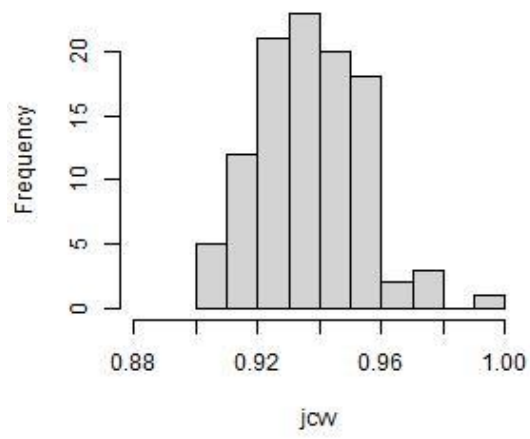
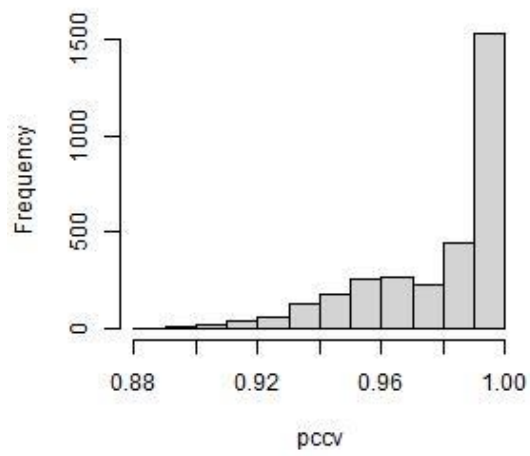


Supplementary Figure 1. Histogram of JC values produced by the GCM, shown in Figure 1 (above), versus histogram of PCC values produced by the WGKS algorithm in Figure 2.

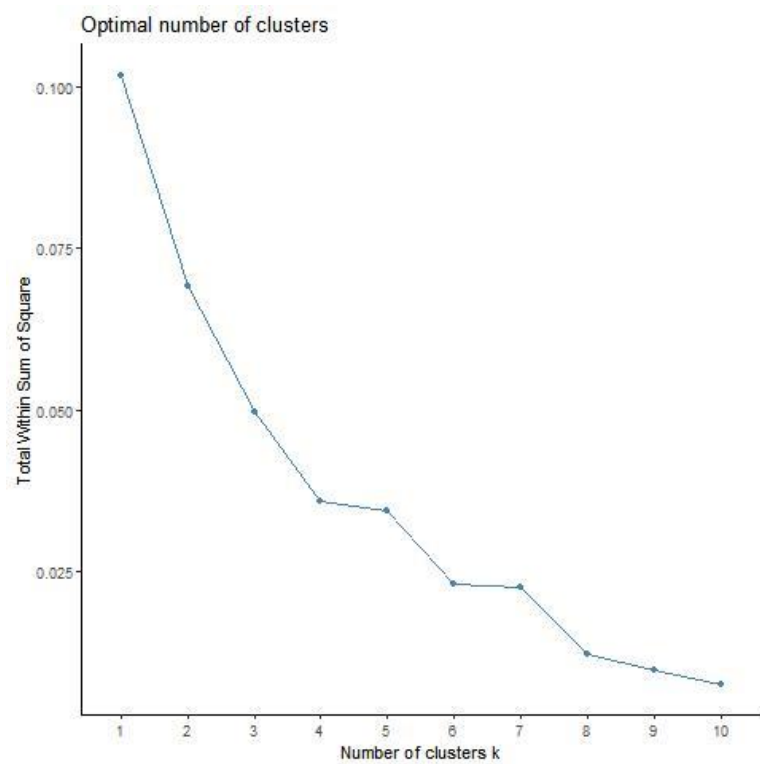
JC values/GCM



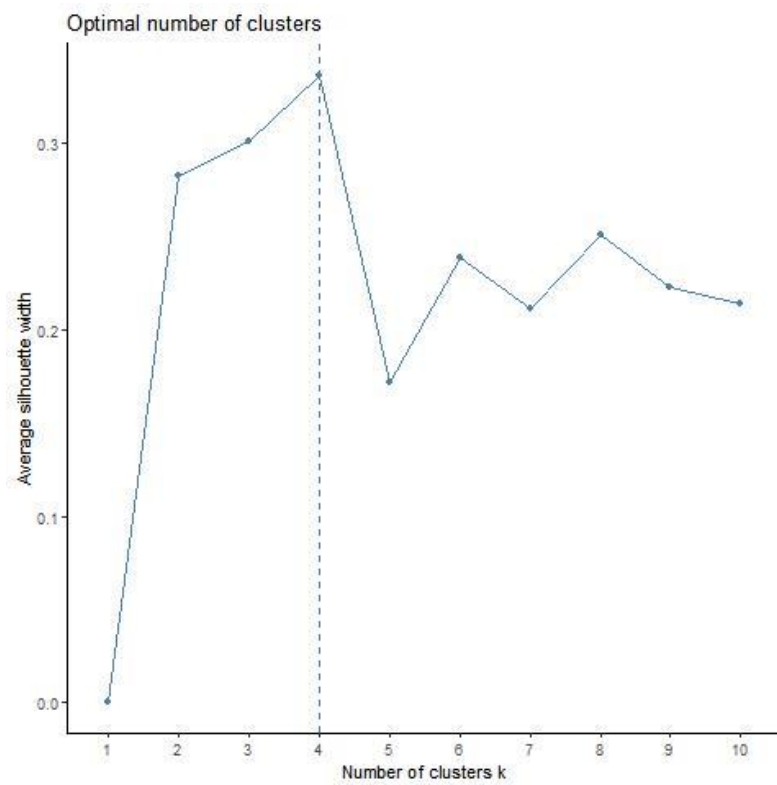
PCC values/WGKS



Supplementary Figure 2. Elbow plot for the Gene Content method showing the total within sum of squares value plotted according to the number of clusters.



Supplementary Figure 3. Silhouette plot showing the average width plotted according to the number of clusters.



Supplementary Table 1. Range of Jaccard Coefficient Values and Pearson Correlation values calculated by the GCM and the WGKS algorithm, depicted in Supplementary Figure 1 above.

	min	mean	mode	max
GCM	0.902	0.937	0.937	1
WGKS	0.882	0.978	0.989	1