**How do we choose the proteins to target:**

For proteins:

we filter out proteins with high variance between versions

we used median to aggregate the versions

then:

1. we are using DEP (R library) to find DE proteins on two sets:
   1. the entire cells populations from Jan experiment (28 cells)
   2. smaller set with lower resolution of cells (9 cells)
2. we are doing the same with entropy based method

for transcriptomic:

1. we take cells from the “Human cell atlas” and using entropy based method
2. we use the analysis from "A genome-wide transcriptomic analysis of protein-coding genes in human blood cells" (which is based on the “Human cell atlas”) to find highly enriched gene:

how we pick:

we keep only genes that are enriched for specific cell/group

we keep genes that not exist in no more than a few cells

we remove Granulocytes

then we split to cell enrichment and group enrichment:

for group enriched - we take the top x% (enrichment score)

for cell enrichment - we take the top N (enrichment score) per cell type

1. we take surface proteins (from the literature – those are the surface proteins we can infer from the mRNA data using the CtpNet project)

additional data :

data from Shiran paper

summary:

“enrichment\_based proteins” : DE on 28 cells + DE on 9 cells

“entropy\_based\_proteins” : entropy analysis on 28 cells + entropy analysis on 9 cells

“surface proteins” : from literature

“enrichment\_based\_mrna” : DE on Human Atlas cells

“entropy\_based \_mrna” : entropy analysis on Human Atlas cells