

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



**Hands-on Workshop  
Answer booklet**

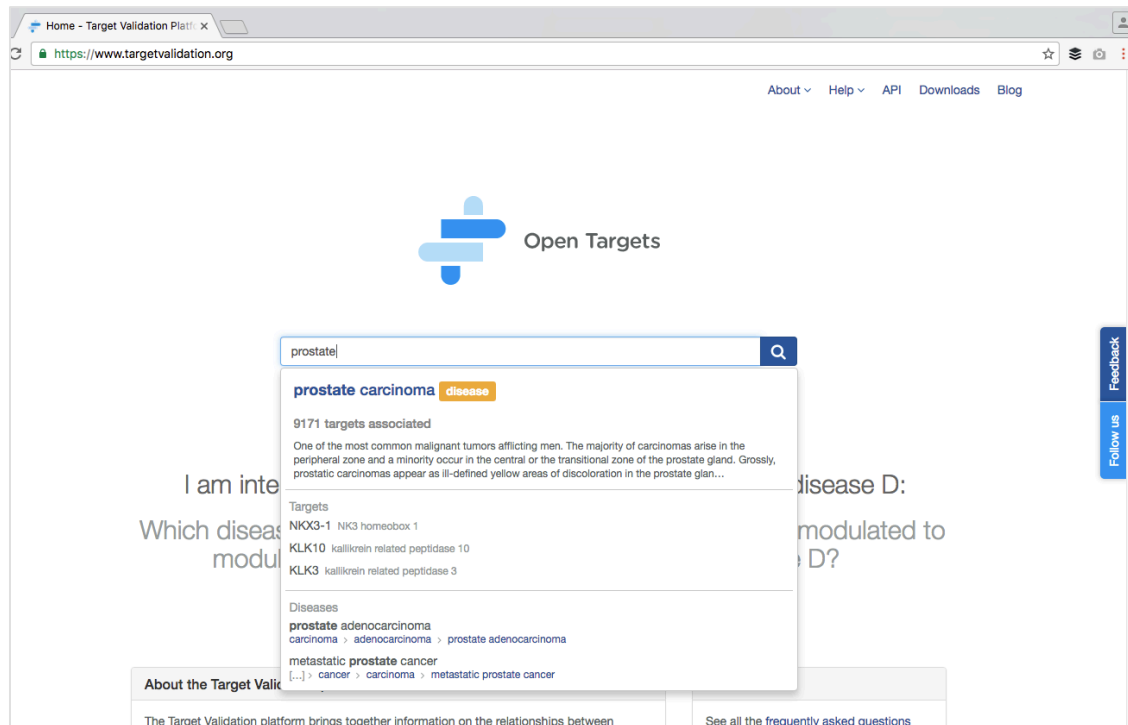
**University of Cambridge  
17<sup>th</sup> October 2016**

**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](https://www.targetvalidation.org) and search for *prostate carcinoma*:

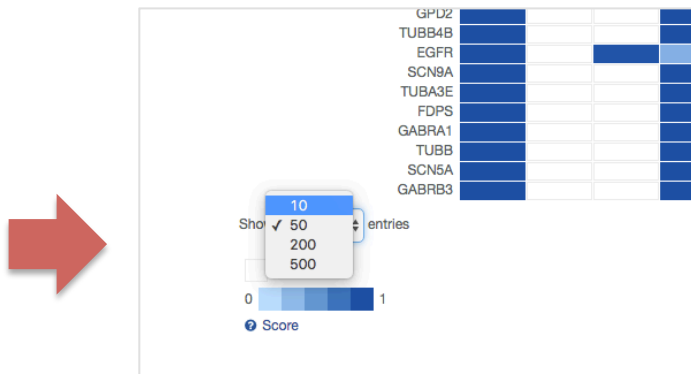


Select the first (best) hit. You will then see a page like this, which lists 9171 targets associated with prostate carcinoma:

The screenshot shows the Open Targets website with the search results for 'prostate carcinoma'. The page title is '9171 targets associated with prostate carcinoma'. On the left, there is a 'Filter by' section with 'Data types' and 'Pathway types'. The main content area shows a table of targets. The table has columns: Target symbol, Association score, Genetic associations, Somatic mutations, Drugs, Affected pathways, RNA expression, Text mining, Animal models, and Target name. The first 10 entries are listed in the table.

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:

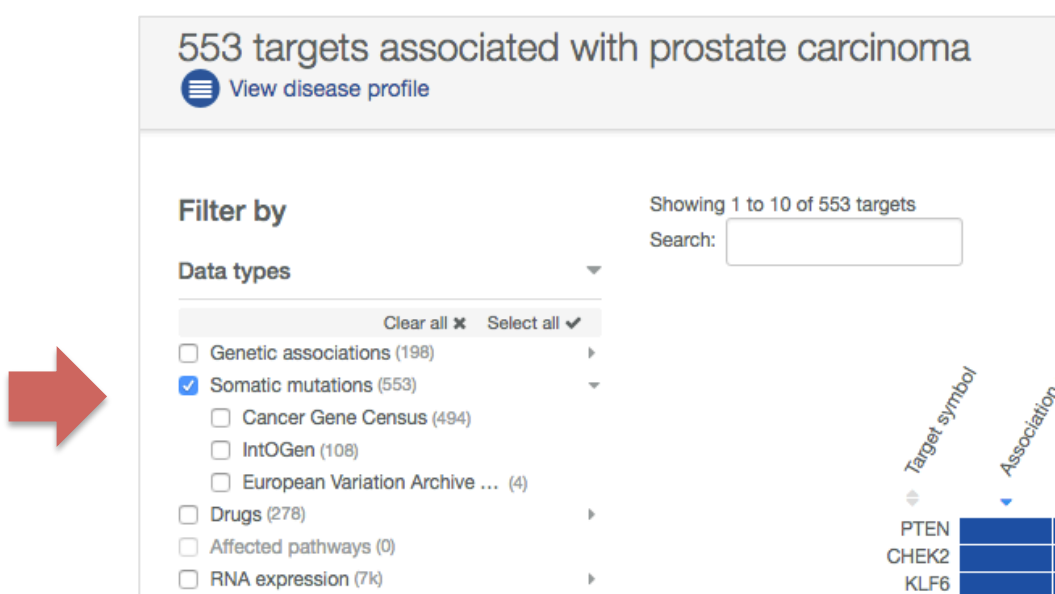


The first 10 rows will show the top 10 targets associated with prostate cancer. These will have the highest score (score of 1): *PTEN*, *CHEK2*, *KLF6*, *FGFR4*, *AR*, *HOXB13*, *ABL1*, *PDGFRB*, *KIT*, and *CACNA1D*.

The confidence on the target-disease association is indicated by the association score, which ranges from 0 to 1 (from no 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:

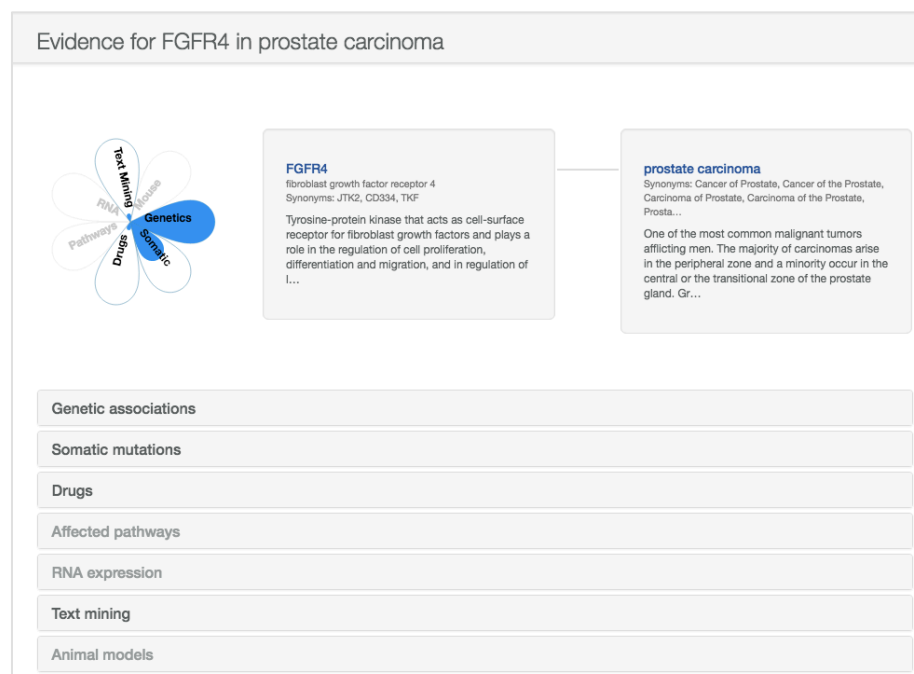
[https://www.targetvalidation.org/data\\_sources](https://www.targetvalidation.org/data_sources)

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:



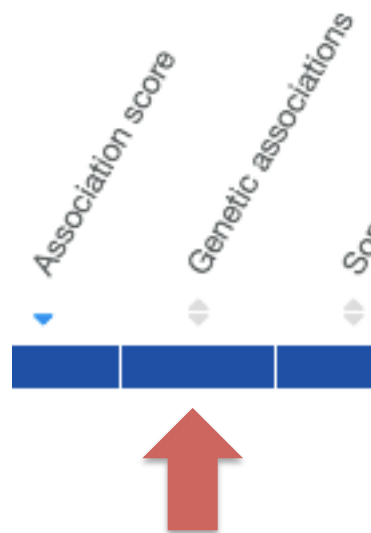
This will take you to a page similar to this:




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 land on a page like this (where the tab containing the Genetic associations) 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	<a href="#">Further details in UniProt database</a>	10 publications
prostate adenocarcinoma	<a href="#">rs351855</a>	missense variant	<a href="#">Further details in UniProt database</a>	7 publications
prostate carcinoma	N/A	Curated evidence	<a href="#">Further details in UniProt database</a>	10 publications
prostate carcinoma	<a href="#">rs351855</a>	missense variant	<a href="#">Further details in UniProt database</a>	7 publications

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:

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

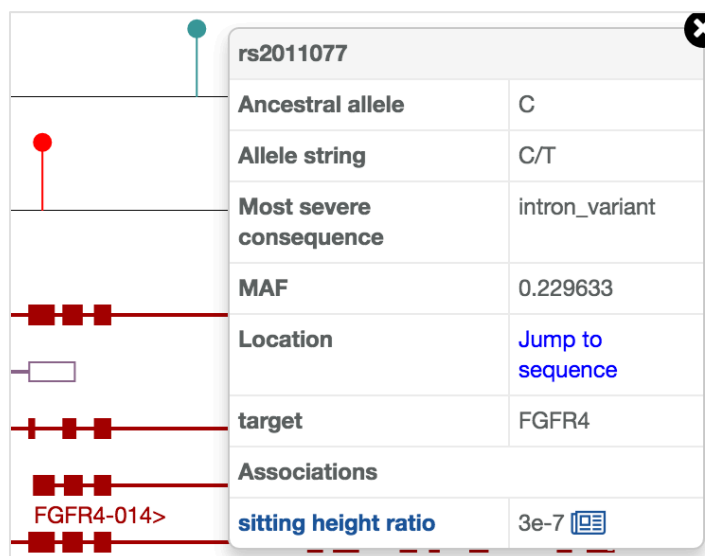
**Genetic associations**

Table Browser

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 mean.



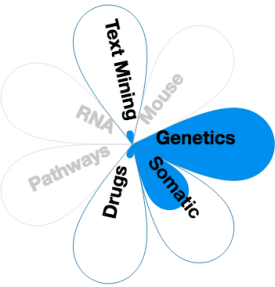
There other few other variants associated with other traits such as body mass index, waist-hip ratio and body height. You may want to zoom out to view more variants, then click on the lollipop for more details:



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:


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

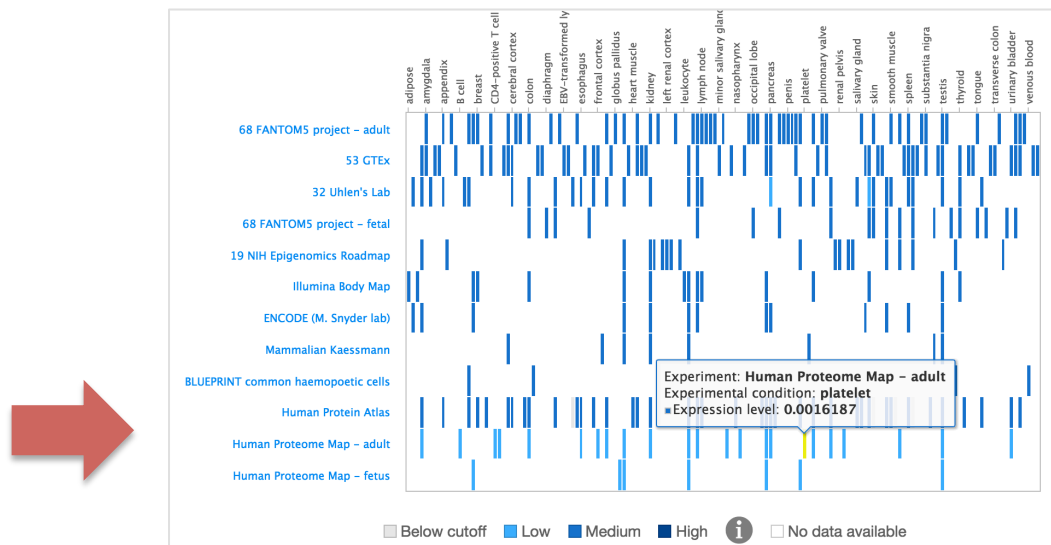
RNA baseline expression

Gene Ontology

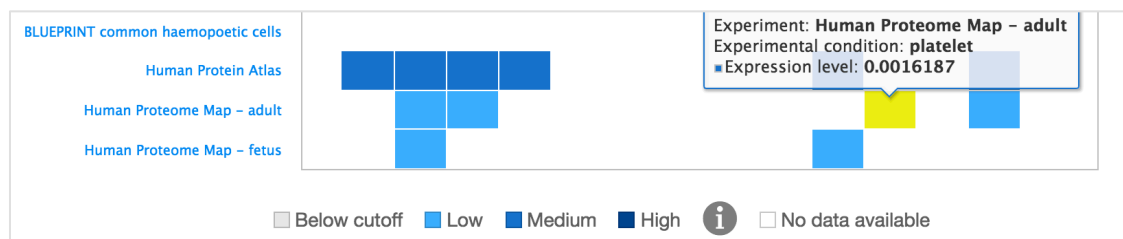
Protein Structure



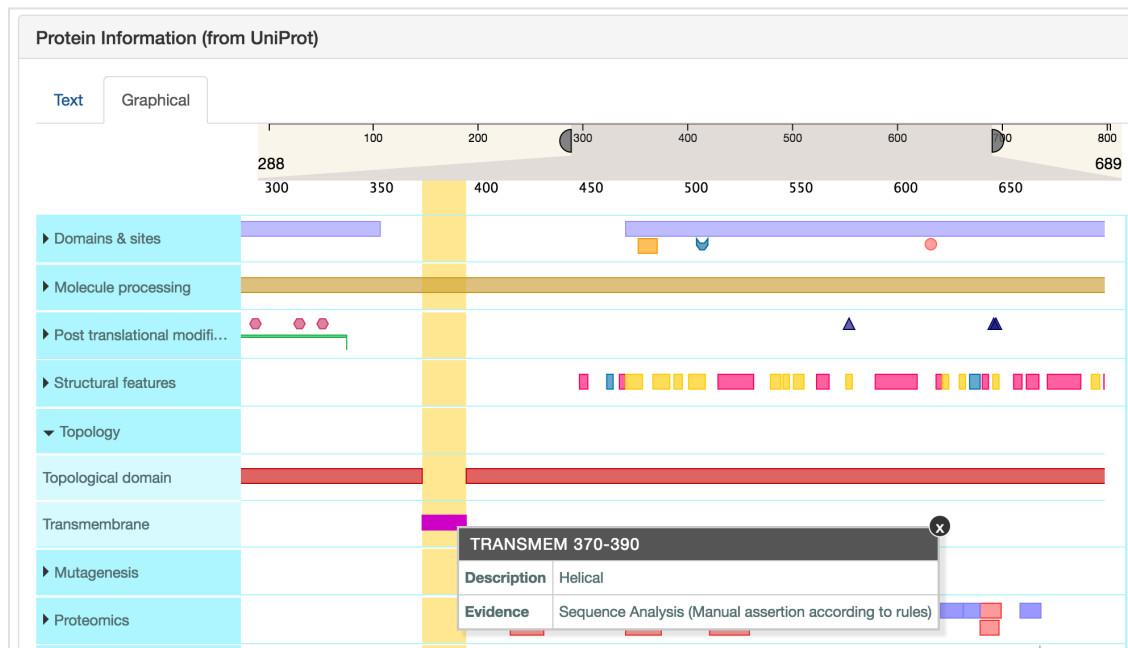
e) Click on RNA baseline expression to find out the tissue with the highest expression level according to Human Proteome Map (in adult tissues) is platelet:



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, 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 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):

Open Targets About Help API Downloads Blog

GLP1R |

**GLP1R target**

glucagon like peptide 1 receptor

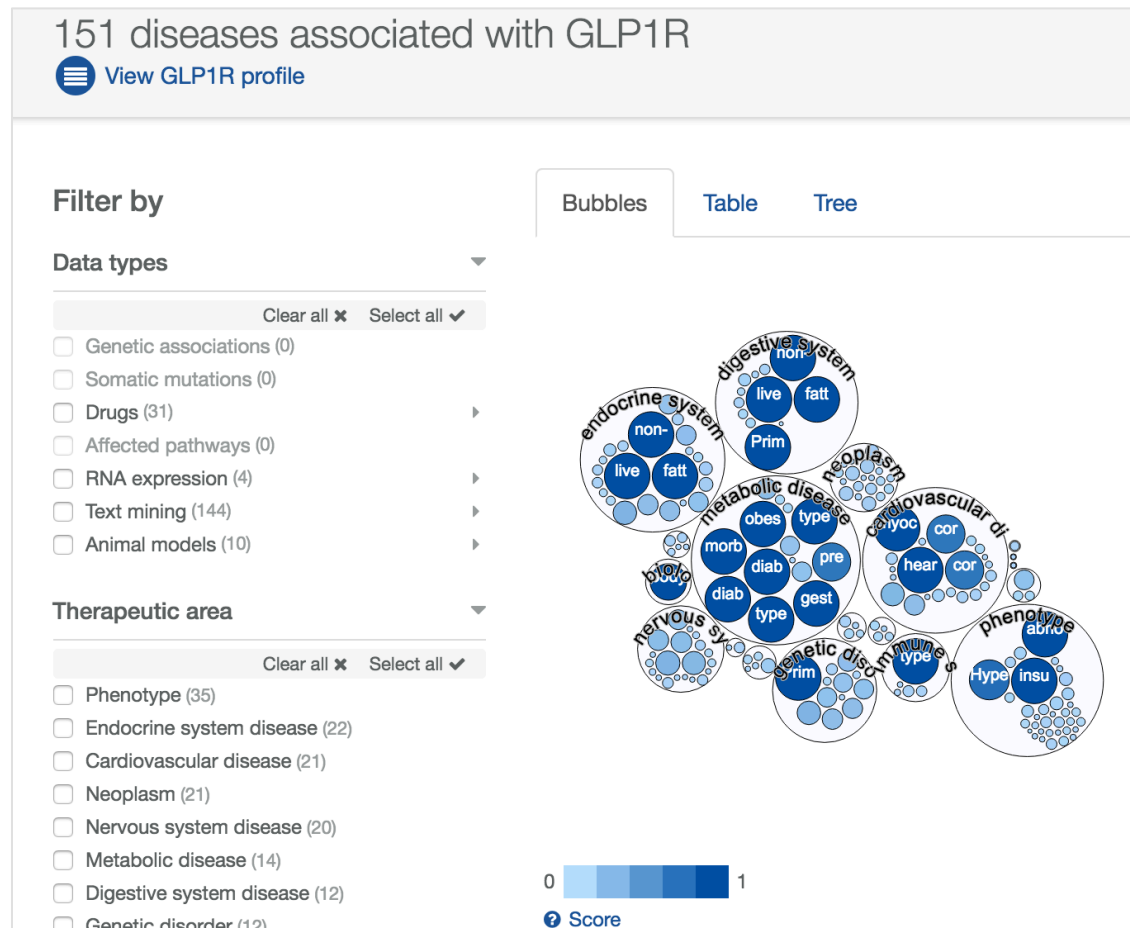
151 diseases associated

This is a receptor for glucagon-like peptide 1. The activity of this receptor is mediated by G proteins which activate adenylyl cyclase.

Targets

- ART1 ADP-ribosyltransferase 1
- GHRHR growth hormone releasing hormone receptor

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:

14 diseases associated with GLP1R  
[View GLP1R profile](#)

**Filter by**

**Data types**

Clear all ✕ Select all ✓

- ☐ Genetic associations (0)
- ☐ Somatic mutations (0)
- ☐ Drugs (9)
- ☐ Affected pathways (0)
- ☐ RNA expression (0)
- ☐ Text mining (13)
- ☐ Animal models (4)

**Therapeutic area**

Clear all ✕ Select all ✓

- ☐ Phenotype (35)
- ☐ Endocrine system disease (22)
- ☐ Cardiovascular disease (21)
- ☐ Neoplasm (21)
- ☐ Nervous system disease (20)
- ☒ Metabolic disease (14)
- ☐ Digestive system disease (12)
- ☐ Genetic disorder (12)

Showing 1 to 14 of 14 entries

Search:

**Bubbles** **Table** **Tree**

Disease	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways
diabetes mellitus					
type I diabetes mellitus					
obesity					
type II diabetes mellitus					
morbid obesity					
diabetic nephropathy					
gestational diabetes					
prediabetes syndrome					
hyperinsulinemic hypoglycemia					
Inborn errors of metabolism					
Hyperlipoproteinemia type 4					
diabetic retinopathy					
metabolic syndrome					


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:

**Bubbles** **Table** **Tree**

Showing 1 to 14 of 14 entries

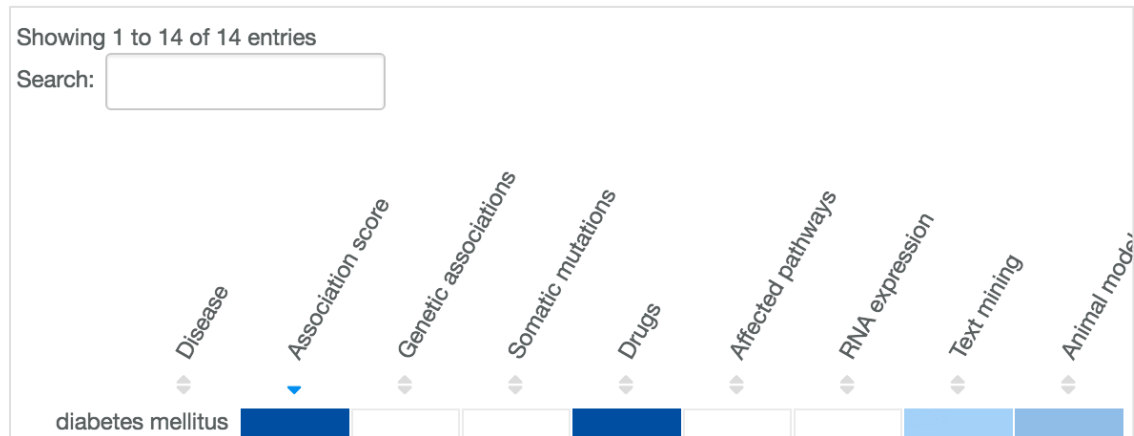
Search:



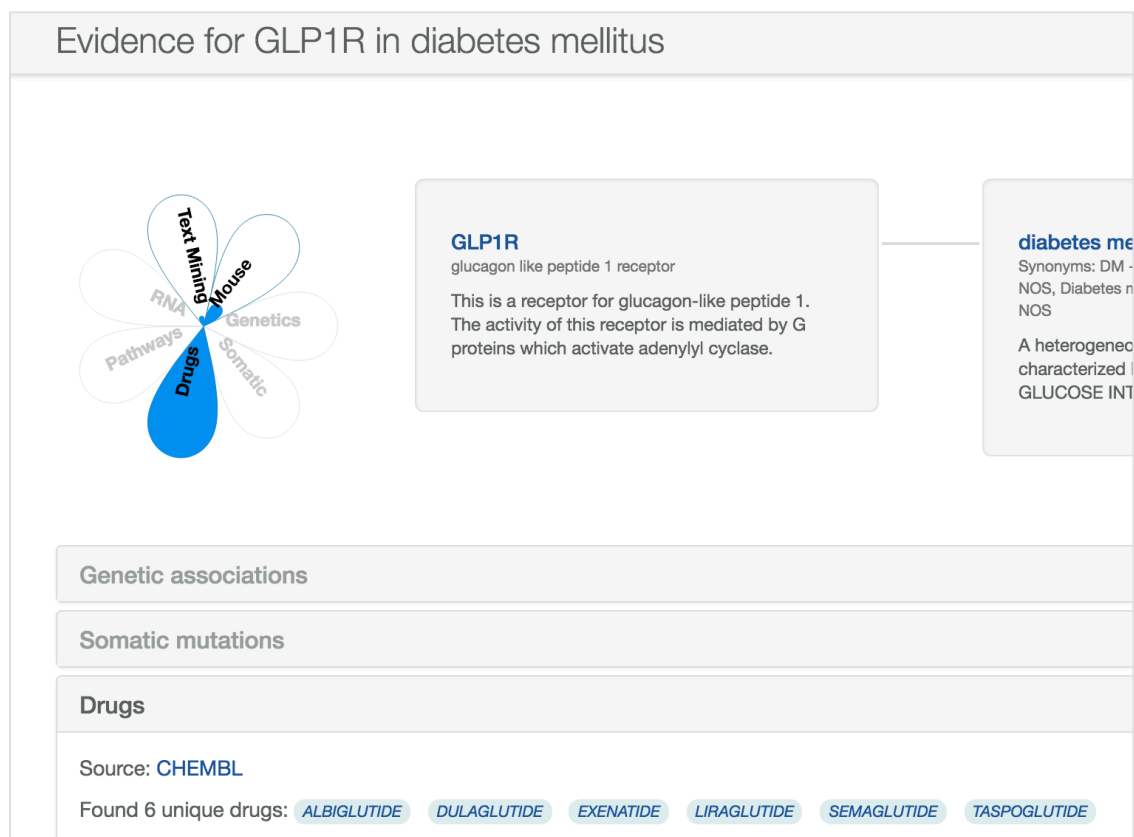
Disease	Association score	Genetic associations	Somatic mutations	Drugs	Affected pathways	RNA expression	Text mining	Animal models	Therapeutic area
diabetes mellitus									metabolic disease
type I diabetes m...									metabolic disease...
obesity									metabolic disease
type II diabetes m...									metabolic disease
morbid obesity									metabolic disease
diabetic nephropa...									metabolic disease

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:



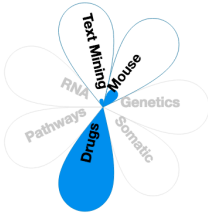
you will go to a page like:



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.

Evidence for GLP1R in diabetes mellitus



**GLP1R**  
glucagon like peptide 1 receptor

This is a receptor for glucagon-like peptide 1. The activity of this receptor is mediated by G proteins which activate adenylyl cyclase.

**diabetes mellitus**

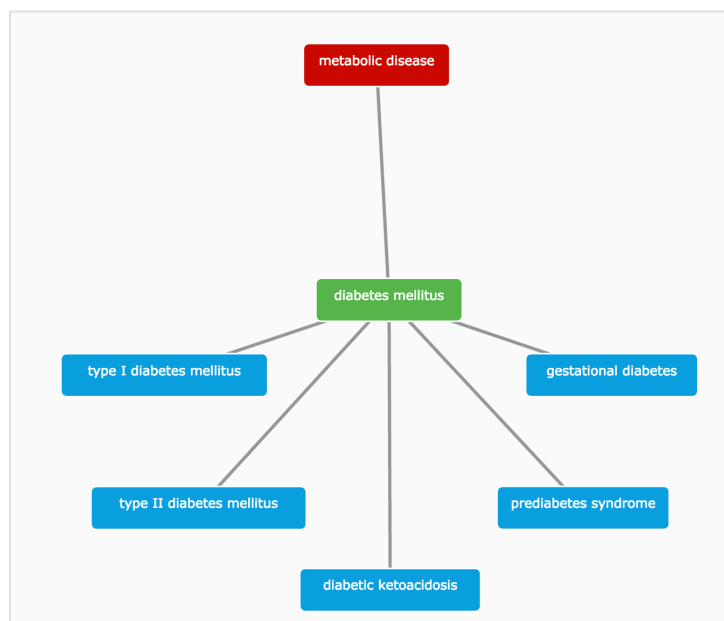
Synonyms: DM - Diabetes mellitus, Diabetes, Diabetes NOS, Diabetes mellitus (disorder), Diabetes mellitus, NOS

A heterogeneous group of disorders characterized by HYPERGLYCEMIA and GLUCOSE INTOLERANCE.

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:

