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

Hands-on Workshop
Answer booklet

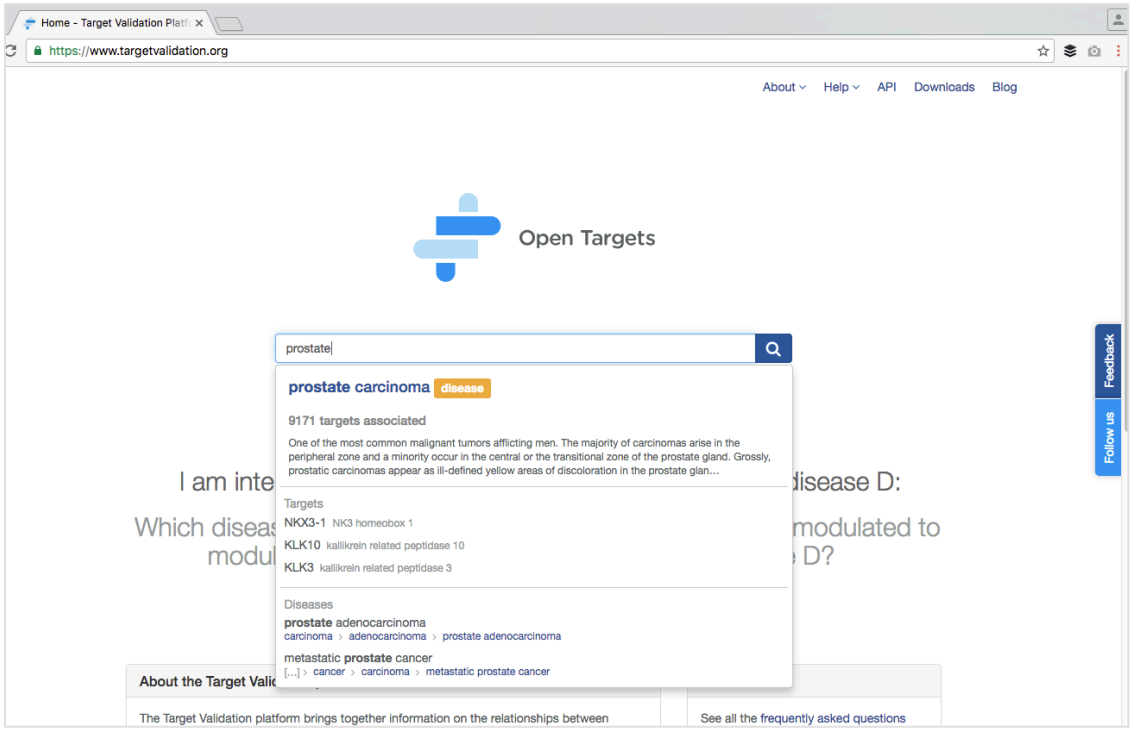
University of Cambridge
17th October 2016

Denise Carvalho-Silva
Open Targets Outreach

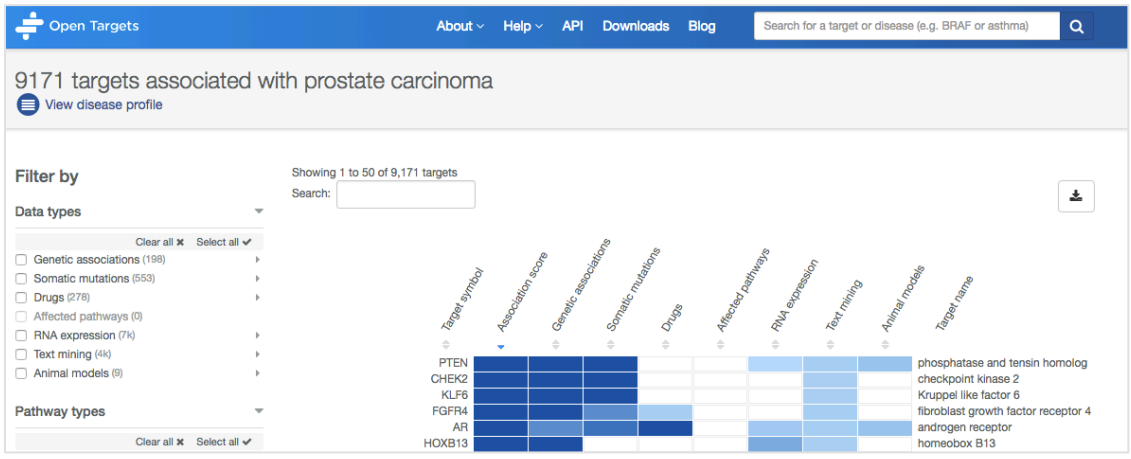
Answers to exercises on pages 21-24 of coursebook

Exercise 1 – Prioritising targets for drug discovery in prostate carcinoma

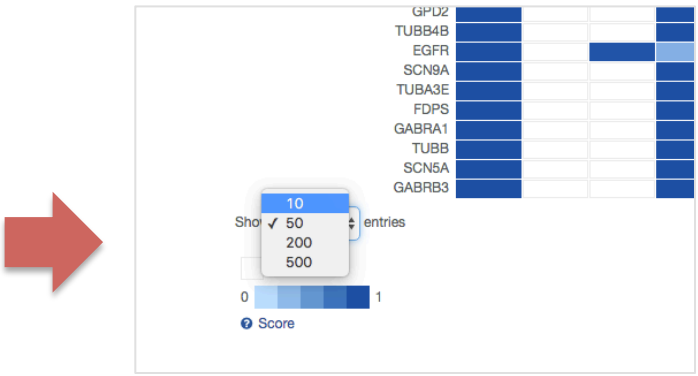
a) Go to www.targetvalidation.org and search for *prostate carcinoma*:



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:

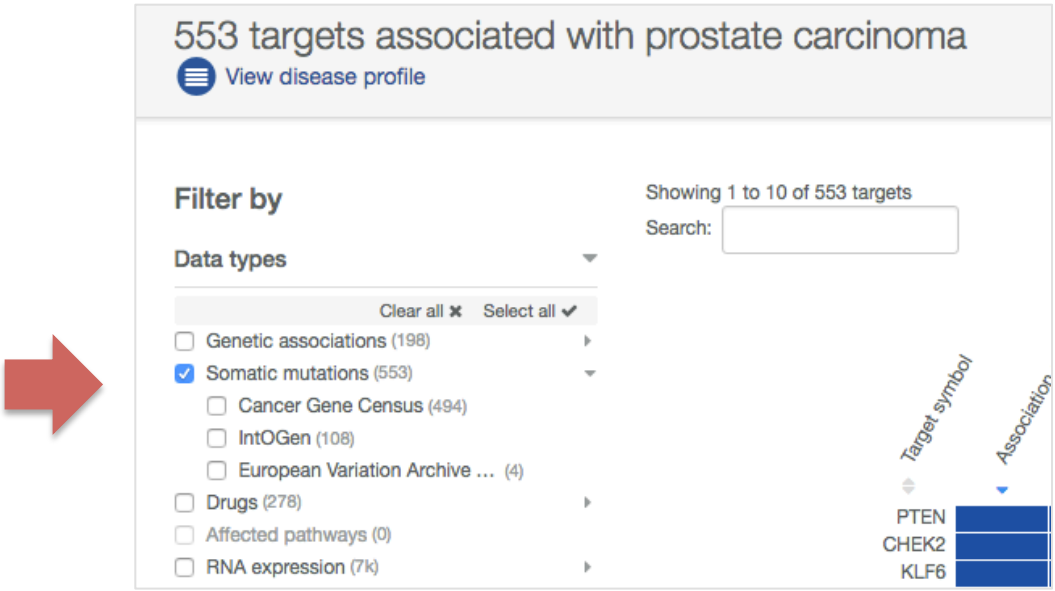


The first 10 rows will show the top 10 targets associated with prostate cancer. These will have the highest score (score of 1): *PTEN*, *CHEK2*, *KLF6*, *FGFR4*, *AR*, *HOXB13*, *ABL1*, *PDGFRB, KIT, and CACNA1D*.

The confidence on the target-disease association is indicated by the association score, which ranges from 0 to 1 (from no to the strongest association). The score is computed individually for each piece of evidence, followed by the score computed for the data sources (e.g. GWAS, 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://www.targetvalidation.org/scoring

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



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

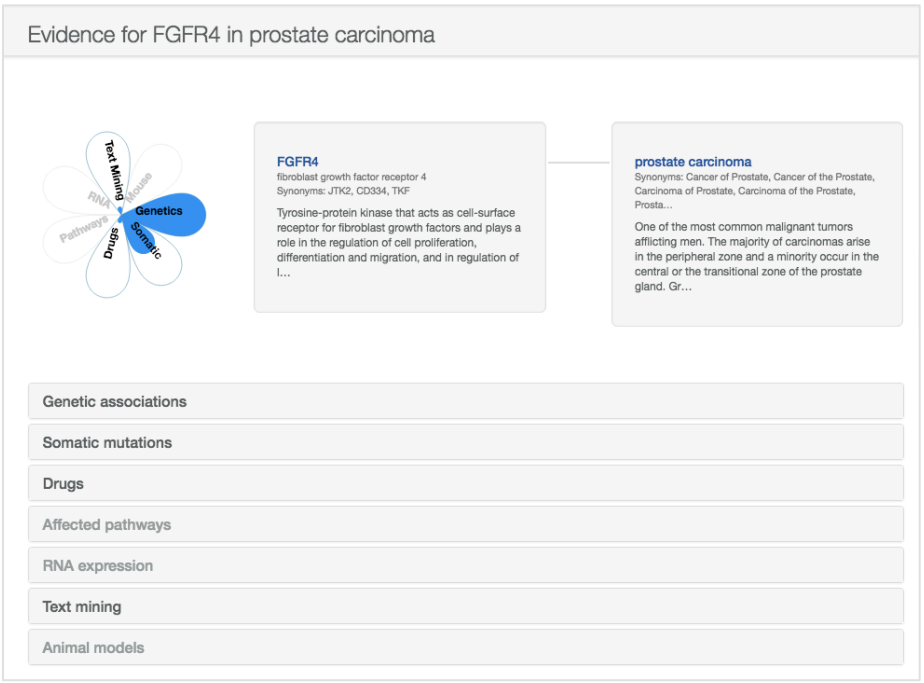
For more details on the data we currently use to associate gene A to disease B can be found below:

https://www.targetvalidation.org/data_sources

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 cancer.

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

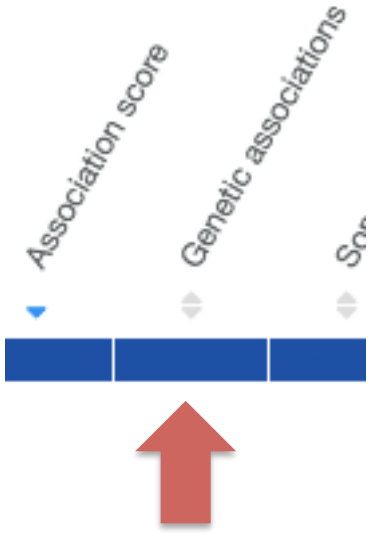
This will take you to a page similar to this:



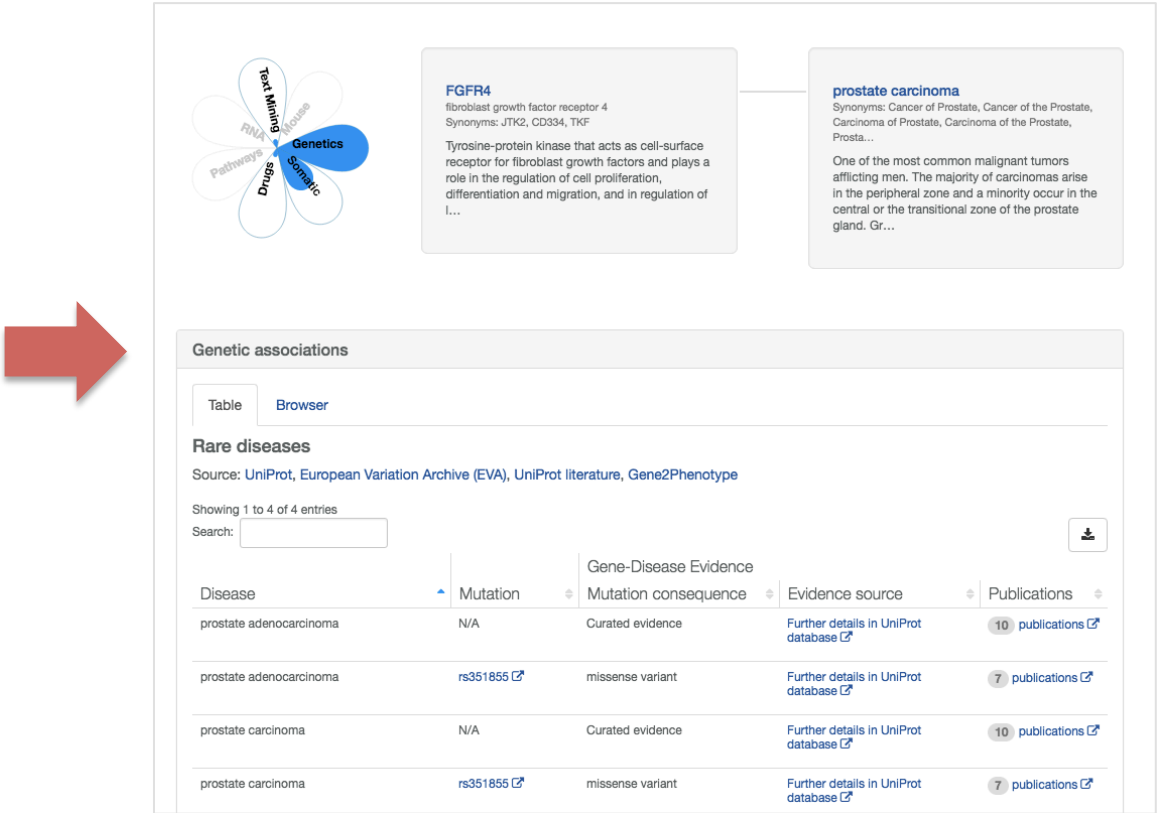
The evidence used to support the association is shown in different tables (the tabs that are greyed out have no data i.e. Affected pathways, RNA expression, Animal models).

c) Expand the ‘Genetic associations’ tab.

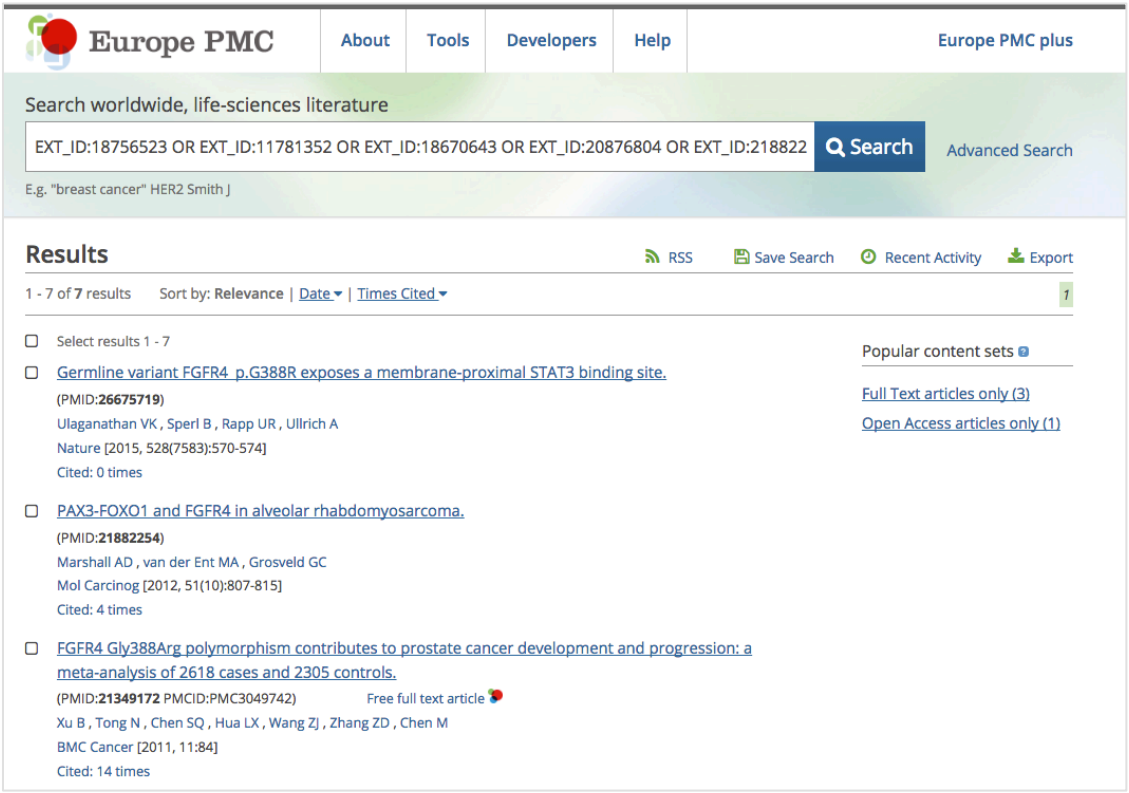
Note: if you click on the cell containing the data relative to Genetic associations (see below):



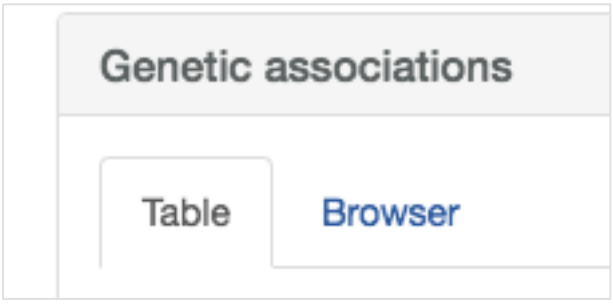
you will automatically lend on a page like this (where the tab containing the Genetic associations) will be already opened:



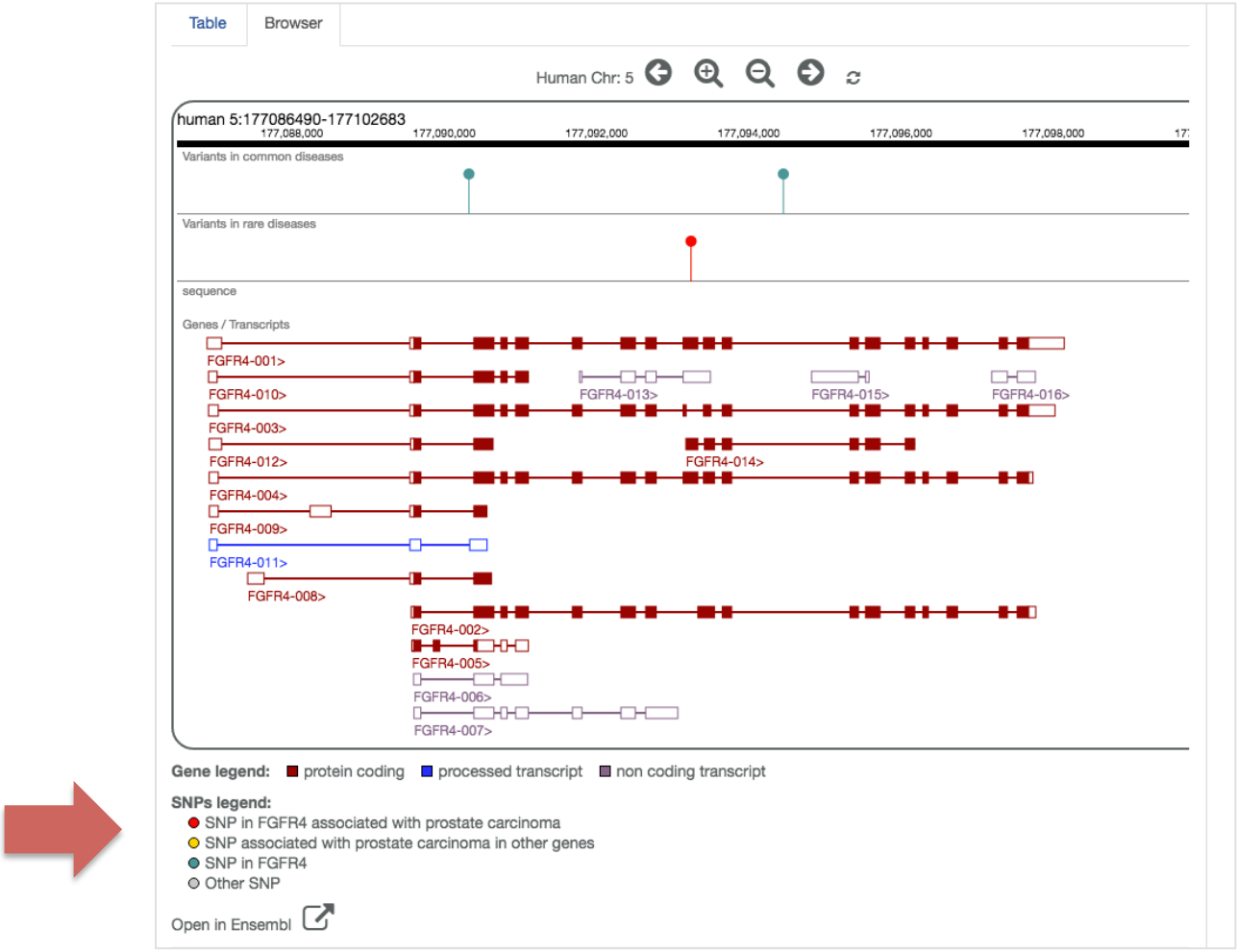
Yes, there is one known mutations (rs351855) in this gene, which is associated with prostate carcinoma. Click on the 7 'publications' link to see the papers supporting the association:



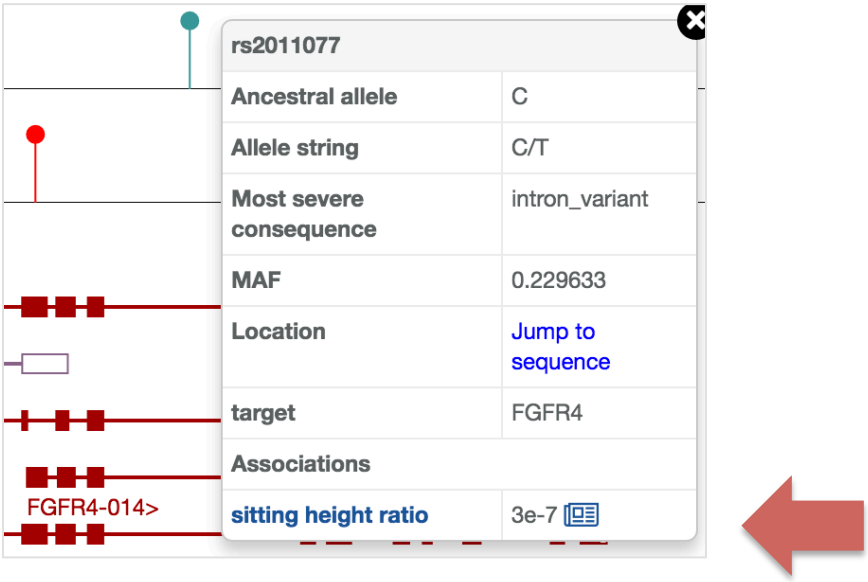
d) This information can also been visualised in a graphical display. Click on the 'Browser' link below:



You will be able to 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 colour smean.

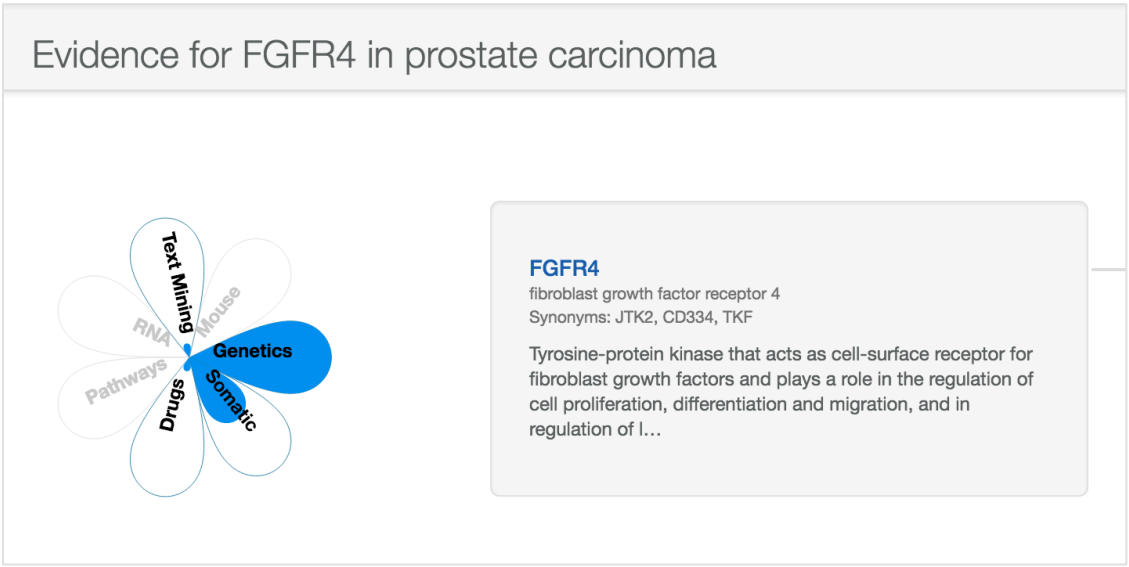


There other few other variants associated with other traits such as body mass index, waist-hip ratio and body height. You may want to zoom out to view more variants, then click on the lollipop for more details:

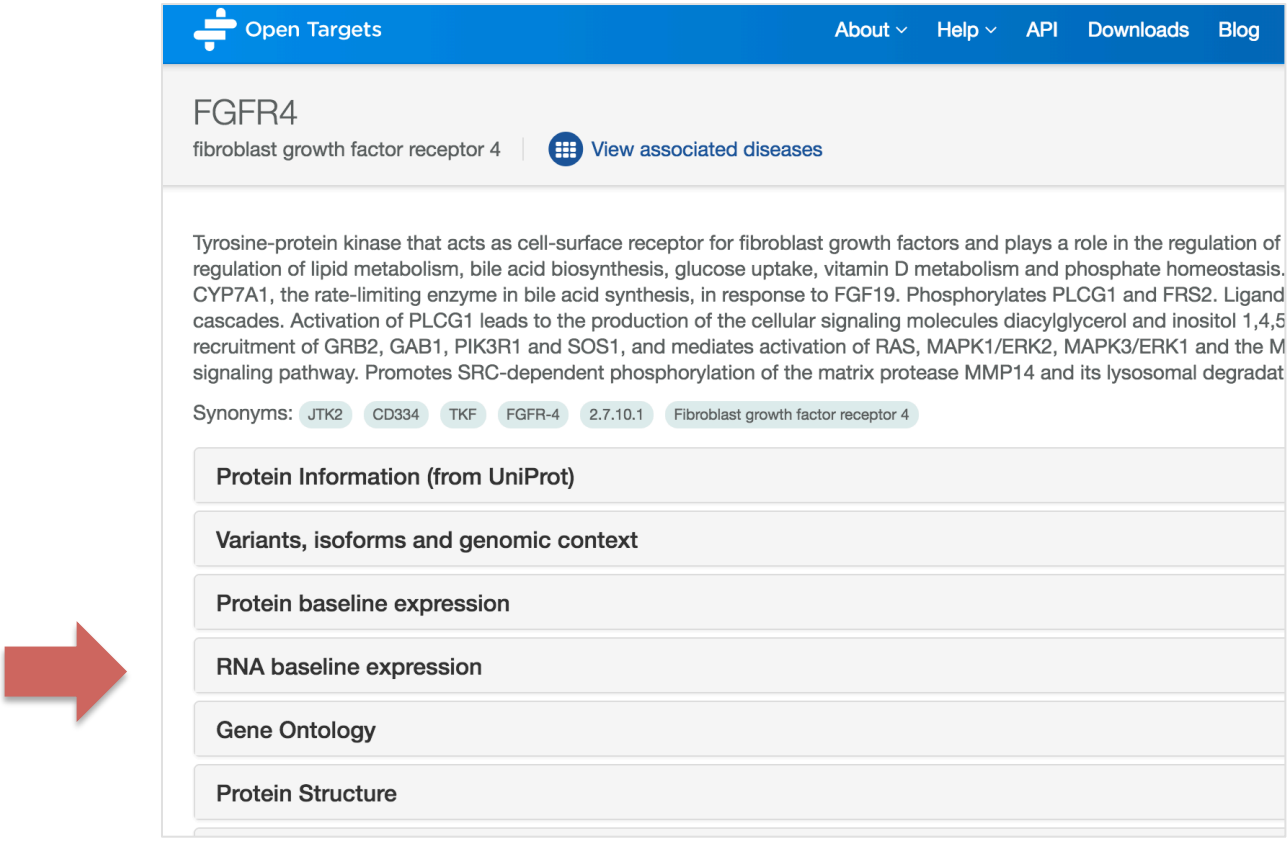


Let's now have a look at the target itself outside the specific context of any disease.

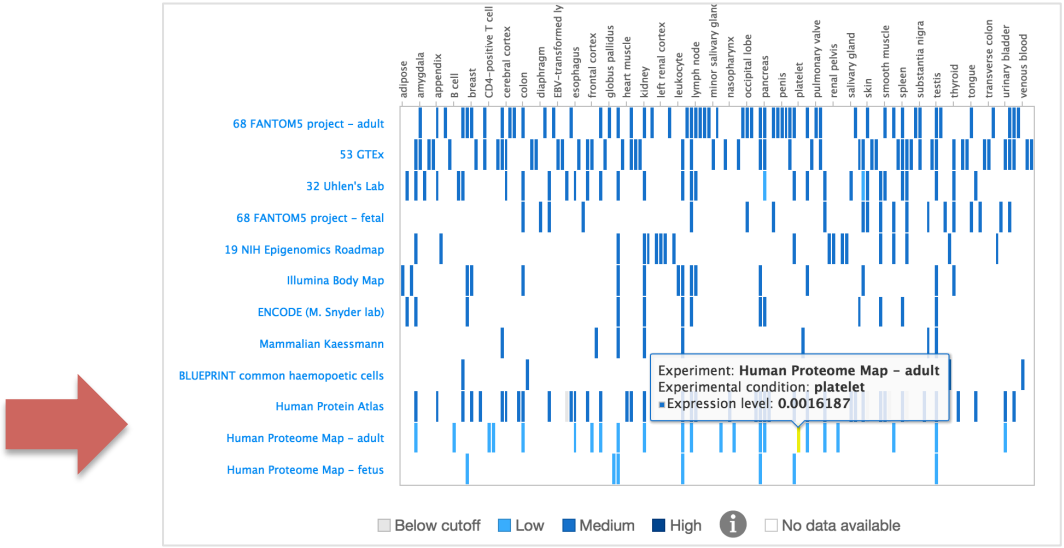
Still on the same page as above, click on the hyperlink FGFR4:



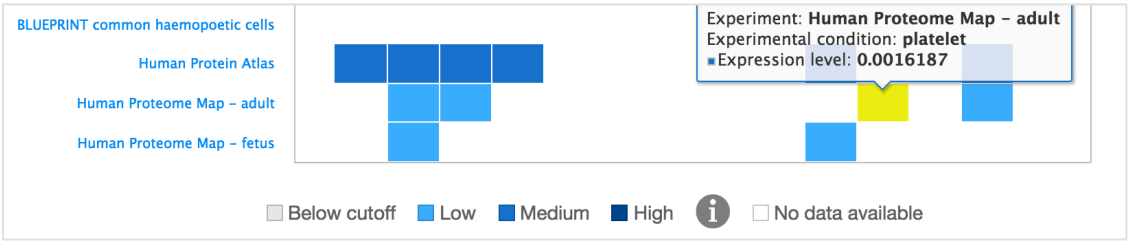
You will end up in a page like this:
https://www.targetvalidation.org/target/ENSG00000160867



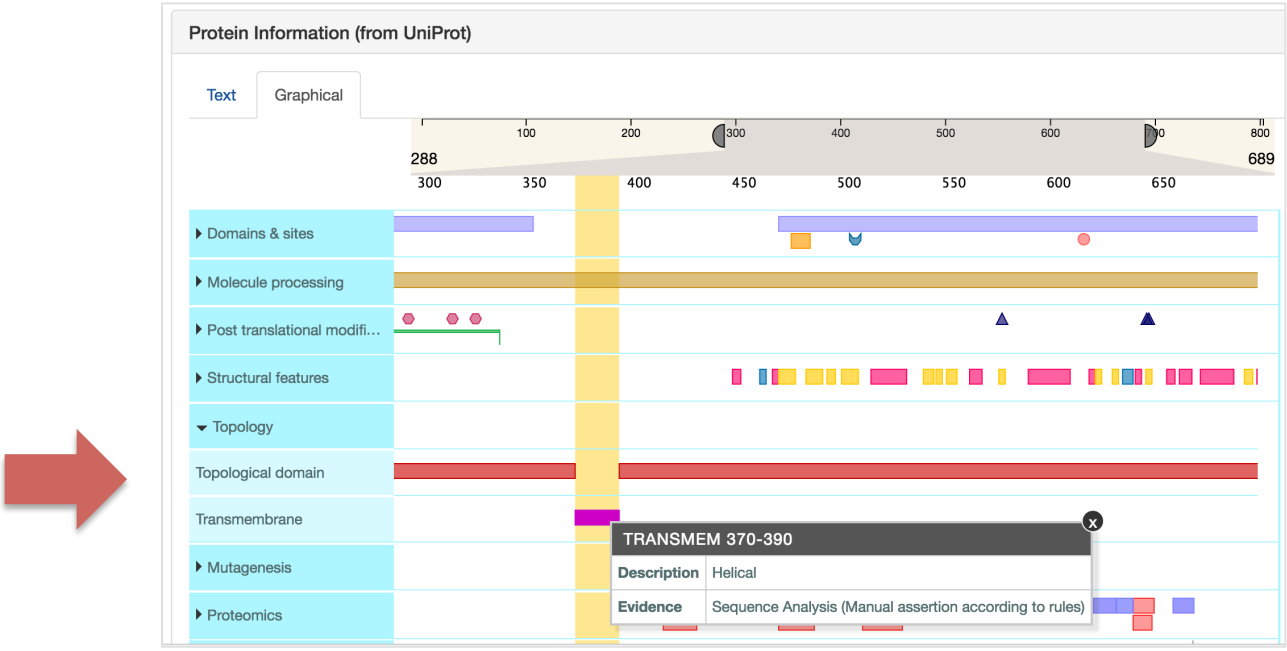
e) Click on RNA baseline expression to find out the tissue with the highest expression level according to Human Proteome Map (in adult tissues) is platelet:



You can also zoom in:

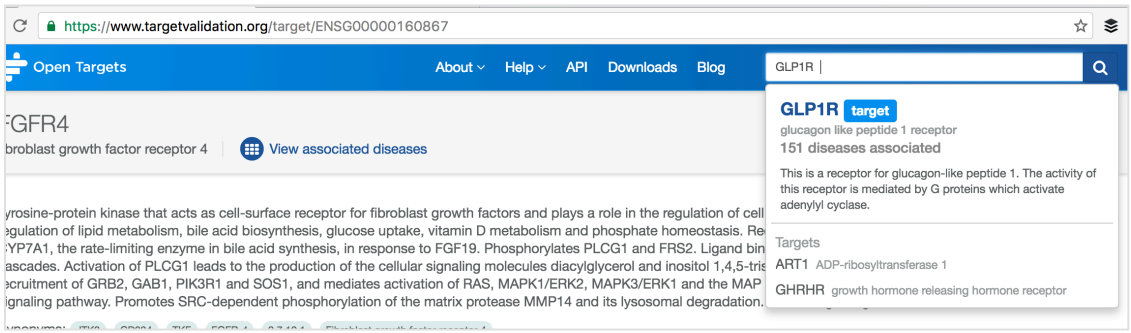


f) When you click on the Graphical view options of the Protein information (from UniProt) tab, you can click on the Topology menu to see the annotated domains: extracellular, transmembrance and intracellular. The transmembrane (TM) domain goes from amino acid 370 to 390. Gene *FGFR4* codes for a receptor,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

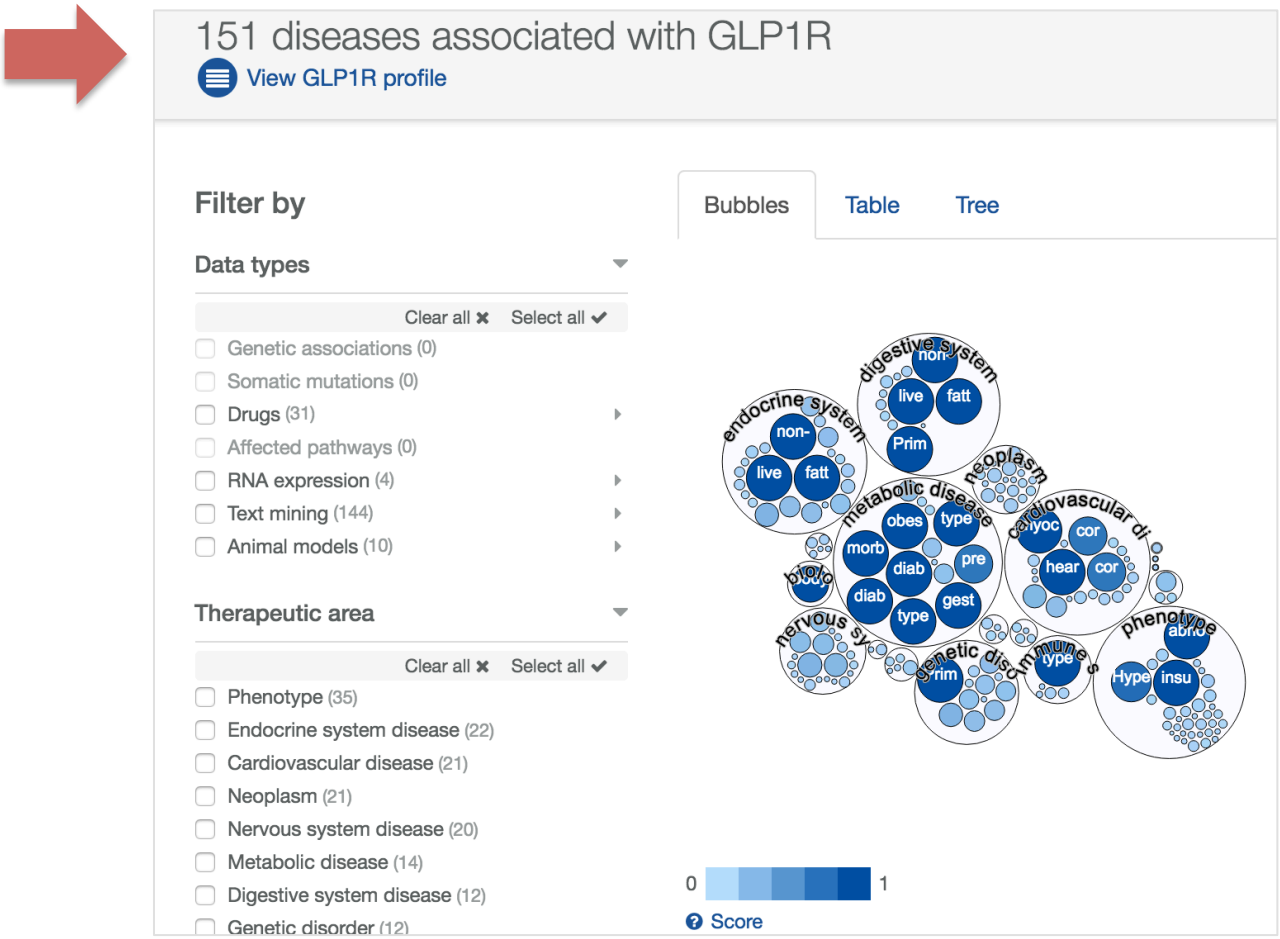


Exercise 2 – *GLP1R* and type II diabetes

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

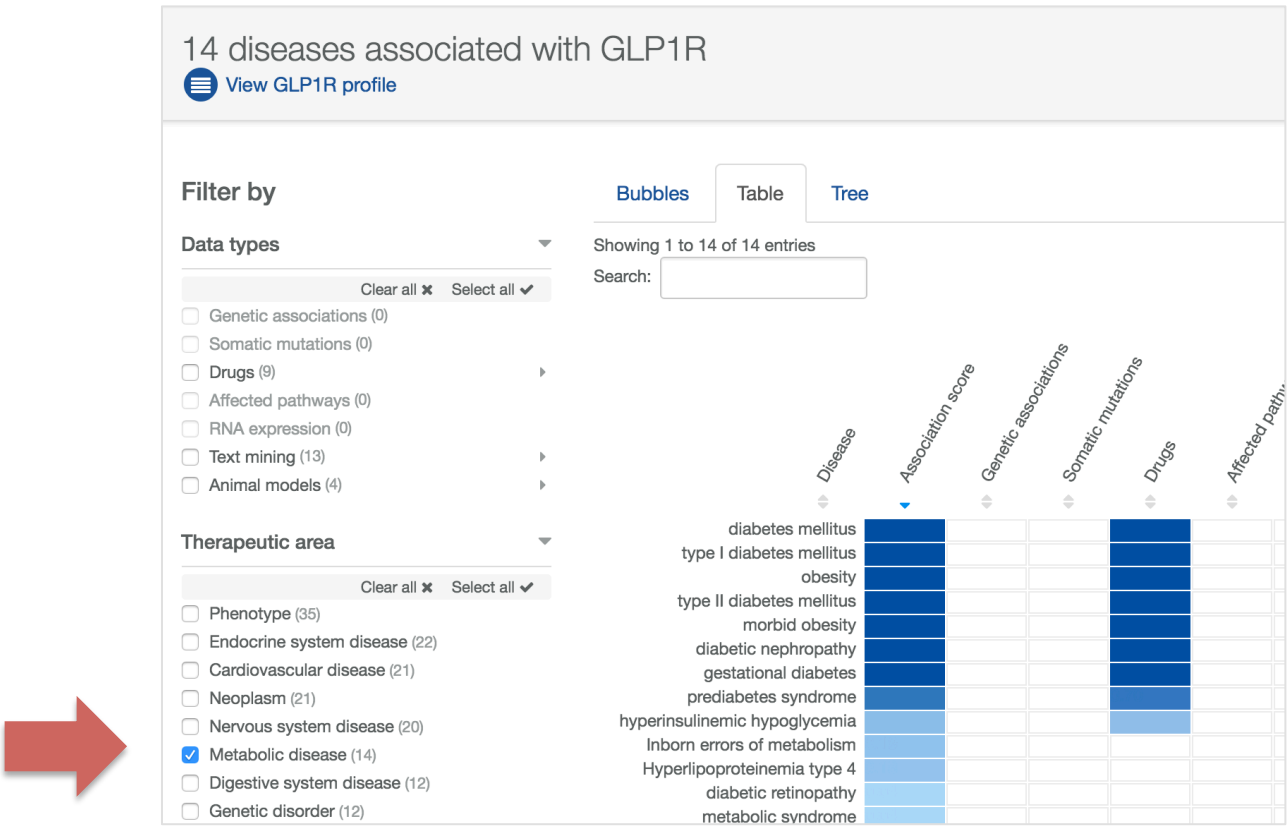


There are 151 diseases associated with target *GLP1R*.



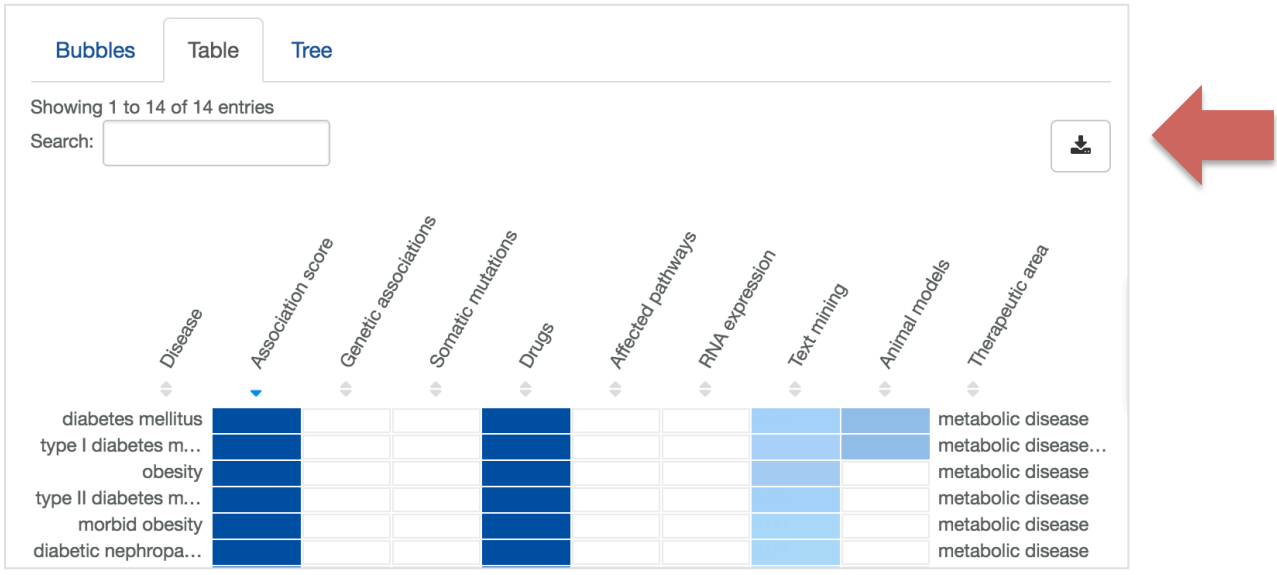
You can filter the results by Therapeutic area, such as '*Metabolic diseases*' (which includes type II diabetes). The number of diseases goes down to 14 such as diabetic nephropathy, type I diabetes, metabolic syndrome, and gestational diabetes.

You can view the results in a Bubbles view, Table view or Tree view:



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

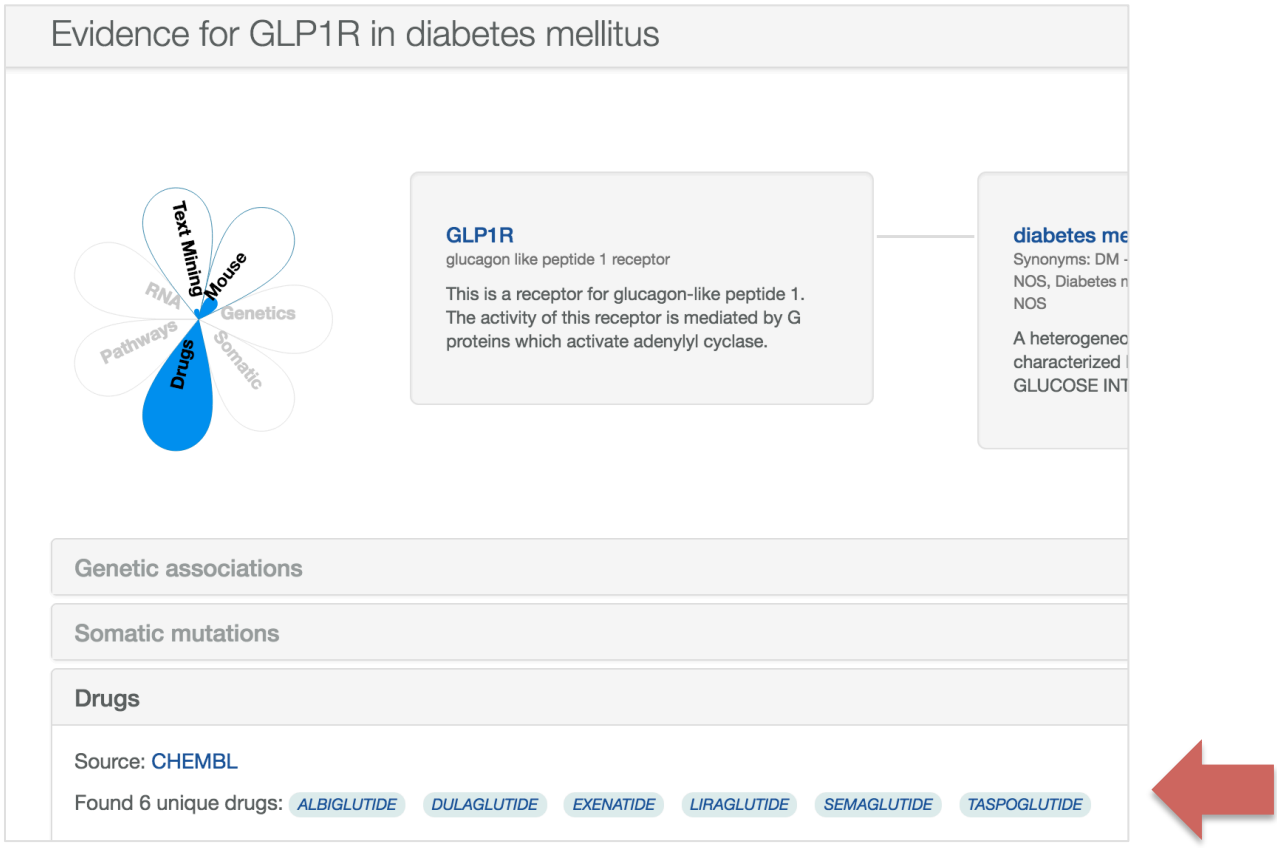
Look for the 'Download' icon to be able to download the table:



b) Drugs (from ChEMBL), Text mining (from EuropePMC) and Animal models (from Phenodigm) are the data types that point to the association of *GLP1R* with type II diabetes. The association based on Drugs information is valued at 1, a strong association.

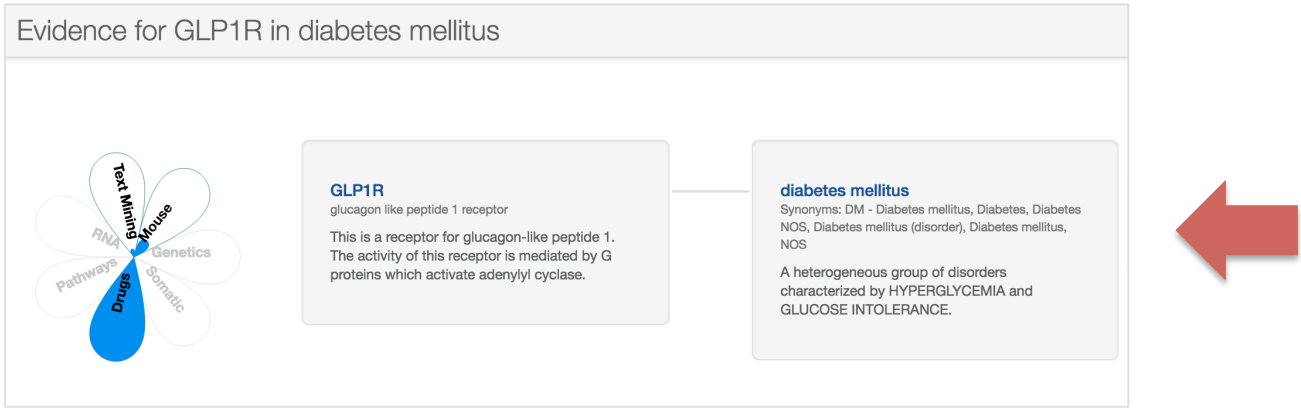
c) If you click on the cell that corresponds to Drugs in the table:

you will go to a page like:



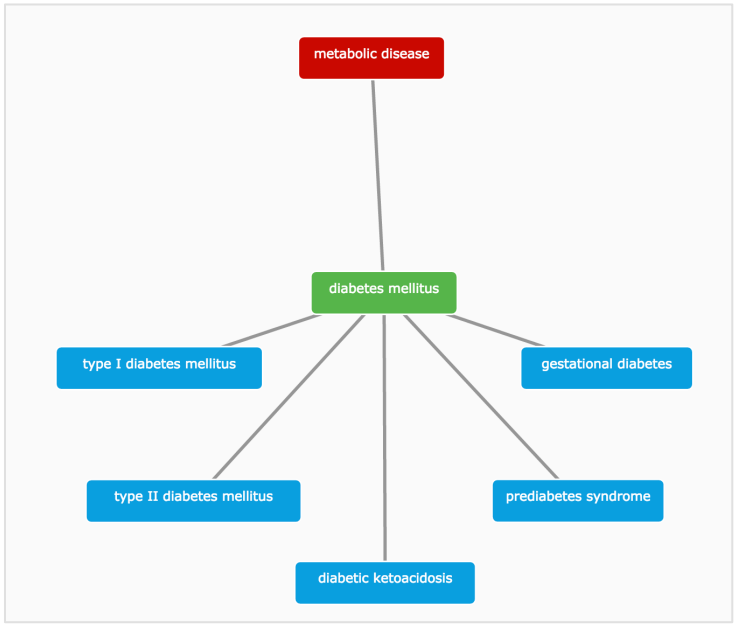
There are six unique drugs are currently mapped to this gene, which could potentially modulate target *GLP1R* in the context of type II diabetes. Two of those are in clinical phase IV (the FDA knows the drug works and there is a licence for it), namely Liraglutide and Exenatide.

d) Still on the same page, click on the box at the right hand side to find out a bit more about the disease.



Click on the disease name above.
Some of the synonyms or type II diabetes are DM and Diabetes NOS. The ID of this disease from the Experimental Factor Ontology (EFO) is EFO_0000400.

This is the diagram of the disease ontology:



It can be downloaded as a PNG format by clicking on the ‘Download’ icon in the top right of the image: