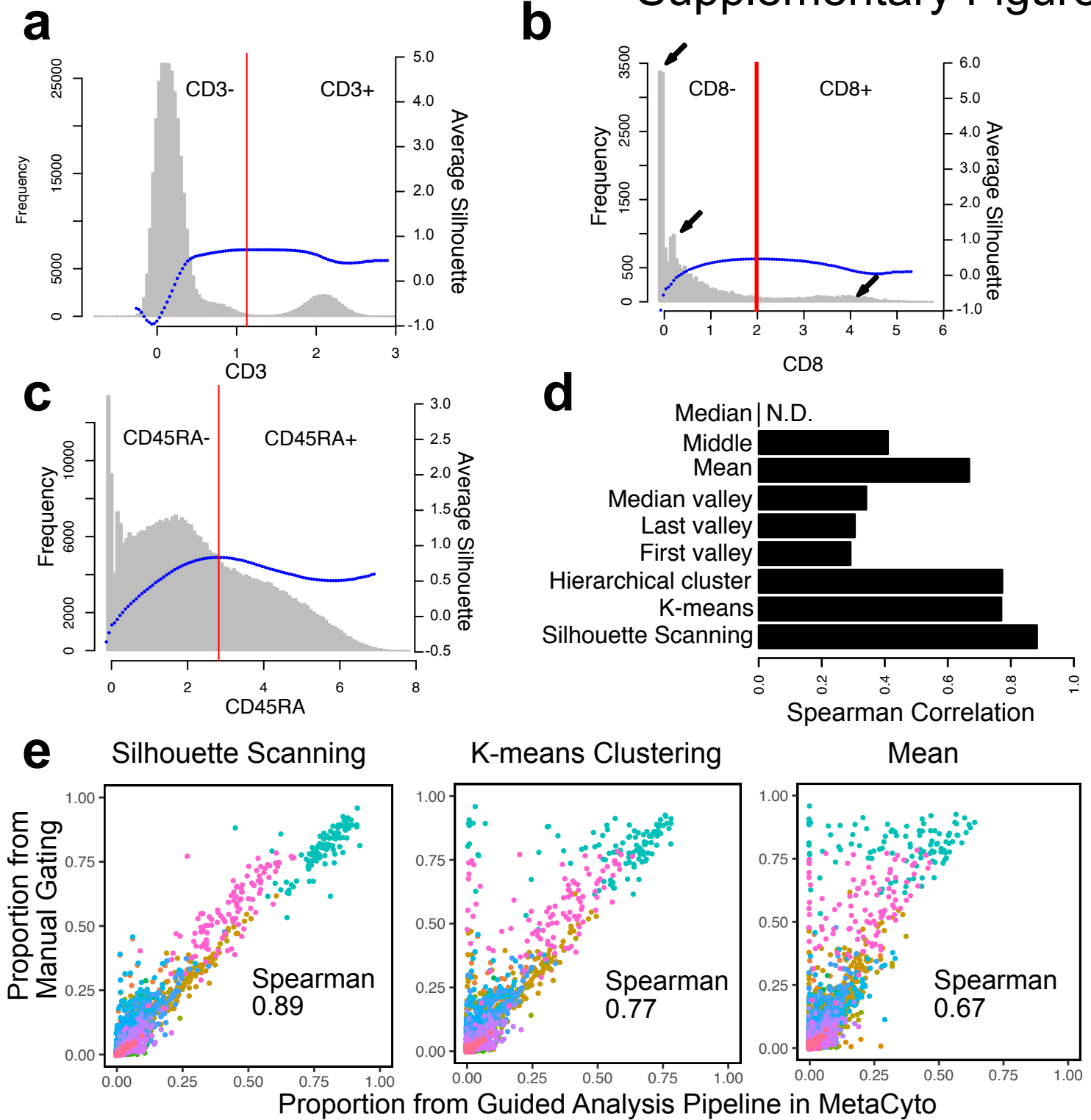


Supplementary Figure 1



Supplementary Figure 1: Silhouette scanning bisects the distribution of each marker in a biologically meaningful way. **(a-c)** Examples illustrating how silhouette scanning bisects markers with bimodal distribution **(a)**, tri-modal distribution **(b)** and a distribution where the positive population does not form a separate peak **(c)**. The range of a marker is divided into 100 intervals using 99 breaks. The distribution is bisected at each break and the corresponding average silhouette is calculated. The break that gives rise to the largest average silhouette is used as the cutoff for bisection. Grey histogram shows the distribution of the markers. Blue dots show the average silhouette at each break. Red line shows the cutoff that maximizes the average silhouette. Black arrows show the position of 3 peaks. **(d-e)** Using different bisection algorithms, each marker in CyTOF data from SDY420 are bisected into positive and negative regions. 24 cell types were identified using the guided analysis pipeline as described in Fig. 1c. The proportion of each cell type in each sample is calculated and compared with manual gating result. **(d)** The Spearman correlation between the estimated proportion and author's proportion are used to measure the performance of each bisection algorithm. **(e)** Scatter plots showing the result generated by using silhouette scanning, k-means clustering and mean as the bisection algorithm. Each dot represents the proportion of a cell type in a sample. Each color represents a cell type. See online methods for a detailed description of the 9 bisection methods tested.