

R-programming of ALL(B-cell)

Stephen Omitogun

February 15, 2016

The library containing the ALL(acute lymphoblastic leukemia) data was downloaded from bioconductor.org through some commands. `biocLite("ALL")` `biocLite("genefilter")`

To open the ALL dataset, the following commands are:

```
library(Biobase)
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
```

```
##
```

```
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
##   clusterExport, clusterMap, parApply, parCapply, parLapply,  
##   parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##   IQR, mad, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##   anyDuplicated, append, as.data.frame, as.vector, cbind,  
##   colnames, do.call, duplicated, eval, evalq, Filter, Find, get,  
##   grep, grepl, intersect, is.unsorted, lapply, lengths, Map,  
##   mapply, match, mget, order, paste, pmax, pmax.int, pmin,  
##   pmin.int, Position, rank, rbind, Reduce, rownames, sapply,  
##   setdiff, sort, table, tapply, union, unique, unlist, unsplit
```

```
## Welcome to Bioconductor
```

```
##
```

```
##   Vignettes contain introductory material; view with  
##   'browseVignettes()'. To cite Bioconductor, see  
##   'citation("Biobase)"', and for packages 'citation("pkgname)"'.
```

```
library(ALL)
```

```
data(ALL)
```

Investigating the ALL dataset

```
class(ALL)
```

```
## [1] "ExpressionSet"  
## attr(,"package")  
## [1] "Biobase"
```

```
ALL1 <- data.frame(ALL)
```

```
dim(ALL)
```

```
## Features Samples  
##      12625      128
```

```
str(ALL)
```

Formal class 'ExpressionSet' [package "Biobase"] with 7 slots ..@ experimentData :Formal class 'MIAME' [package "Biobase"] with 13 slots@ name : chr "Chiaretti et al."@ lab : chr "Department of Medical Oncology, Dana-Farber Cancer Institute, Department of Medicine, Brigham and Women's Hospital, Harvard Med"| **truncated**@ contact : chr ""@ title : chr "Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different respo"| **truncated**@ abstract : chr "Gene expression profiles were examined in 33 adult patients with T-cell acute lymphocytic leukemia (T-ALL). Nonspecific filteri"| **truncated**@ url : chr ""@ pubMedIds : chr [1:2] "14684422" "16243790"@ samples : list()@ hybridizations : list()@ normControls : list()@ preprocessing : list()@ other : list()@ .classVersion:Formal class 'Versions' [package "Biobase"] with 1 slot@ .Data:List of 1@ .\$. : int [1:3] 1 0 0 ..@ assayData : ..@ phenoData :Formal class 'AnnotatedDataFrame' [package "Biobase"] with 4 slots@ varMetadata : 'data.frame': 21 obs. of 1 variable:@ .\$. labelDescription: chr [1:21] " Patient ID" " Date of diagnosis" " Gender of the patient" " Age of the patient at entry"@ data : 'data.frame': 128 obs. of 21 variables:@ .\$. cod : chr [1:128] "1005" "1010" "3002" "4006"@ .\$. diagnosis : chr [1:128] "5/21/1997" "3/29/2000" "6/24/1998" "7/17/1997"@ .\$. sex : Factor w/ 2 levels "F","M": 2 2 1 2 2 2 1 2 2 2@ .\$. age : int [1:128] 53 19 52 38 57 17 18 16 15 40@ .\$. BT : Factor w/ 10 levels "B","B1","B2",...: 3 3 5 2 3 2 2 2 3 3@ .\$. remission : Factor w/ 2 levels "CR","REF": 1 1 1 1 1 1 1 1 1 1@ .\$. CR : chr [1:128] "CR" "CR" "CR" "CR"@ .\$. date.cr : chr [1:128] "8/6/1997" "6/27/2000" "8/17/1998" "9/8/1997"@ .\$. t(4;11) : logi [1:128] FALSE FALSE NA TRUE FALSE FALSE@ .\$. t(9;22) : logi [1:128] TRUE FALSE NA FALSE FALSE FALSE@ .\$. cyto.normal : logi [1:128] FALSE FALSE NA FALSE FALSE FALSE@ .\$. citog : chr [1:128] "t(9;22)" "simple alt." NA "t(4;11)"@ .\$. mol.biol : Factor w/ 6 levels "ALL1/AF4","BCR/ABL",...: 2 4 2 1 4 4 4 4 4 2@ .\$. fusion protein: Factor w/ 3 levels "p190","p190/p210",...: 3 NA 1 NA NA NA NA NA NA 1@ .\$. mdr : Factor w/ 2 levels "NEG","POS": 1 2 1 1 1 1 2 1 1 1@ .\$. kinet : Factor w/ 2 levels "dyploid","hyperd.": 1 1 1 1 1 2 2 1 1 NA@ .\$. ccr : logi [1:128] FALSE FALSE FALSE FALSE FALSE FALSE@ .\$. relapse : logi [1:128] FALSE TRUE TRUE TRUE TRUE TRUE@ .\$. transplant : logi [1:128] TRUE FALSE FALSE FALSE FALSE FALSE@ .\$. f.u : chr [1:128] "BMT / DEATH IN CR" "REL" "REL" "REL"@ .\$. date last seen: chr [1:128] NA "8/28/2000" "10/15/1999" "1/23/1998"@ dimLabels : chr [1:2] "sampleNames" "sampleColumns"@ .classVersion:Formal class 'Versions' [package "Biobase"] with 1 slot@ .Data:List of 1@ .\$. : int [1:3] 1 1 0 ..@ featureData :Formal class 'AnnotatedDataFrame' [package "Biobase"] with 4 slots@ varMetadata : 'data.frame': 0 obs. of 1 variable:@ .\$. labelDescription: logi(0)@ data : 'data.frame': 12625 obs. of 0 variables@ dimLabels : chr [1:2] "featureNames" "featureColumns"@ .classVersion:Formal class 'Versions' [package "Biobase"] with 1 slot@ .Data:List of 1@ .\$. : int [1:3] 1 1 0 ..@ annotation : chr "hgu95av2" ..@ protocolData :Formal class 'AnnotatedDataFrame' [package "Biobase"] with 4 slots@ varMetadata

```

:'data.frame': 0 obs. of 1 variable: .. .. ..$ labelDescription: chr(0) .. .. ..@ data :'data.frame': 128 obs. of
0 variables .. .. ..@ dimLabels : chr [1:2] "sampleNames" "sampleColumns" .. .. ..@ .classVersion:Formal
class 'Versions' [package "Biobase"] with 1 slot .. .. ..@ .Data:List of 1 .. .. .. ..$ : int [1:3] 1 1 0 ..@
.classVersion:Formal class 'Versions' [package "Biobase"] with 1 slot .. .. ..@ .Data:List of 4 .. .. ..$ : int
[1:3] 2 10 0 .. .. ..$ : int [1:3] 2 5 5 .. .. ..$ : int [1:3] 1 3 0 .. .. ..$ : int [1:3] 1 0 0

```

Information about assay and sample data

```
phenoData(ALL)
```

```

## An object of class 'AnnotatedDataFrame'
## sampleNames: 01005 01010 ... LAL4 (128 total)
## varLabels: cod diagnosis ... date last seen (21 total)
## varMetadata: labelDescription

```

```
varLabels(ALL)
```

```

## [1] "cod"           "diagnosis"      "sex"            "age"
## [5] "BT"            "remission"      "CR"             "date.cr"
## [9] "t(4;11)"       "t(9;22)"        "cyto.normal"    "citog"
## [13] "mol.biol"      "fusion protein" "mdr"            "kinet"
## [17] "ccr"           "relapse"        "transplant"     "f.u"
## [21] "date last seen"

```

```
varMetadata(ALL)
```

```

##                                     labelDescription
## cod                                     Patient ID
## diagnosis                             Date of diagnosis
## sex                                   Gender of the patient
## age                                   Age of the patient at entry
## BT                                   does the patient have B-cell or T-cell ALL
## remission       Complete remission(CR), refractory(REF) or NA. Derived from CR
## CR                                           Original remisson data
## date.cr                                   Date complete remission if achieved
## t(4;11)       did the patient have t(4;11) translocation. Derived from citog
## t(9;22)       did the patient have t(9;22) translocation. Derived from citog
## cyto.normal   Was cytogenetic test normal? Derived from citog
## citog         original cytogenetics data, deletions or t(4;11), t(9;22) status
## mol.biol      molecular biology
## fusion protein       which of p190, p210 or p190/210 for bcr/able
## mdr                                           multi-drug resistant
## kinet                                           ploidy: either diploid or hyperd.
## ccr                                           Continuous complete remission? Derived from f.u
## relapse                                           Relapse? Derived from f.u
## transplant       did the patient receive a bone marrow transplant? Derived from f.u
## f.u                                           follow up data available
## date last seen   date patient was last seen

```

```
featureNames(ALL)[1:128]
```

```
## [1] "1000_at" "1001_at" "1002_f_at" "1003_s_at" "1004_at"
## [6] "1005_at" "1006_at" "1007_s_at" "1008_f_at" "1009_at"
## [11] "100_g_at" "1010_at" "1011_s_at" "1012_at" "1013_at"
## [16] "1014_at" "1015_s_at" "1016_s_at" "1017_at" "1018_at"
## [21] "1019_g_at" "101_at" "1020_s_at" "1021_at" "1022_f_at"
## [26] "1023_at" "1024_at" "1025_g_at" "1026_s_at" "1027_at"
## [31] "1028_at" "1029_s_at" "102_at" "1030_s_at" "1031_at"
## [36] "1032_at" "1033_g_at" "1034_at" "1035_g_at" "1036_at"
## [41] "1037_at" "1038_s_at" "1039_s_at" "103_at" "1040_s_at"
## [46] "1041_at" "1042_at" "1043_s_at" "1044_s_at" "1045_s_at"
## [51] "1046_at" "1047_s_at" "1048_at" "1049_g_at" "104_at"
## [56] "1050_at" "1051_g_at" "1052_s_at" "1053_at" "1054_at"
## [61] "1055_g_at" "1056_s_at" "1057_at" "1058_at" "1059_at"
## [66] "105_at" "1060_g_at" "1061_at" "1062_g_at" "1063_s_at"
## [71] "1064_at" "1065_at" "1066_at" "1067_at" "1068_g_at"
## [76] "1069_at" "106_at" "1070_at" "1071_at" "1072_g_at"
## [81] "1073_at" "1074_at" "1075_f_at" "1076_at" "1077_at"
## [86] "1078_at" "1079_g_at" "107_at" "1080_s_at" "1081_at"
## [91] "1082_at" "1083_s_at" "1084_at" "1085_s_at" "1086_at"
## [96] "1087_at" "1088_at" "1089_i_at" "108_g_at" "1090_f_at"
## [101] "1091_at" "1092_at" "1093_at" "1094_g_at" "1095_s_at"
## [106] "1096_g_at" "1097_s_at" "1098_at" "1099_s_at" "109_at"
## [111] "1100_at" "1101_at" "1102_s_at" "1103_at" "1104_s_at"
## [116] "1105_s_at" "1106_s_at" "1107_s_at" "1108_s_at" "1109_s_at"
## [121] "110_at" "1110_at" "1111_at" "1112_g_at" "1113_at"
## [126] "1114_at" "1115_at" "1116_at"
```

The above results are the feature names.

```
sampleNames(ALL)[1:20]
```

The patient id

```
## [1] "01005" "01010" "03002" "04006" "04007" "04008" "04010" "04016"
## [9] "06002" "08001" "08011" "08012" "08018" "08024" "09008" "09017"
## [17] "11005" "12006" "12007" "12012"
```

```
expressionData <- exprs(ALL)
str(expressionData)
```

```
## num [1:12625, 1:128] 7.6 5.05 3.9 5.9 5.93 ...
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:12625] "1000_at" "1001_at" "1002_f_at" "1003_s_at" ...
## ..$ : chr [1:128] "01005" "01010" "03002" "04006" ...
```

We can look at the expression rate at which individual patient gene were expressed. Below two feature names and expresion levels for each patient within the two features are computed.

```
gene.expression <- ALL[c("1000_at", "1001_at"), ]
exprs(gene.expression)
```

```
##           01005    01010    03002    04006    04007    04008    04010
## 1000_at 7.597323 7.479445 7.567593 7.384684 7.905312 7.065914 7.474537
## 1001_at 5.046194 4.932537 4.799294 4.922627 4.844565 5.147762 5.122518
##           04016    06002    08001    08011    08012    08018    08024
## 1000_at 7.536119 7.183331 7.735545 7.591498 7.824284 7.231814 7.879988
## 1001_at 5.016132 5.288943 4.633217 4.583148 4.685951 5.059300 4.830464
##           09008    09017    11005    12006    12007    12012    12019
## 1000_at 7.891793 7.756734 7.640012 7.759599 7.678636 7.464285 7.652719
## 1001_at 5.999496 4.987595 4.967288 4.770481 5.456332 4.785863 5.175609
##           12026    14016    15001    15004    15005    16004    16009
## 1000_at 7.501591 7.570417 7.331509 7.366208 7.455451 7.328875 7.297313
## 1001_at 5.188992 5.258312 4.627955 4.733495 5.125098 5.332775 5.215707
##           19005    20002    22009    22010    22011    22013    24001
## 1000_at 7.563561 7.541133 8.016818 7.862181 7.702580 7.412003 7.916169
## 1001_at 4.858392 4.964424 5.216252 5.135825 4.802946 5.222676 4.790170
##           24005    24008    24010    24011    24017    24018    24019
## 1000_at 7.595848 7.296349 7.506236 7.144425 7.513972 7.815971 7.406135
## 1001_at 4.804743 5.002518 4.218220 5.228892 5.264158 4.899316 4.791335
##           24022    25003    25006    26001    26003    26005    26008
## 1000_at 7.300980 7.845054 7.651229 7.376930 7.663977 7.250353 7.663612
## 1001_at 5.177703 5.250315 4.896195 5.123546 5.078104 4.945670 5.124591
##           27003    27004    28001    28003    28005    28006    28007
## 1000_at 7.329996 7.360754 7.035203 7.705260 7.551734 7.538601 7.501531
## 1001_at 5.438098 4.757900 5.005279 5.009705 4.944978 4.511194 4.888814
##           28019    28021    28023    28024    28028    28031    28032
## 1000_at 7.116676 7.107979 7.427808 6.549926 7.514761 7.377215 6.973861
## 1001_at 5.275964 4.865566 5.057619 5.185277 4.788468 4.778381 4.970430
##           28035    28036    28037    28042    28043    28044    28047
## 1000_at 7.227516 7.407561 7.158049 7.235291 7.589310 7.988476 7.362458
## 1001_at 6.408157 5.042222 5.431469 4.686293 4.851805 4.894379 4.843868
##           30001    31007    31011    33005    36001    36002    37013
## 1000_at 7.508667 7.147843 7.651676 7.486432 7.759074 7.473427 7.627685
## 1001_at 5.587029 4.943857 4.741654 4.642628 4.962544 4.953122 5.358236
##           43001    43004    43007    43012    48001    49006    57001
## 1000_at 7.577529 7.600206 7.776844 7.585928 7.450666 7.004613 7.195206
## 1001_at 5.054157 4.879037 4.949908 5.057530 4.960382 4.836905 4.744006
##           62001    62002    62003    63001    64001    64002    65005
## 1000_at 7.407351 7.756195 7.913324 7.270997 7.694588 7.583071 7.609538
## 1001_at 4.930312 5.238937 5.074681 4.513671 4.928159 4.804083 4.715693
##           68001    68003    84004    LAL5    01003    01007    02020
## 1000_at 7.324502 7.545120 7.679603 7.604093 7.240252 7.676749 7.934247
## 1001_at 5.379102 4.650231 4.795495 4.988922 5.224752 5.129002 5.667907
##           04018    09002    10005    11002    12008    15006    16002
## 1000_at 7.874448 7.404271 7.775253 7.771891 7.355677 7.388882 7.589734
## 1001_at 5.005420 5.127949 4.423445 4.476761 5.461252 5.330129 4.836986
##           16007    17003    18001    19002    19008    19014    19017
## 1000_at 7.675929 7.662426 7.584008 7.840099 7.164922 7.843162 7.695714
## 1001_at 4.959669 5.743215 4.674920 5.208166 4.554529 5.718569 4.498515
##           20005    24006    26009    28008    28009    31015    37001
## 1000_at 7.520867 7.836577 7.470524 7.520806 7.646947 7.727560 7.849455
```

```
## 1001_at 5.135697 5.129836 5.213340 4.690815 4.902946 4.866731 4.959450
##          43006      43015      44001      49004      56007      64005      65003
## 1000_at 7.960842 8.188617 7.399999 7.813474 7.816922 7.913249 7.800199
## 1001_at 4.537677 5.154500 5.071885 4.874525 4.788699 5.403640 5.443827
##          83001      LAL4
## 1000_at 8.030047 7.702217
## 1001_at 5.178633 5.029670
```

```
pData(ALL)[1:15, ]
```

```
##          cod diagnosis sex age BT remission          CR      date.cr t(4;11)
## 01005 1005 5/21/1997   M  53 B2          CR          CR      8/6/1997   FALSE
## 01010 1010 3/29/2000   M  19 B2          CR          CR      6/27/2000   FALSE
## 03002 3002 6/24/1998   F  52 B4          CR          CR      8/17/1998    NA
## 04006 4006 7/17/1997   M  38 B1          CR          CR      9/8/1997    TRUE
## 04007 4007 7/22/1997   M  57 B2          CR          CR      9/17/1997   FALSE
## 04008 4008 7/30/1997   M  17 B1          CR          CR      9/27/1997   FALSE
## 04010 4010 10/30/1997  F  18 B1          CR          CR      1/7/1998    FALSE
## 04016 4016 2/10/2000   M  16 B1          CR          CR      4/17/2000   FALSE
## 06002 6002 3/19/1997   M  15 B2          CR          CR      6/9/1997    FALSE
## 08001 8001 1/15/1997   M  40 B2          CR          CR      3/26/1997   FALSE
## 08011 8011 8/21/1998   M  33 B3          CR          CR      10/8/1998   FALSE
## 08012 8012 10/22/1998  M  55 B3          CR          CR      1/9/1999    FALSE
## 08018 8018 8/27/1999   M   5 B3          CR          CR     10/18/1999   NA
## 08024 8024 7/20/2000   M  18 B2          CR DEATH IN CR      <NA>    FALSE
## 09008 9008 12/17/1999  M  41 B3          CR          CR      2/15/2000   FALSE
##          t(9;22) cyto.normal          citog mol.biol fusion protein mdr
## 01005      TRUE          FALSE          t(9;22) BCR/ABL          p210 NEG
## 01010     FALSE          FALSE simple alt.          NEG          <NA> POS
## 03002      NA           NA          <NA> BCR/ABL          p190 NEG
## 04006     FALSE          FALSE          t(4;11) ALL1/AF4          <NA> NEG
## 04007     FALSE          FALSE del(6q)          NEG          <NA> NEG
## 04008     FALSE          FALSE complex alt.          NEG          <NA> NEG
## 04010     FALSE          FALSE complex alt.          NEG          <NA> POS
## 04016     FALSE          FALSE simple alt.          NEG          <NA> NEG
## 06002     FALSE          TRUE normal          NEG          <NA> NEG
## 08001     FALSE          FALSE del(p15) BCR/ABL          p190 NEG
## 08011     FALSE          FALSE del(p15/p16) BCR/ABL          p190/p210 NEG
## 08012     FALSE          FALSE simple alt.          NEG          <NA> NEG
## 08018      NA           NA          <NA> E2A/PBX1          <NA> NEG
## 08024     FALSE          FALSE simple alt.          NEG          <NA> POS
## 09008      TRUE          FALSE t(9;22)+other BCR/ABL          p190 NEG
##          kinet      ccr relapse transplant          f.u date last seen
## 01005 dyploid FALSE   FALSE          TRUE BMT / DEATH IN CR      <NA>
## 01010 dyploid FALSE   TRUE          FALSE REL      8/28/2000
## 03002 dyploid FALSE   TRUE          FALSE REL     10/15/1999
## 04006 dyploid FALSE   TRUE          FALSE REL      1/23/1998
## 04007 dyploid FALSE   TRUE          FALSE REL      11/4/1997
## 04008 hyperd. FALSE   TRUE          FALSE REL     12/15/1997
## 04010 hyperd. FALSE   TRUE          FALSE REL      3/5/1998
## 04016 dyploid FALSE   TRUE          FALSE REL      9/26/2000
## 06002 dyploid FALSE   TRUE          FALSE REL      3/18/1998
## 08001     <NA> FALSE   TRUE          FALSE REL      7/11/1997
## 08011 dyploid FALSE   FALSE          TRUE BMT / DEATH IN CR      <NA>
```

```
## 08012 dyplloid FALSE      TRUE      FALSE      REL      4/9/1999
## 08018 dyplloid FALSE      TRUE      FALSE      REL      5/23/2000
## 08024 dyplloid  NA        NA        NA        <NA>      <NA>
## 09008 hyperd.  TRUE      FALSE      TRUE      BMT / CCR      00/09/01
```

ALL phenotype data

Subsetting B-cell patients from the ALL dataset

```
ALL$BT
```

```
## [1] B2 B2 B4 B1 B2 B1 B1 B1 B2 B2 B3 B3 B3 B2 B3 B B2 B3 B2 B3 B2 B2 B2
## [24] B1 B1 B2 B1 B2 B1 B2 B B B2 B2 B2 B1 B2 B2 B2 B2 B2 B4 B4 B2 B2 B2
## [47] B4 B2 B1 B2 B2 B3 B4 B3 B3 B3 B4 B3 B3 B1 B1 B1 B1 B3 B3 B3 B3 B3 B3
## [70] B3 B3 B1 B3 B1 B4 B2 B2 B1 B3 B4 B4 B2 B2 B3 B4 B4 B4 B1 B2 B2 B2 B1
## [93] B2 B B T T3 T2 T2 T3 T2 T T4 T2 T3 T3 T T2 T3 T2 T2 T2 T1 T4 T
## [116] T2 T3 T2 T2 T2 T2 T3 T3 T3 T2 T3 T2 T
## Levels: B B1 B2 B3 B4 T T1 T2 T3 T4
```

```
anyB <- grep("^B", ALL$BT)
b.cell <- ALL[, anyB]
```

```
exprsData = exprs(b.cell)
dim(exprsData)
```

```
## [1] 12625 95
```

```
pData = pData(b.cell)
dim(pData)
```

```
## [1] 95 21
```

```
head(pData)
```

```
##      cod diagnosis sex age BT remission CR   date.cr t(4;11) t(9;22)
## 01005 1005 5/21/1997  M  53 B2      CR CR 8/6/1997  FALSE  TRUE
## 01010 1010 3/29/2000  M  19 B2      CR CR 6/27/2000  FALSE  FALSE
## 03002 3002 6/24/1998  F  52 B4      CR CR 8/17/1998    NA    NA
## 04006 4006 7/17/1997  M  38 B1      CR CR 9/8/1997    TRUE  FALSE
## 04007 4007 7/22/1997  M  57 B2      CR CR 9/17/1997  FALSE  FALSE
## 04008 4008 7/30/1997  M  17 B1      CR CR 9/27/1997  FALSE  FALSE
##      cyto.normal      citog mol.biol fusion protein mdr   kinet   ccr
## 01005      FALSE      t(9;22) BCR/ABL      p210 NEG dyplloid FALSE
## 01010      FALSE  simple alt.      NEG      <NA> POS dyplloid FALSE
## 03002      NA      <NA> BCR/ABL      p190 NEG dyplloid FALSE
## 04006      FALSE      t(4;11) ALL1/AF4      <NA> NEG dyplloid FALSE
## 04007      FALSE      del(6q)      NEG      <NA> NEG dyplloid FALSE
## 04008      FALSE complex alt.      NEG      <NA> NEG hyperd. FALSE
##      relapse transplant      f.u date last seen
## 01005      FALSE      TRUE BMT / DEATH IN CR      <NA>
## 01010      TRUE      FALSE      REL      8/28/2000
```

```
## 03002    TRUE    FALSE          REL    10/15/1999
## 04006    TRUE    FALSE          REL    1/23/1998
## 04007    TRUE    FALSE          REL    11/4/1997
## 04008    TRUE    FALSE          REL    12/15/1997
```

```
featureData = featureData(b.cell)
dim(featureData)
```

```
##   featureNames featureColumns
##           12625             0
```

```
table(pData$sex, pData$age)
```

```
##
##      5 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31 32 33 36 37 38 39 40
## F 0  0  1  2  1  3  1  0  0  0  1  0  3  2  0  1  0  1  0  1  1  0  0  0
## M 1  2  4  2  6  3  2  2  2  2  0  1  1  2  1  0  1  1  1  1  1  2  1  1
##
##      41 43 44 45 46 47 48 49 50 51 52 53 54 55 57 58
## F  0  2  0  1  0  1  2  1  1  1  2  1  2  0  0  2
## M  2  0  1  0  1  1  0  1  1  0  1  4  2  1  2  0
```

```
table(pData$mol.biol, pData$BT)
```

```
##
##           B B1 B2 B3 B4  T T1 T2 T3 T4
## ALL1/AF4  0 10  0  0  0  0  0  0  0  0
## BCR/ABL   2  1 19  8  7  0  0  0  0  0
## E2A/PBX1  1  0  0  1  3  0  0  0  0  0
## NEG       2  8 16 14  2  0  0  0  0  0
## NUP-98     0  0  0  0  0  0  0  0  0  0
## p15/p16   0  0  1  0  0  0  0  0  0  0
```

```
table(pData$age, pData$BT)
```

```
##
##      B B1 B2 B3 B4 T T1 T2 T3 T4
## 5  0  0  0  1  0  0  0  0  0  0
## 15 0  0  2  0  0  0  0  0  0  0
## 16 0  1  1  2  1  0  0  0  0  0
## 17 0  2  1  1  0  0  0  0  0  0
## 18 0  2  1  2  2  0  0  0  0  0
## 19 1  2  3  0  0  0  0  0  0  0
## 20 0  1  2  0  0  0  0  0  0  0
## 21 0  0  1  0  1  0  0  0  0  0
## 22 0  0  1  1  0  0  0  0  0  0
## 23 0  0  1  1  0  0  0  0  0  0
## 24 0  0  0  0  1  0  0  0  0  0
## 25 0  1  0  0  0  0  0  0  0  0
## 26 1  0  2  1  0  0  0  0  0  0
## 27 1  0  2  1  0  0  0  0  0  0
```



```

## 28 0 0 1 0 0 0 0 0 0 0
## 29 0 0 1 0 0 0 0 0 0 0
## 31 0 0 0 1 0 0 0 0 0 0
## 32 0 0 1 0 1 0 0 0 0 0
## 33 0 0 0 1 0 0 0 0 0 0
## 36 0 1 0 1 0 0 0 0 0 0
## 37 0 0 1 1 0 0 0 0 0 0
## 38 0 1 1 0 0 0 0 0 0 0
## 39 0 0 1 0 0 0 0 0 0 0
## 40 0 0 1 0 0 0 0 0 0 0
## 41 0 1 0 1 0 0 0 0 0 0
## 43 0 0 2 0 0 0 0 0 0 0
## 44 0 1 0 0 0 0 0 0 0 0
## 45 0 1 0 0 0 0 0 0 0 0
## 46 0 0 0 1 0 0 0 0 0 0
## 47 0 1 0 1 0 0 0 0 0 0
## 48 0 1 1 0 0 0 0 0 0 0
## 49 0 1 0 0 1 0 0 0 0 0
## 50 1 0 0 0 1 0 0 0 0 0
## 51 0 0 1 0 0 0 0 0 0 0
## 52 0 1 0 1 1 0 0 0 0 0
## 53 0 0 3 1 1 0 0 0 0 0
## 54 0 0 0 2 2 0 0 0 0 0
## 55 0 0 0 1 0 0 0 0 0 0
## 57 0 0 2 0 0 0 0 0 0 0
## 58 0 1 1 0 0 0 0 0 0 0

```

```
summary(exprsData)
```

```

##      01005      01010      03002      04006
## Min.   : 2.435   Min.   : 2.423   Min.   : 2.271   Min.   : 2.266
## 1st Qu.: 4.111   1st Qu.: 4.139   1st Qu.: 4.118   1st Qu.: 4.125
## Median : 5.455   Median : 5.532   Median : 5.479   Median : 5.490
## Mean   : 5.630   Mean   : 5.648   Mean   : 5.633   Mean   : 5.631
## 3rd Qu.: 6.826   3rd Qu.: 6.867   3rd Qu.: 6.860   3rd Qu.: 6.828
## Max.   :13.455   Max.   :13.674   Max.   :13.796   Max.   :13.721
##      04007      04008      04010      04016
## Min.   : 2.304   Min.   : 2.402   Min.   : 2.303   Min.   : 2.331
## 1st Qu.: 4.013   1st Qu.: 4.093   1st Qu.: 4.060   1st Qu.: 3.980
## Median : 5.437   Median : 5.501   Median : 5.480   Median : 5.400
## Mean   : 5.630   Mean   : 5.640   Mean   : 5.628   Mean   : 5.597
## 3rd Qu.: 6.902   3rd Qu.: 6.861   3rd Qu.: 6.873   3rd Qu.: 6.850
## Max.   :13.880   Max.   :13.544   Max.   :13.573   Max.   :13.494
##      06002      08001      08011      08012
## Min.   : 2.400   Min.   : 2.382   Min.   : 2.451   Min.   : 2.371
## 1st Qu.: 4.146   1st Qu.: 4.085   1st Qu.: 4.087   1st Qu.: 4.011
## Median : 5.456   Median : 5.456   Median : 5.449   Median : 5.383
## Mean   : 5.621   Mean   : 5.641   Mean   : 5.621   Mean   : 5.575
## 3rd Qu.: 6.788   3rd Qu.: 6.896   3rd Qu.: 6.837   3rd Qu.: 6.804
## Max.   :13.889   Max.   :13.481   Max.   :13.469   Max.   :13.825
##      08018      08024      09008      09017
## Min.   : 2.395   Min.   : 2.310   Min.   : 2.408   Min.   : 2.267
## 1st Qu.: 4.160   1st Qu.: 4.090   1st Qu.: 4.034   1st Qu.: 4.114
## Median : 5.489   Median : 5.443   Median : 5.432   Median : 5.486

```

##	Mean	: 5.620	Mean	: 5.594	Mean	: 5.627	Mean	: 5.631
##	3rd Qu.:	6.786	3rd Qu.:	6.791	3rd Qu.:	6.870	3rd Qu.:	6.864
##	Max.	:13.813	Max.	:13.609	Max.	:13.652	Max.	:13.669
##	11005		12006		12007		12012	
##	Min.	: 2.196	Min.	: 2.451	Min.	: 2.349	Min.	: 2.440
##	1st Qu.:	4.190	1st Qu.:	4.223	1st Qu.:	4.118	1st Qu.:	4.146
##	Median	: 5.484	Median	: 5.484	Median	: 5.471	Median	: 5.462
##	Mean	: 5.626	Mean	: 5.652	Mean	: 5.622	Mean	: 5.630
##	3rd Qu.:	6.814	3rd Qu.:	6.808	3rd Qu.:	6.827	3rd Qu.:	6.796
##	Max.	:13.813	Max.	:13.609	Max.	:13.915	Max.	:13.555
##	12019		12026		14016		15001	
##	Min.	: 2.465	Min.	: 2.271	Min.	: 2.373	Min.	: 2.346
##	1st Qu.:	4.138	1st Qu.:	4.180	1st Qu.:	4.214	1st Qu.:	4.114
##	Median	: 5.486	Median	: 5.502	Median	: 5.508	Median	: 5.477
##	Mean	: 5.633	Mean	: 5.616	Mean	: 5.633	Mean	: 5.616
##	3rd Qu.:	6.823	3rd Qu.:	6.787	3rd Qu.:	6.775	3rd Qu.:	6.803
##	Max.	:13.515	Max.	:14.032	Max.	:14.018	Max.	:13.443
##	15004		15005		16004		16009	
##	Min.	: 2.445	Min.	: 2.424	Min.	: 2.432	Min.	: 2.311
##	1st Qu.:	4.106	1st Qu.:	4.124	1st Qu.:	4.196	1st Qu.:	4.184
##	Median	: 5.452	Median	: 5.480	Median	: 5.465	Median	: 5.562
##	Mean	: 5.584	Mean	: 5.626	Mean	: 5.591	Mean	: 5.649
##	3rd Qu.:	6.787	3rd Qu.:	6.829	3rd Qu.:	6.705	3rd Qu.:	6.837
##	Max.	:13.631	Max.	:13.600	Max.	:13.509	Max.	:13.861
##	19005		20002		22009		22010	
##	Min.	: 2.294	Min.	: 2.316	Min.	: 2.441	Min.	: 2.508
##	1st Qu.:	4.095	1st Qu.:	4.093	1st Qu.:	4.172	1st Qu.:	4.172
##	Median	: 5.435	Median	: 5.499	Median	: 5.529	Median	: 5.468
##	Mean	: 5.601	Mean	: 5.646	Mean	: 5.636	Mean	: 5.615
##	3rd Qu.:	6.800	3rd Qu.:	6.879	3rd Qu.:	6.826	3rd Qu.:	6.811
##	Max.	:14.031	Max.	:13.739	Max.	:13.639	Max.	:13.460
##	22011		22013		24001		24005	
##	Min.	: 2.351	Min.	: 2.431	Min.	: 2.372	Min.	: 2.296
##	1st Qu.:	4.066	1st Qu.:	4.161	1st Qu.:	4.041	1st Qu.:	4.028
##	Median	: 5.438	Median	: 5.479	Median	: 5.414	Median	: 5.422
##	Mean	: 5.602	Mean	: 5.623	Mean	: 5.599	Mean	: 5.621
##	3rd Qu.:	6.831	3rd Qu.:	6.798	3rd Qu.:	6.844	3rd Qu.:	6.882
##	Max.	:13.543	Max.	:13.856	Max.	:13.487	Max.	:13.885
##	24008		24010		24011		24017	
##	Min.	: 2.475	Min.	: 2.375	Min.	: 2.369	Min.	: 2.195
##	1st Qu.:	4.187	1st Qu.:	4.290	1st Qu.:	4.163	1st Qu.:	4.126
##	Median	: 5.434	Median	: 5.389	Median	: 5.410	Median	: 5.524
##	Mean	: 5.637	Mean	: 5.656	Mean	: 5.602	Mean	: 5.644
##	3rd Qu.:	6.779	3rd Qu.:	6.719	3rd Qu.:	6.722	3rd Qu.:	6.884
##	Max.	:13.829	Max.	:13.543	Max.	:14.045	Max.	:13.574
##	24018		24019		24022		25003	
##	Min.	: 2.264	Min.	: 2.470	Min.	: 2.473	Min.	: 2.456
##	1st Qu.:	4.208	1st Qu.:	4.128	1st Qu.:	4.133	1st Qu.:	4.162
##	Median	: 5.501	Median	: 5.503	Median	: 5.512	Median	: 5.492
##	Mean	: 5.635	Mean	: 5.623	Mean	: 5.658	Mean	: 5.629
##	3rd Qu.:	6.771	3rd Qu.:	6.824	3rd Qu.:	6.892	3rd Qu.:	6.795
##	Max.	:13.507	Max.	:13.735	Max.	:13.962	Max.	:13.604
##	25006		26001		26003		26005	
##	Min.	: 2.434	Min.	: 2.422	Min.	: 2.493	Min.	: 2.371

##	1st Qu.: 4.197	1st Qu.: 4.182	1st Qu.: 4.125	1st Qu.: 4.116
##	Median : 5.540	Median : 5.485	Median : 5.479	Median : 5.474
##	Mean : 5.651	Mean : 5.625	Mean : 5.626	Mean : 5.620
##	3rd Qu.: 6.832	3rd Qu.: 6.789	3rd Qu.: 6.846	3rd Qu.: 6.809
##	Max. :13.785	Max. :13.661	Max. :13.690	Max. :13.663
##	26008	27003	27004	28001
##	Min. : 2.419	Min. : 2.416	Min. : 2.338	Min. : 2.377
##	1st Qu.: 4.088	1st Qu.: 4.093	1st Qu.: 4.064	1st Qu.: 4.269
##	Median : 5.463	Median : 5.430	Median : 5.458	Median : 5.435
##	Mean : 5.621	Mean : 5.576	Mean : 5.614	Mean : 5.589
##	3rd Qu.: 6.827	3rd Qu.: 6.756	3rd Qu.: 6.863	3rd Qu.: 6.677
##	Max. :13.453	Max. :13.781	Max. :13.708	Max. :13.556
##	28003	28005	28006	28007
##	Min. : 2.268	Min. : 2.430	Min. : 2.271	Min. : 2.320
##	1st Qu.: 4.236	1st Qu.: 4.038	1st Qu.: 4.242	1st Qu.: 4.006
##	Median : 5.544	Median : 5.416	Median : 5.385	Median : 5.399
##	Mean : 5.665	Mean : 5.605	Mean : 5.645	Mean : 5.600
##	3rd Qu.: 6.839	3rd Qu.: 6.873	3rd Qu.: 6.780	3rd Qu.: 6.873
##	Max. :13.744	Max. :13.593	Max. :13.652	Max. :13.948
##	28019	28021	28023	28024
##	Min. : 2.294	Min. : 2.350	Min. : 2.379	Min. : 2.333
##	1st Qu.: 4.158	1st Qu.: 4.047	1st Qu.: 4.037	1st Qu.: 4.060
##	Median : 5.561	Median : 5.494	Median : 5.455	Median : 5.482
##	Mean : 5.656	Mean : 5.638	Mean : 5.625	Mean : 5.638
##	3rd Qu.: 6.887	3rd Qu.: 6.914	3rd Qu.: 6.913	3rd Qu.: 6.879
##	Max. :13.689	Max. :13.662	Max. :13.631	Max. :13.733
##	28028	28031	28032	28035
##	Min. : 2.444	Min. : 2.385	Min. : 2.422	Min. : 2.365
##	1st Qu.: 4.119	1st Qu.: 4.132	1st Qu.: 4.149	1st Qu.: 4.121
##	Median : 5.489	Median : 5.493	Median : 5.482	Median : 5.538
##	Mean : 5.650	Mean : 5.615	Mean : 5.629	Mean : 5.648
##	3rd Qu.: 6.885	3rd Qu.: 6.836	3rd Qu.: 6.845	3rd Qu.: 6.886
##	Max. :13.759	Max. :13.630	Max. :13.376	Max. :13.914
##	28036	28037	28042	28043
##	Min. : 2.262	Min. : 2.373	Min. : 2.467	Min. : 2.412
##	1st Qu.: 4.123	1st Qu.: 4.141	1st Qu.: 4.178	1st Qu.: 4.099
##	Median : 5.489	Median : 5.554	Median : 5.476	Median : 5.506
##	Mean : 5.627	Mean : 5.663	Mean : 5.583	Mean : 5.630
##	3rd Qu.: 6.839	3rd Qu.: 6.895	3rd Qu.: 6.765	3rd Qu.: 6.876
##	Max. :13.902	Max. :13.702	Max. :13.368	Max. :13.639
##	28044	28047	30001	31007
##	Min. : 2.477	Min. : 2.347	Min. : 2.317	Min. : 2.430
##	1st Qu.: 4.106	1st Qu.: 4.074	1st Qu.: 4.105	1st Qu.: 4.230
##	Median : 5.464	Median : 5.461	Median : 5.513	Median : 5.463
##	Mean : 5.629	Mean : 5.625	Mean : 5.646	Mean : 5.635
##	3rd Qu.: 6.858	3rd Qu.: 6.870	3rd Qu.: 6.869	3rd Qu.: 6.751
##	Max. :13.876	Max. :13.723	Max. :13.871	Max. :13.426
##	31011	33005	36001	36002
##	Min. : 2.427	Min. : 2.355	Min. : 2.320	Min. : 2.331
##	1st Qu.: 4.087	1st Qu.: 4.089	1st Qu.: 4.076	1st Qu.: 4.045
##	Median : 5.432	Median : 5.525	Median : 5.441	Median : 5.406
##	Mean : 5.603	Mean : 5.650	Mean : 5.612	Mean : 5.595
##	3rd Qu.: 6.826	3rd Qu.: 6.898	3rd Qu.: 6.815	3rd Qu.: 6.830
##	Max. :13.503	Max. :13.756	Max. :13.533	Max. :13.950

##	37013	43001	43004	43007
##	Min. : 2.345	Min. : 2.396	Min. : 2.357	Min. : 2.348
##	1st Qu.: 4.135	1st Qu.: 4.181	1st Qu.: 4.102	1st Qu.: 4.080
##	Median : 5.513	Median : 5.460	Median : 5.459	Median : 5.451
##	Mean : 5.634	Mean : 5.643	Mean : 5.583	Mean : 5.607
##	3rd Qu.: 6.872	3rd Qu.: 6.785	3rd Qu.: 6.788	3rd Qu.: 6.805
##	Max. :13.651	Max. :13.624	Max. :13.278	Max. :13.808
##	43012	48001	49006	57001
##	Min. : 2.379	Min. : 2.428	Min. : 2.333	Min. : 2.292
##	1st Qu.: 4.138	1st Qu.: 4.136	1st Qu.: 4.099	1st Qu.: 4.002
##	Median : 5.471	Median : 5.450	Median : 5.416	Median : 5.398
##	Mean : 5.609	Mean : 5.625	Mean : 5.601	Mean : 5.591
##	3rd Qu.: 6.813	3rd Qu.: 6.820	3rd Qu.: 6.776	3rd Qu.: 6.826
##	Max. :13.489	Max. :13.527	Max. :13.638	Max. :13.632
##	62001	62002	62003	63001
##	Min. : 2.382	Min. : 2.416	Min. : 2.374	Min. : 2.380
##	1st Qu.: 4.085	1st Qu.: 4.062	1st Qu.: 4.174	1st Qu.: 4.229
##	Median : 5.455	Median : 5.484	Median : 5.496	Median : 5.382
##	Mean : 5.625	Mean : 5.636	Mean : 5.630	Mean : 5.647
##	3rd Qu.: 6.869	3rd Qu.: 6.897	3rd Qu.: 6.805	3rd Qu.: 6.761
##	Max. :13.825	Max. :13.703	Max. :13.691	Max. :13.727
##	64001	64002	65005	68001
##	Min. : 2.459	Min. : 2.392	Min. : 2.383	Min. : 1.985
##	1st Qu.: 4.044	1st Qu.: 4.081	1st Qu.: 3.974	1st Qu.: 4.153
##	Median : 5.428	Median : 5.465	Median : 5.344	Median : 5.522
##	Mean : 5.618	Mean : 5.620	Mean : 5.580	Mean : 5.638
##	3rd Qu.: 6.839	3rd Qu.: 6.820	3rd Qu.: 6.832	3rd Qu.: 6.860
##	Max. :13.620	Max. :13.527	Max. :13.599	Max. :13.524
##	68003	84004	LAL5	
##	Min. : 2.498	Min. : 2.345	Min. : 2.310	
##	1st Qu.: 4.205	1st Qu.: 3.963	1st Qu.: 4.153	
##	Median : 5.511	Median : 5.377	Median : 5.500	
##	Mean : 5.657	Mean : 5.607	Mean : 5.633	
##	3rd Qu.: 6.847	3rd Qu.: 6.894	3rd Qu.: 6.837	
##	Max. :13.561	Max. :13.568	Max. :13.668	

Reducing genes based on variability

```
tropical = c("darkorange", "dodgerblue", "hotpink", "limegreen", "yellow")
palette(tropical)
par(pch = 19)
library(genefilter)
```

```
##
## Attaching package: 'genefilter'

## The following object is masked from 'package:base':
##
## anyNA
```

```
library(RColorBrewer)
library(devtools)
library(RSkittleBrewer)
library(gplots)
```

```
##
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:Biobase':
##
##     combine

## The following objects are masked from 'package:BiocGenerics':
##
##     combine, intersect, setdiff, union

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(AnnotationDbi)
```

```
## Loading required package: stats4

## Loading required package: IRanges

## Loading required package: S4Vectors

##
## Attaching package: 'S4Vectors'

## The following object is masked from 'package:dplyr':
##
##     rename

##
## Attaching package: 'IRanges'
```

```
## The following objects are masked from 'package:dplyr':  
##  
## collapse, desc, slice
```

```
## The following object is masked from 'package:gplots':  
##  
## space
```

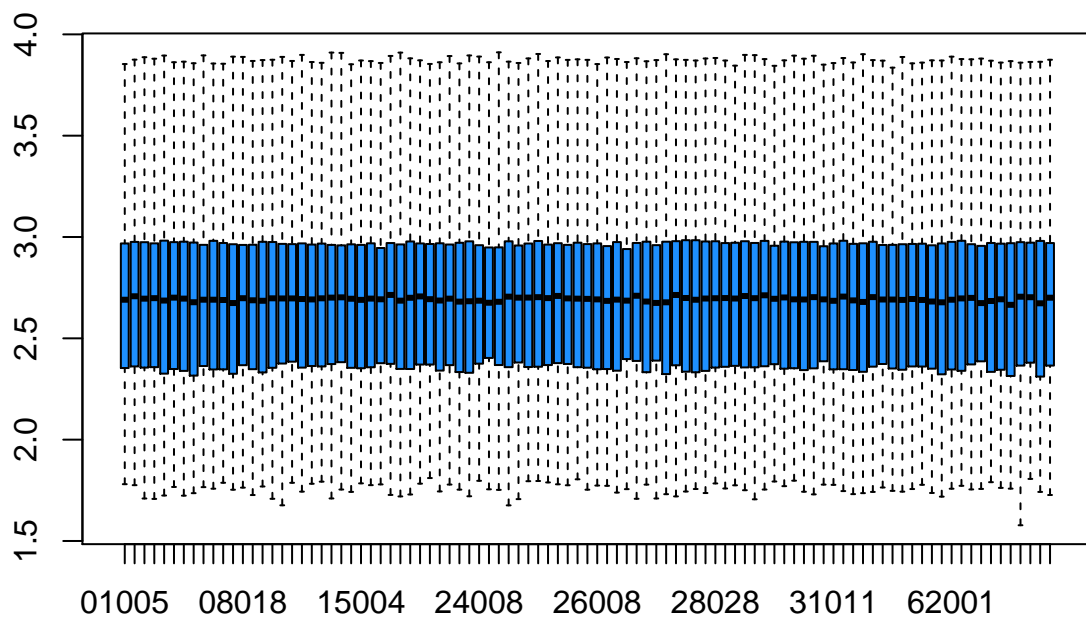
```
##  
## Attaching package: 'AnnotationDbi'
```

```
## The following object is masked from 'package:dplyr':  
##  
## select
```

```
lowgenes = rowMeans(exprsData) < 8.5  
table(lowgenes)  
eData_filt = filter(as.data.frame(exprsData), !lowgenes)  
dim(eData_filt)  
  
lowgenes2 = rowMedians(as.matrix(exprsData)) < 8.5  
  
table(lowgenes2)  
table(lowgenes2, lowgenes)  
  
exprsData_df = data.frame(exprsData)  
  
eData_filt1 = filter(exprsData_df, !lowgenes2)  
  
eData_filt1 = log2(eData_filt1 + 1) ###<- To remove aswell the undefined values.  
dim(eData_filt1)
```

Exploratory Analysis

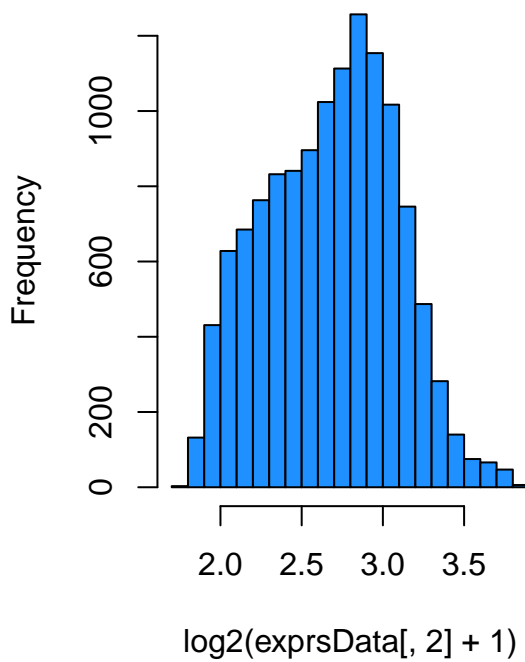
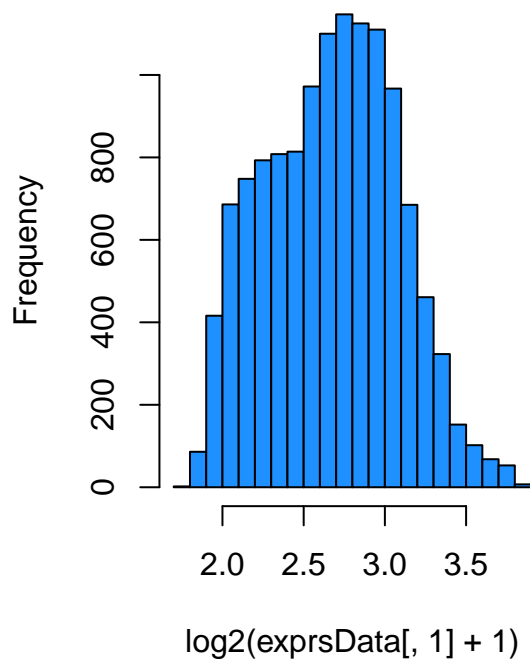
```
##Boxplot applied to entire expression matrix  
boxplot(log2(exprsData+1),col=2,range=0)
```



```
par(mfrow=c(1,2)) ##<- setting up parameter for plots.
hist(log2(exprsData[,1]+1), col=2)

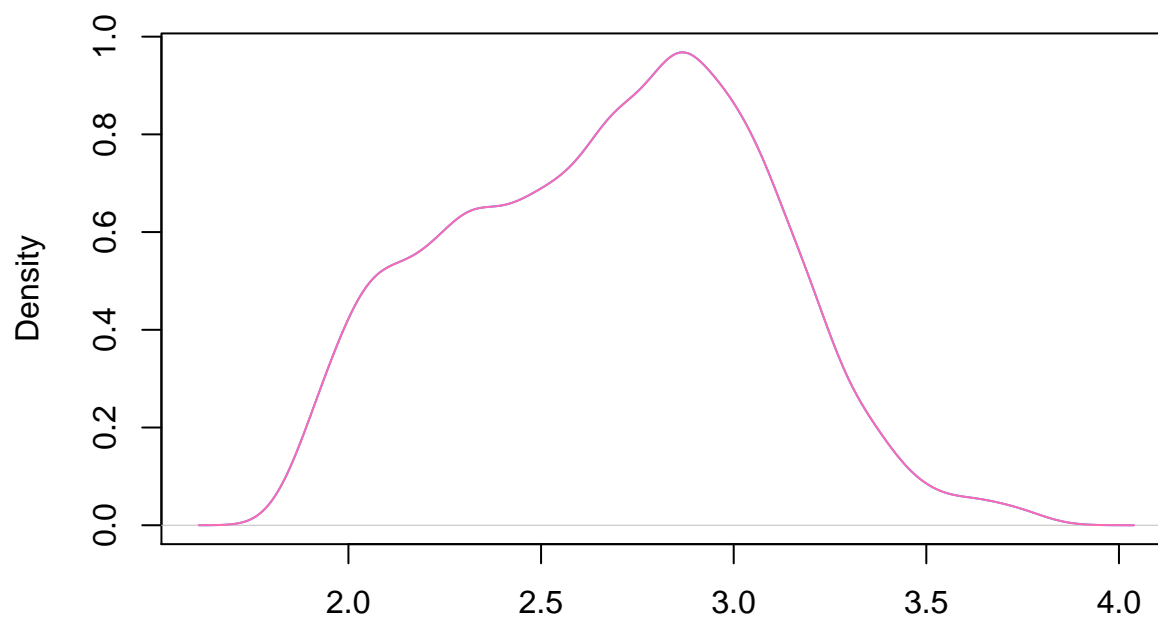
hist(log2(exprsData[,2]+1), col=2)
```

Histogram of $\log_2(\text{exprsData[, 1]} + 1)$ Histogram of $\log_2(\text{exprsData[, 2]} + 1)$



```
par(mfrow=c(1,1))  
plot(density(log2(exprsData[,2]+1)),col=2)  
###The lines command will allow to overlay another plot on top of the plot previously  
lines(density(log2(exprsData[,2]+1)),col=3)
```


density.default(x = log2(exprsData[, 2] + 1))

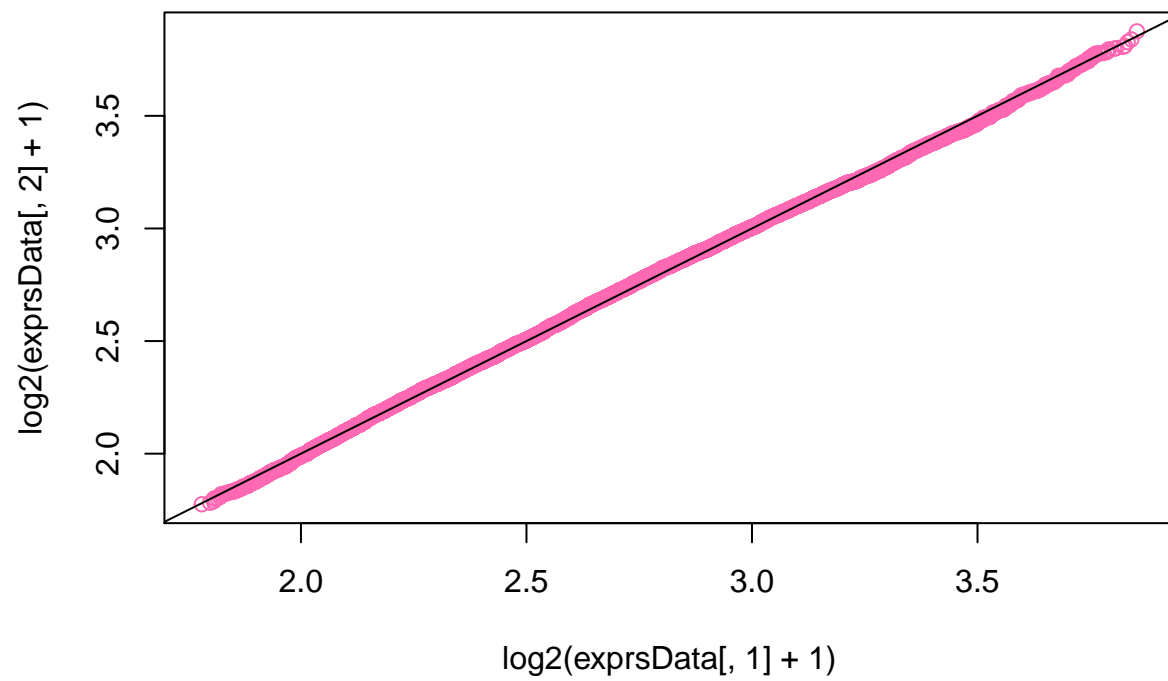


N = 12625 Bandwidth = 0.05458

```
qqplot(log2(exprsData[,1]+1), log2(exprsData[,2]+1), col=3)
```

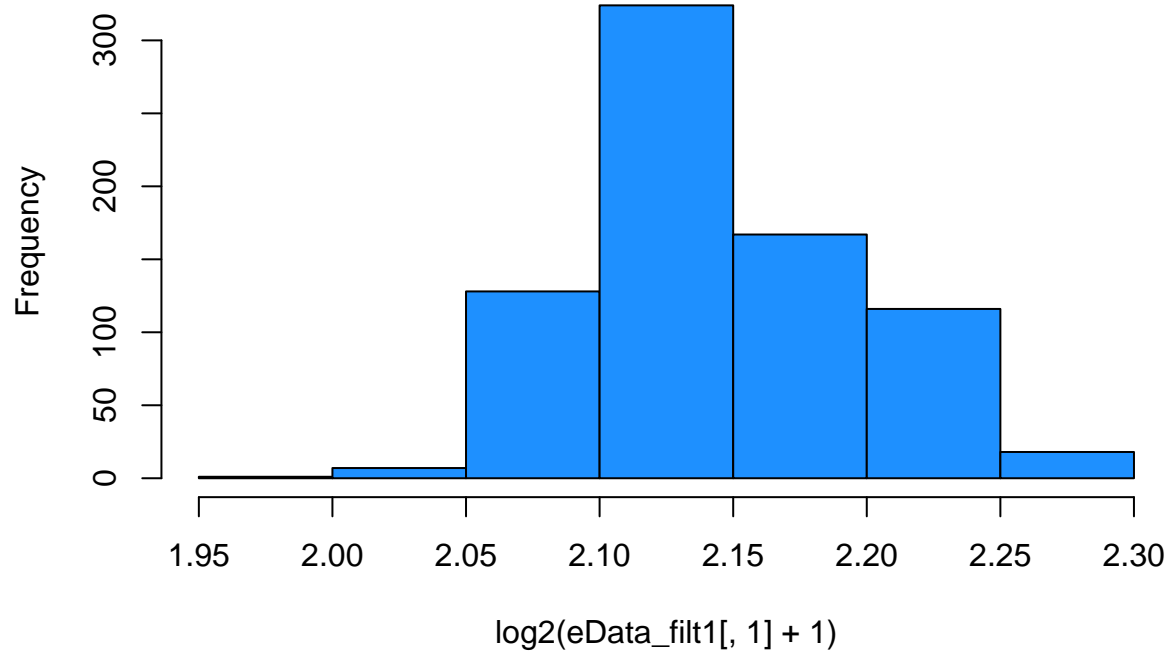
I can also use the qqplot to see if the samples are consistent. qqplot is making one dot for every

```
abline(c(0,1)) ###<- creates a 45 degree line.
```

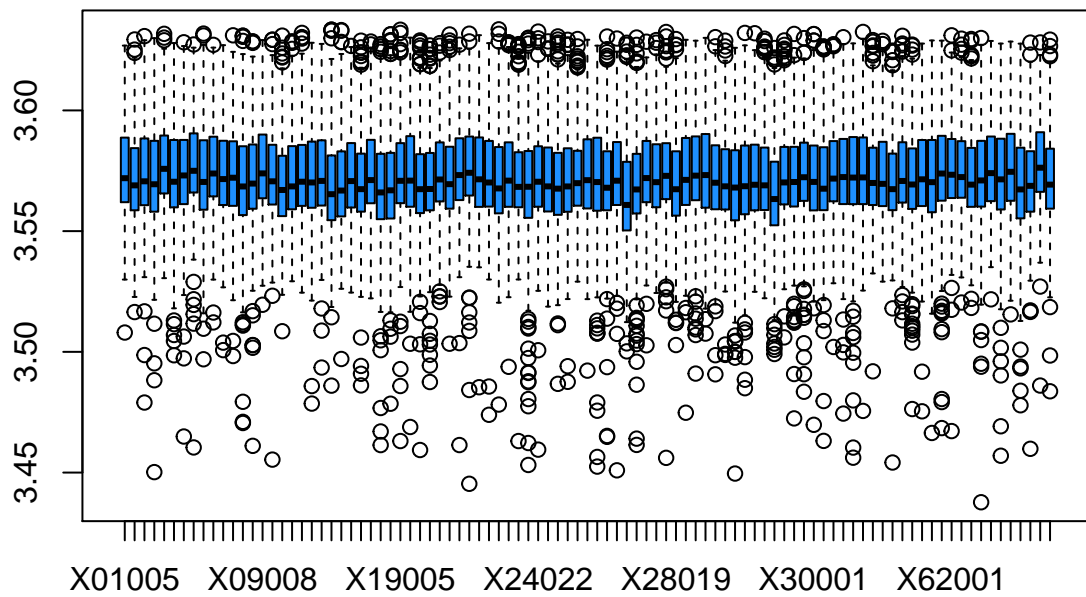


```
hist(log2(eData_filt1[,1]+1), col=2)
```

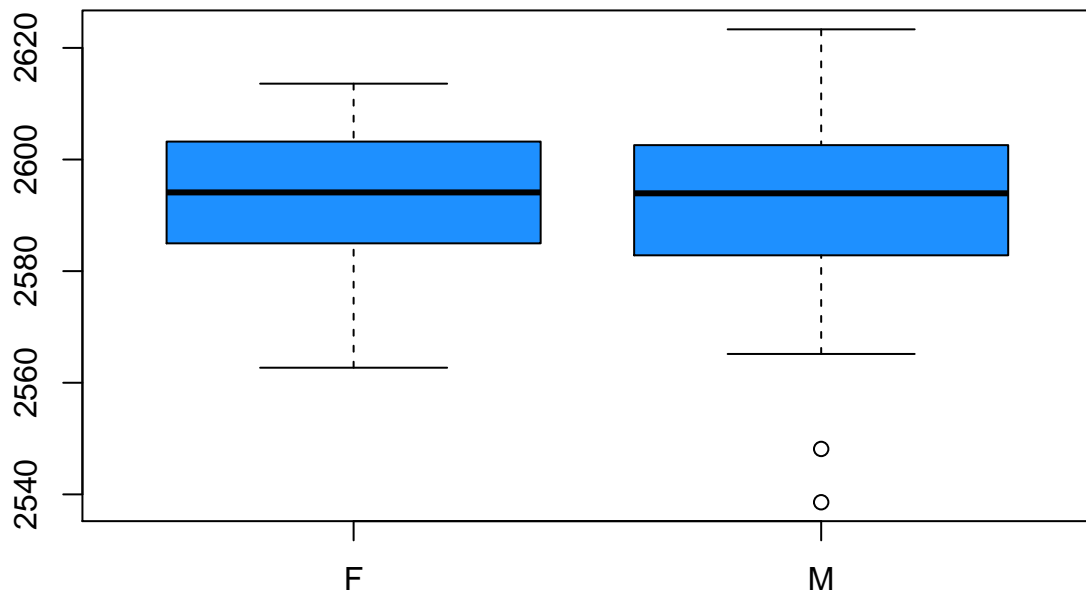
Histogram of $\log_2(\text{eData_filt1[, 1]} + 1)$



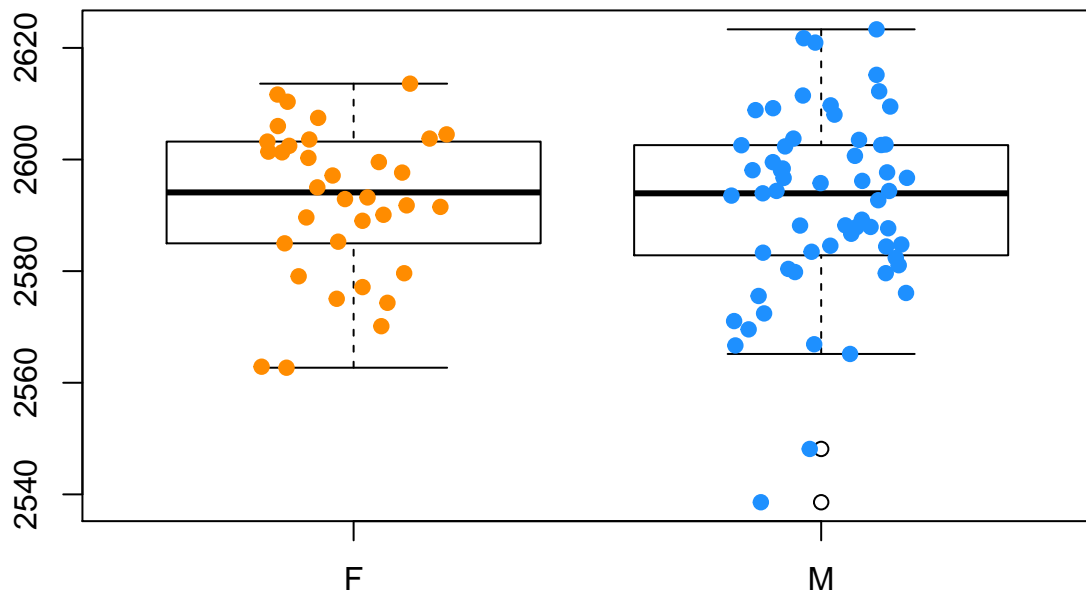
```
boxplot(as.matrix(log2(eData_filt1+8.5)),col=2)
```



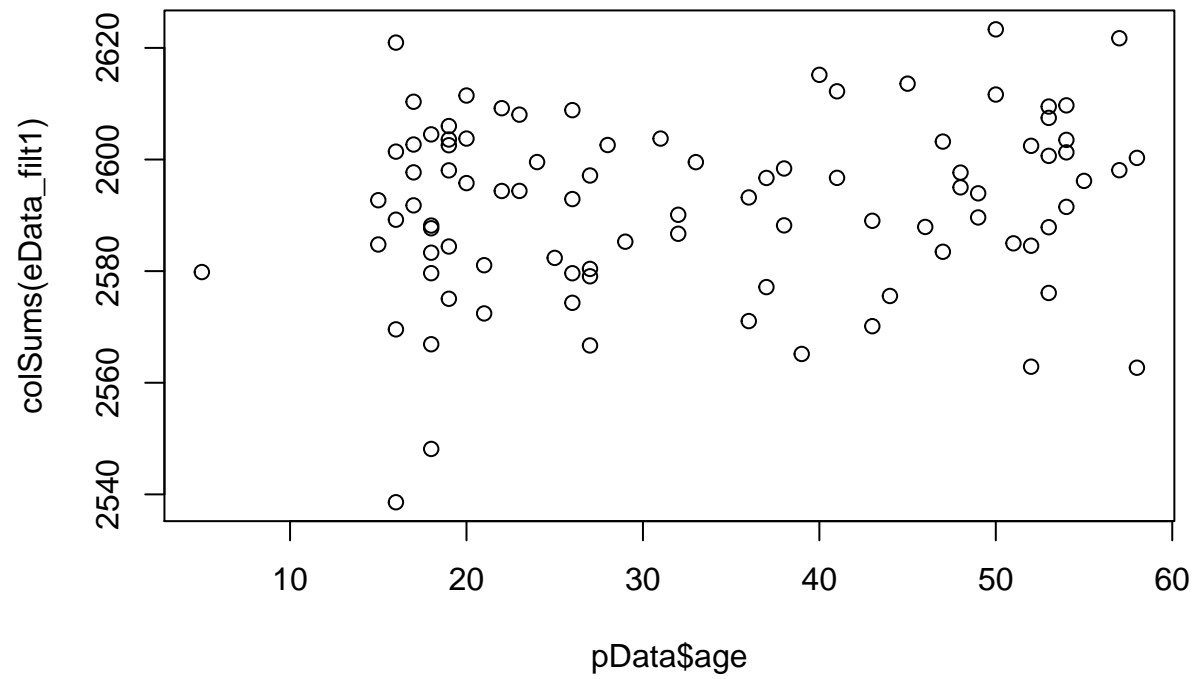
```
boxplot(colSums(eData_filt1) ~ pData$sex, col=2) ###<- comparing the gene expression between male and f
```



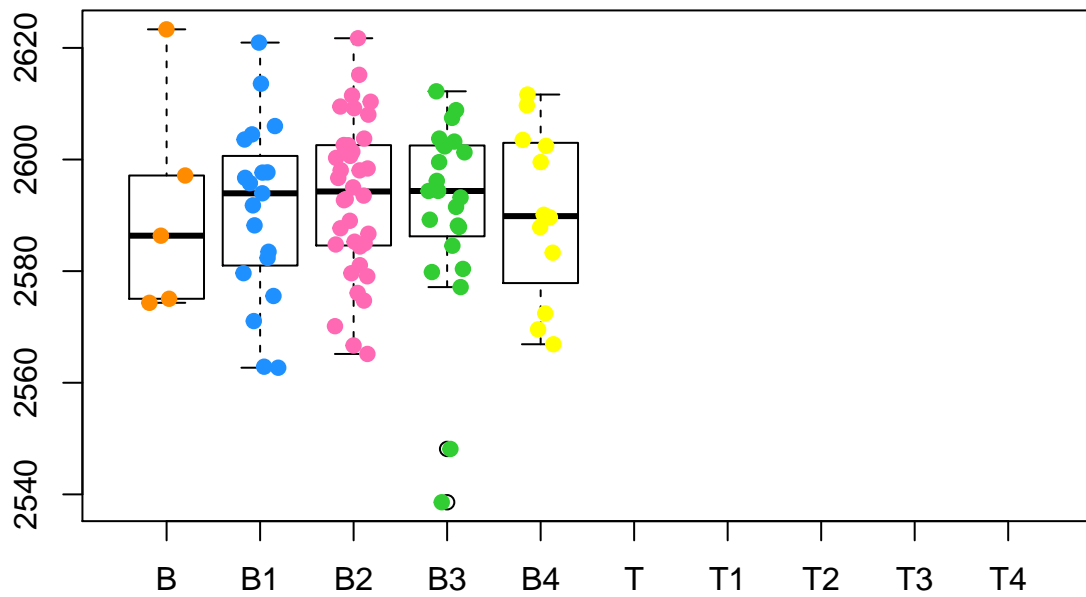
```
boxplot(colSums(eData_filt1) ~ pData$sex)
points(colSums(eData_filt1) ~ jitter(as.numeric(pData$sex)), col=as.numeric(pData$sex), pch=19) ###<- a
```



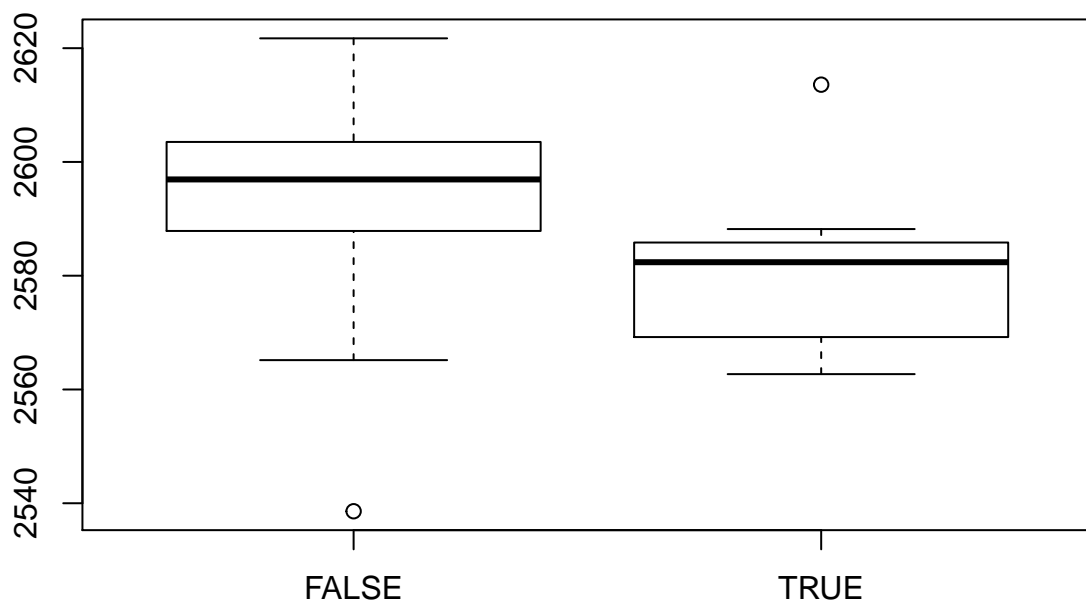
```
plot(colSums(eData_filt1) ~ pData$age)
```



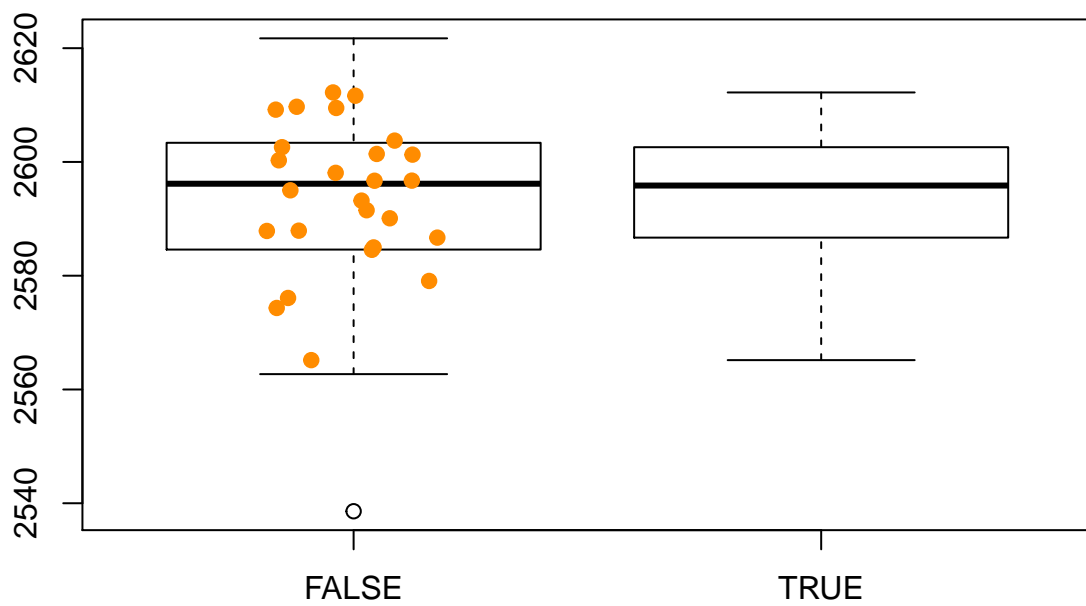
```
boxplot(colSums(eData_filt1) ~ pData$BT)  
points(colSums(eData_filt1) ~ jitter(as.numeric(pData$BT)), col=as.numeric(pData$BT), pch=19)
```



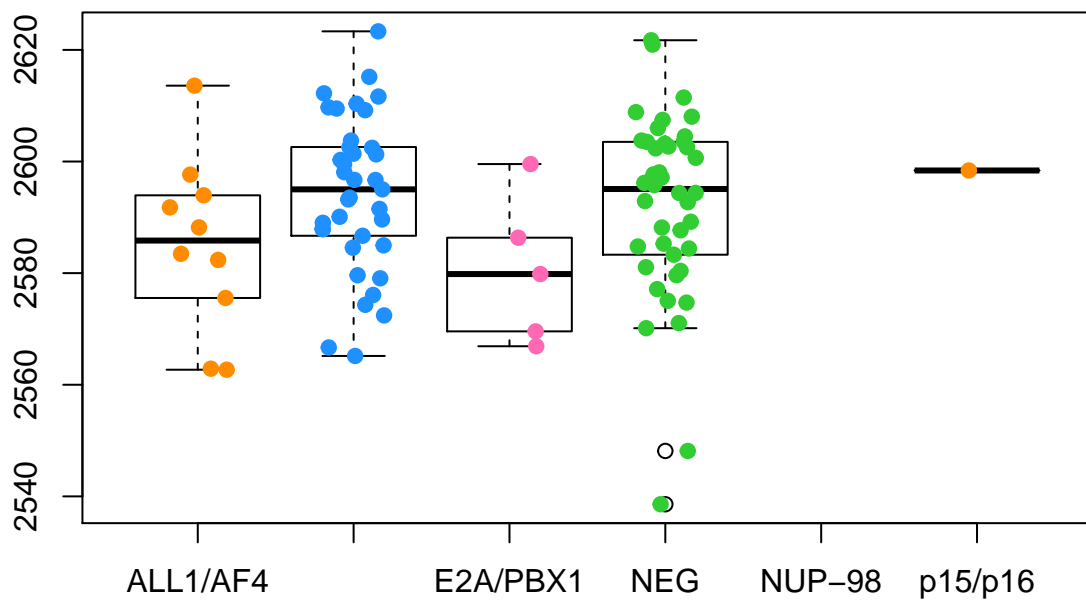
```
eData_filt1_matrix = as.matrix(eData_filt1)
boxplot(colSums(eData_filt1) ~ pData$t(4;11))
```

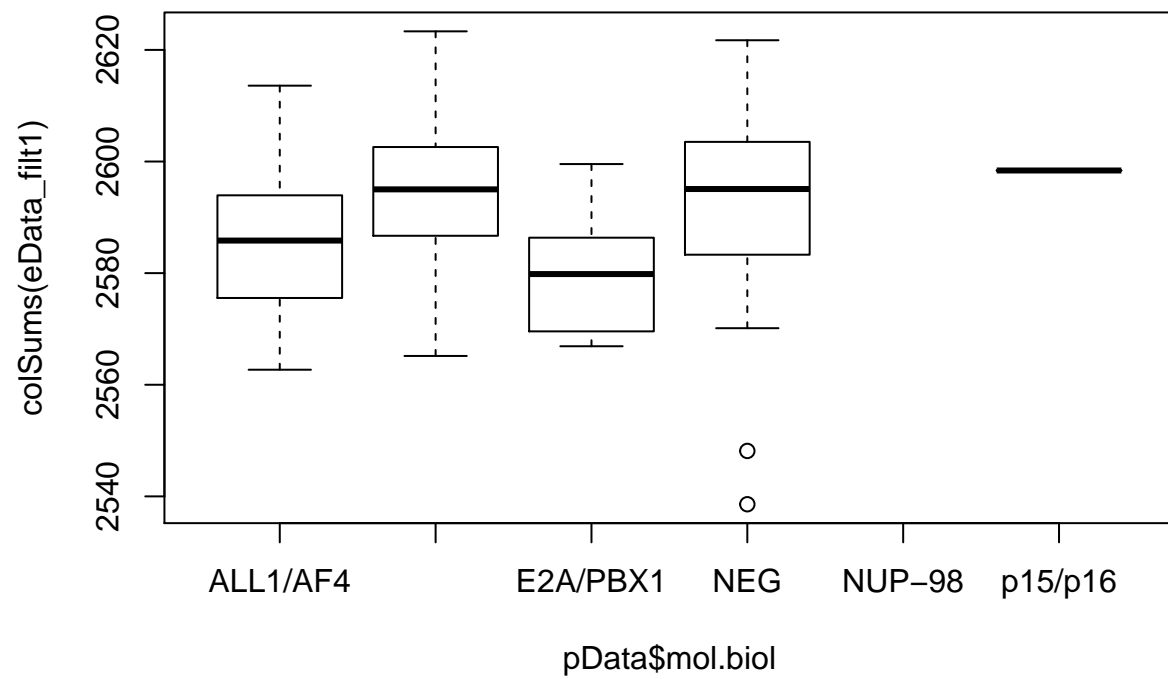
```
boxplot(colSums(eData_filt1) ~ pData$t(9;22))  
points(colSums(eData_filt1) ~ jitter(as.numeric(pData$t(9;22))), col=as.numeric(pData$t(9;22)), pch=
```

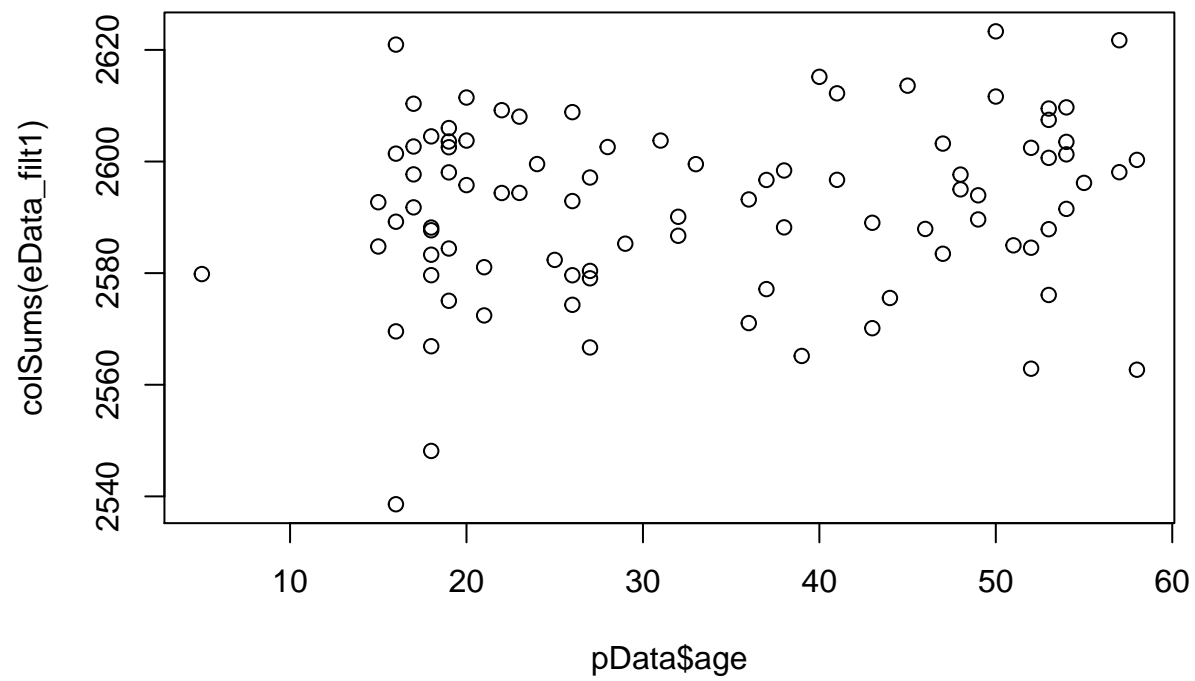


```
##<-
boxplot(colSums(eData_filt1) ~ pData$mol.biol)
points(colSums(eData_filt1) ~ jitter(as.numeric(pData$mol.biol)), col=as.numeric(pData$mol.biol), pch=1)
```

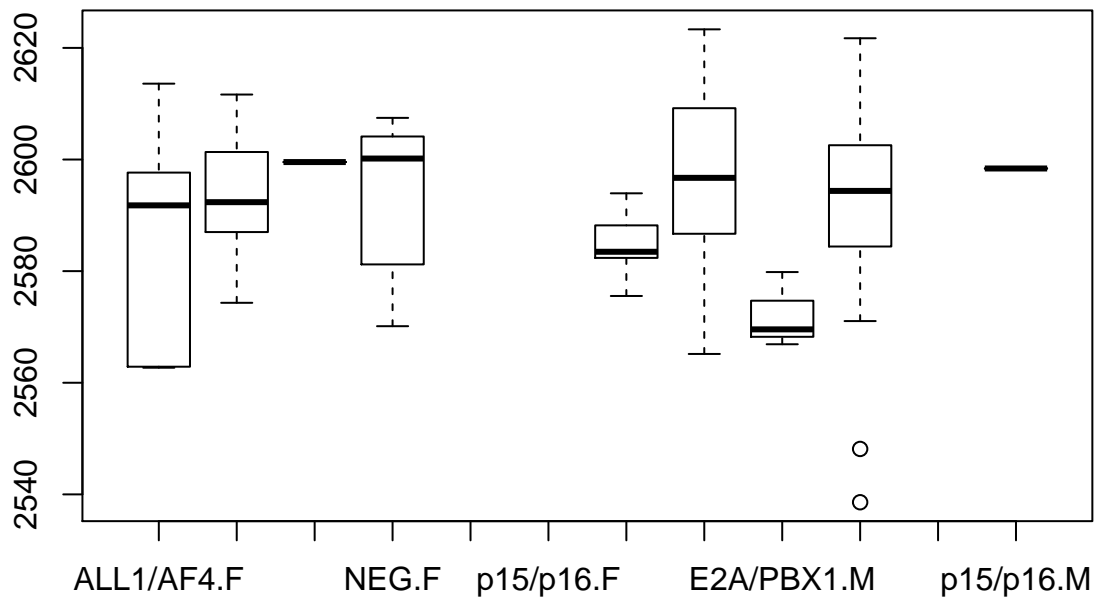


```
plot(colSums(eData_filt1) ~ pData$mol.biol + pData$age)
```

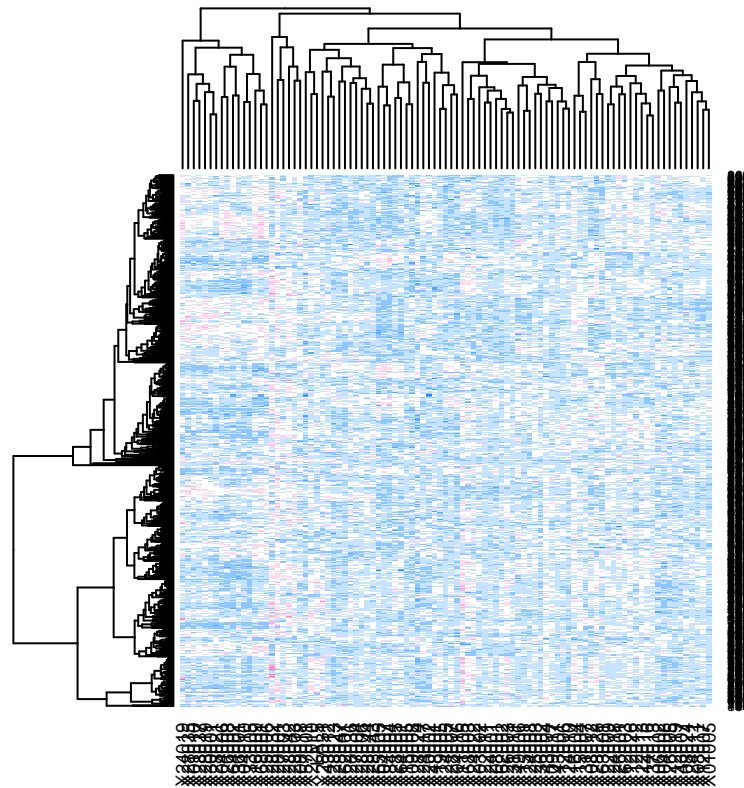




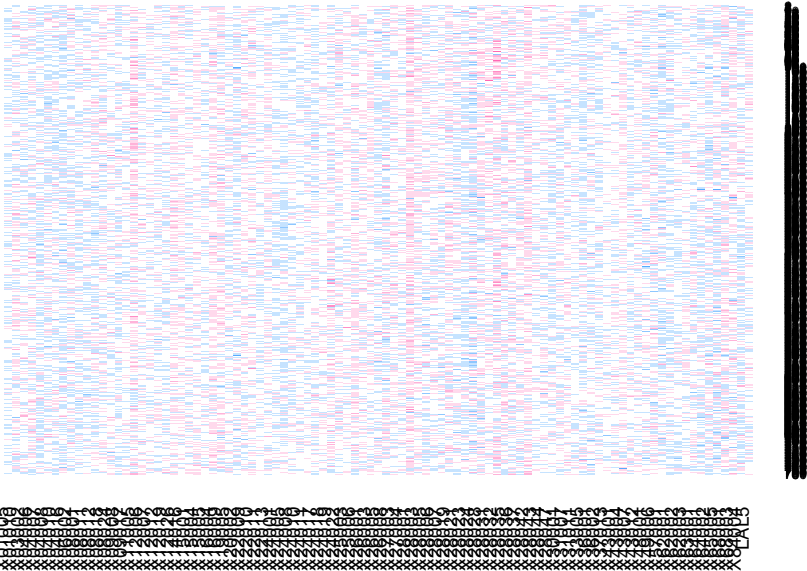
```
boxplot(colSums(eData_filt1) ~ pData$mol.biol + pData$sex)
```



```
eData_filt1_matrix = as.matrix(eData_filt1)
colramp = colorRampPalette(c(3, "white", 2))(9)
heatmap(eData_filt1_matrix, col=colramp)
```



```
heatmap(eData_filt1_matrix, col=colramp, Rowv = NA, Colv = NA)
```

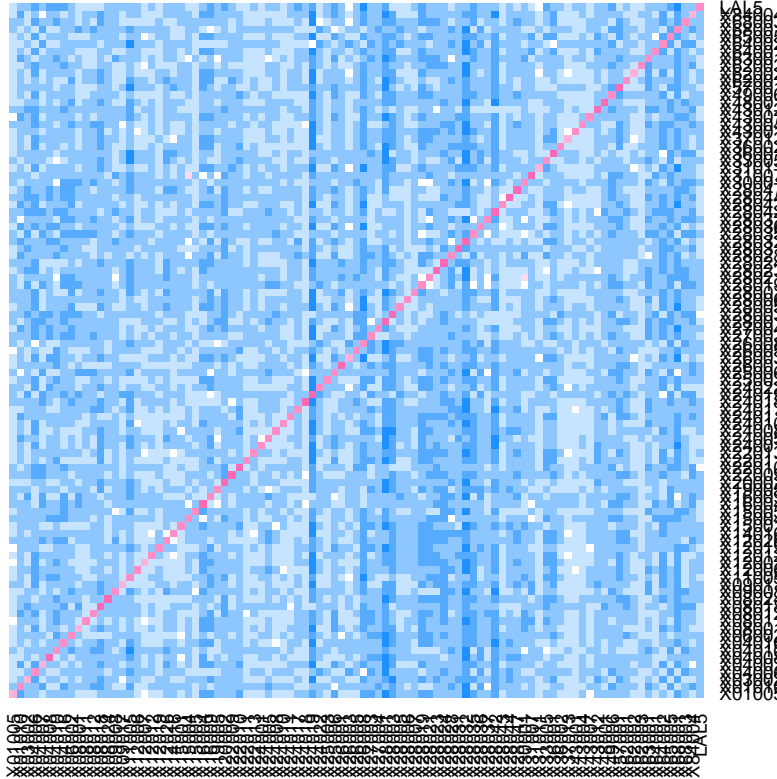
```
library(dendextend)
```

33

```
## The following object is masked from 'package:stats':
##
##      cutree
```

```
dist1 = dist(t(eData_filt1))
```

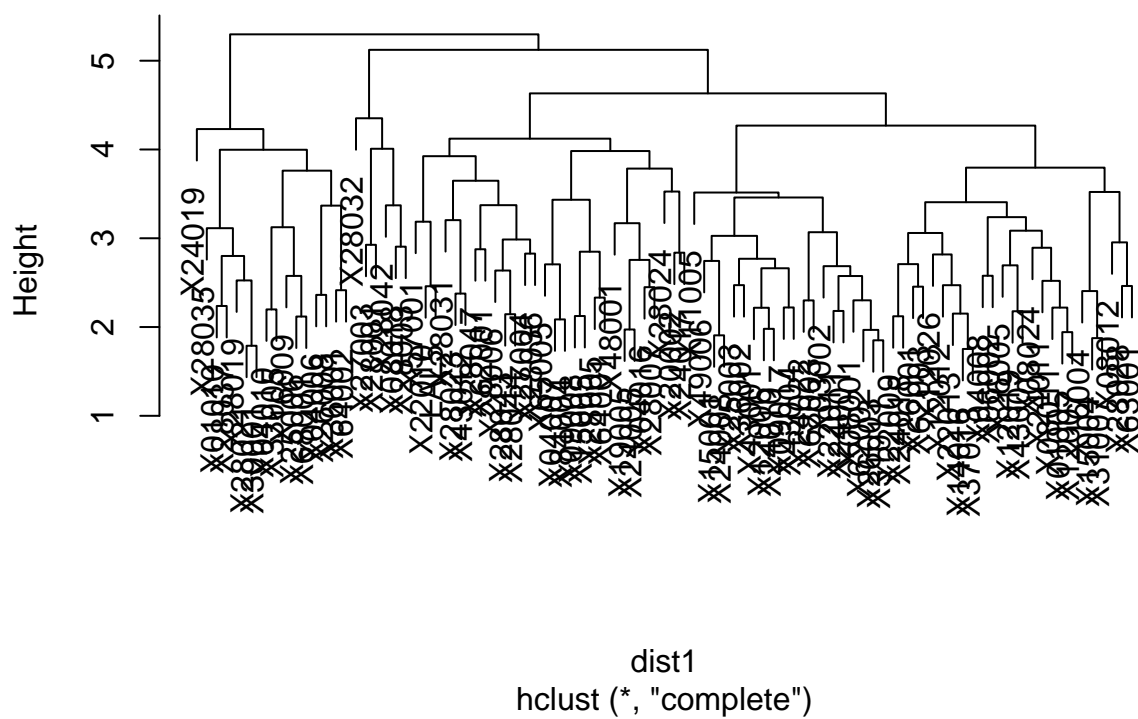
```
heatmap(as.matrix(dist1), col=colramp, Colv = NA, Rowv = NA)
```



```
hclust1 = hclust(dist1)
```

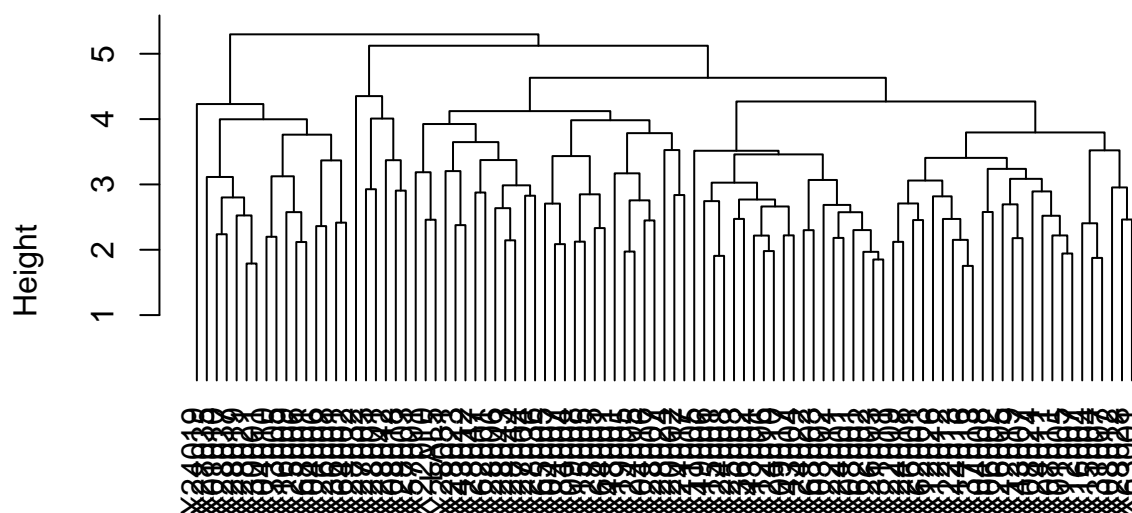
```
plot(hclust1)
```

Cluster Dendrogram



```
plot(hclust1, hang=-1)
```

Cluster Dendrogram



```
dist1
hclust (*, "complete")
```

```
dend = as.dendrogram(hclust1)
```

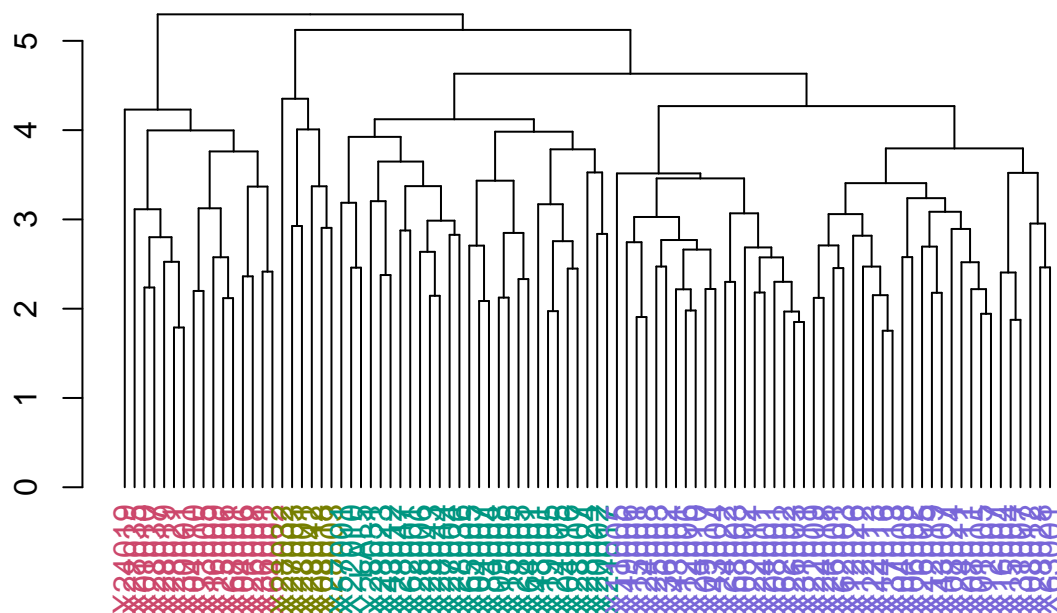
```
dend = color_labels(hclust1,4,1:14)
```

```
## Loading required namespace: colorspace
```

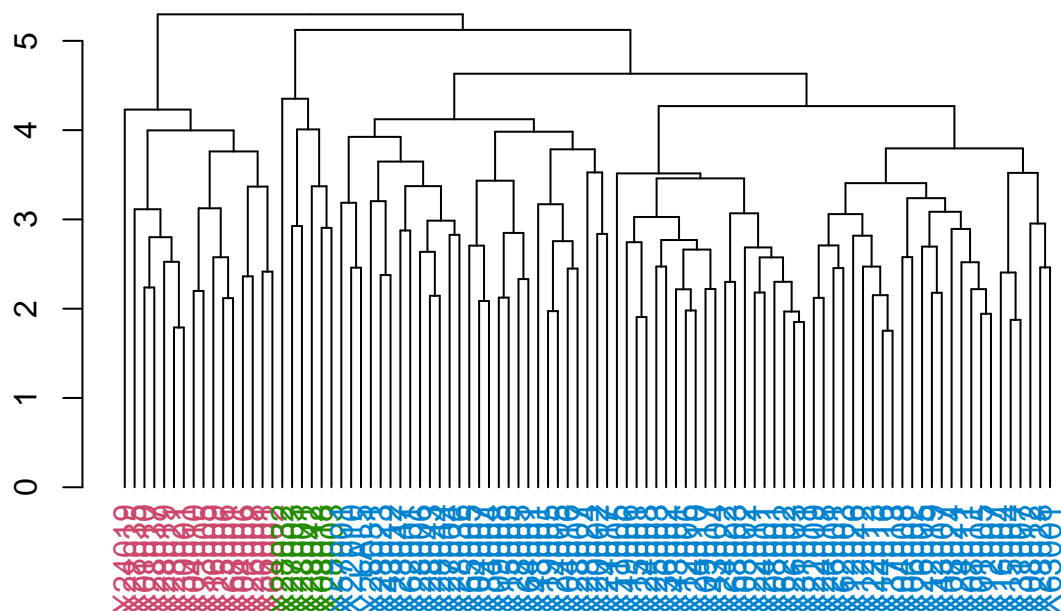
```
dend = color_labels(hclust1,4,1:4)
```

```
dend = color_labels(hclust1,4,1:4)
```

```
plot(dend)
```



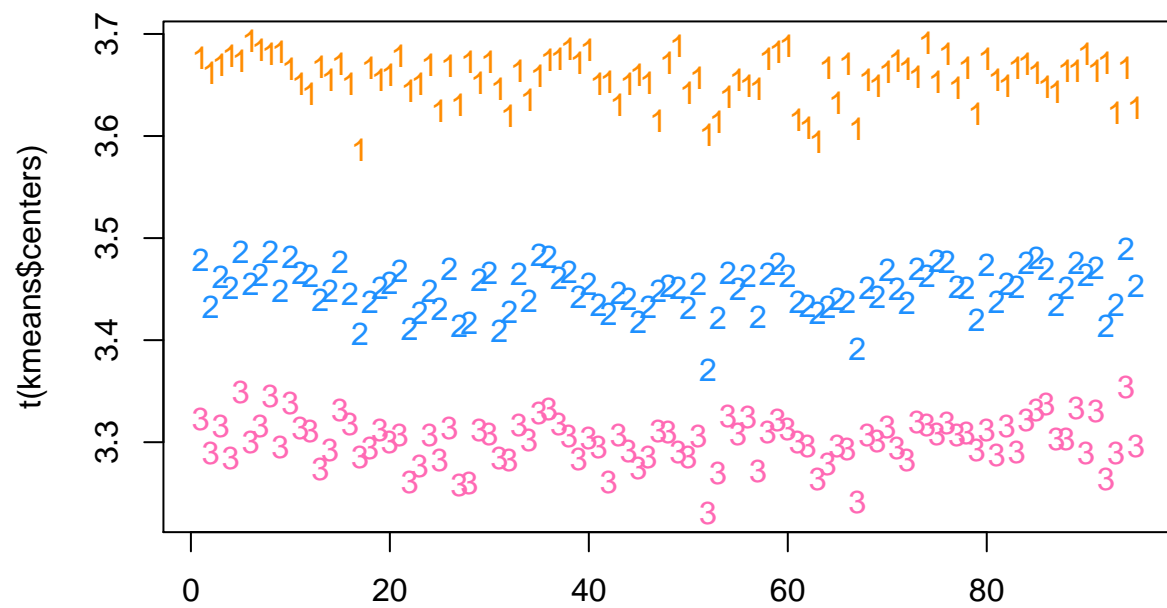
```
dend = color_labels(hclust1,3,1:3)
plot(dend)
```



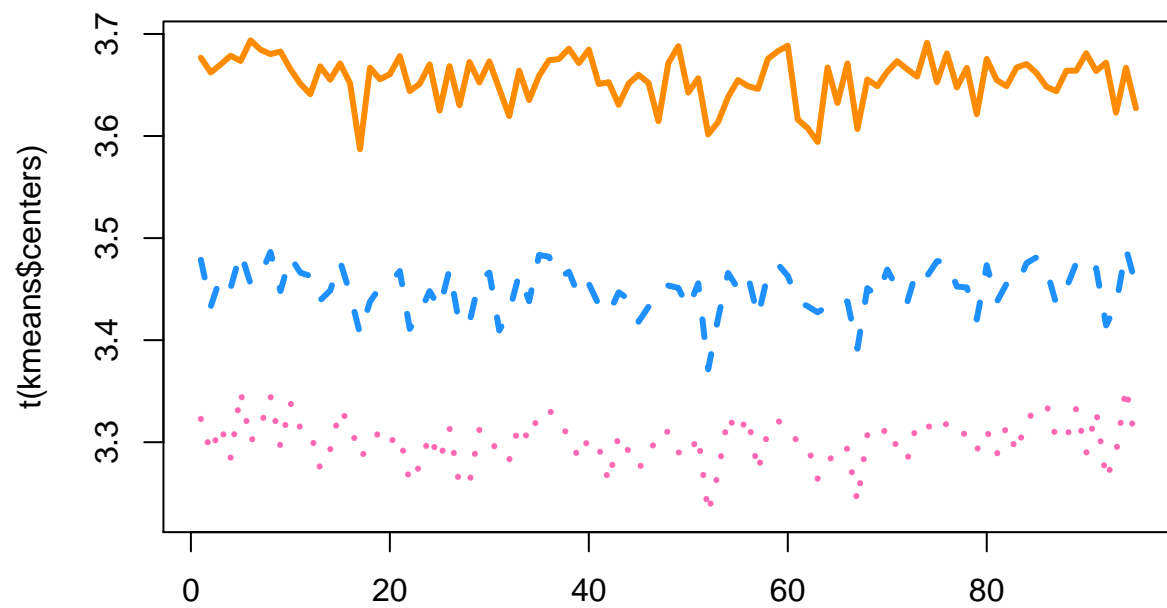
```
kmeans = kmeans(eData_filt1, centers = 3)
names(kmeans)
```

```
## [1] "cluster"      "centers"      "totss"       "withinss"
## [5] "tot.withinss" "betweenss"    "size"        "iter"
## [9] "ifault"
```

```
matplot(t(kmeans$centers))
```



```
matplot(t(kmeans$centers), col=1:3, type = "l", lwd=3)
```



```
table(kmeans$cluster)
```

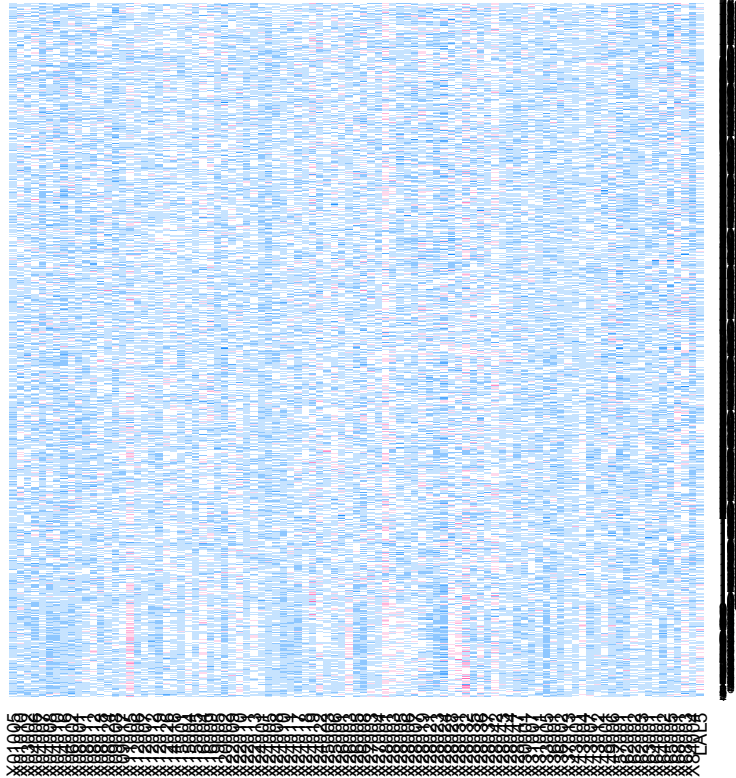
```
##
##  1  2  3
## 141 198 422
```

```
table(kmeans$cluster)
```

```
##
##  1  2  3
## 141 198 422
```

```
newdata = as.matrix(eData_filt1)[order(kmeans$cluster),]
```

```
heatmap(newdata, col = colramp, Colv = NA, Rowv = NA)
```

Singular Decomposition

```
edata_centered = eData_filt1 - rowMeans(eData_filt1) ##<- centering filtered expression data
```

```
edata_sdv1 = svd(edata_centered)
names(edata_sdv1)
```

```
## [1] "d" "u" "v"
```

```
##<- 'd' is the diagonal matrix, returns diagonal matrix , 'v' and 'u' components tells you the variation
dim(eData_filt1)
```

```
## [1] 761 95
```

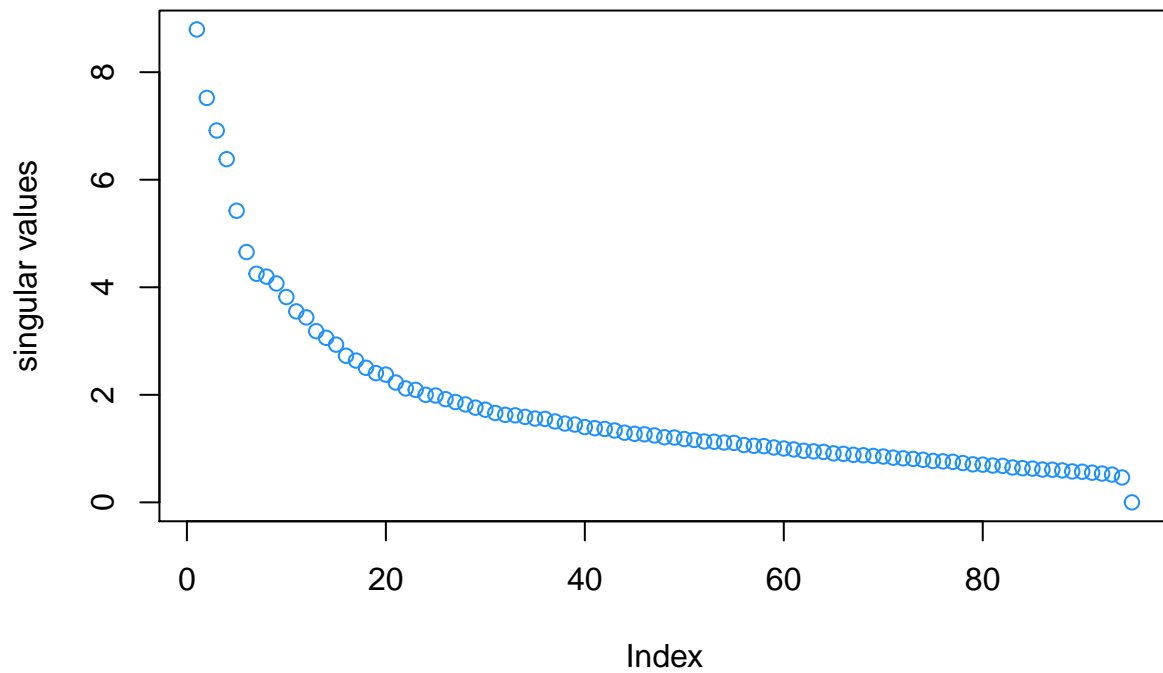
```
dim(edata_sdv1$u)
```

```
## [1] 761 95
```

```
dim(edata_sdv1$v)
```

```
## [1] 95 95
```

```
###<- plotting singular value of our expression data.  
plot(edata_sdv1$d, ylab="singular values", col=2)
```



```
plot(edata_sdv1$d^2/sum(edata_sdv1$d^2), ylab="% Variance explained", col=2)
```

