

The goal of this work was to produce an analysis pipeline to compare distance between pairs of populations over a selection of metrics. A script was written to read in FCS files, calculate the distance between all pairs of populations provided, and generate a heatmap measuring similarity, where 100 represented the exact same population and 0 represented the most dissimilar pairing.

- **A)** A viSNE analysis with two output axes was performed, sampling 5000 events per file across 47 populations. Populations were further downsampled to 1000 events per file and the Earth Mover's Distance (EMD) was calculated between each pair of populations along the viSNE axes to produce the similarity heatmap. The script ordered populations along rows and columns automatically by patient, then by timepoint.
- **B)** A MEM analysis was performed over 140 non-reference populations, with a specified reference population, and the Root Mean Square Distance (RMSD) between each pair of populations was calculated to produce the similarity heatmap. The script automatically detects channels containing phenotype data, and can be easily configured to drop extraneous columns and convert different versions of the same channel name to a single canonical form. The script ordered along rows and columns by hierarchical clustering and produced a dendrogram representing these differences.