

Analysis of a correlation between EHMT expression and drug sensitivity in human PDO lines

Florian Heigwer

2021-06-08

Abstract

This document contains code to analyze a micro-array experiment performed by Johannes Bedge and colleagues at the University Clinic Mannheim Bedge Group. The data comprises multiple Affymetrix chips performed by the DKFZ core facility. Purpose is to find if EHMT1/2 expression correlates with the PDO's sensitivity to HDAC inhibitor treatments. Data was produced in replicates sampling each donor derived PDO line in 2 different passages.

Contents

1	Dependencies.	2
1.1	B110 Theme	2
1.2	annotation and data	4
2	Analyze correlations	7
3	EHMT2 expression drug sensitivity linear dependence test. . .	7
4	EHMT2 expression drug sensitivity linear dependence test. . .	8
5	Session info	8
6	Bibliography	10

1 Dependencies

We load a number of packages whose functions are needed throughout the analysis

```
library(limma)
library(ggplot2)
library(patchwork)
library(fgsea)
library("hgu133plus2.db")
library(tidyverse)
library(ggplots)
library(RColorBrewer)
library(eulerr)
library(oligo)
library(arrayQualityMetrics)
library(ggsignif)
library(ggpubr)
```

#Lets define some themes

This we do as a quality of life step to give all figures that we produce a common look and feel.

1.1 B110 Theme

#Read in data

Data is used as provided by the DKFZ core-facility. Data is given as CEL raw data files, next to be quantile normalized per experiment already log-transformed and summarized by averaging on gene level. Gene - Level View is presented.

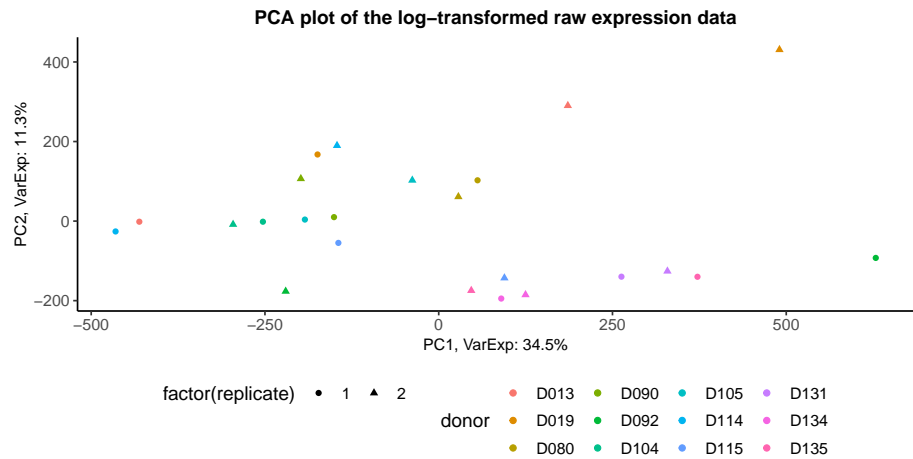
Cel files were collected and meta-data derived from files names.

#QC

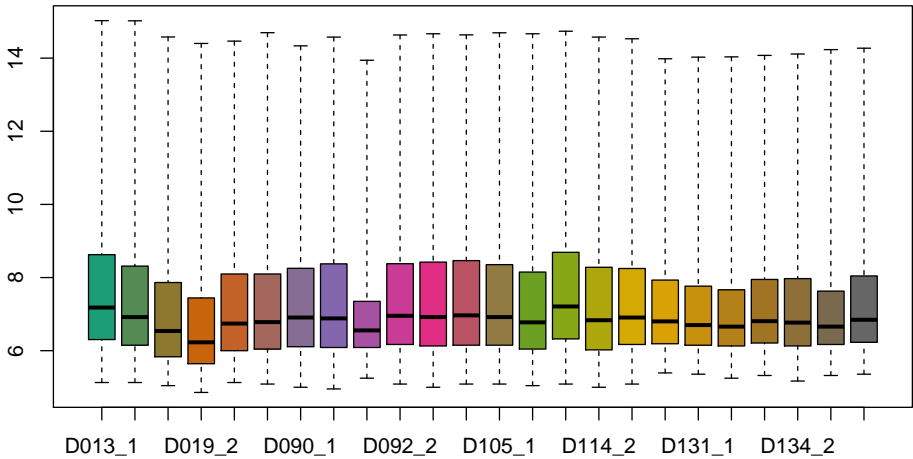
#inspect the data

```
##   D013_1 D019_1 D013_2 D019_2 D080_1
## 1    216    147    113     96     96
## 2  10741   9914   5608   4995   7393
## 3    200    238    175    122    123
## 4  10601  10100   5576   5062   7389
## 5    267    273    190     94    127
```

Analysis of a correlation between EHMT expression and drug sensitivity in human PDO lines



Boxplot of log2-intensities for the raw data

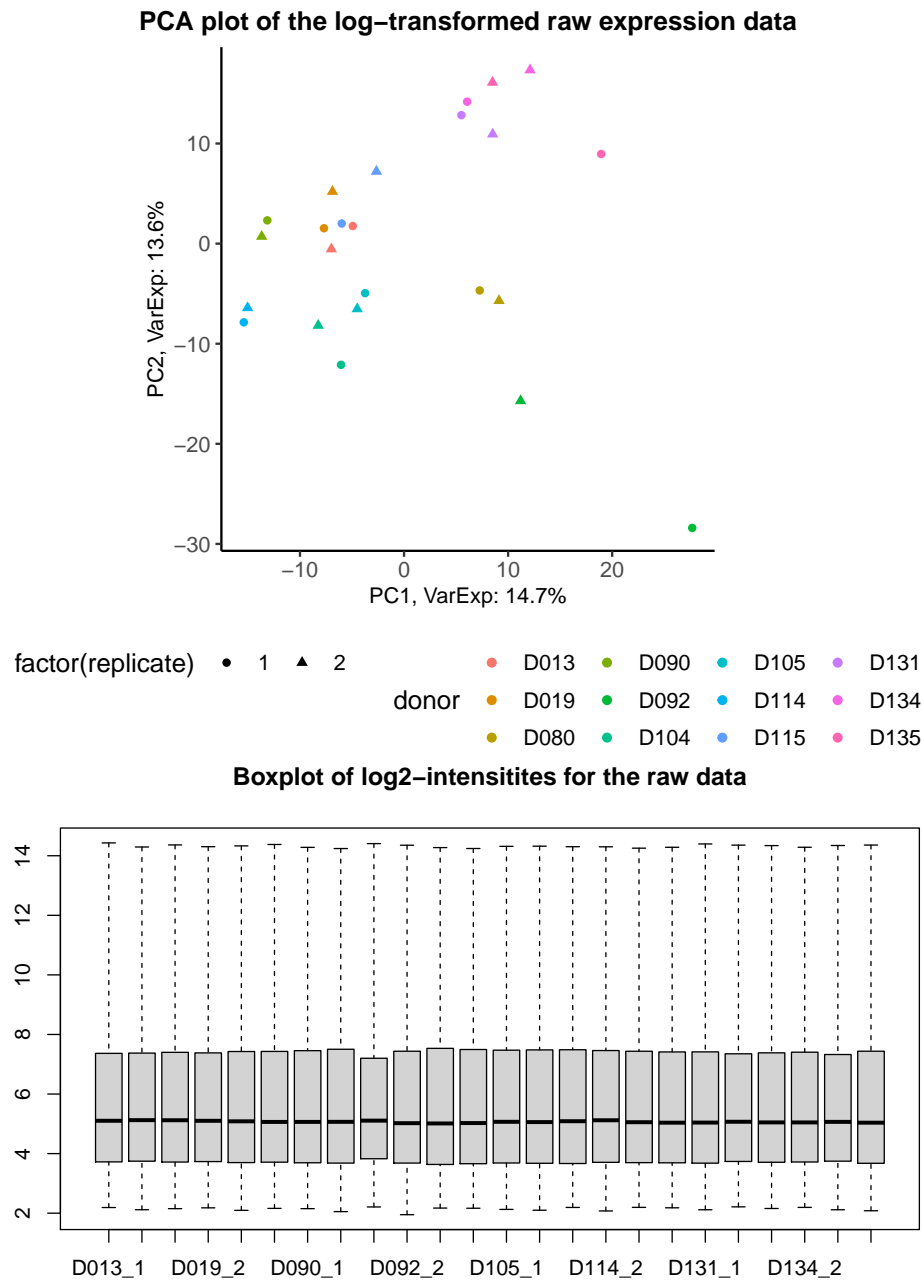


#Normalization

```
norm_data <- oligo::rma(raw_data)
## Background correcting
## Normalizing
## Calculating Expression
```

#inspect the normed data

##	D013_1	D019_1	D013_2	D019_2	D080_1
## 1007_s_at	10.206202	10.258460	10.149864	10.160429	10.470672
## 1053_at	8.884593	9.790716	8.827939	9.632787	7.456577
## 117_at	4.764373	4.796270	4.985429	5.153804	4.985915
## 121_at	8.712680	8.379036	8.587126	8.245918	8.055976
## 1255_g_at	2.968861	2.969611	2.880114	3.041247	2.901949



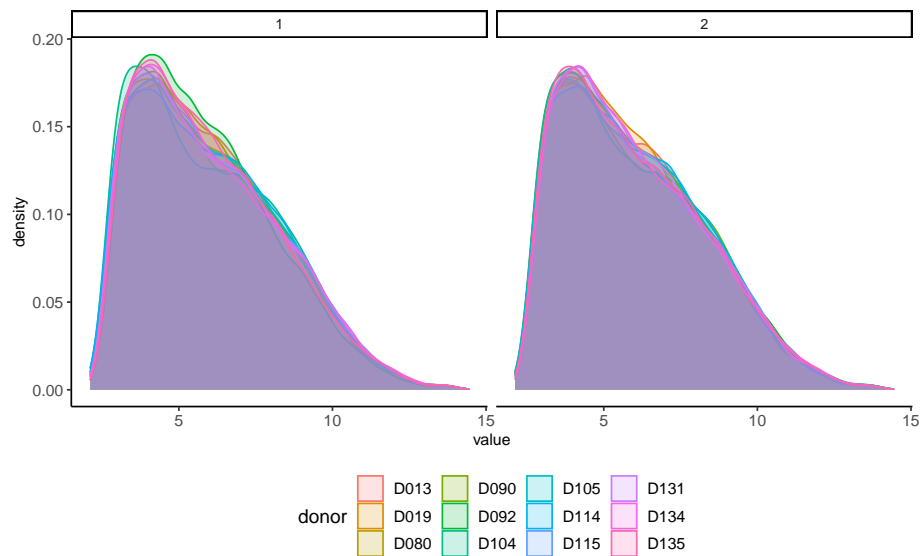
1.2 annotation and data

We assume the data is indeed already normalized and non-log transformed. So we do the log-transform of the expression data here.

##Overview density plots

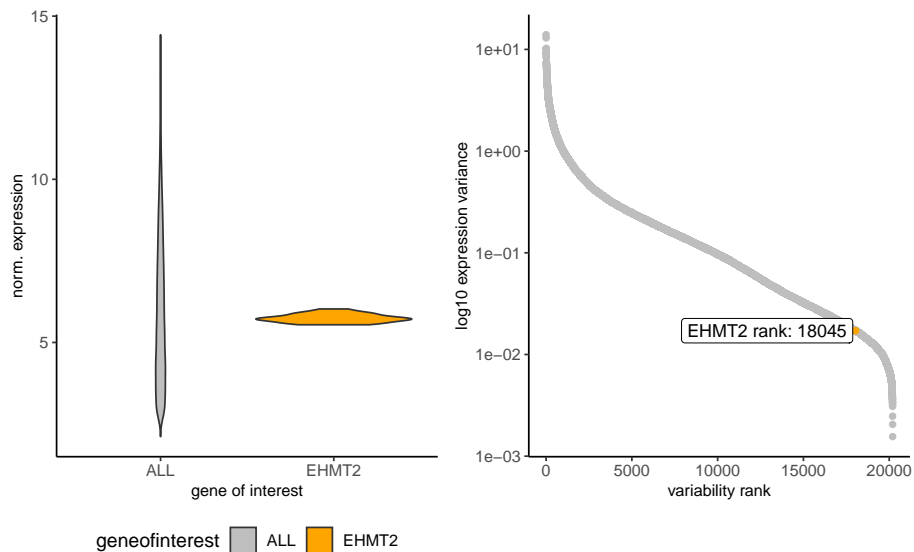
So we vizualize a common density plot.

Analysis of a correlation between EHMT expression and drug sensitivity in human PDO lines



#EHMT2 expression variance in human colorectal cancer oganoids

```
## # A tibble: 1 x 3
##   `FALSE` `TRUE` percentVar
##   <dbl>   <dbl>   <dbl>
## 1    0.258 0.0171    6.64
```



#EHMT2 expression variance compared to all genes

```
## # A tibble: 1 x 3
##   `FALSE` `TRUE` perc_range
##   <dbl>   <dbl>   <dbl>
## 1    5.22 0.0171    0.328
```

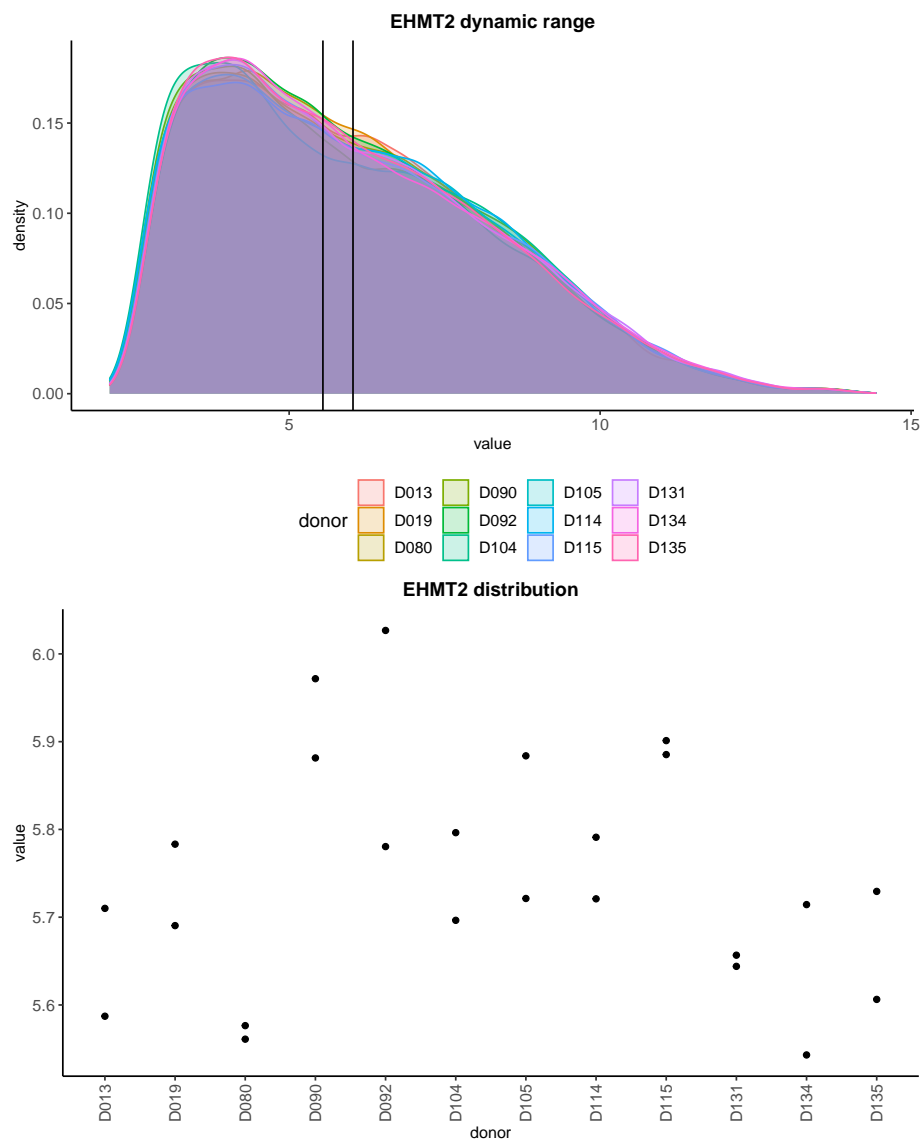
#EHMT2 expression range compared to all genes

Analysis of a correlation between EHMT expression and drug sensitivity in human PDO lines

```
## # A tibble: 1 x 3
##   `FALSE` `TRUE` perc_range
##   <dbl>   <dbl>   <dbl>
## 1    12.3    0.484    3.93
```

Indeed the data seems sufficiently normalized as all distributions looks very similar and overlap to the extend expected from normalized data.

Next we check for EHMT expression of the different donors and get the drug treatment result data.



`#get treatment data`

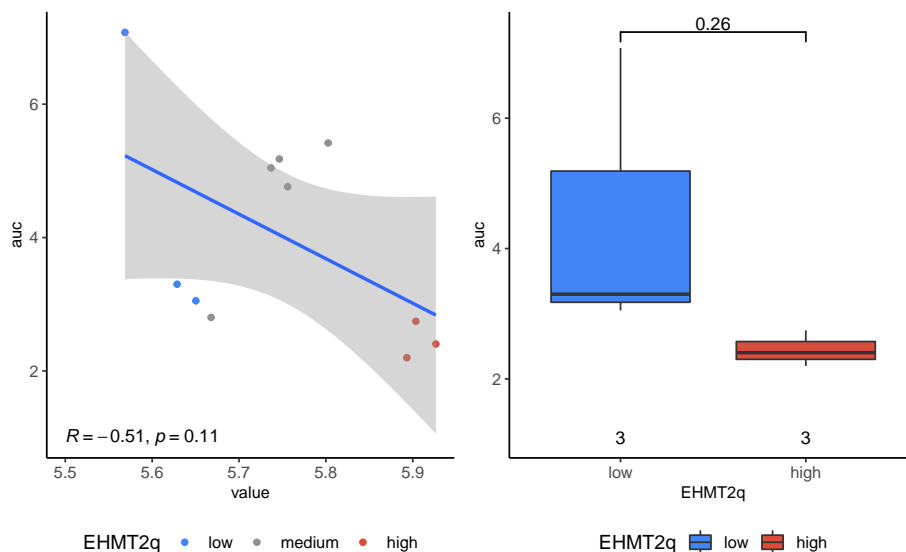
I extracted this data from Source Data main FiguresV2.xlsx (Fig7) and reformatted it to be amchine readable.

Analysis of a correlation between EHMT expression and drug sensitivity in human PDO lines

```
##  
## -- Column specification -----  
## cols(  
##   replicate = col_double(),  
##   vorinostat_uM = col_double(),  
##   response = col_double(),  
##   donor = col_character()  
## )  
## `summarise()` has grouped output by 'donor'. You can override using the `.groups` argument.
```

2 Analyze correlations

```
## Joining, by = "donor"  
## Warning: `fun.ymin` is deprecated. Use `fun.min` instead.  
## `geom_smooth()` using formula 'y ~ x'  
## `geom_smooth()` using formula 'y ~ x'
```



3 EHMT2 expression drug sensitivity linear dependence test

```
##  
## Call:  
## lm(formula = auc ~ EHMT2, data = .)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -1.7631 -1.1942 -0.2467  1.0385  1.8487   
##  
## Coefficients:
```

Analysis of a correlation between EHMT expression and drug sensitivity in human PDO lines

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  42.431      21.785   1.948  0.0833 .
## EHMT2        -6.681       3.786  -1.765  0.1115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.429 on 9 degrees of freedom
## Multiple R-squared:  0.257, Adjusted R-squared:  0.1745
## F-statistic: 3.114 on 1 and 9 DF, p-value: 0.1115
```

4 EHMT2 expression drug sensitivity linear dependence test

```
##
## Welch Two Sample t-test
##
## data: auc by EHMT2q
## t = 1.5453, df = 2.0597, p-value = 0.2588
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.463574  7.518174
## sample estimates:
## mean in group low mean in group high
##           4.476083           2.448783
```

5 Session info

```
writeLines(capture.output(sessionInfo()), "results/SessionInfo.txt")
sessionInfo()
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] parallel stats4 stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
## [1] pd.hg.u133.plus.2.3.12.0 DBI_1.1.1
```


Analysis of a correlation between EHMT expression and drug sensitivity in human PDO lines

```
## [3] RSQLite_2.2.5          ggpubr_0.4.0.999
## [5] ggsignif_0.6.1         arrayQualityMetrics_3.46.0
## [7] oligo_1.54.1           Biostrings_2.58.0
## [9] XVector_0.30.0         oligoClasses_1.52.0
## [11] eulerr_6.1.0           RColorBrewer_1.1-2
## [13] gplots_3.1.1           forcats_0.5.1
## [15] stringr_1.4.0          dplyr_1.0.5
## [17] purrr_0.3.4            readr_1.4.0
## [19] tidyr_1.1.3            tibble_3.1.0
## [21] tidyverse_1.3.0        hgu133plus2.db_3.2.3
## [23] org.Hs.eg.db_3.12.0    AnnotationDbi_1.52.0
## [25] IRanges_2.24.1         S4Vectors_0.28.1
## [27] Biobase_2.50.0         BiocGenerics_0.36.0
## [29] fgsea_1.16.0           patchwork_1.1.1
## [31] ggrepel_0.9.1          ggplot2_3.3.3
## [33] limma_3.46.0           BiocStyle_2.18.1
##
## loaded via a namespace (and not attached):
## [1] readxl_1.3.1           backports_1.2.1
## [3] Hmisc_4.5-0            fastmatch_1.1-0
## [5] systemfonts_1.0.1      plyr_1.8.6
## [7] splines_4.0.3          BiocParallel_1.24.1
## [9] GenomeInfoDb_1.26.4    digest_0.6.27
## [11] foreach_1.5.1          htmltools_0.5.1.1
## [13] fansi_0.4.2            magrittr_2.0.1
## [15] checkmate_2.0.0        memoise_2.0.0
## [17] affyPLM_1.66.0         cluster_2.1.1
## [19] gcrma_2.62.0           openxlsx_4.2.3
## [21] annotate_1.68.0        modelr_0.1.8
## [23] matrixStats_0.58.0     beadarray_2.40.0
## [25] svglite_2.0.0          askpass_1.1
## [27] jpeg_0.1-8.1           colorspace_2.0-0
## [29] blob_1.2.1             rvest_1.0.0
## [31] haven_2.3.1            xfun_0.22
## [33] hexbin_1.28.2          crayon_1.4.1
## [35] RCurl_1.98-1.3         jsonlite_1.7.2
## [37] genefilter_1.72.1      survival_3.2-10
## [39] iterators_1.0.13       glue_1.4.2
## [41] gtable_0.3.0           zlibbioc_1.36.0
## [43] DelayedArray_0.16.3    BeadDataPackR_1.42.0
## [45] car_3.0-10             abind_1.4-5
## [47] scales_1.1.1           setRNG_2013.9-1
## [49] vsn_3.58.0            rstatix_0.7.0
## [51] Rcpp_1.0.6            xtable_1.8-4
## [53] htmlTable_2.1.0        foreign_0.8-81
## [55] bit_4.0.4              preprocessCore_1.52.1
## [57] Formula_1.2-4          htmlwidgets_1.5.3
## [59] httr_1.4.2            ellipsis_0.3.1
## [61] ff_4.0.4              farver_2.1.0
## [63] pkgconfig_2.0.3        XML_3.99-0.6
## [65] nnet_7.3-15           dbplyr_2.1.0
```

```
## [67] utf8_1.2.1           labeling_0.4.2
## [69] reshape2_1.4.4       tidyselect_1.1.0
## [71] rlang_0.4.10         munsell_0.5.0
## [73] cellranger_1.1.0     tools_4.0.3
## [75] cachem_1.0.4         cli_2.4.0
## [77] generics_0.1.0       broom_0.7.6
## [79] evaluate_0.14        fastmap_1.1.0
## [81] yaml_2.2.1          knitr_1.31
## [83] bit64_4.0.5         fs_1.5.0
## [85] zip_2.1.1           caTools_1.18.2
## [87] nlme_3.1-152        xml2_1.3.2
## [89] compiler_4.0.3      rstudioapi_0.13
## [91] curl_4.3            png_0.1-7
## [93] affyio_1.60.0       reprex_2.0.0
## [95] stringi_1.5.3       lattice_0.20-41
## [97] Matrix_1.3-2        vctrs_0.3.7
## [99] pillar_1.5.1        lifecycle_1.0.0
## [101] BiocManager_1.30.12 data.table_1.14.0
## [103] bitops_1.0-6        GenomicRanges_1.42.0
## [105] R6_2.5.0            latticeExtra_0.6-29
## [107] affy_1.68.0         hwriter_1.3.2
## [109] bookdown_0.21       rio_0.5.26
## [111] KernSmooth_2.23-18  gridSVG_1.7-2
## [113] gridExtra_2.3       affxparser_1.62.0
## [115] codetools_0.2-18    gtools_3.8.2
## [117] assertthat_0.2.1    SummarizedExperiment_1.20.0
## [119] openssl_1.4.3       withr_2.4.1
## [121] GenomeInfoDbData_1.2.4 mgcv_1.8-34
## [123] hms_1.0.0           grid_4.0.3
## [125] rpart_4.1-15        base64_2.0
## [127] rmarkdown_2.7       carData_3.0-4
## [129] illuminaio_0.32.0   MatrixGenerics_1.2.1
## [131] lubridate_1.7.10    base64enc_0.1-3
```

6 Bibliography
