

# 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

## Welcome to Bioconductor
## 
## Vignettes contain introductory material; view with
## 'browseVignettes()'. To cite Bioconductor, see
## 'citation("Bioconductor")', 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”, .. . . . .@ remission : Factor w/ 2 levels “CR”, “REF”: 1 1 1 1 1 1 1 1 1 1 .. . . . .@ CR : chr [1:128] “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”, .. . . . .@ fusion protein: Factor w/ 3 levels “p190”, “p190/p210”, .. . . . .@ 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 remission 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 citogenetics 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] "1110_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 computated.

```

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

```

ALL phenotype data

```

ALL1 <- data.frame(ALL)
length(colnames(ALL1))

```

Converting ALL data into a data frame

```

## [1] 12646

head(colnames(ALL1))

## [1] "X1000_at"    "X1001_at"    "X1002_f_at"  "X1003_s_at"  "X1004_at"
## [6] "X1005_at"

tail(colnames(ALL1))

## [1] "kinet"          "ccr"            "relapse"        "transplant"
## [5] "f.u"            "date.last.seen"

```

Subsetting B-cell patients from the ALL dataset

```

ALL$BT

## [1] B2 B2 B4 B1 B2 B1 B1 B2 B2 B3 B3 B2 B3 B 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 B4 B4 B2 B2 B2
## [47] B4 B2 B1 B2 B2 B3 B4 B3 B3 B4 B3 B3 B1 B1 B1 B1 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 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]

```

```

table(pData(b.cell)$sex)

```

```

##
## F M
## 34 59

```

```



```

### Reducing genes based on variability

```

library(genefilter)

##
## Attaching package: 'genefilter'

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

```

```

library(RColorBrewer)

B.cell1 = 2^exprs(b.cell)
mean_gene = rowMeans(B.cell1)
sd_gene = rowSds(B.cell1)
cv = sd_gene/mean_gene
blues.ramp = colorRampPalette(brewer.pal(9, "Reds")[-1])
dCol = densCols(log(mean_gene), log(sd_gene), colramp = blues.ramp)
par(mfrow=c(2,2))

plot(mean_gene, sd_gene, log='xy', col=dCol, pch = 16, cex=0.1)
abline(v=100, lwd=3, col='blue')

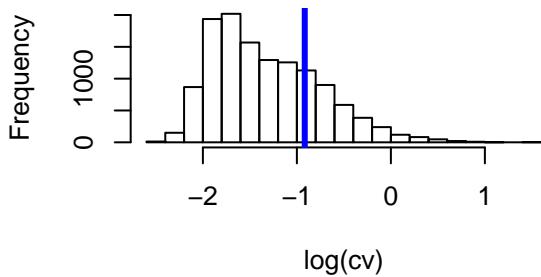
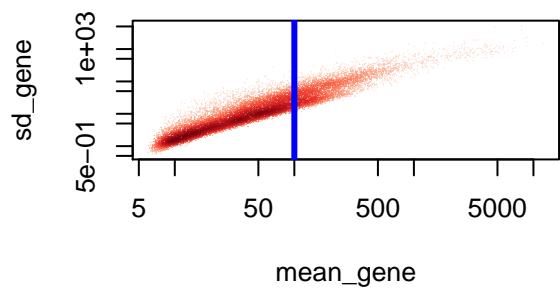
hist(log(cv), main = NA)
abline(v=log(0.4), lwd=3, col='blue')

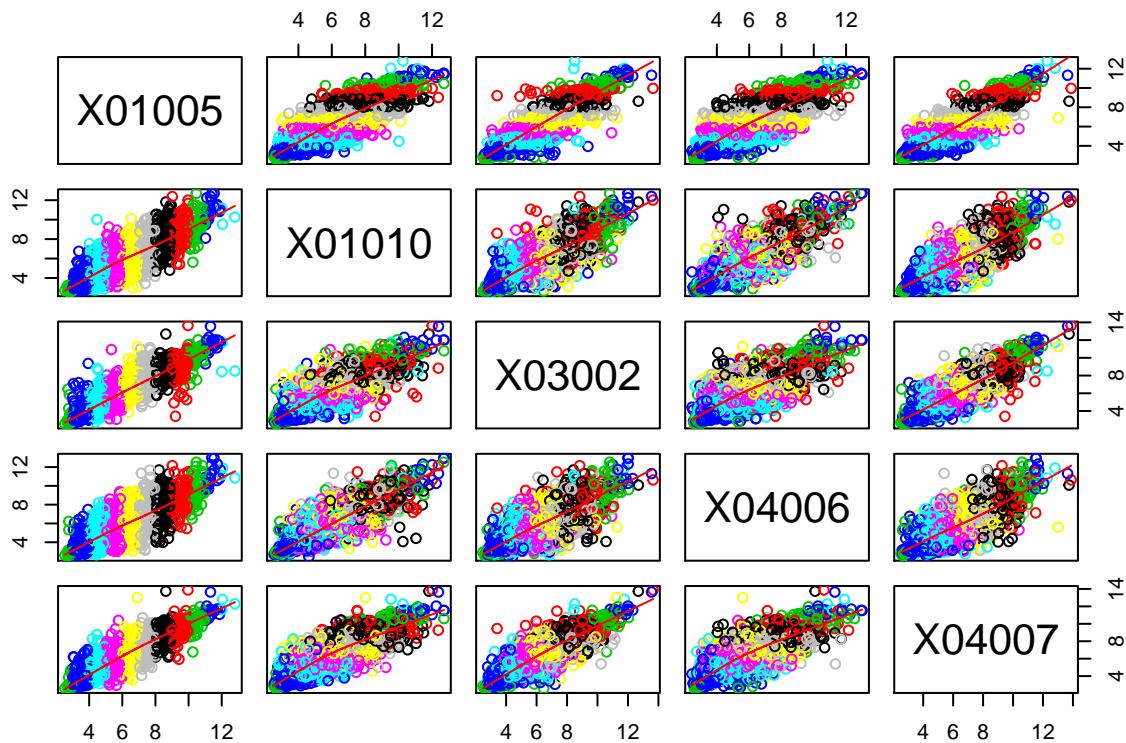
B.genes = filterfun(cv(0.4,10))
B.gene.fil = genefilter(B.cell1, B.genes)
B.cell2 = log2(B.cell1[B.gene.fil,])

##correlation of first 5 reduced genes starting with x01005

B.genes = filterfun(cv(0.4,10))
B.gene.fil = genefilter(B.cell1, B.genes)
B.cell2 = log2(B.cell1[B.gene.fil,])
B.cell2_df = data.frame(B.cell2)
B.cell2_df_vars = B.cell2_df[, c(1:5)]
mar.orig = par()$mar
par(mar=c(1,1,1,1))
pairs(B.cell2_df_vars,panel = panel.smooth, col= 9 + B.cell2_df$X01005)

```





```

BT = which(as.character(ALL$BT) %in% c("B", "T"))
Bcell = ALL[, intersect(anyB, BT)]
Bcell$BT = factor(Bcell$BT)
str(Bcell$BT)

## Factor w/ 1 level "B": 1 1 1 1 1

B.expressionData = exprs(Bcell)
str(B.expressionData)

## num [1:12625, 1:5] 7.76 4.99 4.05 6.1 6.21 ...
## - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:12625] "1000_at" "1001_at" "1002_f_at" "1003_s_at" ...
##   ..$ : chr [1:5] "09017" "22009" "22010" "84004" ...

B.sds = rowSds(B.expressionData)

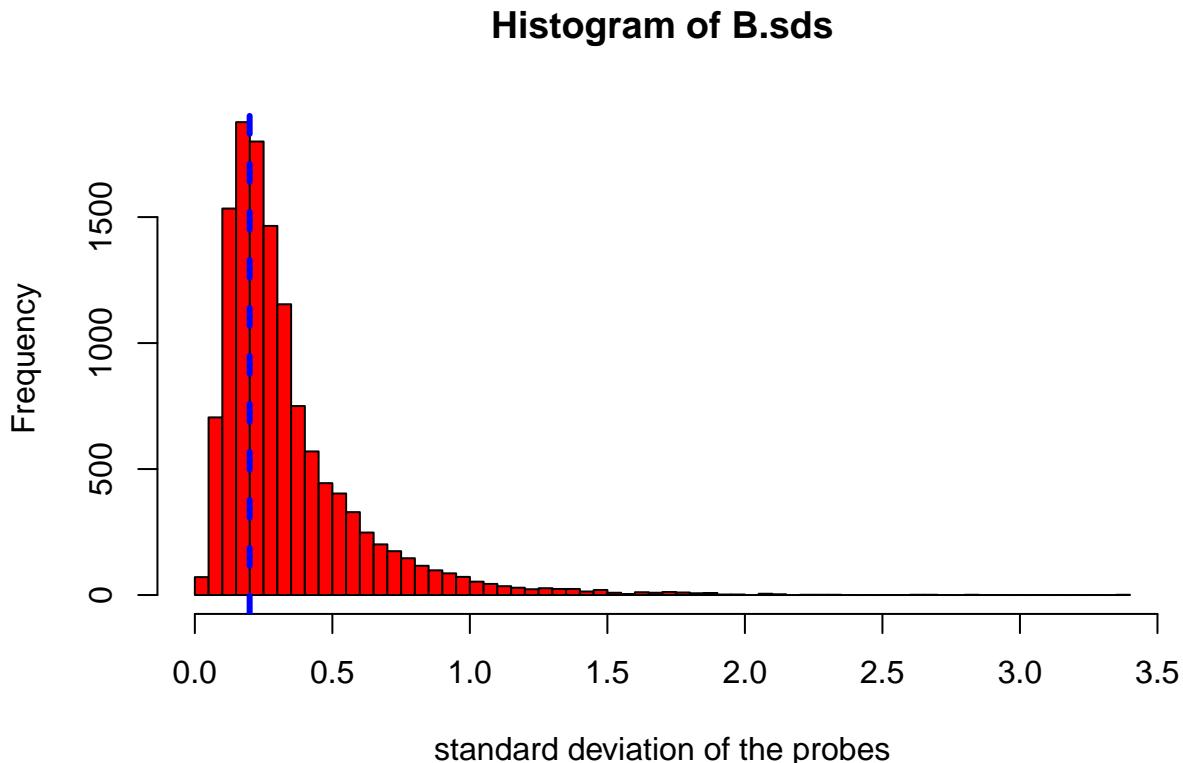
hist(B.sds, breaks=50, col="red", xlab="standard deviation of the probes")

B.shorthen = shorth(B.sds)
B.shorthen

## [1] 0.1992332

```

```
abline(v=B.shorthen, col="blue", lwd=3, lty=2)
```



```
Bcellfiltered = Bcell[B.sds>=B.shorthen]
```

```
dim(exprs(Bcellfiltered))
```

```
## [1] 8467      5
```

```
B.sds_new = rowSds(exprs(Bcellfiltered))
```

```
hist(B.sds_new, breaks=50, col="blue", xlab = "new distribution of standard deviation")
```

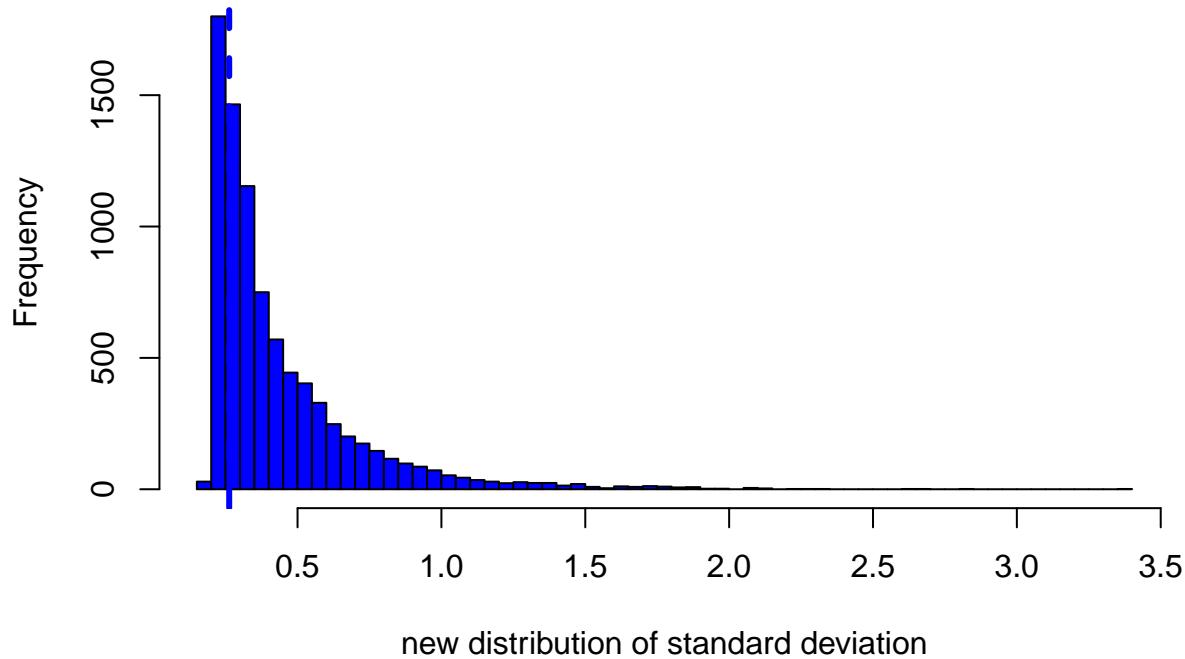
```
B.shorthen_new = shorth(B.sds_new)
```

```
B.shorthen_new
```

```
## [1] 0.2624267
```

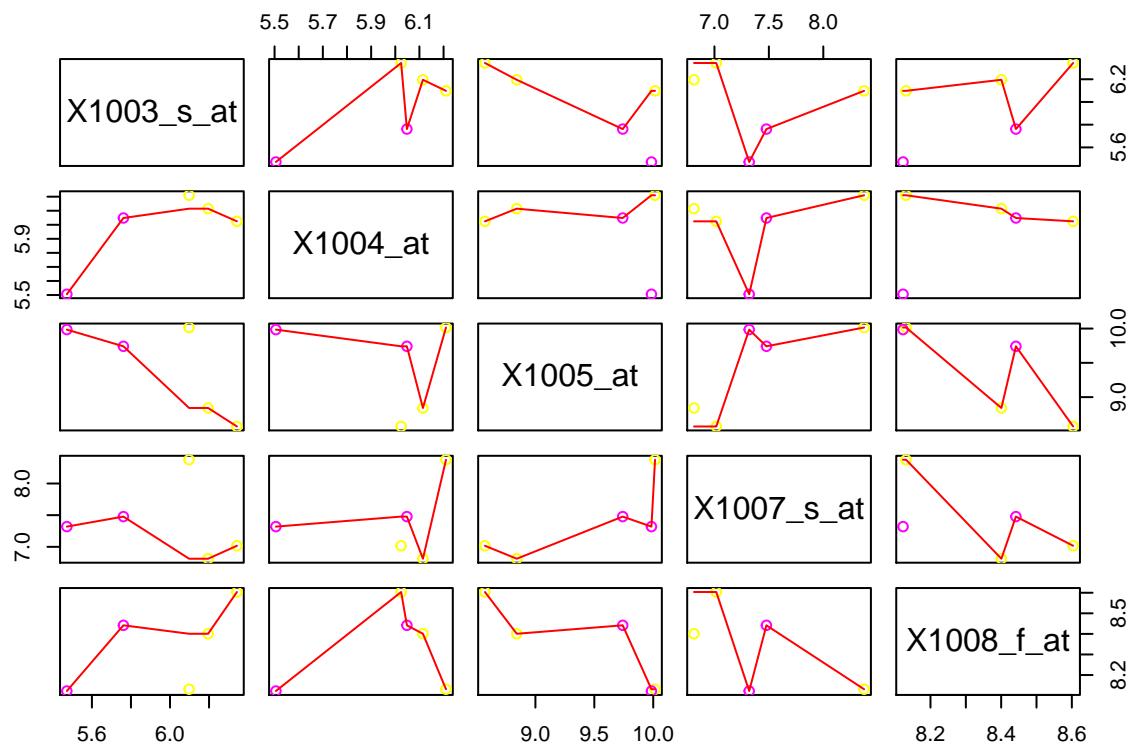
```
abline(v=B.shorthen_new, col="blue", lwd=3, lty=2)
```

## Histogram of B.sds\_new

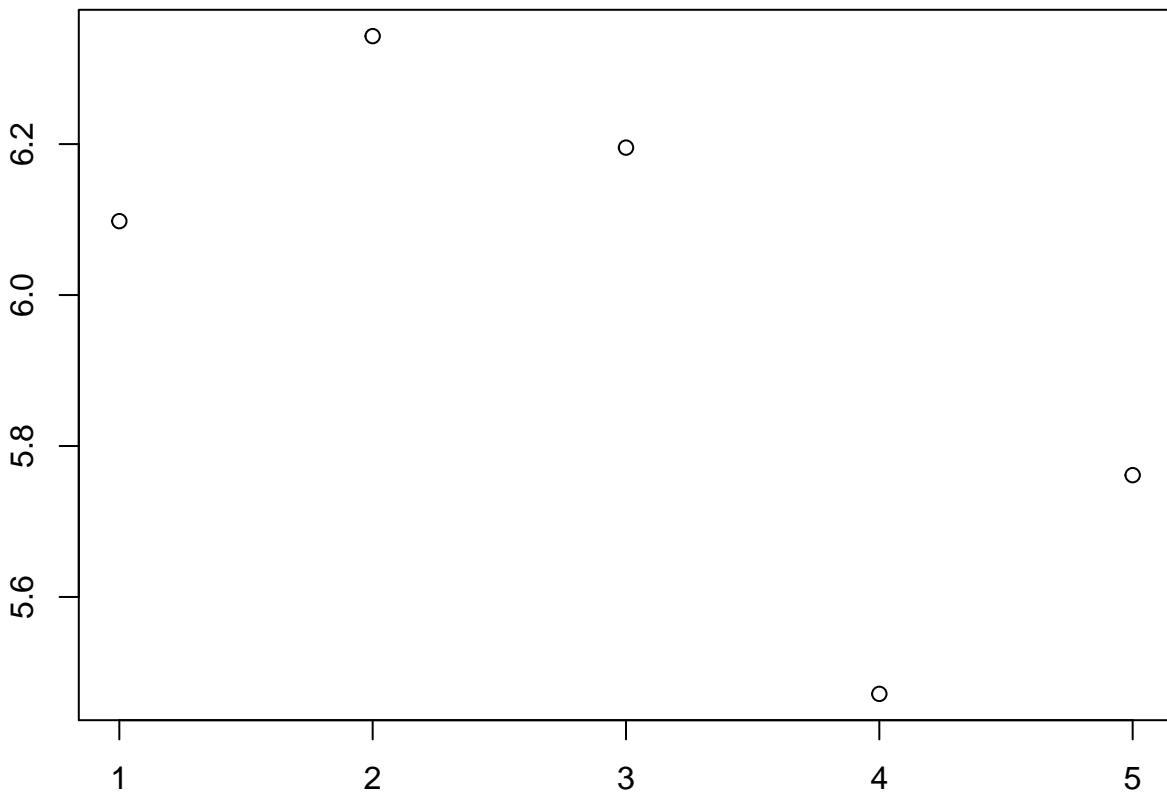


Correlation of the first 5 furthered reduced genes

```
Bcellfiltered_df = data.frame(Bcellfiltered)
Bcellfiltered_vars = Bcellfiltered_df[, c(1:5)]
mar.orig = par()$mar
par(mar = c(1,1,1,1))
pairs(Bcellfiltered_vars, panel = panel.smooth, col= 9 + Bcellfiltered_df$X1003_s_at)
```



```
par(mar = rep(2,4))
plot(Bcellfiltered_df$X1003_s_at)
```



```
summary(lm(Bcellfiltered_df$X1003_s_at~Bcellfiltered_df$X1004_at))

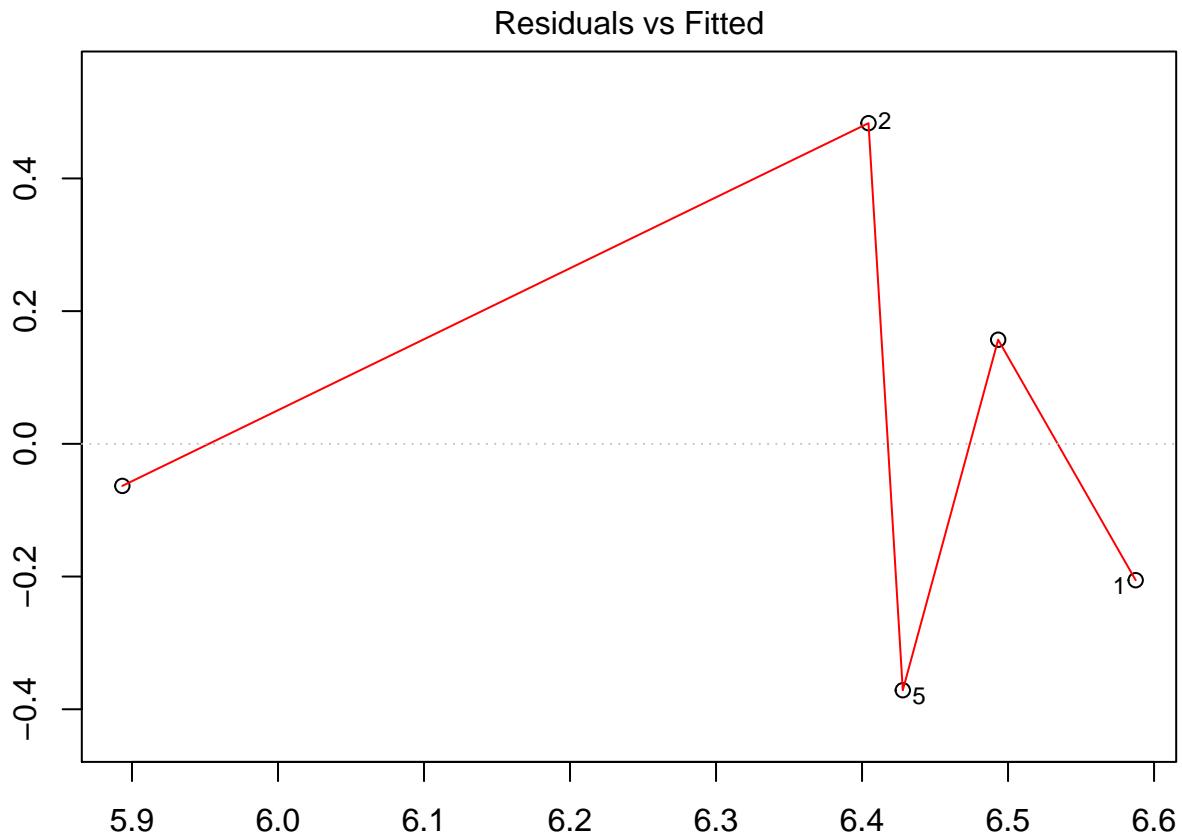
##
## Call:
## lm(formula = Bcellfiltered_df$X1003_s_at ~ Bcellfiltered_df$X1004_at)
##
## Residuals:
##      1       2       3       4       5 
## -0.10229  0.32601  0.08937 -0.03392 -0.27917
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 0.0772    2.8354   0.027   0.980    
## Bcellfiltered_df$X1004_at 0.9860    0.4737   2.081   0.129    
## 
## Residual standard error: 0.2607 on 3 degrees of freedom
## Multiple R-squared:  0.5908, Adjusted R-squared:  0.4545 
## F-statistic: 4.332 on 1 and 3 DF,  p-value: 0.1288
```

```
Bcell_fit = lm(Bcellfiltered_df$X101_at~Bcellfiltered_df$X1004_at)
anova(Bcell_fit)
```

```
## Analysis of Variance Table
##
## Response: Bcellfiltered_df$X101_at
```

```
##          Df  Sum Sq Mean Sq F value Pr(>F)
## Bcellfiltered_df$X1004_at    1 0.29377 0.29377  1.9926 0.2529
## Residuals                      3 0.44230 0.14743

par(mar = rep(2,4))
plot(Bcell_fit)
```



Normal Q–Q

