads to the activation of PLCG1. Activation of PLCG1 leads to the production of the cellular signaling molecule 1,4,5-trisphosphate. Phosphorylation of FRS2 triggers recruitment of GRB2, which mediates activation of RAS, MAPK1/ERK2, MAPK3/ERK1 and the MAP kinase signaling pathway, as well as

There are 411 diseases associated with target *MS4A1*.



You can filter the results by Therapeutic area, such as 'Hematological system' (which includes non-Hodgkin's lymphoma) and by Data type such as 'Drugs'. The number of diseases associated with *MS4A1* for which there is Drug data for the disease-target association is now 16:



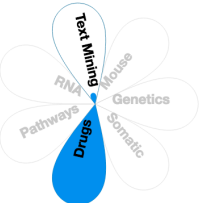
b) The default display for the diseases associated with a target is the Bubbles view. But you can the same results as a Table or Tree. Not sure what does views mean? Check our help page:

https://targetvalidation.org/about#target_assoc

Some of diseases with the overall association score of 0.90 or above (for the filters selected) are neoplasm of mature B-cells, chronic lymphocytic leukemia, and Hodgkins lymphoma

The results displayed in a Table format can be downloaded as CSV (comma separated value) and opened up in Excel or other:

Evidence for MS4A1 in non-Hodgkins lymphoma



MS4A1
 membrane spanning 4-domains A1
 Synonyms: B1, Bp35, MS4A2, CD20
 This protein may be involved in the regulation of B-cell activation and proliferation.

non-Hodgkins lymphoma
 Synonyms: NHL, NHL, NOS, lymphoma, non-Hodgkins, lymphoma, nonhodgkin
 Distinct from Hodgkin lymphoma biologically, non-Hodgkin lymphoma by the absence of Reed-Sternberg cells, and usually presents as...

Genetic associations

Somatic mutations

Drugs

Source: [ChEMBL](#)

Found 4 unique drugs: [OBINUTUZUMAB](#) [OFATUMUMAB](#) [RITUXIMAB](#) [TOSITUMOMAB](#)

Showing 1 to 10 of 41 entries

And find the four different drugs under clinical trials for the treatment of non-Hodgkin's lymphoma, when targeting and modulating *MS4A1*.

Search for 'phase III' to limit the number of rows and find that all drugs but OBINUTUZUMAB are in clinical phase III.

Drugs

Source: [ChEMBL](#)

Found 4 unique drugs: [OBINUTUZUMAB](#) [OFATUMUMAB](#) [RITUXIMAB](#) [TOSITUMOMAB](#)

Showing 1 to 10 of 17 entries (filtered from 41 total entries)

Search:

Drug Information					
Disease	Drug	Phase	Status	Type	Mechanism of action
non-Hodgkins lymphoma	RITUXIMAB	Phase III	Terminated	Antibody	B-lymphocyte antigen CD20 inhibitor DailyMed
non-Hodgkins lymphoma	RITUXIMAB	Phase III	Active, not recruiting	Antibody	B-lymphocyte antigen CD20 inhibitor DailyMed

e) Still on the same page, click on the disease name in the box at the right hand side to jump to the disease page and find out a bit more about the disease including all drugs under investigation and its disease ontology.

non-Hodgkins lymphoma

Synonyms: NHL, NHL, NOS, lymphoma, non-Hodgkin's, lymphoma, non-Hodgkins, lymphoma, nonhodgkin, lymphoma, nonh...

Distinct from Hodgkin lymphoma both morphologically and biologically, non-Hodgkin lymphoma (NHL) is characterized by the absence of Reed-Sternberg cells, can occur at any age, and usually presents as...

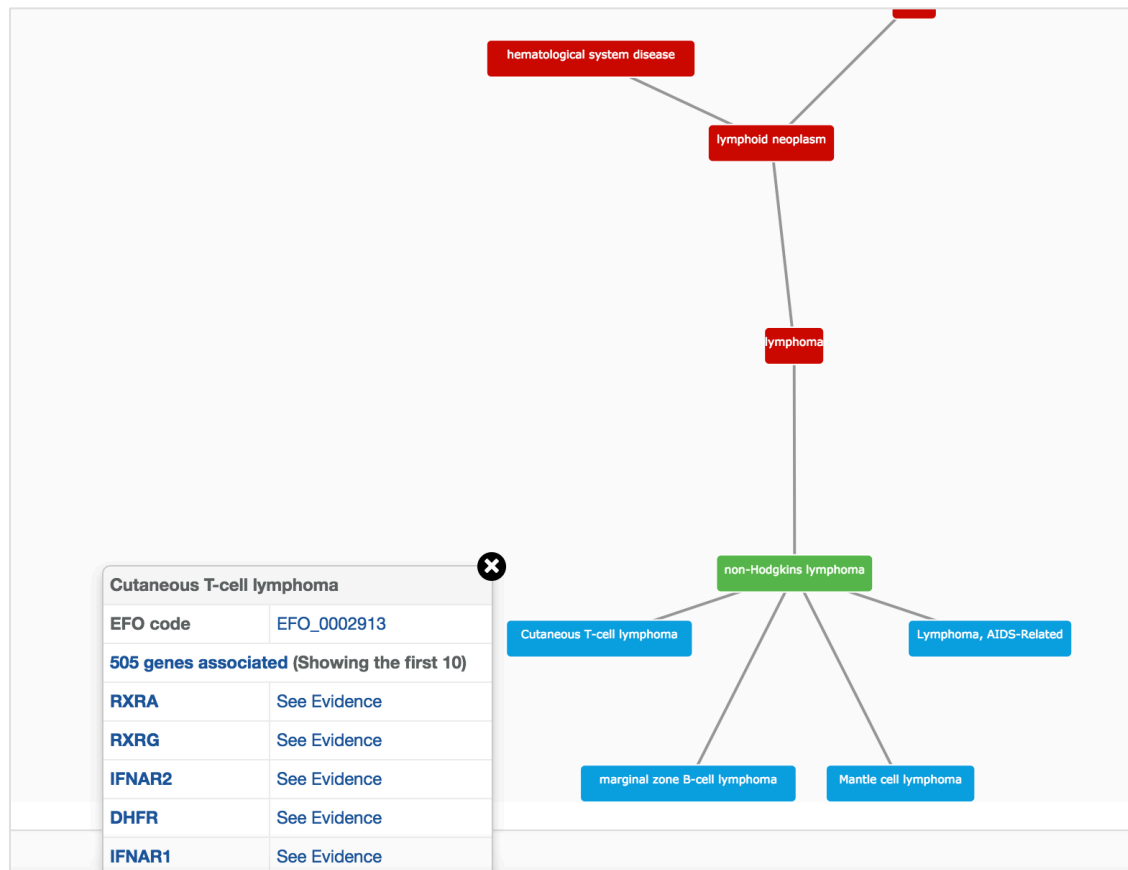
Click on the disease name above.

These are all drugs linked to this disease (therefore targeting other genes, not only *MS4A1*):

Drugs	
Source: ChEMBL	
Found 42 unique drugs: ABEMACICLIB ALEMTUZUMAB APITOLISIB BECAROTENE BLINATUMOMAB BORTEZOMIB BRENTUXIMAB VEDOTIN BUPARLISIB CLOFARABINE CYCLOSPORINE DASATINIB DEXAMETHASONE DEXAMETHASONE PHOSPHORIC ACID DOXORUBICIN ENZASTAURIN EVEROLIMUS FILGRASTIM IBRUTINIB IDELALISIB INTERFERON ALFA-2B IXAZOMIB CITRATE METHOTREXATE NILOTINIB NIVOLUMAB OBINUTUZUMAB OFATUMUMAB PALBOCICLIB PANOBINOSTAT PEGFILGRASTIM PICTILISIB PREDNISOLONE PREDNISONE RITUXIMAB ROMIDEPSIN ROMIPLOSTIM SIROLIMUS TASELISIB TEMSIROLIMUS THIIOGUANINE TOSITUMOMAB VALSARTAN VORINOSTAT	

f) Note some drugs you've seen in step e) have been investigated in sub-types of non-Hodgkin's lymphoma, such as Cutaneous T-cell lymphoma and Mantle cell lymphoma. This is easier to see when viewing the relationship of the parent disease (non-Hodgkin's lymphoma) and its children diseases (e.g. Cutaneous T-cell lymphoma and Mantle cell lymphoma).

This is part of the diagram showing the ontology:



Click on the nodes for more information, such as the EFO ID and genes associated with the diseases in the ontology.

The disease ontology can be downloaded as a PNG format. Click on the 'Download' icon in the top right of the image:

