S9 Fig. Comparison between GoM model and hierarchical in terms of power to separate samples from pairs of tissues.
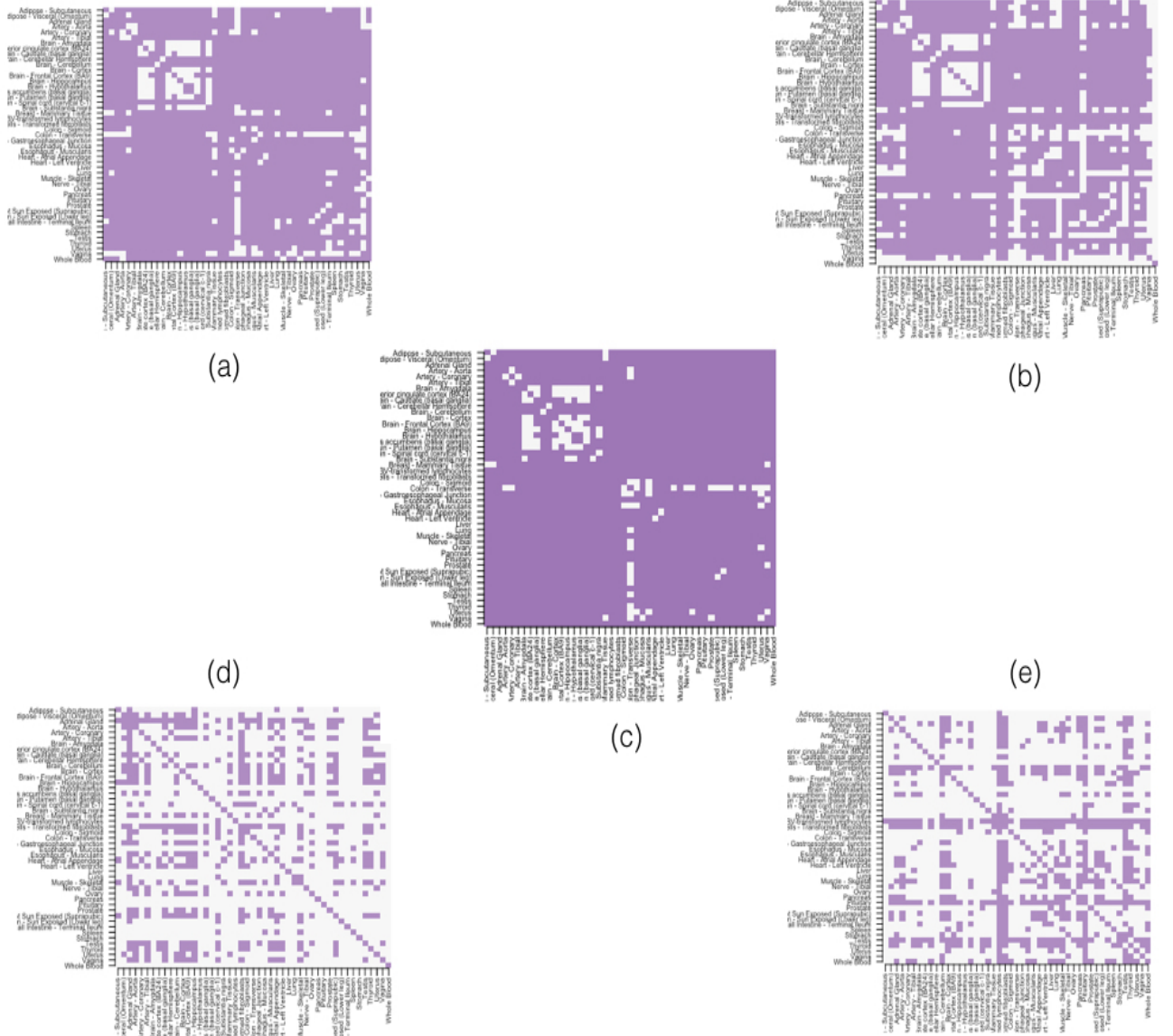
Fig 1. A comparison of accuracy of GoM model vs hierarchical clustering. Image plots to compare the GoM model with 4 different hierarchical clustering models on various transformations of the data. For each pair of tissues from the GTEx data we assessed whether or not each method (with $K = 2$ clusters) separated the samples precisely according to their actual tissue of origin, with successful separation indicated by a filled square. Very clearly, the GoM model seems to be more successful in separating pairs of tissues compared to any of the hierarchical clustering approaches. In SubFig (a), hierarchical clustering was performed on log counts per million (cpm) data using Euclidean distance. In SubFig(b), the log cpm data data was mean and scale transformed for each gene and then the hierarchical clustering was performed on the transformed data using the Euclidean distance. In SubFig (d), the hierarchical clustering was performed on counts data with the assumption the counts c_{ng} for each gene have a variance $\bar{c}_g + 1$, which we used to scale while computing distance matrix. In SubFig (e), we took the same scaled data as in SubFig(c), but we additionally performed mean and scale adjustments further so that all genes have expression of mean 0 and variance 1. In SubFig(c), GoM model is used to separate the tissues. Very clearly, GoM model seems to be performing better than any of the hierarchical methods.