**First of all, thank you to offer these challenges. I myself need more time to understand the science of Cancer Immunotherapy particularly your presentation. So I had short time to develop the code. Even though, I did what I can do to participate these challenges.**

**## CHALLANGE 1 is in "dataset.py"**

The dataset is formatted anndata in the file **sc\_training.h5ad**, but from the presentation the format data is **Tcell\_challange.h5ad**. Both should be the same.

Unfortunately, I didn’t get one of the both of files. Actually, I expected you to send the data file testing.

The condition information ('Unperturbed' or the target gene name if perturbed) is provided in .obs[‘condition’]

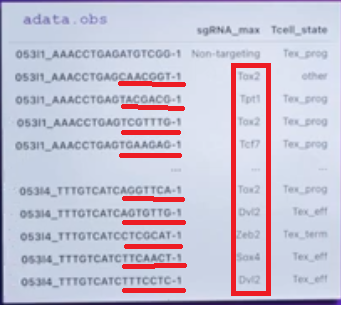
The cell state of each single cell is stored in .obs[‘state’]

**## Requirements ##**

Predict the cell-state proportions of the held-out genes: For each of the 7 held-out knockouts (targeting genes 'Aqr', 'Bach2', 'Bhlhe40', 'Ets1', 'Fosb', 'Mafk', 'Stat3')

Cell state proportions should add to be a+b+c+d+e=1.

Regarding to this requirements, more need information of data categories. I got to identify the categories of genes.



I didn’t see how to identify of your targeting genes such as (targeting genes 'Aqr', 'Bach2', 'Bhlhe40', 'Ets1', 'Fosb', 'Mafk', 'Stat3')

**## Result ##**

validation\_output.csv: This file should contain the 3×5 matrix in csv format, as well as the name of the gene in the first column.

test\_output.csv: This file should contain the 4×5 matrix in csv format. the 5-demensional vector prediction for the 4 held-out test set genes.

**## 3 parts of Therapy :**

**## (A) Immune Checkpoint Blockade Therapy**

The first goal is thus to identify the gene (out of all 15,077 possible genes that passed quality control) that when knocked out would lead to the largest proportion of progenitor cells

cell state proportion vector to be P=(1,0,0,0,0), cell state proportion vector of Pi=(ai,bi,ci,di,ei). the minimum cycling percentage constraint (i.e.,di≥0.05)

**## (B) CAR T-Cell Therapy**

define the objective function

the goal is thus to identify the gene (out of all 15,077 possible genes that passed quality control).

The top 20 submissions of Challenge 1 will have their top predicted genes.

each group is scored using the AUC computed similarly as in (A) but using the predicted and experimental rankings obtained based on the objective function used in (B).

**## (C) Other Therapeutic Strategies**

**## Scoring**

Propose a statistic s(⋅) that summarizes the gene expression distribution contained from knocking out gene

i as well as a scoring function that takes in Q and the predicted statistic ) and outputs the score of knocking out gene i (where a larger score indicates a better perturbation). For example, Challenge 2 (A) and (B) use simple scoring functions that do not depend on and use as the statistic to summarize the cell state proportion vector.