

## 13. Übung „Angewandte Bioinformatik mit Perl und R“

1. Laden Sie wieder die `golub`-Daten.

- (a) Filtern Sie erneut die Gene mit den Kriterien (a), (b) und (c) aus der Aufgabe 4 der letzten Serie. Benutzen Sie die Funktion `hclust` um die Proben der Golubdaten mit diesen Genen zu clustern. Verwenden Sie unterschiedliche Distanzfunktionen (z.B. aus dem package `bioDist`) und vergleichen Sie die Ergebnisse (jeweils visuell).  
(2 Punkte)
- (b) Wiederholen Sie dies für unterschiedliche Teilmengen signifikanter Gene und wiederum unterschiedliche Distanzen. Vergleichen Sie wiederum die Ergebnisse, was beeinflusst die Ergebnisse am stärksten.  
(2 Punkte)
- (c) Verwenden Sie nun das k-means Verfahren (Funktion `kmeans`) für das Clustern. Vergleichen Sie erneute die Ergebnisse.  
(2 Punkte)
- (d) Testen Sie einige weitere der genannten Clusterverfahren.  
(2 Punkte)

2. Versuchen Sie nun die Gene zu clustern. Überlegen Sie sich sinnvolle Einschränkungen.  
(2 Punkte)