**Report on choosen CAR-T targets**

EGFR, CD7, MSLN, and CD19 have been evaluated as potential CAR-T targets.

For each target, the properties that have been evaluated are:

* Isoforms expression in cancer and healthy tissues
* Protein topology
* Cellular localization
* Epitope colocalization, when possible (missing)

The script (Rmarkdown) generating the results is in: cartcontent/scr/ 02\_Targets\_evaluation\_figures\_paper\_2024.Rmd.

**Isoforms expression in cancer and healthy tissues**

EGFR 🡪 higher in cancer, but not absent in healthy tissues.

CD7 🡪 it is normally expressed on T cells and NK cells (source [the protein atlas](https://www.proteinatlas.org/ENSG00000173762-CD7/single+cell+type)). However, in TCGA the only blood related cancers are acute myeloid leukemia and diffuse large B cell lymphoma. So we cannot verify the expression of CD7 in T cell related malignancies.

A screenshot of a computer

Description automatically generated

Based on the GTEX data, CD7 is also expressed in lung, small intestine, spleen, stomach. Most likely, the reason why it is expressed in such tissues is that there are infiltrating T cells in the tissue. We could check the literature, to see if there is agreement with the fact that CD7 is only expressed on T cells or NK cells. In addition, we could check how the expression of CD7 correlates with CD3 or CD56 espression (T cell and NK markers)?

MSLN 🡪 in both cancer and healthy tissues.

CD19 🡪 it is used as target for both hematological and solid malignancies.

It is expressed on healthy small intestine, spleen, stomach, testis.

Same as for CD7: the reason why it’s expressed on non-haematological tissues is most likely because there are B cells infiltrating in such tissues. Also in this case, we should check if literature is in agreement with CD19 being only expressed by B cells. We can also check how the expression of CD19 correlates with CD20 or other known B cell markers.

Conclusion: it is hard to find a target that is not expressed in healthy tissues. Maybe we can focus on bringing CD19 or CD7 as example?

**Protein topology & cellular localization**

Protein topology is line with cellular localization: for CD7, CD19, and EGFR, variants with a TM domain are also predicted to be on the cell membrane by DeepLoc2. Regarding MSLN, none of the variants has a TM domain. DeepLoc2 still predicts them as membrane bound. This can be explained by the fact that MSLN is anchored to the cell membrane by glycosylphosphatidylinositol (GPI) anchors, and therefore does not need a TM domain. It would be interesting to predict possible GPI anchors (tool [NetGPI](https://services.healthtech.dtu.dk/services/NetGPI-1.1/)).

Example: EGFR (the canonical isoform is highlighted in the plots below).

Both ENST00000275493 and ENST00000455089 have a TM domain. They are also predicted to be localized on the cell membrane by DeepLoc2. Other variants do not have the TM domain, and are predicted to be extracellular.

A colorful squares with white text

Description automatically generated



A graph with different colored bars

Description automatically generated with medium confidence



**Further observations on CD19 canonical vs CD19 delta exon2 variant**

* After running AlphaFold3 for CD19 and CD19 delta exon2, both in complex with the Ab specific for FCM63, I imported the two resulting structures in PyMol.

The PyMol sessions can be found in “cartcontent/results/CD19 delta exon 2 analysis/Alphafold3\_results”, and are called “cd19\_canonical\_AgAb\_highligths.pse” and “cd19\_deltaexon2\_AgAb\_highligths.pse”.

Left: CD19 canonical. Right: CD19 delta exon2

Grey: Ab. Blue: CD19. Green: exon2. Orange: epitope FMC63.

A structure of a protein

Description automatically generated A close-up of a blue and white structure

Description automatically generated

The epitope position/orientation seems to be quite different in how the loops forming the epitope are oriented in the two structures, also according to Magnus. It could be enough to impair the binding. However, looking at the figures only gives a qualitative observation.

* I aligned the two epitopes using PyMOL. Using “super epitope\_name1, epitope\_name2” one can align the two structures and get the RMSD for the sequences.

As seen from the figure below, the RMSD is 1.841 for the epitope sequences in the two structures. Again, it could be enough to prevent the correct binding of the Ab in delta exon2.

The PyMOL file/session is called “superimposed\_ cd19\_AgAb.pse”.

A computer screen with colorful objects

Description automatically generated

* Moreover: in the canonical CD19 structure, the predicted structural quality (pLDDT) is high around the epitope. That means that AF is confident about the structure. However, in CD19 delta exon2, the quality is between low and very low (see figure below for delta exon2 bound to Ab). The region of the epitope is the contact region between the two proteins. This can indicate that (1) the loops forming the epitope are quite flexible and therefore could influence the ability of the Ab to bind theepitope FMC63, or that (2) AF is not sure about the location of the two loops forming the epitope, also in this case the binding could be affected.

A screenshot of a computer

Description automatically generated

* Another thing that one can look at is the PAE ([predicted alignment error](https://www.alphafold.ebi.ac.uk/faq)). How does the average PAE change between the bottom part of the epitope loop and the top? If AlphaFold is very uncertain about their relative placement, we can infer the loop might be very flexible.

The PAE measurements are in the .json file in the AF3 result folder. The file of interest is “Alphafold3\_results/cd19\_deltaexon2\_AgAb/fold\_cd19\_deltaexon2\_agab\_full\_data\_0.json”. This file can be imported in python (maybe also in R) and the matrix containing the PAE values can be visualized. However, I decided to have a look at the graph from AF3 to have a preliminary assessment of the PAE.

A screenshot of a graph

Description automatically generated🡨 We know that in CD19 delta exon2 the epitope goes from residue ~44 to 56 and the other one is from ~104 to 114 (these positions are approximate, do not trust this, check!).  
So, in the graph it would be in this area (red square) more or less, which has a high expected position error, so it could mean that the position of the two loops relative to each other is (very) uncertain. Which can mean that the loops forming the epitope are quite flexible, and consequently that the epitope is not stable in the structure and that can impair the binding.

Please note that these are not strong conclusions. These are just preliminary observations. Talk to Magnus for more insigths.