

11. Übung „Angewandte Bioinformatik mit Perl und R“

1. Im Linux-Pool ist Bioconductor unter `/lehre/agprbio/Bioconductor/lib` installiert. Um Pakete mittels der Funktion `library()` von diesen Nicht-Standard-Pfad zu laden, können Sie z.B. in Ihrer shell die Environment-Variable `R_LIBS` entsprechend setzen, in der `tcsh` also

```
setenv R_LIBS /lehre/agprbio/Bioconductor/lib
```

setzen.

Sie können auch in *R* den Aufruf

```
.libPaths("/lehre/agprbio/Bioconductor/lib")
```

verwenden.

Für die folgenden Aufgaben können Sie die Folien des Bioconductor nutzen, diese sind unter www.bioconductor.org und auch auf der Web-Seite der Vorlesung zu finden.

2. Lesen Sie den Dilution Datensatz ein. Wieviele Gene enthält jedes array? Wieviele und welche Gen-Namen enthalten als Teilstring "Lys". Wieviele probes enthalten die zugehörigen probesets, geben Sie textuell deren Intensitäten aus.
3. Erstellen Sie ein Histogramm über die Größe der probesets, d.h. wieviele probesets gibt es für die verschiedenen Größen von probesets. Da eine probeset-Größe sehr oft vorkommt, können wir in diesem Histogramm nicht viel sehen. Überlegen (und realisieren) Sie eine bessere Darstellung Welche Größe kommt am häufigsten vor, wie oft?
Welches/welche Gen/Gene hat/haben die größte Anzahl an probes in ihrem probeset? Plotten Sie die Intensitäten und versuchen Sie möglichst viel Informationen über das/die Gen/Gene zu finden.
4. Plotten Sie für alle Gene mit dem Teilstring "Lys" im Namen die Intensitäten über alle probes, und zwar einmal die perfect matches, dann die mis-matches
5. Wenden Sie mindestens zwei verschiedene Folgen von Hintergrundkorrektur, Normalisierung und Zusammenfassen durch und veranschaulichen Sie die Ergebnisse in Scatterplots und MA-Plots. Diskutieren Sie die Ergebnisse.