BIOENG 245
Project Part 1

Catalina Villouta

May 2022

In this project I implemented an autoencoder in TensorFlow that obtains an embedding of cells based on scRNA-seq data. After training the autoencoder, we can recover the encoder with trained parameters which successfully achieves meaningful dimensionality reduction of the original 765 features into 32 latent variables. Given that the sample is of only 700 samples, I added dropout in between the hidden fully connected layers to reduce memorization of the training set and promote generalizable pattern discovery. In Figure 1 I compare different dimensionality reduction techniques including PCA, t-SNE, and the trained encoder. Since the embedding has 32 dimensions I had to further reduce dimensions for plotting purposes. This gives some choices since both PCA and t-SNE can be used to reduce the encoder's output to 2 dimensions. In the figure we can see that both t-SNE applied after PCA and t-SNE applied after the encoder give visually separable groups. This gives confidence that the encoder is finding useful patterns that can be used for predicting the cell type, which will be discussed in Part 2.
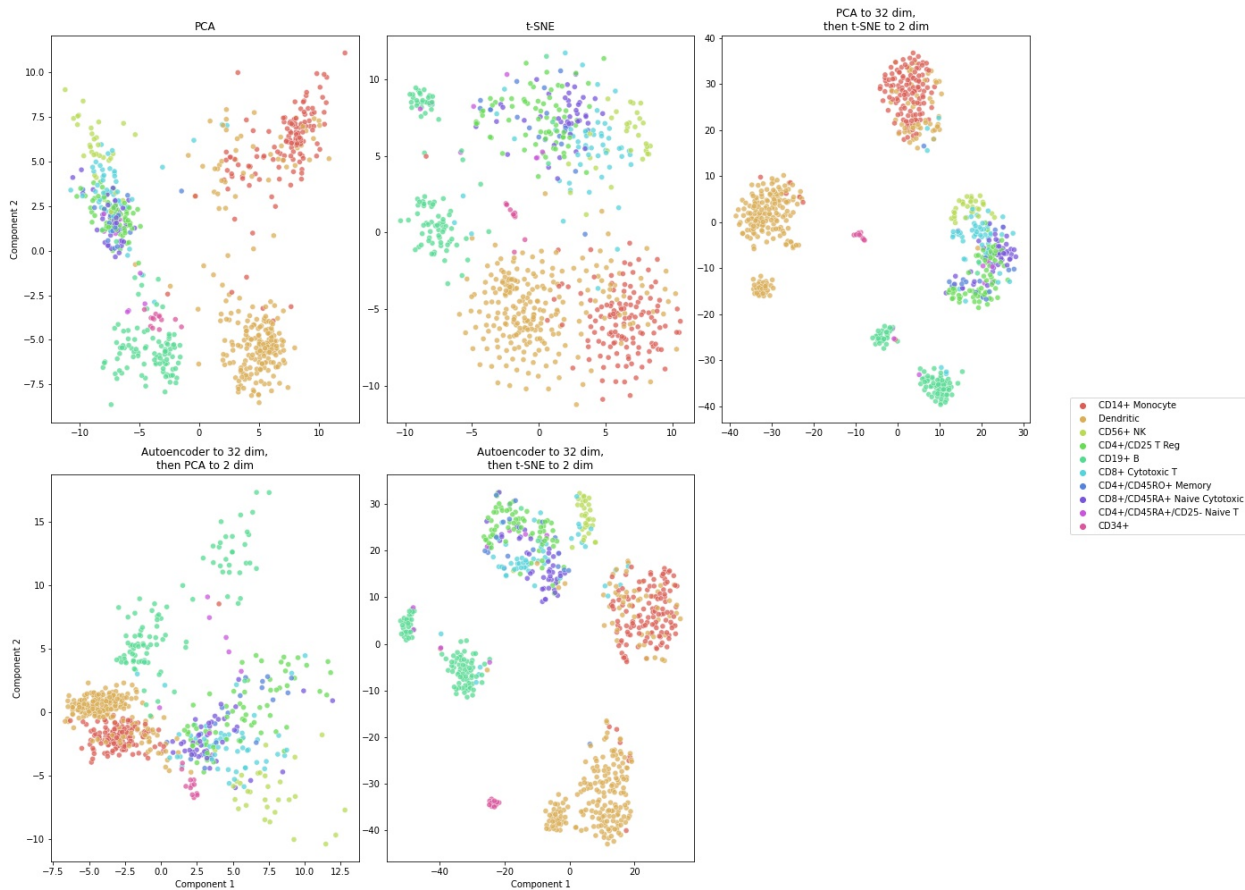

Figure 1: Projections