**Latent dimension space in single-cell RNA auto-encoders**

**Supervisor: Anne Hartebrodt**

Autoencoders are popular tools for single-cell RNASeq data analysis. The autoencoder takes as input a single cell expression data set and is trained to reconstruct the input data. Different architectures have been proposed, including standard autoencoders, some variational autoencoders, some are using MSELoss, some using specialized losses designed based on the data properties of single cell data sets.

The project consists of a systematic evaluation of the latent space dimensionality. Based on the result of this analysis, the goal is to give a selection criterion for the hidden dimensionality in gene expression autoencoders. The working theory is that the dimensionality of the latent space should somehow depend on the number of classes. Therefore, a formula should be proposed based on the results.

Required:
- Solid python skills
- Command line / Unix tools
- Any DL framework
- Interest in molecular biology/ single-cell OMICS.

Literature:
- https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04150-3
- https://www.sciencedirect.com/science/article/pii/S2001037023002751
- https://academic.oup.com/bioinformatics/article/36/16/4415/5838187
- https://www.nature.com/articles/s41592-018-0229-2