

Mining gene-disease associations and drug target validation with Open Targets



**Hands-on Workshop
Answer booklet**

**CRUK Cambridge Institute
16th December 2016**

**Denise Carvalho-Silva
Open Targets Outreach**

Answers to exercises 1 and 2, pages 25-27 of coursebook

Exercise 1 – Prioritising targets for drug discovery in prostate carcinoma

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

The screenshot shows the Open Targets Platform interface. At the top, the logo and name 'Open Targets Platform' are displayed. Below this, the text 'Find new targets for drug discovery' is shown. A search bar contains the text 'prostate'. Below the search bar, a dropdown menu displays the search results for 'prostate carcinoma'. The results include the title 'prostate carcinoma', the number of associated targets '9116 targets associated', a 'Disease' category with a description, a list of 'Targets' (NKX3-1, SPDEF, KLK3), and a list of 'Diseases' (prostate adenocarcinoma, metastatic prostate cancer).

Open Targets Platform

Find new targets for drug discovery

prostate

prostate carcinoma
9116 targets associated

Disease

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. Grossly, prostatic carcinomas appear as ill-defined yellow areas of discoloration in the prostate gland...

Targets

NKX3-1 NK3 homeobox 1

SPDEF SAM pointed domain containing ETS transcription factor

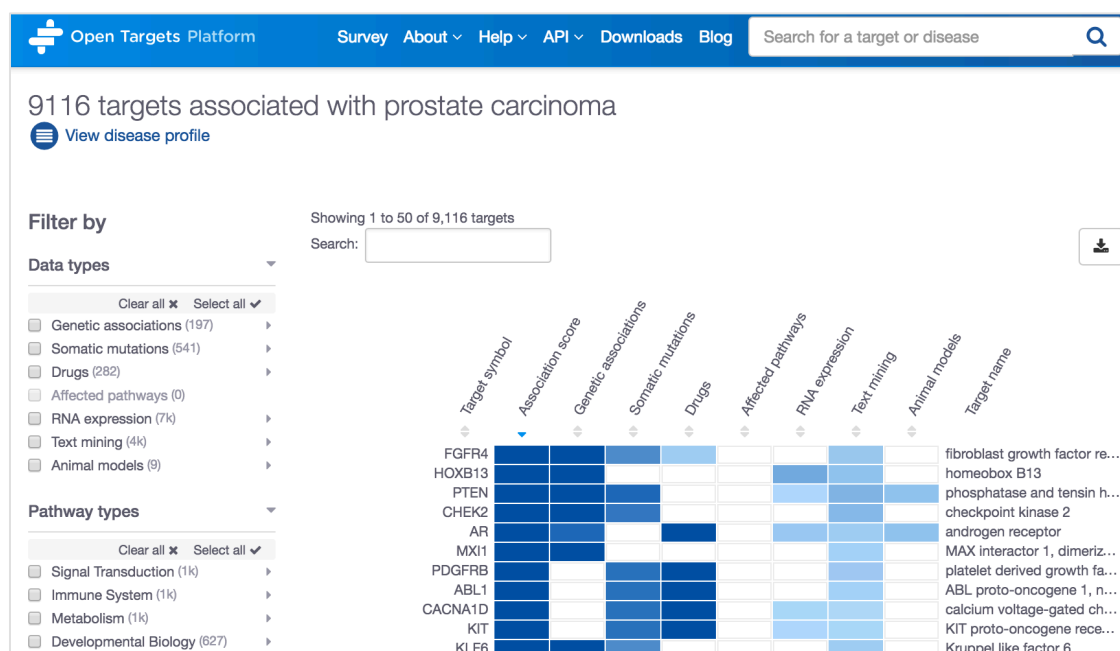
KLK3 kallikrein related peptidase 3

Diseases

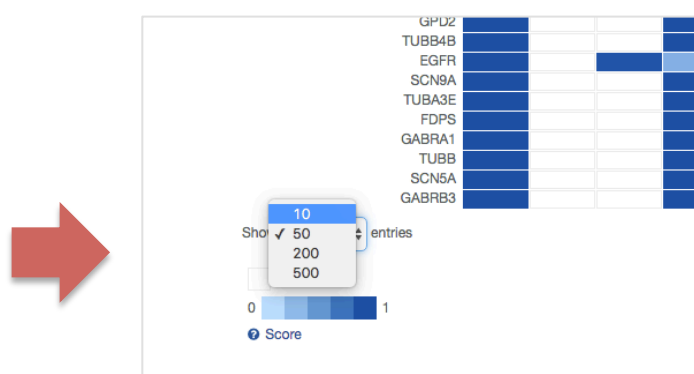
prostate adenocarcinoma
urogenital neoplasm > prostate carcinoma > prostate adenocarcinoma

metastatic prostate cancer
neoplasm > urogenital neoplasm > metastatic prostate cancer

Select the first (best) hit. You will then see a page like this, which lists 9116 targets associated with prostate carcinoma:



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



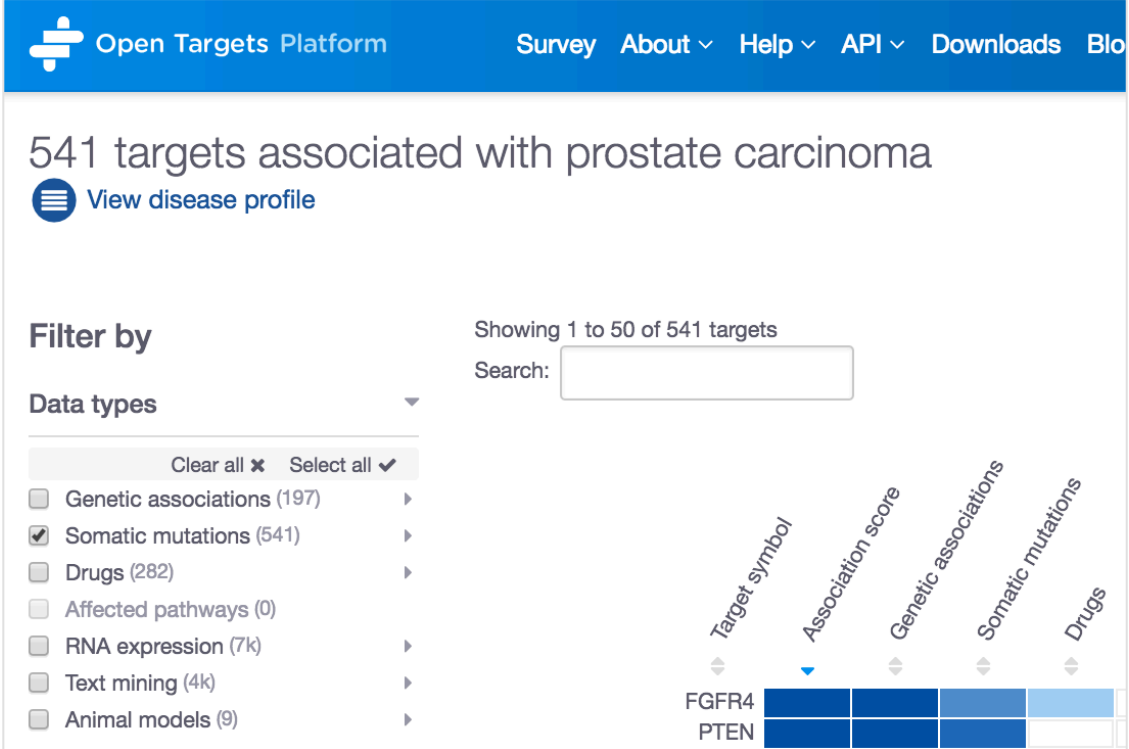
The first 10 rows will show the top 10 targets associated with prostate cancer. These will have the highest score (score of 1): *FGFR4*, *HOXB13*, *PTEN*, *CHEK2*, *AR*, *MXI1*, *PDGFRB*, *ABL1*, *CACNA1D*, and *KIT*.

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 of the table shown above. More details on the scoring can be found below:

https://github.com/CTTV/association_score_methods
nar.oxfordjournals.org/content/early/2016/11/29/nar.gkw1055

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



The screenshot shows the Open Targets Platform interface. A red arrow points to the header '541 targets associated with prostate carcinoma'. Below this, there is a 'Filter by' section with 'Data types' expanded. Under 'Data types', 'Somatic mutations (541)' is selected. To the right, a search bar is visible with the text 'Showing 1 to 50 of 541 targets'. Below the filter section, a table of targets is displayed. The table has five columns: 'Target symbol', 'Association score', 'Genetic associations', 'Somatic mutations', and 'Drugs'. The first two rows are 'FGFR4' and 'PTEN'. The 'Somatic mutations' column for both 'FGFR4' and 'PTEN' is highlighted in blue, indicating a high association score.

This filtered list (restricted to somatic mutations only) shows different targets than the list resulting from step (a) above.

There are no somatic mutations described in the *HOXB13* and *MXI1* genes, therefore *HOXB13* and *MXI1* are no longer in the top 10 targets based on somatic mutations.

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

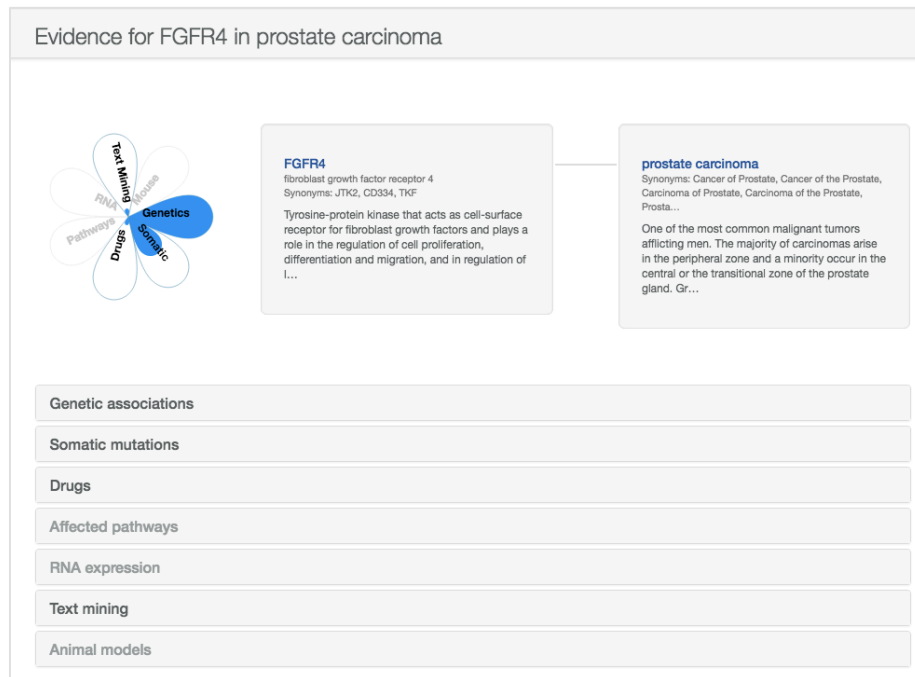
https://www.targetvalidation.org/data_sources

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:




The evidence used to support the association is shown in different tabs (the grey tabs have no data: there is no data for Affected pathways, RNA expression and Animal models to support FGFR4-prostate carcinoma association).


Expand the 'Genetic associations' tab.

Tip: 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 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

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 genetic variant that is known in public databases to be associated with prostate carcinoma. Its ID is rs351855.

Note that we aggregate evidence from highly specific terms of the disease ontology (e.g. prostate adenocarcinoma) to broader, parent terms (e.g. 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

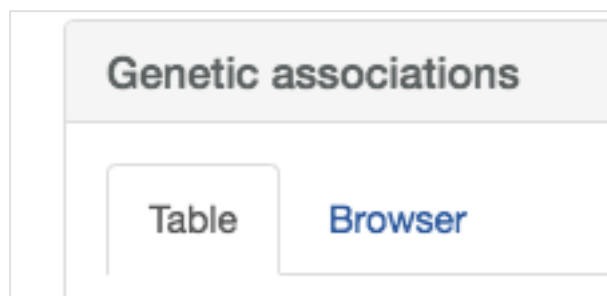
☐ [Germline variant FGFR4 p.G388R exposes a membrane-proximal STAT3 binding site.](#)
(PMID:26675719)
Ulaganathan VK, Sperl B, Rapp UR, Ullrich A
Nature [2015, 528(7583):570-574]
Cited: 0 times

☐ [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 \(1\)](#)

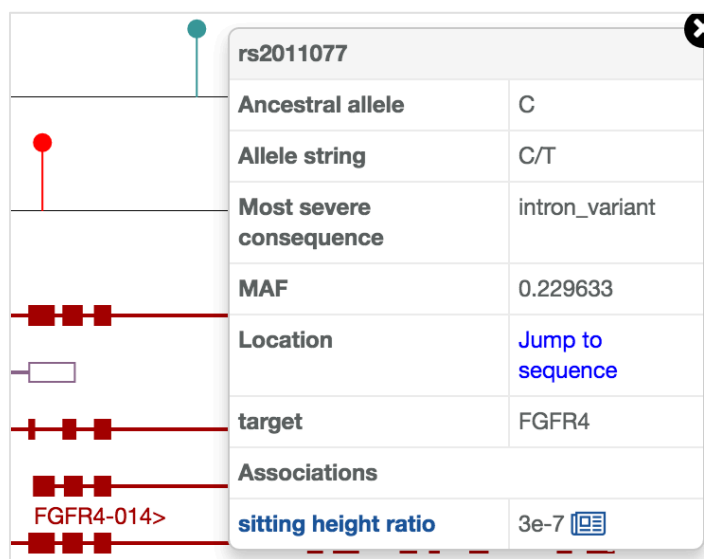
d) The Genetic associations data can also be visualised in a graphical display. Click on the 'Browser' link:



You will see the transcripts mapped to that gene and the variants (SNPs or mutations) in the region. Check 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.



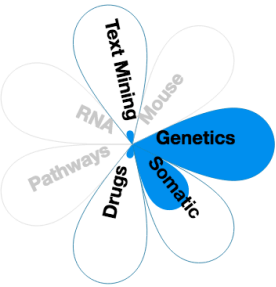
In this genomic region, you can see there are two mutations associated with other traits (i.e. sitting height ratio and body height). You may want to zoom out to view more variants up or downstream of the gene, and then click on the lollipop (the variant) for more details:



e) Let's now have a look at the target itself. This will give us information outside the 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>

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 cell proliferation, differentiation and migration, and in regulation of lipid metabolism, bile acid biosynthesis and homeostasis. Required for normal down-regulation of the expression of CYP7A1, the rate-limiting enzyme in bile acid synthesis. Phosphorylates PLCG1 and FRS2. Ligand binding leads to the activation of several signaling molecules, including diacylglycerol and inositol 1,4,5-trisphosphate. Phosphorylates PIK3R1 and SOS1, and mediates activation of RAS, MAPK1/ERK2, MAPK3/ERK1 and MAPK4/ERK3 signaling pathway. Promotes SRC-dependent phosphorylation of the matrix protease MMP13.

[\[show more\]](#)

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

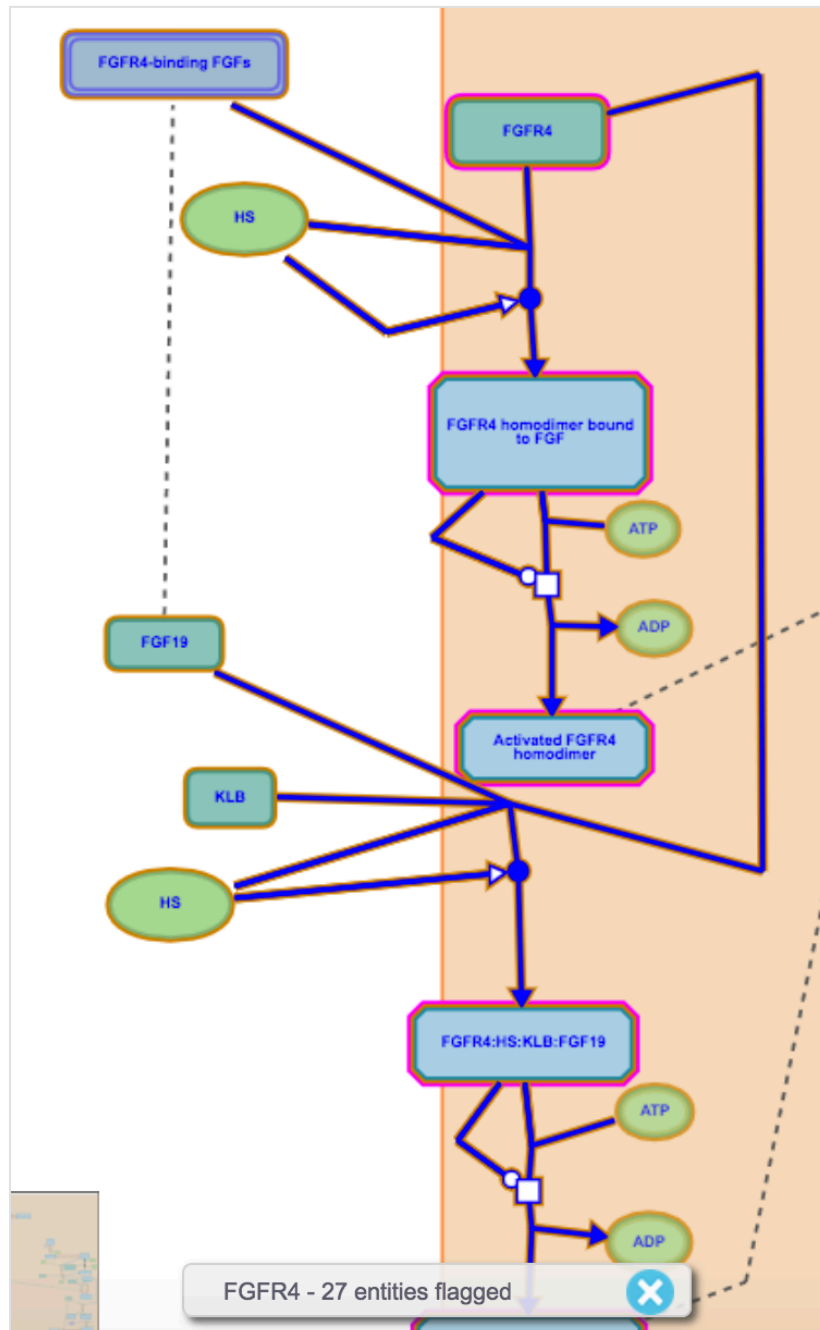
Protein Structure

Pathways

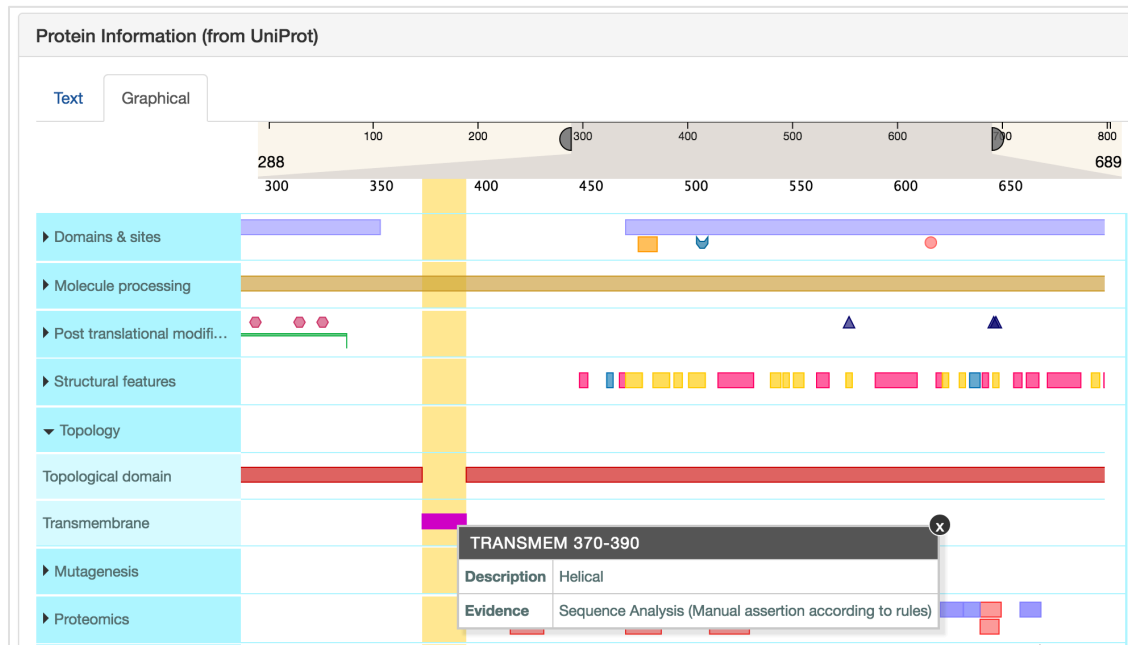


Click on Pathways to find out cellular pathways and biochemical process this gene is involved in e.g. PI3K Cascade, Constitutive Signaling by Aberrant PI3K in Cancer and few others.

You can visualise FGFR4 ligand binding and activation pathway in an interactive display:



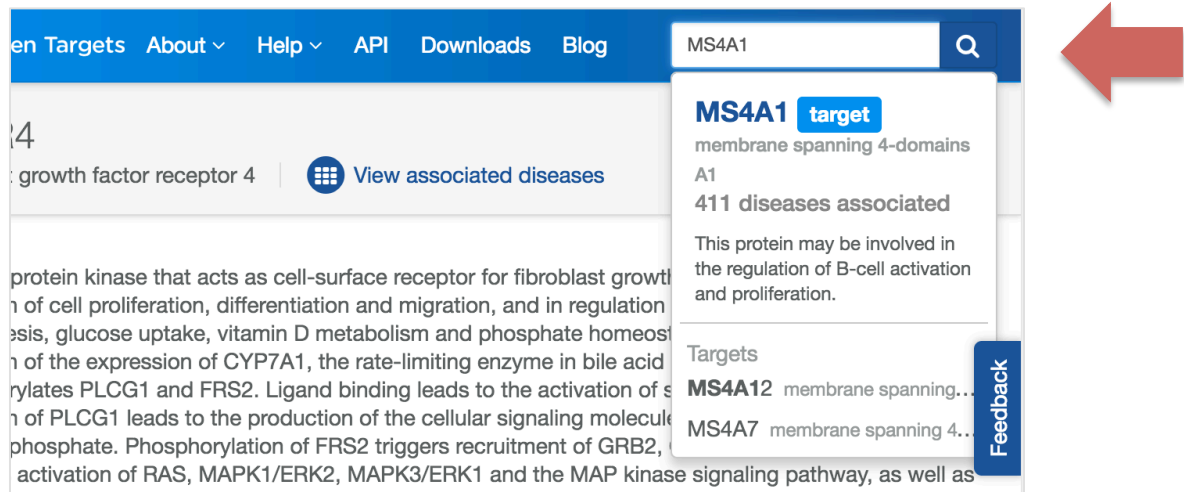
f) Click on the Protein information (from UniProt) tab. Now, click on the Graphical view option, then click on the Topology menu to see the annotated domains: extracellular, transmembrane and intracellular. The transmembrane (TM) domain goes from amino acid 370 to 390. Gene *FGFR4* codes for a receptor,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



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

Search for MS4A1.

Tip: you do not need to go back to the homepage: you can use the search box at the top right corner of any pages in the Platform:



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.

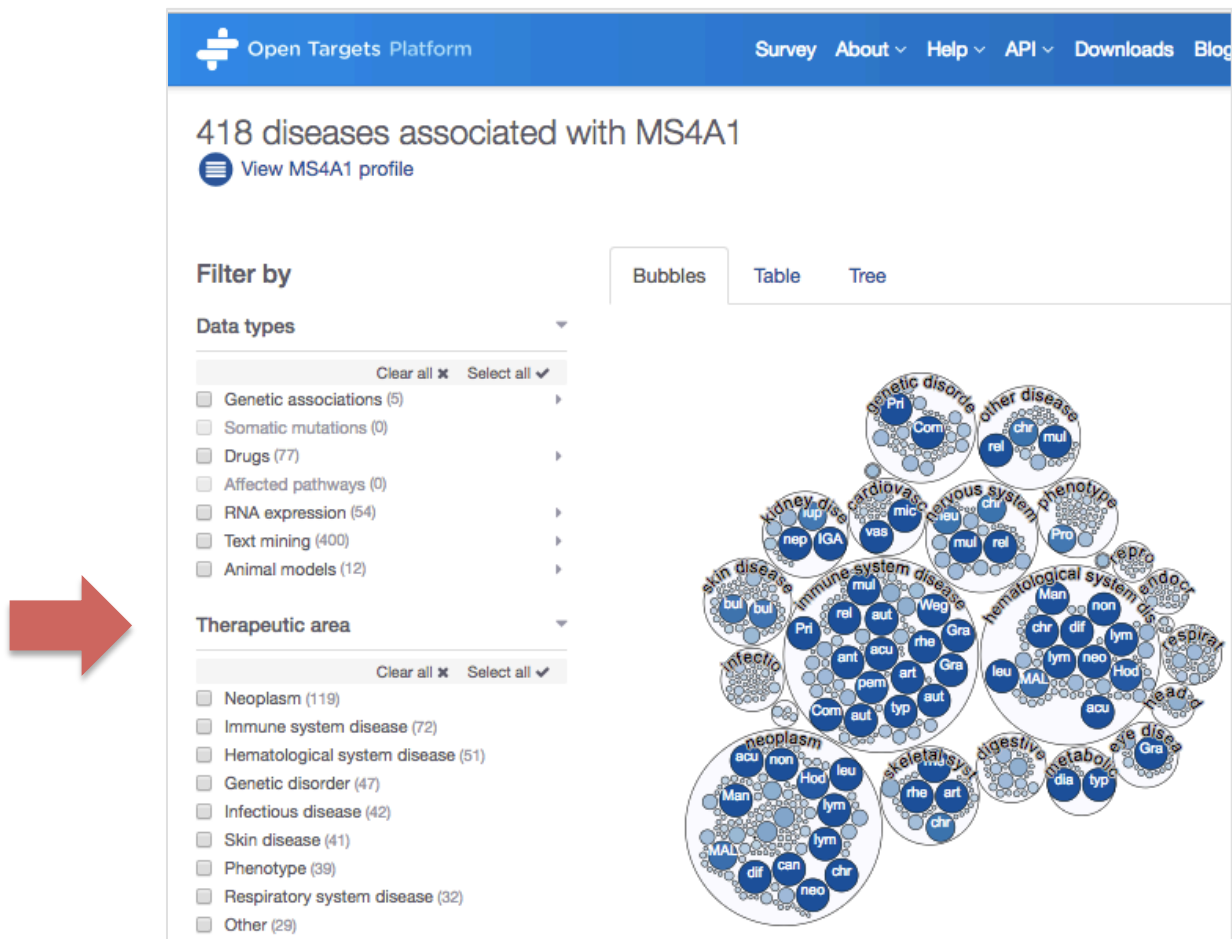
Targets
MS4A12 membrane spanning...
MS4A7 membrane spanning 4...

Feedback

growth factor receptor 4 | View associated diseases

protein kinase that acts as cell-surface receptor for fibroblast growth...
n of cell proliferation, differentiation and migration, and in regulation...
esis, glucose uptake, vitamin D metabolism and phosphate homeost...
n of the expression of CYP7A1, the rate-limiting enzyme in bile acid...
rylates PLCG1 and FRS2. Ligand binding leads to the activation of s...
n of PLCG1 leads to the production of the cellular signaling molecu...
phosphate. Phosphorylation of FRS2 triggers recruitment of GRB2, ...
activation of RAS, MAPK1/ERK2, MAPK3/ERK1 and the MAP kinase signaling pathway, as well as

There are 418 diseases associated with target *MS4A1*.



Open Targets Platform Survey About ▾ Help ▾ API ▾ Downloads Blog

418 diseases associated with MS4A1
View MS4A1 profile

Filter by

Data types

Clear all ✕ Select all ✓

- ☐ Genetic associations (5)
- ☐ Somatic mutations (0)
- ☐ Drugs (77)
- ☐ Affected pathways (0)
- ☐ RNA expression (54)
- ☐ Text mining (400)
- ☐ Animal models (12)

Therapeutic area

Clear all ✕ Select all ✓

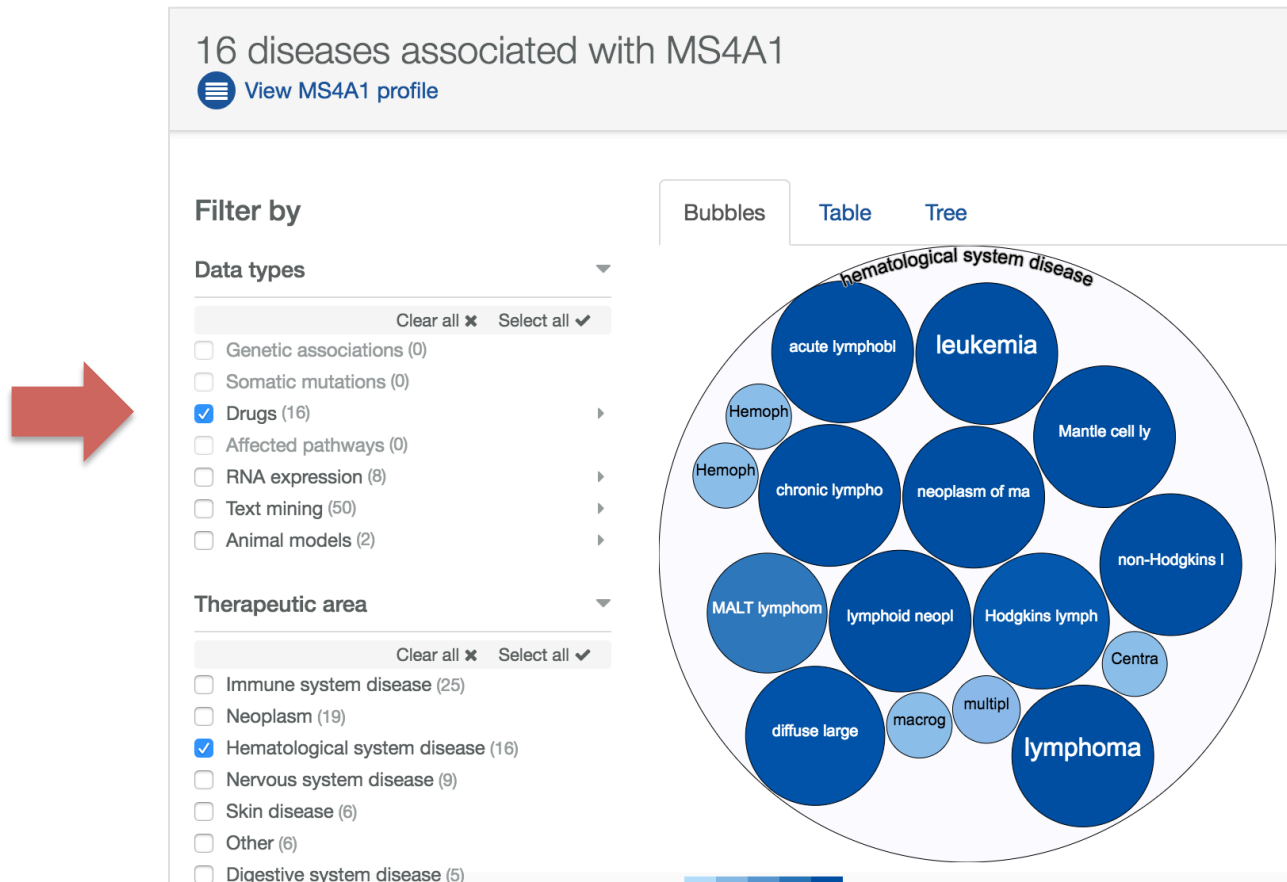
- ☐ Neoplasm (119)
- ☐ Immune system disease (72)
- ☐ Hematological system disease (51)
- ☐ Genetic disorder (47)
- ☐ Infectious disease (42)
- ☐ Skin disease (41)
- ☐ Phenotype (39)
- ☐ Respiratory system disease (32)
- ☐ Other (29)

Bubbles Table Tree

Genetic disorder other disease
Pri Com chr mul
Kidney disease cardiovascular disease nervous system phenotype
tup IGA vas mul rel chr
skin disease immune system disease hematological system disease
bul bul mul aut Weg Pro
Pri ant acu rha Gra
infectio pem art Gra
Com aut typ aut
neoplasm
acu non Hod leu
Man lym
MAL dif can chr
skeletal system
rhe art chr
digestive metabolic eye disease
dia typ Gra

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 in the

Hematological system associated with *MS4A1* for which there is Drug information is 16:



b) The default display for the diseases associated with a target is the Bubbles view.

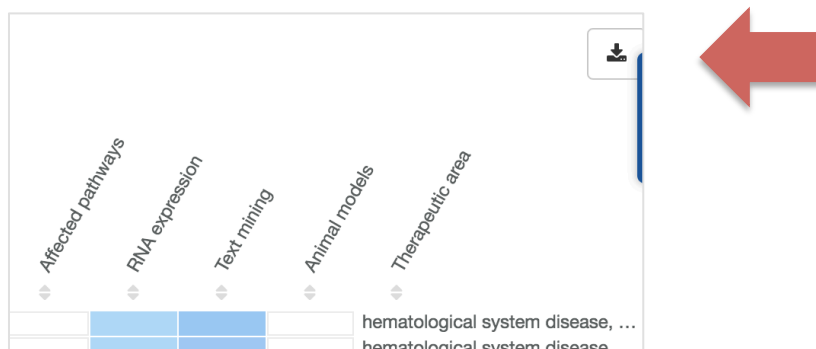
You can view the same information as a Table or Tree.

Tip: Not sure what those views mean? Check our help page:

https://targetvalidation.org/about#target_assoc

Diseases with the (overall) association score of 0.90 or above (for the filters selected) are neoplasm of mature B-cells (score of 1), chronic lymphocytic leukemia (score of 1), and Hodgkins lymphoma (0.92).

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



Look for the 'Download' icon:



Download the table in CSV.

c) Filter the table with 'non-Hod' for 'non-Hodgkin's lymphoma'. Text mining also supports the association between the disease and *MS4A1*.

Note: the data coming from mining the literature is given a lower weight in our analysis; therefore it gets a lower score (than Drugs coming from ChEMBL for example).

Click on the 'Text mining' cell in the table to see the 778 research articles mined from EuropePMC. These articles are flagged if they have the co-occurrence between the gene (or its synonym i.e. CD20) and the disease name in the same sentence.

d) Still in the same page as c) scroll up and expand the option 'Drugs':

Genetic associations

Somatic mutations

Drugs

Source: ChEMBL
Found 5 unique drugs: [OBINUTUZUMAB](#) [OCRELIZUMAB](#) [OFATUMUMAB](#) [RITUXIMAB](#) [TOSITUMOMAB](#)


Showing 1 to 10 of 40 entries

Search:

Drug Information							Gene-Drug Evidence	
Disease	Drug	Phase	Status	Type	Mechanism of action	Activity	Target class	Evidence source
non-Hodgkins lymphoma	RITUXIMAB	Phase IV	Completed	Antibody	B-lymphocyte antigen CD20 inhibitor DailyMed	antagonist	CD20 Ca ²⁺ -channel family	Curated from Clinical Trials Information ↗

Five drugs that target and modulate *MS4A1* are currently under clinical trials for the treatment of non-Hodgkin's lymphoma.

Search for 'IV comp' to limit the number of rows and find that RITUXIMAB is the only drug currently in phase IV, status completed.



Drugs

Source: [CHEMBL](#)

Found 5 unique drugs: [OBINUTUZUMAB](#) [OCRELIZUMAB](#) [OFATUMUMAB](#) [RITUXIMAB](#) [TOSITUMOMAB](#)

Showing 1 to 2 of 2 entries (filtered from 40 total entries)

Search:

Drug Information							Gene-Drug Evidence
Disease	Drug	Phase	Status	Type	Mechanism of action	Activity	Target class
non-Hodgkins lymphoma	RITUXIMAB	Phase IV	Completed	Antibody	B-lymphocyte antigen CD20 inhibitor DailyMed	antagonist	CD20 Ca2+ channel family
non-Hodgkins lymphoma	RITUXIMAB	Phase IV	Completed	Antibody	B-lymphocyte antigen CD20 inhibitor DailyMed	antagonist	CD20 Ca2+ channel family

Why do you see two different rows with the same information in all the columns?

Some rows may look like identical at a first glance. If you click on the link in the last column (Evidence source) you will find out the drug is in phase IV completed in two different studies, NCT00090038 and NCT00430352).

e) Still on the same page, scroll up till you see the flower and click on the disease name in the box at the right hand side.

	<p>non-Hodgkins lymphoma</p> <p>Synonyms: NHL, NHL, NOS, lymphoma, non-Hodgkin's, lymphoma, non-Hodgkins, lymphoma, nonhodgkin, lymphoma, nonh...</p> <p>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...</p>	
--	--	--

Click on the disease name above.

This will take you the disease page with more info on the disease including all drugs under investigation, its ontology and its phenotypes (if available).

There are 52 drugs linked to this disease (therefore targeting other genes, not only *MS4A1*):

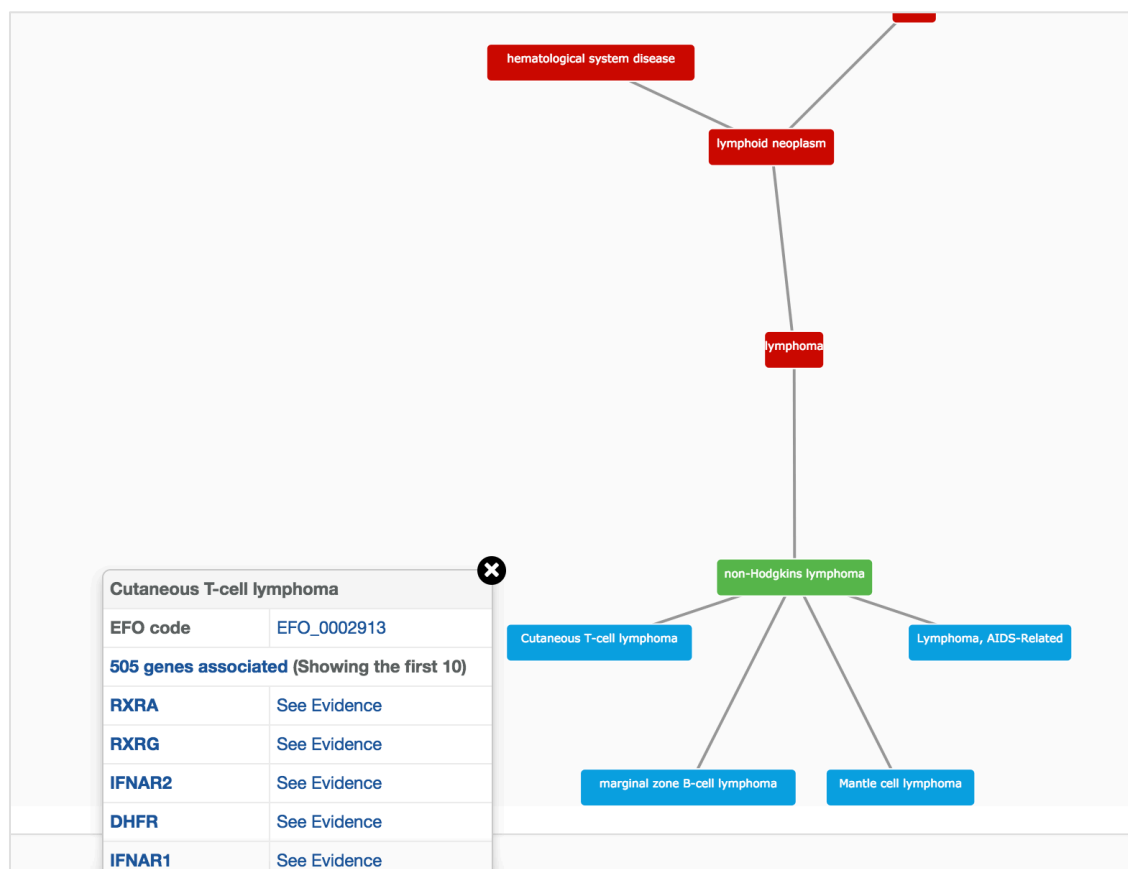
Drugs

Source: ChEMBL

Found 52 unique drugs: [ABEMACICLIB](#) [ALEMTUZUMAB](#) [APITOLISIB](#) [AT-7519](#) [BEXAROTENE](#) [BLINATUMOMAB](#) [BORTEZOMIB](#) [BRENTUXIMAB VEDOTIN](#) [BUPARLISIB](#) [CLOFARABINE](#) [COPANLISIB](#) [CYCLOSPORINE](#) [DASATINIB](#) [DEXAMETHASONE](#) [DEXAMETHASONE PHOSPHORIC ACID](#) [DOXORUBICIN](#) [DUVELISIB](#) [ENZASTAURIN](#) [EVEROLIMUS](#) [FILGRASTIM](#) [FLUDARABINE PHOSPHATE](#) [GALIXIMAB](#) [GSK-461364](#) [IBRUTINIB](#) [IDELALISIB](#) [INTERFERON ALFA-2B](#) [IXAZOMIB CITRATE](#) [METHOTREXATE](#) [NIVOLUMAB](#) [OBINUTUZUMAB](#) [OCRELIZUMAB](#) [OFATUMUMAB](#) [PALBOCICLIB](#) [PANOBINOSTAT](#) [PEGFILGRASTIM](#) [PICTILISIB](#) [PREDNISOLONE](#) [PREDNISONE](#) [RITUXIMAB](#) [ROMIDEPSIN](#) [ROMIPLOSTIM](#) [RP-6530](#) [SIROLIMUS](#) [TASELISIB](#) [TEMSIROLIMUS](#) [THIOGUANINE](#) [TOSITUMOMAB](#) [VALSARTAN](#) [VENETOCLAX](#) ...

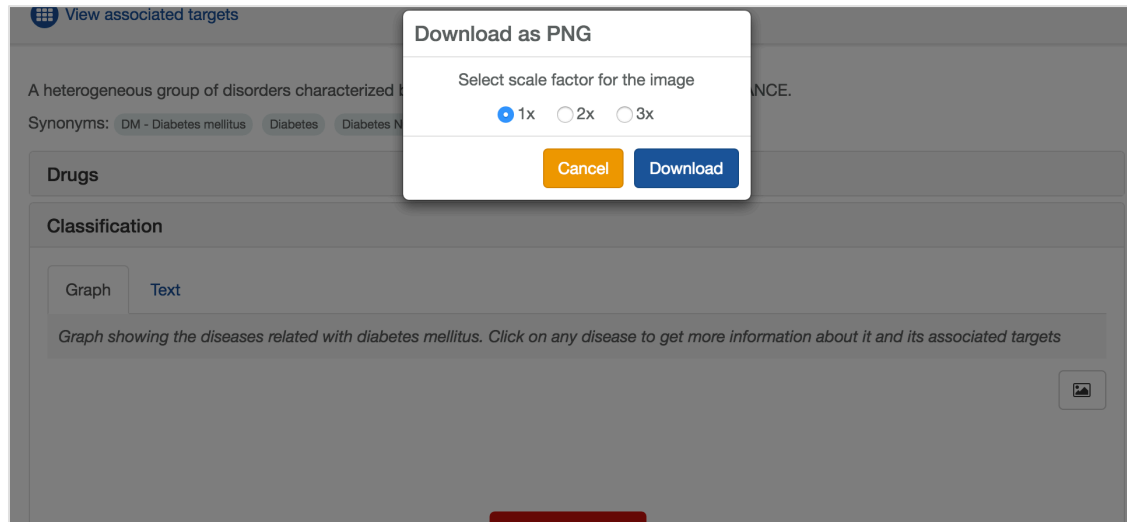
[\[Show more\]](#)

f) 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:



Note some drugs you've seen in step e) have been investigated in subtypes 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).