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

**Hands-on Workshop
Answer booklet**
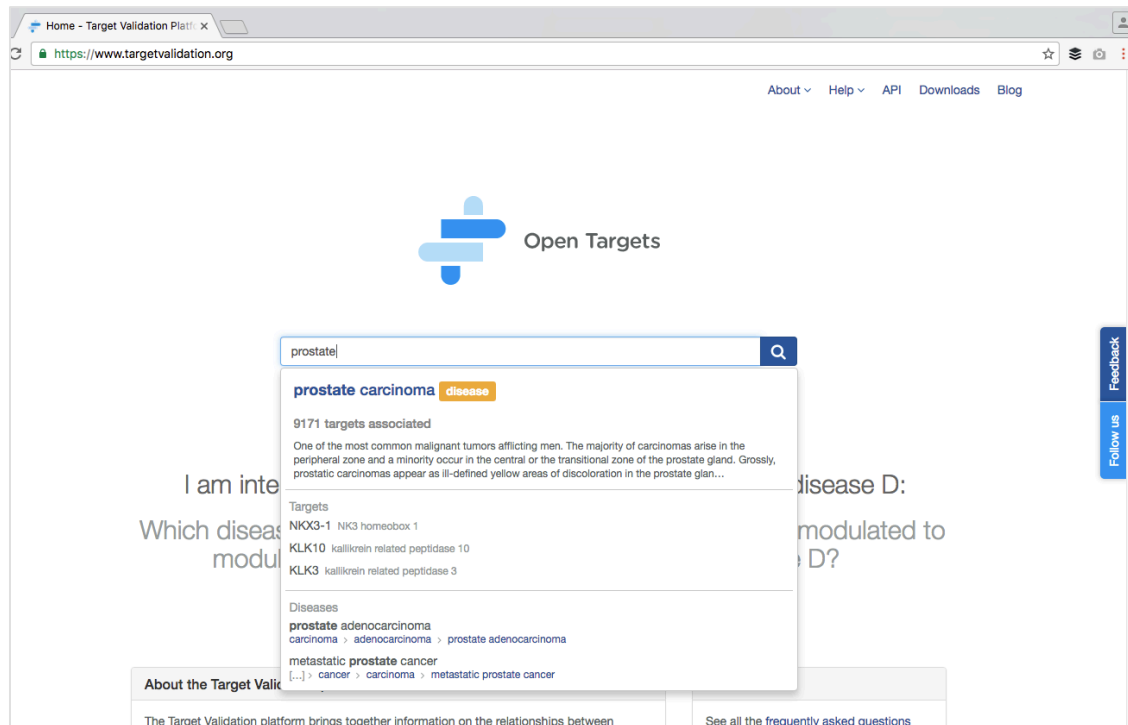
**CRUK Manchester Institute
22nd November 2016**

**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

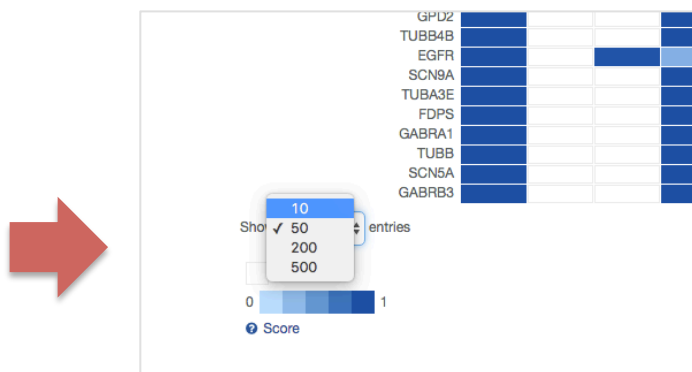a) Go to www.targetvalidation.org and search for *prostate carcinoma*:

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

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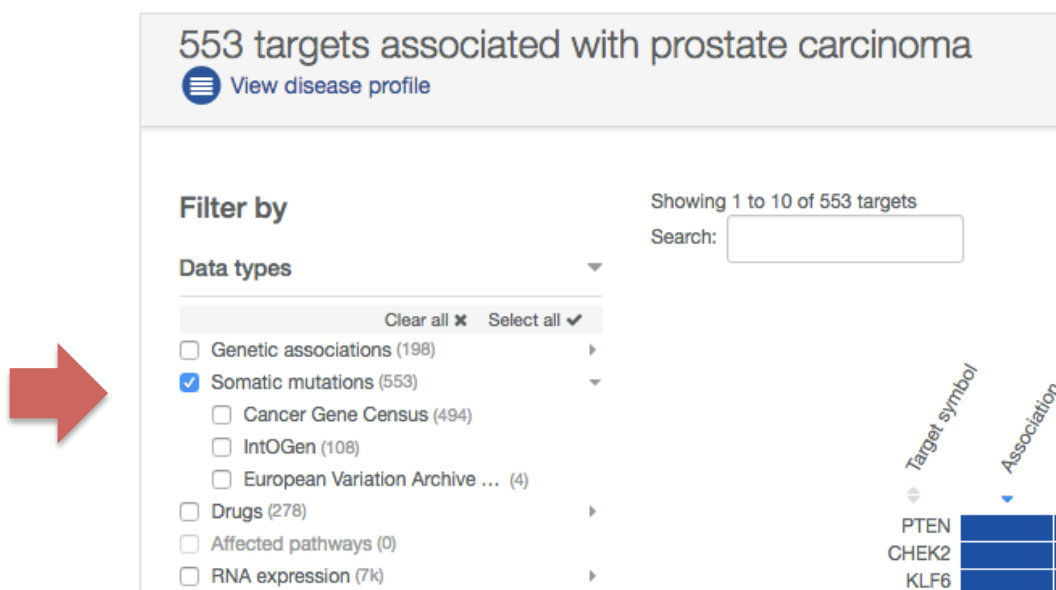
This refined list restricted to somatic mutations only does not match 100% the list resulting from step (a) above. There are no somatic mutations described in the *HOXB13* gene, so *HOXB13* is no longer in the top 10 genes when restricting the data based on somatic mutations.

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:

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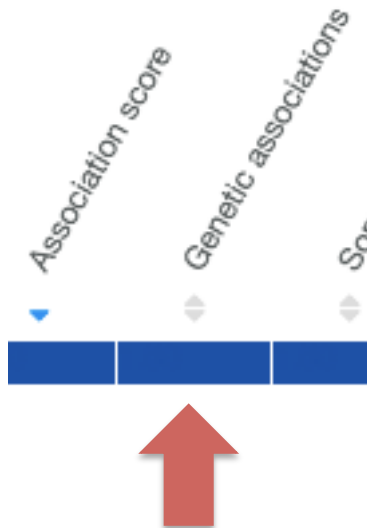
you will automatically lend on a page where the tab containing the Genetic association will be already opened:

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:

d) The Genetic associations datacan also been visualised in a graphical display. Click on the 'Browser' link:
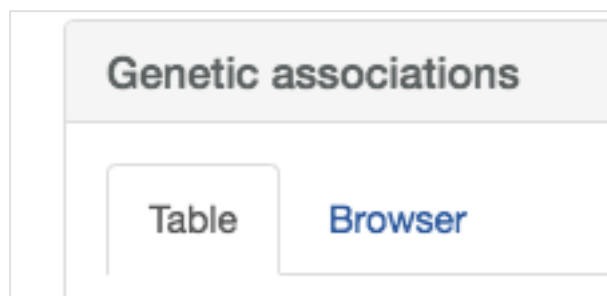
You will see the transcripts annotated in that gene and the variants (SNPs or mutations) that map to the region. Look at the legend to find out what the colours mean. This browser view is interactive and dynamic: you can zoom in and out and scroll along the genomic region.
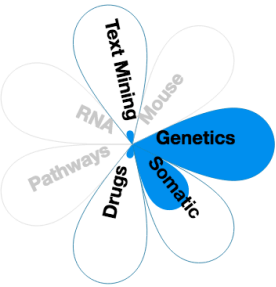
There are few variants associated with other traits such as body mass index, waist-hip ratio and body height. You may want to zoom out to view more variants up or downstream of the gene, and then click on the lollipop (the variant) for more details:
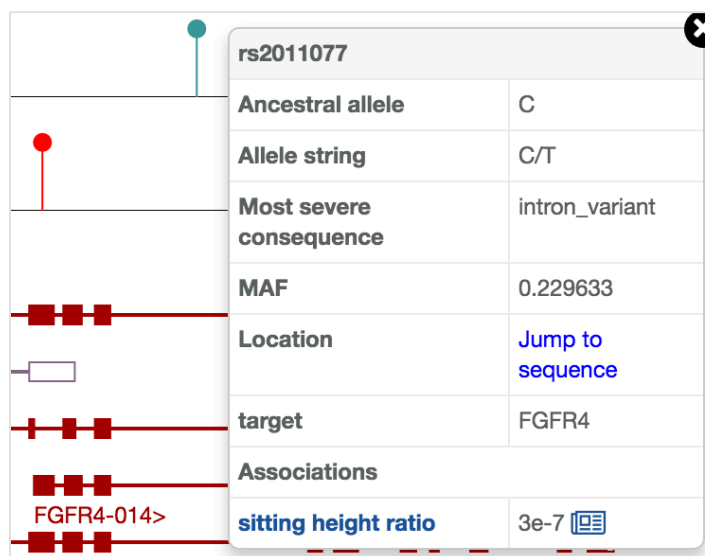
e) Let's now have a look at the target itself outside the specific context of any disease.

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

You will end up in a page like this:
https://www.targetvalidation.org/target/ENSG00000160867

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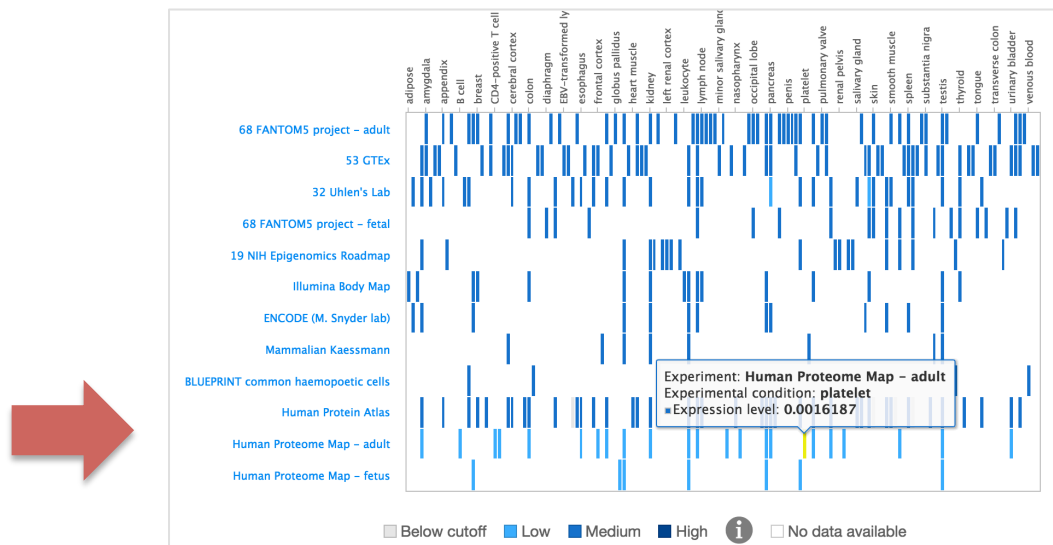
You can also zoom in to get an image like this:
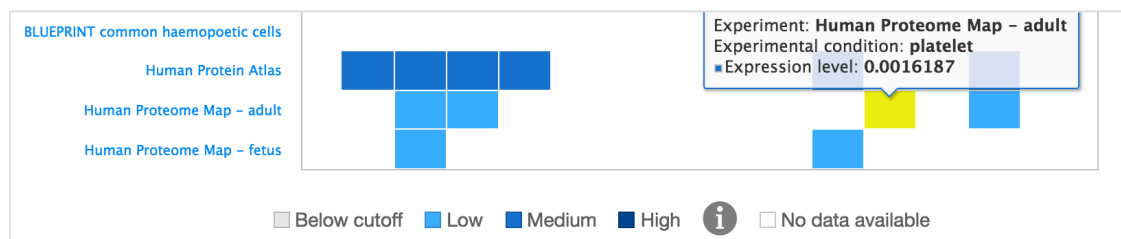
f) Click on the Protein information (from UniProt) tab. Now, click on the Graphical view option, then click on the Topology menu to see the annotated domains: extracellular, transmembrance and intracellular. The transmembrane (TM) domain goes from amino acid 370 to 390. Gene *FGFR4* codes for a receptor,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

Protein Information (from UniProt)
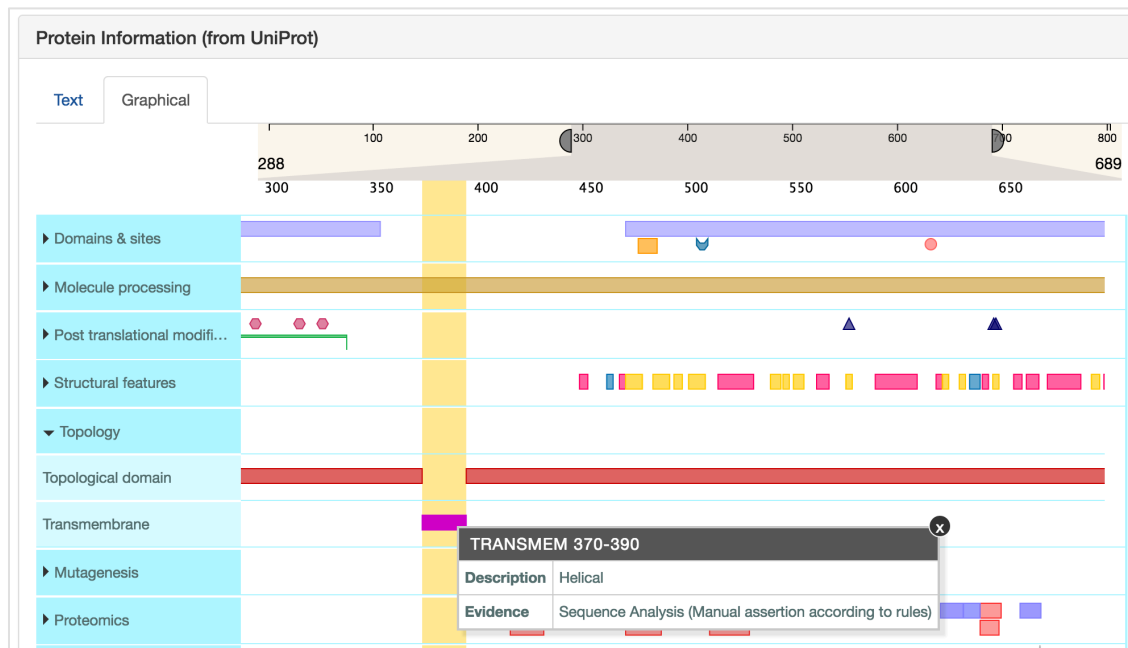
Exercise 2 – *MS4A1* as a possible drug target in the treatment of non-Hodgkin's lymphoma

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):

There are 411 diseases associated with target *MS4A1*.
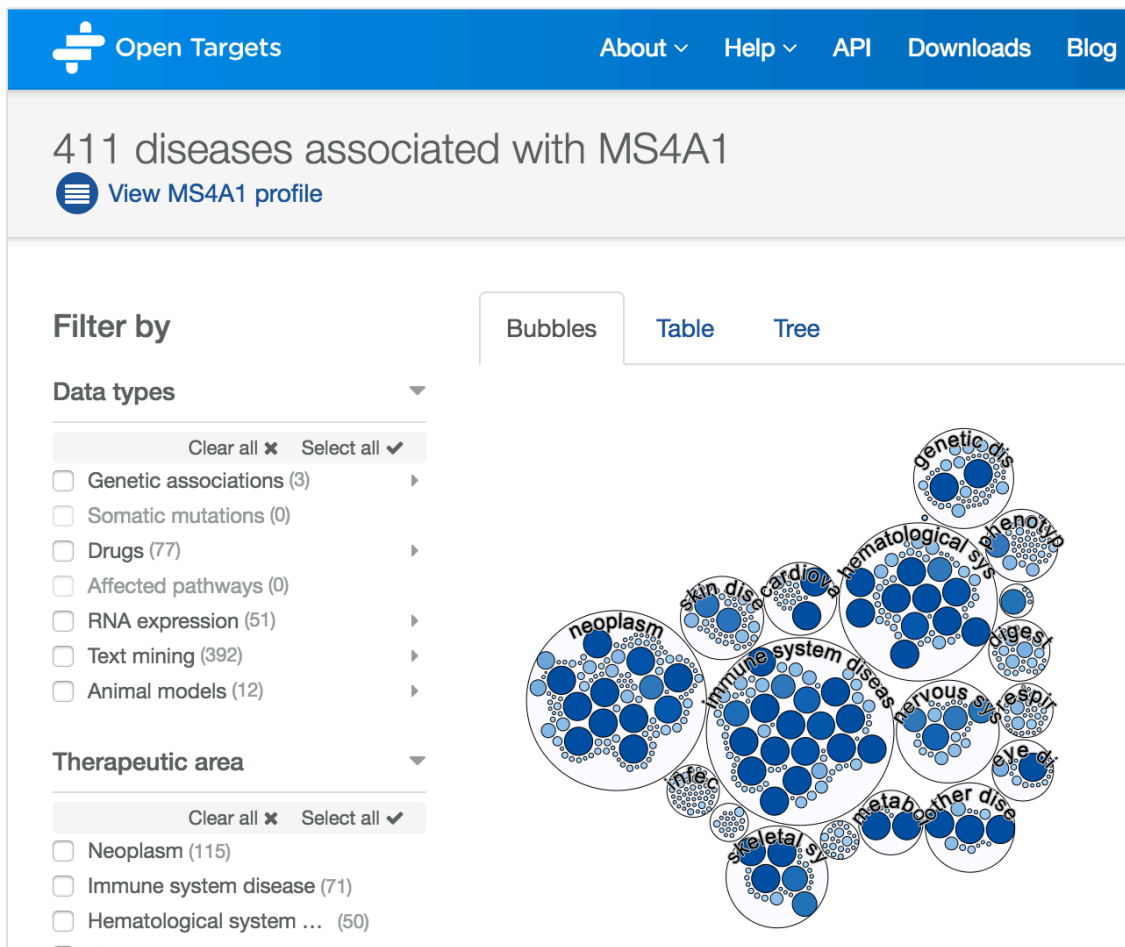
You can filter the results by Therapeutic area, such as 'Hematological system' (which includes non-Hodgkin's lymphoma) and by Data type such as 'Drugs'. The number of diseases associated with *MS4A1* for which there is Drug data for the disease-target association is now 16:
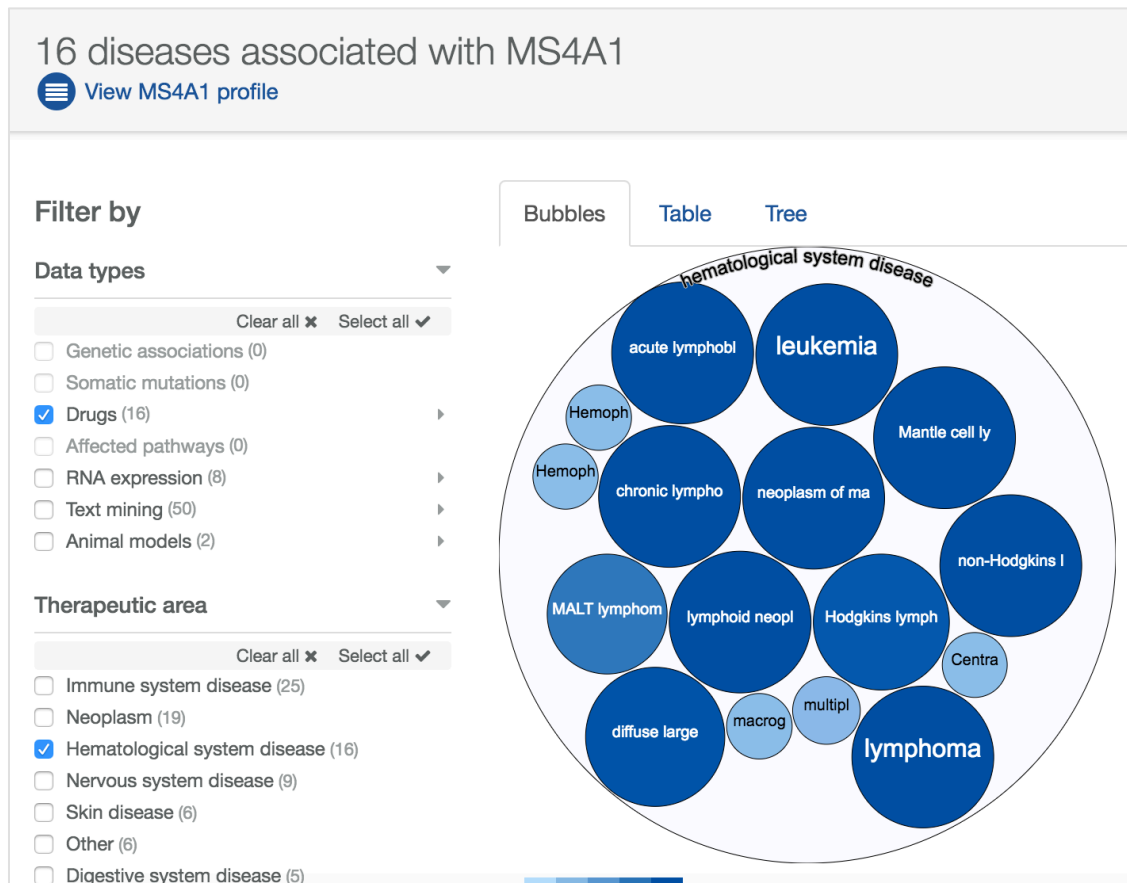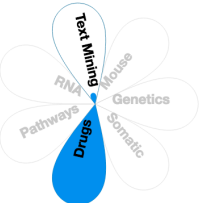
b) The default display for the diseases associated with a target is the Bubbles view. But you can the same results as a Table or Tree. Not sure what does views mean? Check our help page:
https://targetvalidation.org/about#target_assoc

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:

Look for the 'Download' icon and download the table in CSV:

c) There are over 700 papers mined by EuropePMC, which suggest the *MS4A1* and non-Hodgkin's lymphoma association. Note the low score given to data from text mining, which is in sharp contrast given to pieces of evidence coming from ChEMBL (i.e. drugs).

d) Click on the cell that corresponds to Drugs in the table:

You will go land on a page like:

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, lymphom non-Hodgkins, lymphoma, nonhodgki

Distinct from Hodgkin lymphoma biologically, non-Hodgkin lympho by the absence of Reed-Sternber age, 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: phase III

| Disease | Drug Information | | | | |
	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…
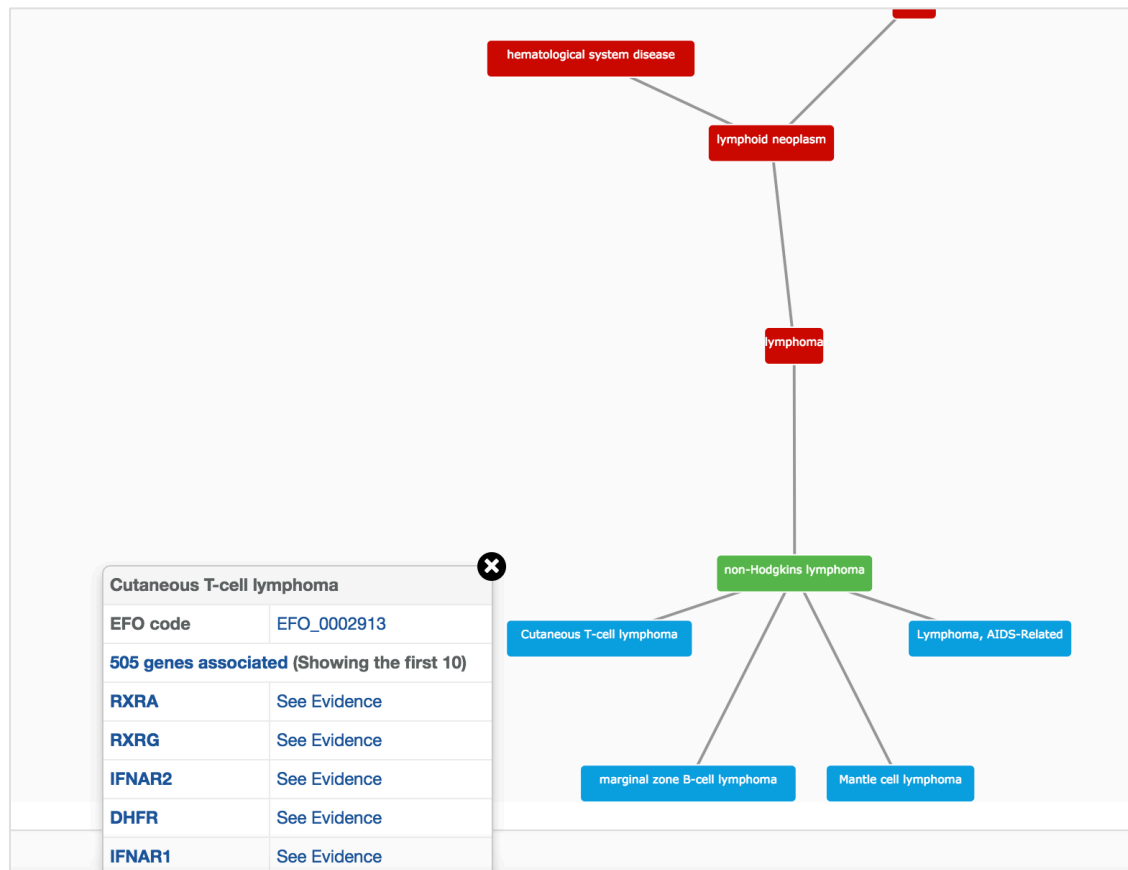
Click on the disease name above.

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 BEXAROTENE 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 THIOGUANINE 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: