**ReadMe**

This folder contains the code files of the paper ***“An optimal transport-based low-dimensional visualization framework for high-parameter flow cytometry”.*** In this paper, we build a low dimensional visualization method for high-parameter flow cytometry, and we demonstrate our methodology using two datasets. We mainly focus on a **Malignant Peritoneal Mesothelioma** dataset, but we also work with a comparatively small public dataset of **Acute Myeloid Leukemia** for comparison purposes.

The **Preprocessing** folder has all the code files that were used for preprocessing the raw .fcs files of the MPM dataset and it contains its own **readme** file for detailed description.

Because we analyze three panels (**CoInhibition, CoStimulation, and Cytokine**) from the MPM dataset, we have three separate folders for codes to analyze them separately. Though the basic workflow is the same for all of them, because the cell state markers are different for those three panels, we made separate code files for their analysis.

There are some individual R code files in this folder.

* **MPMUtilities.R** contain all the common computational functions that are called from the panel specific code files.
* **MPMMarkerChannelMap.R** was used to generate panel-specific marker-fluorochrome mapping for the three panels. Those mappings are stored in supporting metadata files used in the codes.
* **MPMDimRedExample.R, and MPMTcellGEDDistribution.R** have codes used for mostly result-summarizing figure generation for the paper.

**Folders: CoInhibition, CoStimulation, Cytokine**

* **MPM…PhenoGraph.R** computes and saves the Sinkhorn distance matrix for the given FC sample, which is to be used for building the visual comparisons.
* **MPM…ComparePlot.R** generates individual and compare visuals for the FC samples.
* **MPM…ALLGEDForMDS.R** and **MPMCoInhibitionGEDSurvival.R** mainly generates plots for the paper that summarizes GED between the three time points per patient and helps correlating them with other biological features known from the paper the dataset came from.

While the main coding is done using R, the optimal transport framework (Sinkhorn distance computation) and Graph Edit Distance computations were done using Python libraries, being called in from R codes. There are two python files in this folder, **TestOTSinkhorn.py and TestGED.py**, that were used for that purpose. Inside **MPMUtilities.R** we do the R-Python integration by setting up a virtual python environment of a python project that has those two Python files. So, remember to do that on your machine.