**scUnifrac Report**

**Overview**

**scUnifrac is to quantify cell subpopulation diversity between two single-cell transcriptome profiles**, which calculates the distance, estimates the statistical significance of the distance, identifies the subpopulations that are significant different between two profiles, finds genes that mark subpopulations, and predicts cell types of subpopulations. If two single-cell RNA-seq profiles have identical cell subpopulation structures (the same subpopulations with the same proportion), the distance is zero, whereas, the distance is one if completely different subpopulations.

* **Results Summary** - The distance, the statistical significance (p-value), the subpopulations that are different in two profiles, and phylogenetic tree plot and tSNE plot to show the subpopulation structures.
* **Subpopulation view** – heatmap of gene expression that mark the statistical significant subpopulations, and predicted cell types of these subpopulations based on Mouse Cell Atlas [Han et al., 2018, Cell, 1091-1107] if scRNA-seq are derived from mouse samples.

**Results Summary**

**Command**: scUnifrac(data1=colon1,data2=pan1,genenum=500,ncluster=10,nDim=4, samplenames=c("Colon","Pan"))

**Distance= 1**

**p-value= 0**

**Subpopulation structure:**

Table 1. The number of cells in each subpopulation in two scRNA-seq profiles

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subpopulation | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Colon | 411 | 113 | 272 | 213 | 64 | 46 | 28 | 7 | 8 | 0 |
| Pan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 534 |

Table 2. Significantly different subpopulations and their proportion difference between two scRNA-seq profiles

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Subpopulation | 10 | 1 | 3 | 4 | 2 | 5 | 6 |
| Proportion difference | 1 | 0.35 | 0.23 | 0.18 | 0.10 | 0.06 | 0.04 |



Figure 1. Phylogenetic tree plot of subpopulation structures. Cell subpopulations that are enriched in the first scRNA-seq profile labeled as red, while subpopulations enriched in the second profile are labeled as blue. The line width is proportion to the enrichment. Thicker lines suggest more enriched in one profile. Cell subpopulations that are similar in both samples are labeled as black.



Figure 2. tSNE plot of subpopulation structures. Each subpopulation that are different between two profiles are denoted as different colors, and subpopulations that are similar between two profiles are denotes as gray. Cells from the first profile are denoted by circle, while cell from the second profile are denoted by star.

**Subpopulation View**

* **The cell subpopulation 10**
  + **Marker genes**



Figure 3. Heatmap of gene expressions that mark the cell subpopulation

* + **Predicted cell type based on Mouse Cell Atlas**



Figure 4. Predicted cell types of the cell subpopulation

* **The cell subpopulation 1**
  + **Marker genes**

****

Figure 5. Heatmap of gene expressions that mark the cell subpopulation

* + **Predicted cell type based on Mouse Cell Atlas**



Figure 6. Predicted cell types of the cell subpopulation