

# Intro to ALL data for Bioc monograph

VJ Carey

June 6, 2007

## 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)
> data(ALL)
> show(ALL)
```

```
ExpressionSet (storageMode: lockedEnvironment)
assayData: 12625 features, 128 samples
  element names: exprs
phenoData
  sampleNames: 01005, 01010, ..., LAL4 (128 total)
  varLabels and varMetadata:
    cod: Patient ID
    diagnosis: Date of diagnosis
    ...: ...
    date last seen: date patient was last seen
    (21 total)
featureData
  featureNames: 1000_at, 1001_at, ..., AFFX-YEL024w/RIP1_at (12625 total)
  varLabels and varMetadata: none
experimentData: use 'experimentData(object)'
pubMedIds: 14684422 16243790
Annotation [1] "hgu95av2"
```

### 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

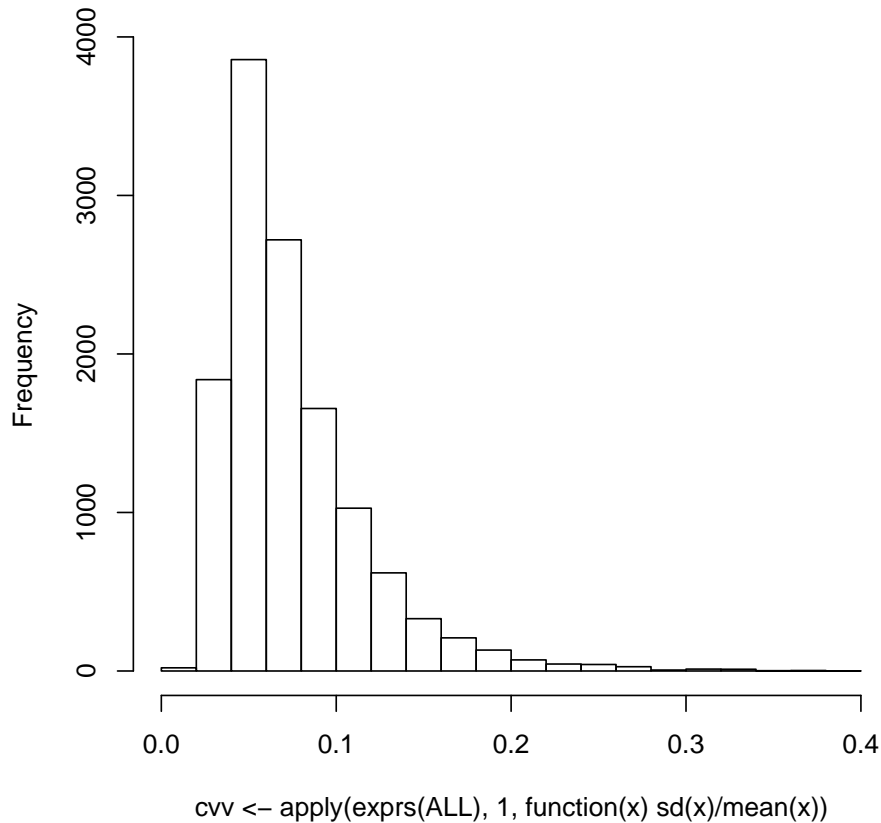
t(9;22)	cyto.normal	citog	mol.biol
Mode :logical	Mode :logical	Length:128	ALL1/AF4:10
FALSE:67	FALSE:69	Class :character	BCR/ABL :37
TRUE :26	TRUE :24	Mode :character	E2A/PBX1: 5
NA's :35	NA's :35		NEG :74
			NUP-98 : 1
			p15/p16 : 1

fusion protein	mdr	kinet	ccr	relapse
p190 :17	NEG :101	dyploid:94	Mode :logical	Mode :logical
p190/p210: 8	POS : 24	hyperd.:27	FALSE:74	FALSE:35
p210 : 8	NA's: 3	NA's : 7	TRUE :26	TRUE :65
NA's :95			NA's :28	NA's :28

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

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

### Histogram of `cvv <- apply(exprs(ALL), 1, function(x) sd(x)/mean`



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

```
ExpressionSet (storageMode: lockedEnvironment)
assayData: 3841 features, 128 samples
  element names: exprs
phenoData
  sampleNames: 01005, 01010, ..., LAL4 (128 total)
  varLabels and varMetadata:
    cod: Patient ID
    diagnosis: Date of diagnosis
    ....: ...
    date last seen: date patient was last seen
    (21 total)
featureData
  rowNames: 1005_at, 1007_s_at, ..., AFFX-YEL024w/RIP1_at (3841 total)
```

```

varLabels and varMetadata: none
experimentData: use 'experimentData(object)'
pubMedIds: 14684422 16243790
Annotation [1] "hgu95av2"

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

> library(randomForest)

randomForest 4.5-18
Type rfNews() to see new features/changes/bug fixes.

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

```

Call:

```

randomForest(formula = 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: 42.52%

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	2	31	3	0	0	0	0	0	0	0.1388889
B3	0	1	9	12	1	0	0	0	0	0	0.4782609
B4	0	0	6	6	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	0	0	0	NaN
T2	0	0	0	0	0	0	0	13	2	0	0.1333333
T3	0	0	0	0	0	0	0	4	6	0	0.4000000
T4	0	0	0	0	0	0	0	2	0	0	1.0000000