**Overview:**  
In this task, you will reproduce the unsupervised clustering of gliomas (n = 516 LGG) based on methylation levels to classify the six IDH statuses as reported in a specified paper. You (the data scientist) will employ the K-Nearest Neighbors (KNN) method for the classification. The goal is to demonstrate your ability to perform unsupervised clustering using methylation data and to evaluate the clustering performance. While the biomarker hunter will develop expression biomarkers that can distinguish any 2 IDH classes.  
:- You can also get the dataset for LGG from [XENA Browser](https://xenabrowser.net/datapages/?dataset=TCGA-LGG.methylation450.tsv&host=https%3A%2F%2Fgdc.xenahubs.net&removeHub=http%3A%2F%2F127.0.0.1%3A7222).**Steps**

1. Read the paper: <https://www.cell.com/cell/fulltext/S0092-8674(15)01692-X>
2. Carefully read the paper that reports the clustering of gliomas using methylation levels to classify the six IDH statuses.
3. Understand the methodology and the specific findings related to unsupervised clustering and the use of methylation data.
4. Obtain the dataset
5. You can compile from the TCGA (more up to date) using TCGABiolinks in R or the XENA Browser link above
6. Ensure you have IDH status for each sample. You can get this only from the TCGA
7. Data Preprocessing
8. Clean and preprocess according to your pipeline and needs.
9. For ML engineers, feature selection is important
10. For Biomarker-Devs, normalization is important
11. Implement the same pipeline as last week
12. For ML, perform kNN and random forest
13. For BDevs, perform DEA
14. Generate final report
15. Introduction to gliomas, IDH status, and the significance of methylation levels.
16. Description of the dataset and preprocessing steps.
17. Methodology for implementing and applying the KNN algorithm.
18. Results of the kNN, including visualizations and evaluation metrics as presented in the paper.
19. Comparison with the findings reported in the target paper. (Do you think we need more clusters based on newer datasets)
20. Conclusion and insights gained from the analysis.