Übungsblatt 7  
Genexpressionsanalyse  
Deadline: 22.6.2022 um 20:00 MEZ  
Bioinformatik für Biochemiestudierende  
Dr. Florian Klimm  
Sommersemester 2022

**Aufgabe 7.1 Quantil-Normalisierung (30 %)**  
**Gegeben seien folgende Expressionsdaten von Genen (G1, G2, . . . G7) in drei  
verschiedenen Experimenten (A, B, und C):**  
  
 Table

Description automatically generated  
**Normalisieren Sie per Hand die Experimente mit Hilfe der Quantil-Normalisierung,  
wie sie in der Vorlesung vorgestellt wurde.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 22 | 20 | 18 | 19 | 19 | 23 | 17.5 |
| 2 | 1 | 0.5 | 1 | 2.5 | 1.5 | 0.5 |
| 3 | 1 | 0 | 5 | 1 | 7 | 0 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 2 | 1 | 0 | 1 | 1 | 1.5 | 0 |
| 3 | 1 | 0.5 | 5 | 2.5 | 7 | 0.5 |
| 22 | 20 | 18 | 19 | 19 | 23 | 17.5 |

|  |
| --- |
| 0.93 |
| 2.79 |
| 19.79 |

|  |
| --- |
| i |
| ii |
| iii |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| iii | iii | iii | iii | iii | iii | iii |
| i | i | ii | i | ii | i | ii |
| ii | ii | i | ii | i | ii | i |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 19.79 | 19.79 | 19.79 | 19.79 | 19.79 | 19.79 | 19.79 |
| 0.93 | 0.93 | 2.79 | 0.93 | 2.79 | 0.93 | 2.79 |
| 2.79 | 2.79 | 0.93 | 2.79 | 0.93 | 2.79 | 0.93 |

**Aufgabe 7.2 Student t-Test (30 %)**

**Sie wollen die Genexpression eines bestimmten Genes zwischen zwei Patienten-  
gruppen vergleichen. Gruppe A ist an einer Krankheit erkrankt. Gruppe B ist nicht  
an der Krankheit erkrankt. Die Genexpressionswerte der beiden Gruppen sind:**

A picture containing text

Description automatically generated

**Bestimmen Sie mit Hilfe des t-Test ob sich die Expression des Genes signifikant an  
der Grenze 0.05 zwischen den beiden Patientengruppen unterschiedet. Interpretie-  
ren Sie ihr Ergebnis.**

Hinweis: Die Werte der t-Verteilung finden Sie zum Beispiel unter https://www.  
statistik.tu-dortmund.de/fileadmin/user\_upload/Lehrstuehle/Oekonometrie/  
Lehre/WiSoOekoSS17/tabelletV.pdf

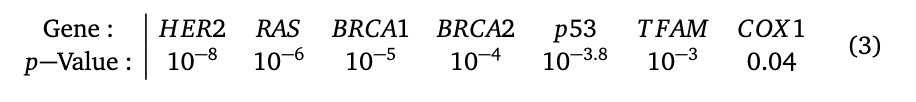
Mit der Varianz

Hier:

Vergleich des Ergebnisses mit der t-Verteilung auf dem 5% Niveau, unter Benutzung der Definition der Freiheitsgrade:

🡪 Die Genexpression zwischen den beiden Patientenpopulationen ist signifikant unterschiedlich

**Aufgabe 7.3 Multiple-testing Korrektur (10 %)**  
**Sie haben Genexpressionsdaten zweier Gewebe verglichen. Mit einer Signifikanz-  
schwelle von 0.05 haben Sie die folgenden Gene identifiziert und unkorrigierte  
p-Werte erhalten**



**Insgesamt haben sie 100 Gene getestet. Warum ist eine multiple-testing Korrektur  
notwendig? Wenden Sie die Bonferroni-Korrektur an. Welche der Gene sind nach  
der Korrektur signifikant? Haben Sie eine Vermutung was die Gewebe unterschei-  
det?**

Kern des Problems der multiplen Vergleiche ist, dass mit steigender Anzahl von Vergleichen (hier: Gene) die Wahrscheinlichkeit ansteigt, dass sich die Daten unterscheiden und falsch positive Ergebnisse (hier: DEGs) erlangt werden.

🡪 Was ist die Wahrscheinlichkeit, mind. 1 falsch positives DEG zu bekommen, wenn n Gene getestet werden?

daraus folgt

Hier:

Wie sieht der Graph dazu aus?

Chart, line chart

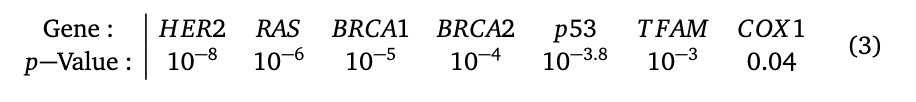
Description automatically generated

🡪 die Funktion nähert sich y=1 an

🡪 je mehr Gene getestet werden, desto größer ist die Wahrscheinlichkeit, ein falsch positives Ergebnis zu bekommen

Um für diese Fehler-Kumulierung zu korrigieren, kann die Bonferroni-Korrektur eingesetzt werden. wird durch die Anzahl der multiplen Tests (hier: 100) geteilt, sodass diese Korrektur den p-Wert verringert.

In dem gegebenen Beispiel gelten demnach nur noch solche Gene als signifikant unterschiedlich exprimiert, wenn der p-Wert kleiner ist als .



Demnach sind die Gene HER2, RAS, BRCA1, BRCA2 und p53 signifikant reguliert. Bei diesen Genen handelt es sich um Tumorsuppressoren oder Proto-Onkogene, weshalb sich vermuten lässt, dass es sich bei dem einen Gewebe um entartete Zellen handeln könnte.

**Aufgabe 7.4 Computergestützte Genexpressionsanalyse (30 %)**  
Nutzen Sie den Code in DESEQ2\_EXAMPLE.R um eine Genexpressionsanalyse der  
Daten in DF\_SC.CSV durchzuführen. Beachten Sie, dass Sie die auskommentierten  
Zeilen nur einmal zum Installieren des DESeq2 Packets ausführen müssen.  
Verwenden Sie eine Signifikanzschwelle von 0.01 und beantworten Sie die folgen-  
den Fragen.  
1. Wieviele Gene sind in ihrer Expression erhöht?  
2. Beschreiben Sie wie sich der Volcono Plot verändert wenn Sie die Signifi-  
kanzschwelle verändern.  
3. Welches sind die 3 Gene die in ihrer Expression am stärksten erhöht bzw.  
unterdrückt sind?  
4. Ab welcher Signifikanschwelle gibt es keine signifikanten Gene mehr?

Ich habe diese Aufgabe wegen der Fehlermeldungen nicht lösen können. Das Problem lag vermutlich an der fehlgeschlagenen Installation des Packages „DESeq2“. Allerdings habe ich online keine gute Lösung für die Fehlermeldung finden können. Hier die gesamte Meldung:

> if (!require("BiocManager", quietly = TRUE))

+ BiocManager::install(version = "3.15")

> install.packages("BiocManager")

trying URL 'https://cran.rstudio.com/bin/macosx/big-sur-arm64/contrib/4.1/BiocManager\_1.30.18.tgz'

Content type 'application/x-gzip' length 357729 bytes (349 KB)

==================================================

downloaded 349 KB

The downloaded binary packages are in

/var/folders/lc/zgw605bs7t13yfsqx74hp6hm0000gn/T//Rtmpn7HRI2/downloaded\_packages

> BiocManager::install("DESeq2")

'getOption("repos")' replaces Bioconductor standard repositories, see '?repositories' for details

replacement repositories:

CRAN: https://cran.rstudio.com/

Bioconductor version 3.14 (BiocManager 1.30.18), R 4.1.3 (2022-03-10)

Installing package(s) 'DESeq2'

also installing the dependencies ‘Biostrings’, ‘zlibbioc’, ‘KEGGREST’, ‘GenomeInfoDb’, ‘XVector’, ‘DelayedArray’, ‘AnnotationDbi’, ‘annotate’, ‘S4Vectors’, ‘IRanges’, ‘GenomicRanges’, ‘SummarizedExperiment’, ‘Biobase’, ‘BiocParallel’, ‘genefilter’, ‘geneplotter’

Warning: unable to access index for repository https://bioconductor.org/packages/3.14/bioc/bin/macosx/big-sur-arm64/contrib/4.1:

cannot open URL 'https://bioconductor.org/packages/3.14/bioc/bin/macosx/big-sur-arm64/contrib/4.1/PACKAGES'

Warning: unable to access index for repository https://bioconductor.org/packages/3.14/data/annotation/bin/macosx/big-sur-arm64/contrib/4.1:

cannot open URL 'https://bioconductor.org/packages/3.14/data/annotation/bin/macosx/big-sur-arm64/contrib/4.1/PACKAGES'

Warning: unable to access index for repository https://bioconductor.org/packages/3.14/data/experiment/bin/macosx/big-sur-arm64/contrib/4.1:

cannot open URL 'https://bioconductor.org/packages/3.14/data/experiment/bin/macosx/big-sur-arm64/contrib/4.1/PACKAGES'

Warning: unable to access index for repository https://bioconductor.org/packages/3.14/workflows/bin/macosx/big-sur-arm64/contrib/4.1:

cannot open URL 'https://bioconductor.org/packages/3.14/workflows/bin/macosx/big-sur-arm64/contrib/4.1/PACKAGES'

Warning: unable to access index for repository https://bioconductor.org/packages/3.14/books/bin/macosx/big-sur-arm64/contrib/4.1:

cannot open URL 'https://bioconductor.org/packages/3.14/books/bin/macosx/big-sur-arm64/contrib/4.1/PACKAGES'

Packages which are only available in source form, and may need compilation of C/C++/Fortran: ‘Biostrings’ ‘zlibbioc’

‘XVector’ ‘DelayedArray’ ‘S4Vectors’ ‘IRanges’ ‘GenomicRanges’ ‘Biobase’ ‘BiocParallel’ ‘genefilter’ ‘DESeq2’

Do you want to attempt to install these from sources? (Yes/no/cancel) Yes

installing the source packages ‘Biostrings’, ‘zlibbioc’, ‘KEGGREST’, ‘GenomeInfoDb’, ‘XVector’, ‘DelayedArray’, ‘AnnotationDbi’, ‘annotate’, ‘S4Vectors’, ‘IRanges’, ‘GenomicRanges’, ‘SummarizedExperiment’, ‘Biobase’, ‘BiocParallel’, ‘genefilter’, ‘geneplotter’, ‘DESeq2’

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/Biostrings\_2.62.0.tar.gz'

Content type 'application/x-gzip' length 12645827 bytes (12.1 MB)

==================================================

downloaded 12.1 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/zlibbioc\_1.40.0.tar.gz'

Content type 'application/x-gzip' length 257409 bytes (251 KB)

==================================================

downloaded 251 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/KEGGREST\_1.34.0.tar.gz'

Content type 'application/x-gzip' length 25778 bytes (25 KB)

==================================================

downloaded 25 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/GenomeInfoDb\_1.30.1.tar.gz'

Content type 'application/x-gzip' length 3457147 bytes (3.3 MB)

==================================================

downloaded 3.3 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/XVector\_0.34.0.tar.gz'

Content type 'application/x-gzip' length 67783 bytes (66 KB)

==================================================

downloaded 66 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/DelayedArray\_0.20.0.tar.gz'

Content type 'application/x-gzip' length 676428 bytes (660 KB)

==================================================

downloaded 660 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/AnnotationDbi\_1.56.2.tar.gz'

Content type 'application/x-gzip' length 4330396 bytes (4.1 MB)

==================================================

downloaded 4.1 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/annotate\_1.72.0.tar.gz'

Content type 'application/x-gzip' length 1723060 bytes (1.6 MB)

==================================================

downloaded 1.6 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/S4Vectors\_0.32.4.tar.gz'

Content type 'application/x-gzip' length 668558 bytes (652 KB)

==================================================

downloaded 652 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/IRanges\_2.28.0.tar.gz'

Content type 'application/x-gzip' length 455647 bytes (444 KB)

==================================================

downloaded 444 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/GenomicRanges\_1.46.1.tar.gz'

Content type 'application/x-gzip' length 1171939 bytes (1.1 MB)

==================================================

downloaded 1.1 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/SummarizedExperiment\_1.24.0.tar.gz'

Content type 'application/x-gzip' length 1524667 bytes (1.5 MB)

==================================================

downloaded 1.5 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/Biobase\_2.54.0.tar.gz'

Content type 'application/x-gzip' length 1641111 bytes (1.6 MB)

==================================================

downloaded 1.6 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/BiocParallel\_1.28.3.tar.gz'

Content type 'application/x-gzip' length 883999 bytes (863 KB)

==================================================

downloaded 863 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/genefilter\_1.76.0.tar.gz'

Content type 'application/x-gzip' length 807639 bytes (788 KB)

==================================================

downloaded 788 KB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/geneplotter\_1.72.0.tar.gz'

Content type 'application/x-gzip' length 1437803 bytes (1.4 MB)

==================================================

downloaded 1.4 MB

trying URL 'https://bioconductor.org/packages/3.14/bioc/src/contrib/DESeq2\_1.34.0.tar.gz'

Content type 'application/x-gzip' length 2088431 bytes (2.0 MB)

==================================================

downloaded 2.0 MB

\* installing \*source\* package ‘zlibbioc’ ...

\*\* using staged installation

configure: creating ./config.status

config.status: creating src/Makevars

\*\* libs

xcode-select: note: no developer tools were found at '/Applications/Xcode.app', requesting install. Choose an option in the dialog to download the command line developer tools.

ERROR: compilation failed for package ‘zlibbioc’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/zlibbioc’

\* installing \*source\* package ‘S4Vectors’ ...

\*\* using staged installation

\*\* libs

xcode-select: note: no developer tools were found at '/Applications/Xcode.app', requesting install. Choose an option in the dialog to download the command line developer tools.

ERROR: compilation failed for package ‘S4Vectors’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/S4Vectors’

\* installing \*source\* package ‘Biobase’ ...

\*\* using staged installation

\*\* libs

xcode-select: note: no developer tools were found at '/Applications/Xcode.app', requesting install. Choose an option in the dialog to download the command line developer tools.

ERROR: compilation failed for package ‘Biobase’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/Biobase’

\* installing \*source\* package ‘BiocParallel’ ...

\*\* using staged installation

checking for gcc... gcc

checking whether the C compiler works... no

configure: error: in `/private/var/folders/lc/zgw605bs7t13yfsqx74hp6hm0000gn/T/RtmpH1sBpM/R.INSTALL1123789f3d93/BiocParallel':

configure: error: C compiler cannot create executables

See `config.log' for more details

ERROR: configuration failed for package ‘BiocParallel’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/BiocParallel’

ERROR: dependency ‘S4Vectors’ is not available for package ‘IRanges’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/IRanges’

ERROR: dependencies ‘S4Vectors’, ‘IRanges’ are not available for package ‘GenomeInfoDb’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/GenomeInfoDb’

ERROR: dependencies ‘S4Vectors’, ‘IRanges’, ‘zlibbioc’ are not available for package ‘XVector’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/XVector’

ERROR: dependencies ‘S4Vectors’, ‘IRanges’ are not available for package ‘DelayedArray’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/DelayedArray’

ERROR: dependencies ‘S4Vectors’, ‘IRanges’, ‘XVector’, ‘GenomeInfoDb’ are not available for package ‘Biostrings’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/Biostrings’

ERROR: dependencies ‘S4Vectors’, ‘IRanges’, ‘GenomeInfoDb’, ‘XVector’ are not available for package ‘GenomicRanges’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/GenomicRanges’

ERROR: dependency ‘Biostrings’ is not available for package ‘KEGGREST’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/KEGGREST’

ERROR: dependencies ‘GenomicRanges’, ‘Biobase’, ‘S4Vectors’, ‘IRanges’, ‘GenomeInfoDb’, ‘DelayedArray’ are not available for package ‘SummarizedExperiment’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/SummarizedExperiment’

ERROR: dependencies ‘Biobase’, ‘IRanges’, ‘S4Vectors’, ‘KEGGREST’ are not available for package ‘AnnotationDbi’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/AnnotationDbi’

ERROR: dependencies ‘AnnotationDbi’, ‘Biobase’ are not available for package ‘annotate’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/annotate’

ERROR: dependencies ‘AnnotationDbi’, ‘annotate’, ‘Biobase’ are not available for package ‘genefilter’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/genefilter’

ERROR: dependencies ‘Biobase’, ‘annotate’, ‘AnnotationDbi’ are not available for package ‘geneplotter’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/geneplotter’

ERROR: dependencies ‘S4Vectors’, ‘IRanges’, ‘GenomicRanges’, ‘SummarizedExperiment’, ‘Biobase’, ‘BiocParallel’, ‘genefilter’, ‘geneplotter’ are not available for package ‘DESeq2’

\* removing ‘/Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/DESeq2’

The downloaded source packages are in

‘/private/var/folders/lc/zgw605bs7t13yfsqx74hp6hm0000gn/T/Rtmpn7HRI2/downloaded\_packages’

There were 17 warnings (use warnings() to see them)

> library( "DESeq2" )

Error in library("DESeq2") : there is no package called ‘DESeq2’