

S8. Pseudo-code for SPADE analysis

```
1: Gather user input: List of fcs files (i.e. 30 fcs files in the human bone marrow data)
2: Gather user input: List of markers to use (i.e. 13 core surface markers)
3: Gather user input: Outlier density, target density, desired number of clusters
4:
5: % Step 1: density-dependent downsampling
6: for {each fcs file to be analyzed} do
7:   read data matrix from one fcs file, transform data if needed
8:   estimate the median minimum distance between a cell and its nearest neighbor, med_min_dist (randomly pick
     2000 cells, compute the distance of each one to its nearest neighbor, and take the median of these distances)
9:   define dist_threshold = med_min_dist *  $\alpha$  (default  $\alpha = 5$ )
10:  compute local_density for each cell, which is the number of cells within a neighborhood defined by dist_threshold
11:  downsample the data according to Equation (1) in main text
12: end for
13:
14: % Step 2: pool downsampled data together
15: read all the cells in the downsampled data files created above
16: pool all these cells into one single dataset, the “pooled downsampled data”
17:
18: % Step 3: clustering
19: apply the agglomerative clustering method in the main text to group pooled downsampled cells into clusters
20:
21: % Step 4: minimum spanning tree (MST) construction
22: represent each cluster by its center (average intensities of measure markers, averaging across all cells in this cluster)
23: construct the MST that connects the centers, using the algorithm in the main text
24:
25: % Step 5: upsampling
26: for {each original fcs file} do
27:   read data matrix from this original fcs file, transform data if needed
28:   for {each cell in this original fcs file} do
29:     compute its distance to all the cells in the pooled downsampled data
30:     find its nearest neighbor
31:     assign it to the cluster to which the nearest neighbor belongs
32:   end for
33: end for
34:
35: % Step 6: visualization
36: compute the average intensity for each marker, each cluster, and each file
37: display the MST and color-code each node using the average intensity of one marker (i.e. Figures 2b-e and 4 in the
     main text, and Figures S12 and S13 here)
38: display the MST and color-code each node using the ratio between average intensity of one marker across different
     files (i.e. Figure 5 in the main text)
```