Intro to ALL data for Bioc monograph

VJ Carey

May 2, 2004

1 Introduction

This document is for authors of the Bioc monograph, it just goes over various aspects of the ALL data. Example analyses can be added here for illustration.

2 Attachment and data list

```
> library(ALL)

Loading required package: Biobase
Welcome to Bioconductor
         Vignettes contain introductory material.  To view,
         simply type: openVignette()
         For details on reading vignettes, see
         the openVignette help page.

> data(ALL)
> show(ALL)

Expression Set (exprSet) with
        12625 genes
        128 samples
                 phenoData object with 21 variables and 128 cases
         varLabels
                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:  Original remisson data
```

```
date.cr:  Date complete remission if achieved
t(4;11):  did the patient have t(4;11) translocation. Derived from cito
t(9;22):  did the patient have t(9;22) translocation. Derived from cito
cyto.normal:  Was cytogenetic test normal? Derived from citog
citog:  original citogenetics data, deletions or t(4;11), t(9;22) statu
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
f.u:  follow up data available
date last seen:  date patient was last seen
```

3 Tables and graphs for phenodata

> print(summary(pData(ALL)))

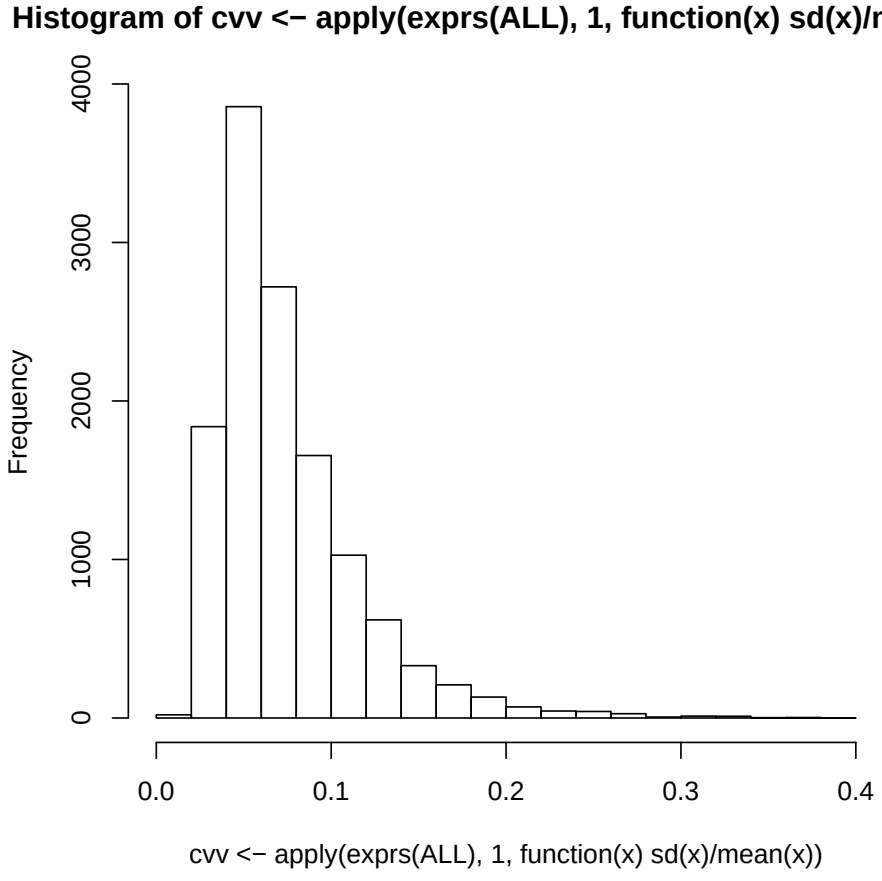
```
    cod             diagnosis           sex          age               BT
Length:128         Length:128         F   :42   Min.   : 5.00   B2     :36
Class :character   Class :character   M   :83   1st Qu.:19.00   B3     :23
Mode  :character   Mode  :character   NA's: 3   Median :29.00   B1     :19
                                                Mean   :32.37   T2     :15
                                                3rd Qu.:45.50   B4     :12
                                                Max.   :58.00   T3     :10
                                                NA's   : 5.00   (Other):13
remission      CR              date.cr          t(4;11)
CR  :99   Length:128         Length:128         Mode :logical
REF :15   Class :character   Class :character   FALSE:86
NA's:14   Mode  :character   Mode  :character   TRUE :7
                                                NA's :35


 t(9;22)        cyto.normal        citog              mol.biol
Mode :logical   Mode :logical   Length:128         ALL1/AF4:10
FALSE:67        FALSE:69        Class :character   BCR/ABL :37
TRUE :26        TRUE :24        Mode  :character   E2A/PBX1: 5
NA's :35        NA's :35                           NEG     :74
                                                   NUP-98  : 1
                                                   p15/p16 : 1
```

```
  fusion protein     mdr                kinet       ccr
p190     :17     Length:128         dyploid:94   Mode :logical
p190/p210: 8     Class :character   hyperd.:27   FALSE:74
p210     : 8     Mode  :character   NA's   : 7   TRUE :26
NA's     :95                                     NA's :28


 relapse        transplant          f.u            date last seen
Mode :logical   Mode :logical   Length:128         Length:128
FALSE:35        FALSE:91        Class :character   Class :character
TRUE :65        TRUE :9         Mode  :character   Mode  :character
NA's :28        NA's :28
```

```
> hist(cvv <- apply(exprs(ALL), 1, function(x) sd(x)/mean(x)))
```



```
> ok <- cvv > 0.08 & cvv < 0.18
> fALL <- ALL[ok, ]
> show(fALL)

Expression Set (exprSet) with
        3841 genes
        128 samples
                 phenoData object with 21 variables and 128 cases
         varLabels
                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:  Original remisson data
                date.cr:  Date complete remission if achieved
                t(4;11):  did the patient have t(4;11) translocation. Derived from cito
                t(9;22):  did the patient have t(9;22) translocation. Derived from cito
                cyto.normal:  Was cytogenetic test normal? Derived from citog
                citog:  original citogenetics data, deletions or t(4;11), t(9;22) statu
                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
                f.u:  follow up data available
                date last seen:  date patient was last seen

> allx2 <- data.frame(t(exprs(fALL)), class = ALL$BT)

> library(randomForest)

Attaching package 'randomForest':


        The following object(s) are masked from package:Biobase :

         combine

> rf1 <- randomForest(class ~ ., data = allx2)
> print(rf1)
```

```
Call:
 randomForest.formula(x = class ~ ., data = allx2)
               Type of random forest: classification
                     Number of trees: 500
No. of variables tried at each split: 61

        OOB estimate of  error rate: 43.75%
Confusion matrix:
   B B1 B2 B3 B4 T T1 T2 T3 T4 class.error
B  0  0  3  2  0 0  0  0  0  0   1.0000000
B1 0 11  5  3  0 0  0  0  0  0   0.4210526
B2 0  1 32  3  0 0  0  0  0  0   0.1111111
B3 0  1  8 12  2 0  0  0  0  0   0.4782609
B4 0  0  5  7  0 0  0  0  0  0   1.0000000
T  0  0  0  0  0 0  0  5  0  0   1.0000000
T1 0  0  0  0  0 0  0  1  0  0   1.0000000
T2 0  0  0  0  0 0  0 12  3  0   0.2000000
T3 0  0  0  0  0 0  0  5  5  0   0.5000000
T4 0  0  0  0  0 0  0  2  0  0   1.0000000
```