

Prof. Dr. Knut Reinert
Dr. Roland Krause
Matthias Winkelmann
Patrick Pett

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmische Bioinformatik

8. Übungsblatt WS 11/12

Abgabe am 19. Januar 2012

Name:

Matrikelnummer:

Übungsgruppe:

1 2 3

Ich kann Aufgabe nicht vorrechnen.

Zur Durchführung

Alle Aufgaben sollten in R¹ durchgeführt werden. R-Studio² bietet ein einfach zu installierendes GUI.

Aufgabe 1: Lowess-Normalisierung

Ziel der Aufgabe ist es, den Effekt der Lowess-Normalisierung zu bewerten.

- Laden Sie die Daten³ und lesen Sie sie in R ein⁴. Beachten Sie, dass die ersten beiden Spalten Gen- und Reportername enthalten. Die weiteren Spalten sind nach der Array ID benannt und dem entsprechenden Farbkanal (R und G).
- Erstellen Sie für jeden Array einen MvA-Plot inkl. einer Geraden für den Median. Dabei kann es passieren, dass für einzelne Gene ungültige Werte herausbekommt. Warum? Am besten man ignoriert diese einfach. Bewerten Sie die MvA Plots!

¹<http://r-project.org>

²<http://rstudio.org/>

³http://lectures.molgen.mpg.de/Algorithmische_Bioinformatik_WS1011/dokumente/wood_data.txt

⁴`read.delim('wood_data.txt')`

- c) Führen Sie eine LOESS Normalisierung (R Befehl `lowess`) für alle Arrays durch. Experimentieren Sie mit der Fenstergröße. Wie macht sich der Einfluss bemerkbar? Stellen Sie sowohl die nichtnormalisierten Daten mit der Regressionslinie als auch die normalisierten Daten dar.
- d) *Freiwilliger Zusatz: Verwenden Sie verschiedene Kernel-Functions in der Lowess-Normalisierung und interpretieren Sie das Ergebnis.*

Aufgabe 2: Quantilnormalisierung

Eine Möglichkeit zur Normalisierung von Daten ist die Quantil-Normalisierung.

- a) Ziehen Sie 1000 Proben einer Normalverteilung mit Erwartungswert 1 und Varianz 1, sowie 1000 Proben aus einer γ -Verteilung mit shape $k = 2$ und scale $\theta = 1$.
- b) Erzeugen Sie QQ-Plots gegen eine Standard-Normalverteilung und der beiden Verteilungen gegeneinander. Was beobachten Sie und warum?
- c) Benutzen Sie die Quantil-Normalisierung, um beide Verteilungen miteinander zu normalisieren. Erstellen Sie wiederum einen QQ-plot gegen eine Standard-Normalverteilung. Was wird beobachtet ?
- d) Was kann man daraus für eine Gefahr dieser Normalisierung schliessen? Warum macht es keinen Sinn, die beiden normalisierten Daten gegeneinander mit einem QQ-Plot zu plotten? Was würden Sie beobachten?

Aufgabe 3: Explorative Analyse und Normalisierung mit vsn

In der Analyse von Microarrays gehen explorative Datenanalyse und Wahl der Normalisierung Hand in Hand. Das ist der Inhalt dieser Aufgabe. Die Aufgabe sollte in R mit installiertem `Biobase` ausgeführt werden, da die meisten Funktionen bereits implementiert sind.

A Benutzen Sie den `lymphoma` Datensatz, um eine kurze explorative Datenanalyse durchzuführen. Der Datensatz ist Teil des `vsn` Pakets und wird folgendermaßen eingelesen:

```
library(vsn);data(lymphoma);e <- exprs(lymphoma)
```

Dabei enthält `e` am Ende die Expressionsmatrix. Es genügt, wenn 4 Chips analysiert werden. Für die explorative Analyse sind folgende Werkzeuge hilfreich, die auf die logarithmierten Daten (Gene mit ungültigen Daten wieder entfernen!) angewandt werden sollten:

- a) Verteilung der Expressionswerte pro Chip (boxplot, qq-plot)
- b) Scatterplot: Führen Sie dies sowohl für die logarithmisch transformierten Daten wie auch für die noch nicht transformierten Daten durch und vergleichen Sie.
- c) Matrix of Absolute Deviation (MAD)
- d) Plot der Standardabweichung gegen den Mittelwert (sdm plots), dafür eignet sich der Befehl `meanSdPlot`. Führe dies sowohl für die logarithmisch transformierten Daten wie auch für die noch nicht transformierten Daten durch und vergleiche.
- e) *Vorgriff für ganz Interessierte: Korrespondenzanalyse*

B Normalisierung und Vergleich

- a) Was ist die wichtigste Annahme, um eine Normalisierung z.B. mit `vsn` durchzuführen? Warum ist das wichtig?
- b) Verwenden Sie `vsn` zur Normalisierung der Daten (keine Hintergrundkorrektur und Summarizing mit *median polish*). Das muss man nicht selbst implementieren, sondern man kann einfach das Paket `vsn` in R benutzen.
- c) Vergleichen Sie die transformierten Daten mit den originalen Daten. Wie soll der Vergleich aussehen? Boxplots, Scatterplots zwischen den normalisierten Daten und diese Plots vergleichen mit den Rohdaten und den logarithmierten Daten. Gleiches auch für *sdm* Plots.