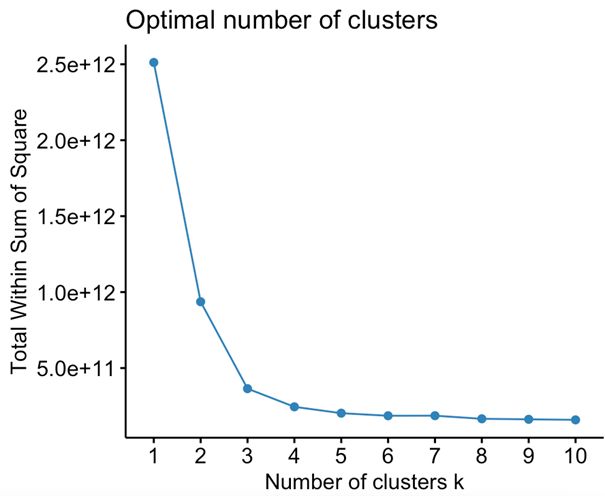
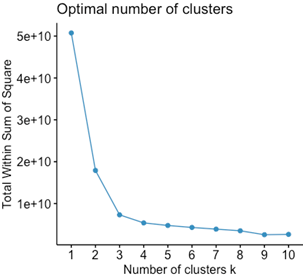
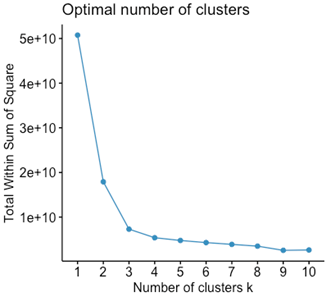
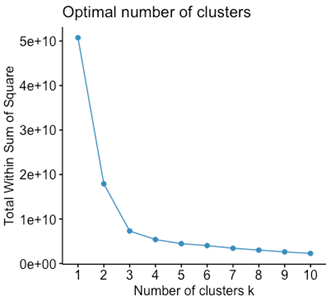
**K-means**

As we plot the total within sum of square vs. the number of clusters, we found that as k reaches 4, the decrease in the curve starts to elbow. Thus, the optimal k should be 4.



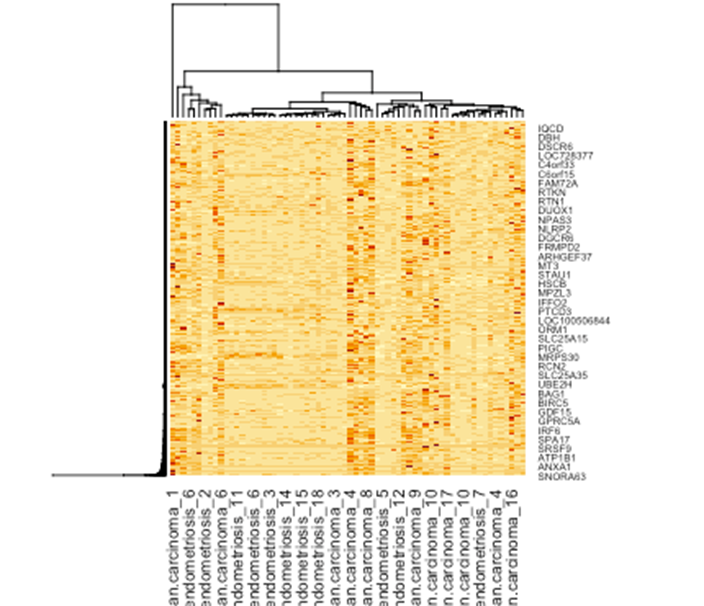
**Different # of genes & effect on clustering**

**Kmeans: n = 100, 1000, 10000**

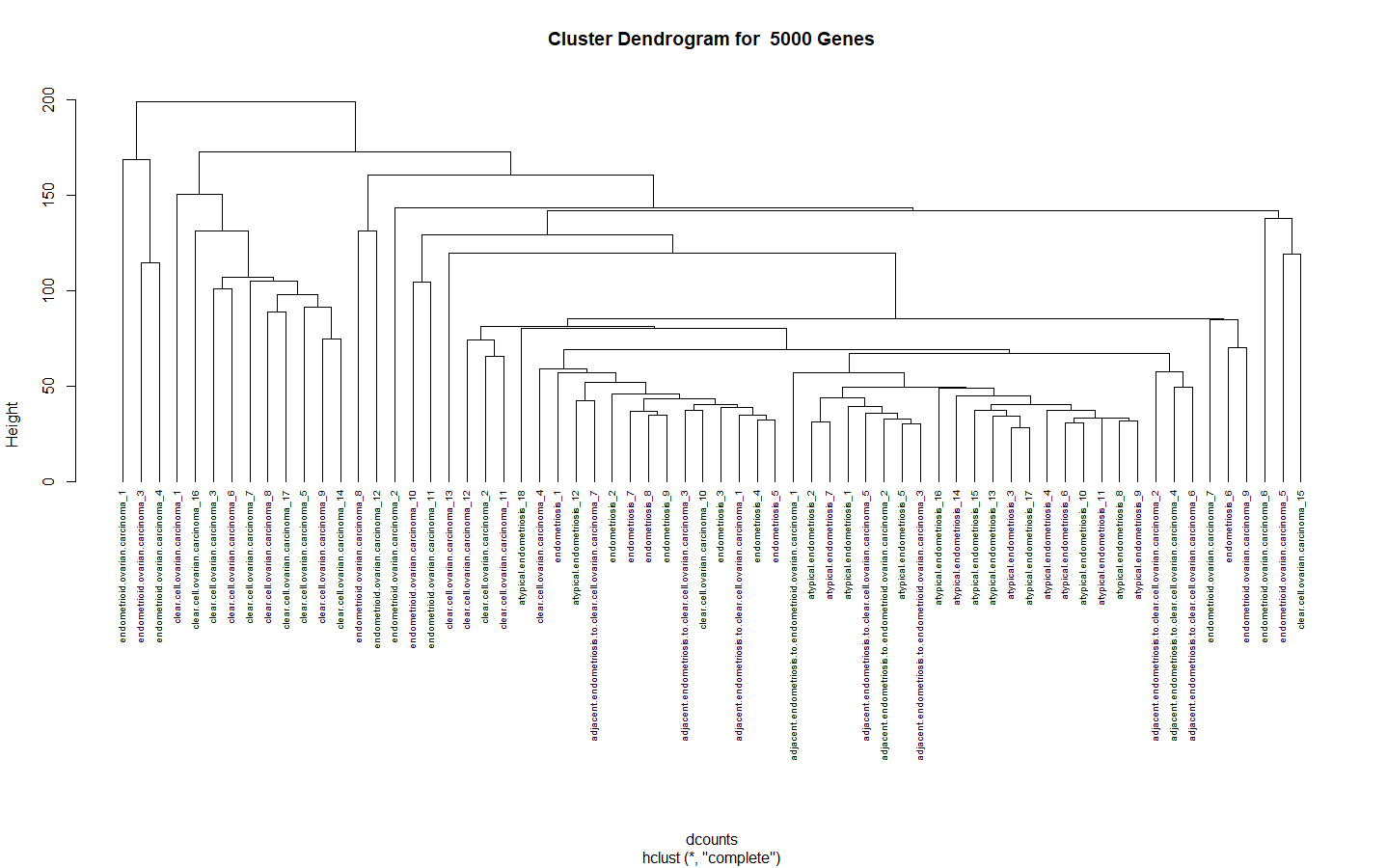


As we increase n, the number of clusters did not change significantly for each clustering method

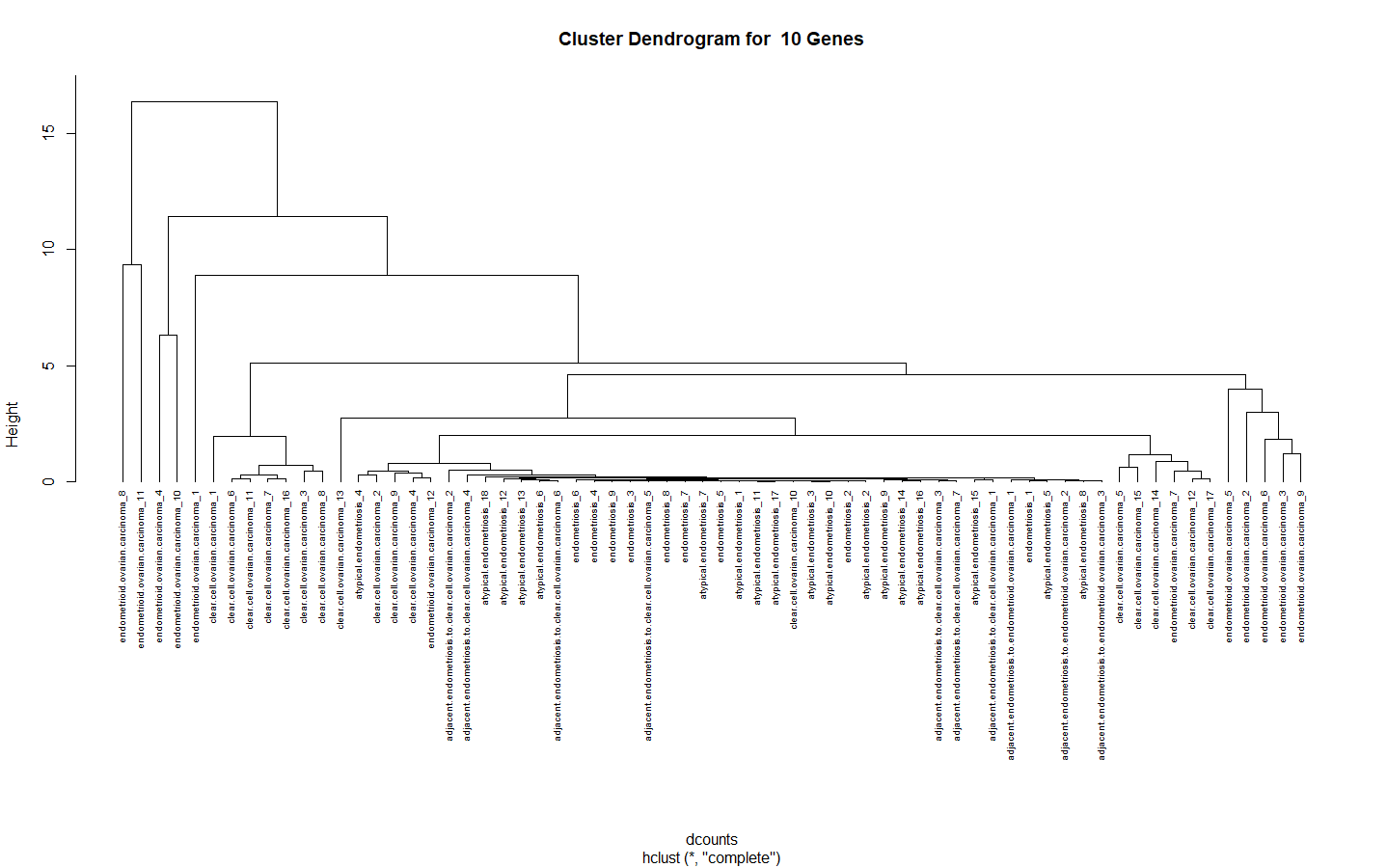
**Gene Heatmap**

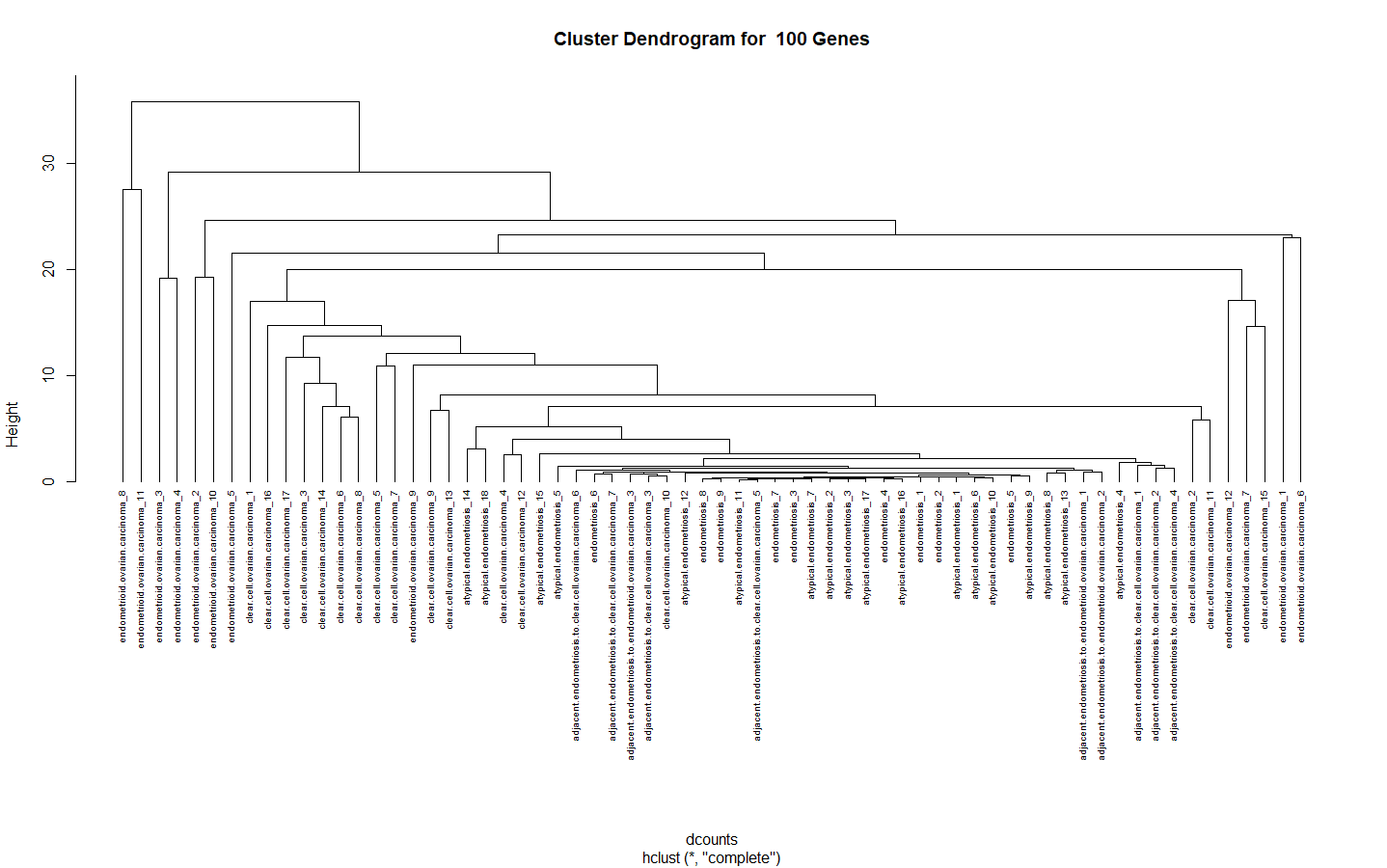
****

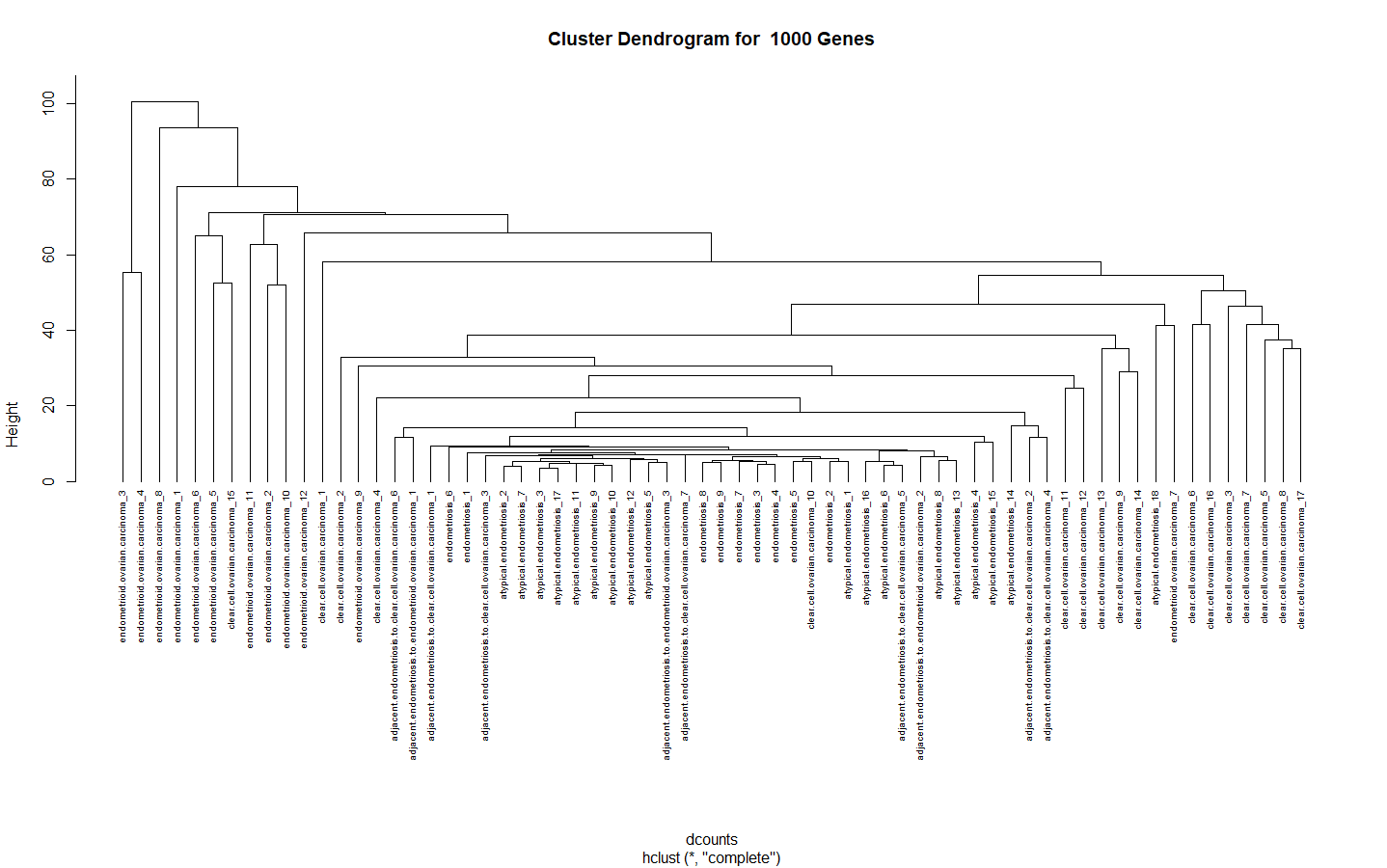
**Hierarchical Clusters**

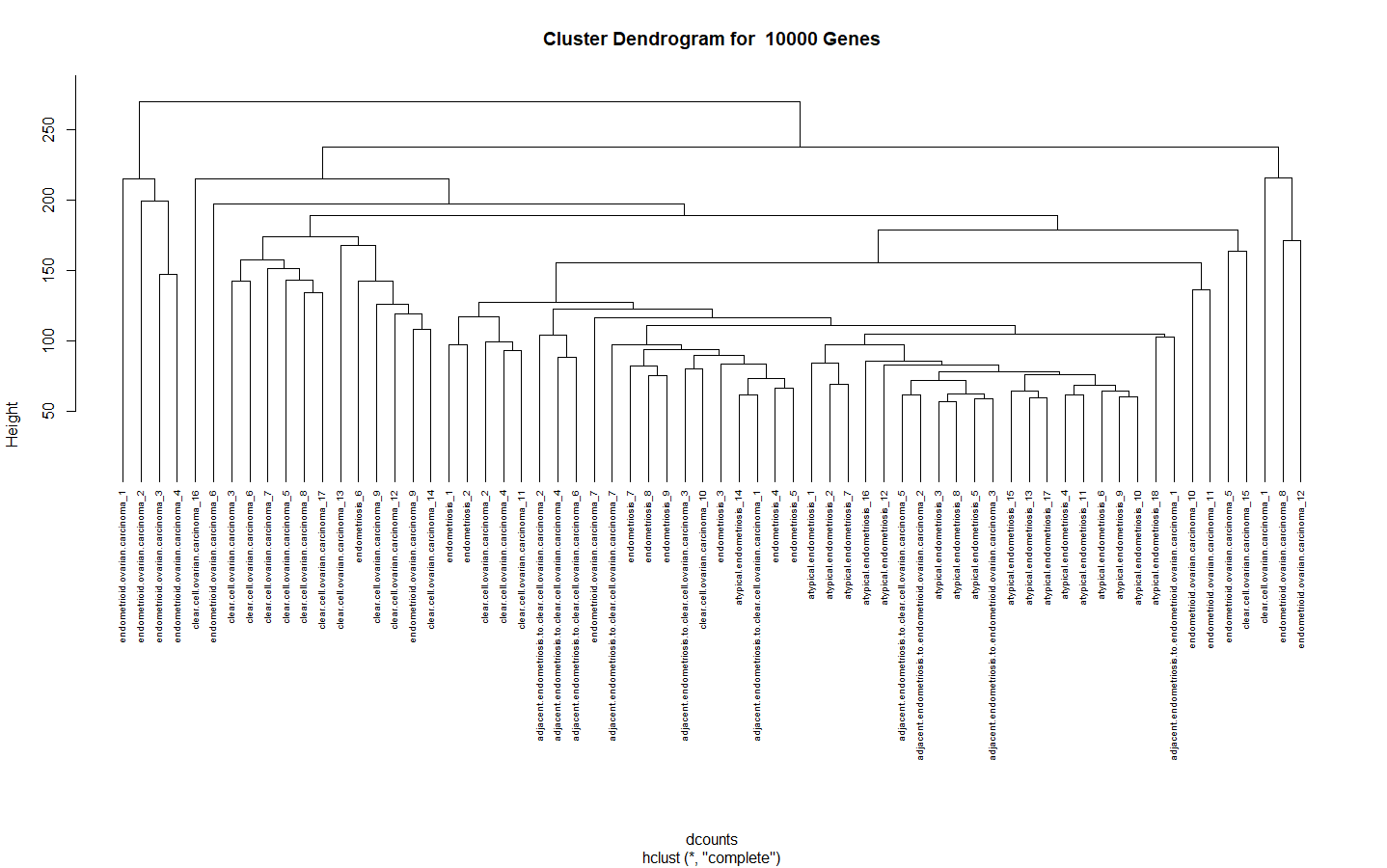


Hierarchical cluster dendrograms as varying gene sample sizes (10, 100, 1000, 10000)









**Pam Clustering**