

Übungsblatt zu den Vorlesungen A.3 und A.4

Aufgabe 1: Ein bereits aufbereiteter Datensatz: die Lungenkrebs-Genexpressionsdaten

1. Laden Sie den „michigan“ Datensatz aus dem Bioconductor Paket „lungExpression“.
2. Schauen Sie sich die klinischen Variablen an.
3. Extrahieren Sie die Expressionsdaten als Matrix und zeigen Sie die Dimensionen, die ersten Zeilen und Spalten usw.
4. Berechnen Sie die $p \times p$ Korrelationsmatrix und führen für die zwei Gene mit der höchsten Korrelation einen Korrelationstest nach Pearson durch.

Aufgabe 2: Preprocessing von Genexpressionsdaten aus Affymetrix-Microarrays: der Östrogen-Datensatz

1. Laden Sie das Bioconductor Paket „affy“.
2. Laden Sie das Bioconductor Paket „estrogen“, das die Östrogen-Rohdaten enthält. Gehen Sie in das Verzeichnis wo die Östrogen-Daten sich befinden und schauen Sie sich seinen Inhalt an. Machen Sie dieses Verzeichnis zu Ihrem Arbeitsverzeichnis.
3. Lesen Sie die phenotypischen Daten aus der Datei „estrogen.txt“ mithilfe der Funktion `read.table` aus. Nennen Sie das entsprechende Objekt „pd“.
4. Laden Sie mithilfe der Funktion „`ReadAffy`“ die Array-Rohdaten, die den in „estrogen.txt“ beschriebenen Experimenten entsprechen. Als Argument „phenoData“ geben Sie dabei „pd“ an.
5. Schauen Sie sich zur Qualitätskontrolle die Intensitäten des ersten Arrays mithilfe der Funktion „`image`“ an. Laden Sie dann die Array-Rohdaten aus der CEL-Datei „bad.cel“ (Beispiel für Daten schlechter Qualität) und schauen Sie sich für dieses Array die Intensitäten an. Wie interpretieren Sie diese Bilder?
6. Bereiten Sie die Daten mit der RMA Methode vor (Funktion „`rma`“).
7. Bereiten Sie die Daten mit der MAS5 Methode von Affymetrix vor (Funktion „`mas5`“).
8. Bereiten Sie die Daten mit der flexiblen Funktion „`expresso`“ vor, in dem Sie verschiedene Optionen ausprobieren (Argumente `bg.correct`, `normalize.method`, `summary.method`). Siehe dazu die Beschreibung des affy-Paketes für mehr Details:

<http://www.bioconductor.org/packages/release/bioc/vignettes/affy/inst/doc/affy.pdf>

9. Exportieren Sie den erhaltenen Datensatz als txt Datei und schauen Sie sich die erhaltene Datei an.

Aufgabe 3: ProData Datensatz (SELDI-TOF Proteomische Daten)

1. Laden Sie das Paket „ProData“. Schauen Sie sich den Inhalt des Verzeichnisses, das die rohen Daten enthält, an.
2. Lesen Sie das erste Spektrum aus diesem Verzeichnis mithilfe der Funktion `read.csv` aus. Nennen Sie das entsprechende Objekt „spektrum1“.
3. Plotten Sie das erste Spektrum.
4. Laden Sie den Datensatz „f45cbmk“ und extrahieren Sie die phänotypischen Daten zu den 167 Proben. Zeigen Sie die Anzahl der Proben in jeder der vier Gruppen.
5. Extrahieren Sie die aufbereiteten Daten mithilfe der Funktion „`exprs`“.

Für mehr Details zum Processing siehe das Paket PROcess:

<http://www.bioconductor.org/packages/bioc/1.8/vignettes/PROcess/inst/doc/howtoprocess.pdf>