Lab 1: Bioconductor Basics

June 4, 2003

In this laboratory we will introduce some of the basic interactions with Bioconductor.

> library(Biobase)

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

Creating a new generic function for "summary" in package reposTools

> library(annotate)
> library(golubEsets)

The package golubEsets contains three data sets that were obtained from the web and slightly massaged. They represent the data analysed in ? to perform class prediction using microarray data. The data were collected on Affymetrix Hu 6800 chip and which contains probes for 7129 genes.

An exprSet basically consists of the gene expression matrix (optionally a set of standard errors for those estimates), the related experimental metadata (who did what when and to what), and the phenotypic data. Here phenotype is interpreted quite broadly – it represents any physical characteristics of the sample.

> data(golubTrain)
> golubTrain

Expression Set (exprSet) with
7129 genes
38 samples
phenoData object with 11 variables and 38 cases
varLabels
Samples: Sample index
ALL.AML: Factor, indicating ALL or AML
BM.PB: Factor, sample from marrow or peripheral blood
T.B.cell: Factor, T cell or B cell leuk.
FAB: Factor, FAB classification
Date: Date sample obtained
Gender: Factor, gender of patient
pctBlasts: pct of cells that are blasts
Treatment: response to treatment
PS: Prediction strength
Source: Source of sample

> golubTrain[, 1:10]

Expression Set (exprSet) with
  7129 genes
  10 samples

  phenoData object with 11 variables and 10 cases

varLabels
Samples: Sample index
ALL.AML: Factor, indicating ALL or AML
BM.PB: Factor, sample from marrow or peripheral blood
T.B.cell: Factor, T cell or B cell leuk.
FAB: Factor, FAB classification
Date: Date sample obtained
Gender: Factor, gender of patient
pctBlasts: pct of cells that are blasts
Treatment: response to treatment
PS: Prediction strength
Source: Source of sample

> golubTrain[1:100, ]

Expression Set (exprSet) with
  100 genes
  38 samples

  phenoData object with 11 variables and 38 cases

varLabels
Samples: Sample index
ALL.AML: Factor, indicating ALL or AML
BM.PB: Factor, sample from marrow or peripheral blood
T.B.cell: Factor, T cell or B cell leuk.
FAB: Factor, FAB classification
Date: Date sample obtained
Gender: Factor, gender of patient
pctBlasts: pct of cells that are blasts
Treatment: response to treatment
PS: Prediction strength
Source: Source of sample

Notice that when subsetting we have arranged it so that the rows correspond to genes and the columns correspond to samples.

The phenotypic data are stored in a separate, but linked, data frame. You can obtain it and interact with it using specific methods.

```r
> pD <- phenoData(golubTrain)
> pD

phenoData object with 11 variables and 38 cases

varLabels
   Samples: Sample index
   ALL.AML: Factor, indicating ALL or AML
   BM.PB: Factor, sample from marrow or peripheral blood
   T.B.cell: Factor, T cell or B cell leuk.
   FAB: Factor, FAB classification
   Date: Date sample obtained
   Gender: Factor, gender of patient
   pctBlasts: pct of cells that are blasts
   Treatment: response to treatment
   PS: Prediction strength
   Source: Source of sample

> pd <- pData(pD)
> pd

        Samples ALL.AML BM.PB T.B.cell   FAB Date Gender pctBlasts Treatment
1         1     ALL     BM   B-cell <NA> 9/4/1996   M    NA          <NA>
2         2     ALL     BM   T-cell <NA> <NA>    M    NA          <NA>
3         3     ALL     BM   T-cell <NA> <NA>    M    NA          <NA>
4         4     ALL     BM   B-cell <NA> <NA>    <NA>  NA          <NA>
5         5     ALL     BM   B-cell <NA> <NA>    <NA>  NA          <NA>
6         6     ALL     BM   T-cell <NA> <NA>    M    NA          <NA>
7         7     ALL     BM   B-cell <NA> 3/25/1983 F    NA          <NA>
8         8     ALL     BM   B-cell <NA> <NA>    F    NA          <NA>
9         9     ALL     BM   T-cell <NA> <NA>    M    NA          <NA>
10        10    ALL     BM   T-cell <NA> 7/23/1987 M    NA          <NA>
```
<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>12</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>9/17/1985</td>
<td>F</td>
<td>NA</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>7/27/1985</td>
<td>F</td>
<td>NA</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>ALL</td>
<td>BM</td>
<td>T-cell</td>
<td>&lt;NA&gt;</td>
<td>11/27/1987</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>3/25/1989</td>
<td>F</td>
<td>NA</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>2/12/1990</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>9/26/1990</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>F</td>
<td>NA</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>NA</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>NA</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>1/24/1984</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>5/27/1988</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>23</td>
<td>23</td>
<td>ALL</td>
<td>BM</td>
<td>T-cell</td>
<td>&lt;NA&gt;</td>
<td>7/9/1991</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>5/19/1981</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>2/18/1982</td>
<td>M</td>
<td>NA</td>
</tr>
<tr>
<td>26</td>
<td>26</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>F</td>
<td>NA</td>
</tr>
<tr>
<td>27</td>
<td>27</td>
<td>ALL</td>
<td>BM</td>
<td>B-cell</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>F</td>
<td>NA</td>
</tr>
<tr>
<td>34</td>
<td>34</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M2</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>77</td>
</tr>
<tr>
<td>35</td>
<td>35</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M1</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>67</td>
</tr>
<tr>
<td>36</td>
<td>36</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M5</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>76</td>
</tr>
<tr>
<td>37</td>
<td>37</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M2</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>44</td>
</tr>
<tr>
<td>38</td>
<td>38</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M1</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>80</td>
</tr>
<tr>
<td>28</td>
<td>28</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M2</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>79</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M2</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>34</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M5</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>93</td>
</tr>
<tr>
<td>31</td>
<td>31</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M4</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>77</td>
</tr>
<tr>
<td>32</td>
<td>32</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M1</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>86</td>
</tr>
<tr>
<td>33</td>
<td>33</td>
<td>AML</td>
<td>BM</td>
<td>&lt;NA&gt;</td>
<td>M2</td>
<td>&lt;NA&gt;</td>
<td>&lt;NA&gt;</td>
<td>70</td>
</tr>
</tbody>
</table>

PS Source
1 1.00  DFCI
2 0.41  DFCI
3 0.87  DFCI
4 0.91  DFCI
5 0.89  DFCI
6 0.76  DFCI
7 0.78  DFCI
8 0.77  DFCI
9 0.89  DFCI
10 0.56  DFCI
11 0.74  DFCI
12 0.20  DFCI
13 1.00  DFCI
An object of class **phenoData** is a combination of a dataframe containing the various data elements and a list that explains what each variable represents. This information is usually relegated to a help page but we felt that it was important to keep it more closely associated with the data.

The `$` operator performs the job of extracting particular variables from an object of class **phenoData**. It also can be used directly on the `exprSet`.

```r
> table(pD$ALL.AML)

ALL  AML
  27   11

> data(golubTest)
> table(golubTest$ALL.AML)

ALL  AML
  20   14
```
The S4 methods package has introduced substantial new capabilities into R. To obtain
the manual pages for S4 classes you should use the following syntax `class?exprSet`.
Please do that now and we will look at help page.

Almost all R functions have a set of runnable examples that are shown at the bottom
of the manual page. You can either scroll down to them and cut-and-paste them across
or use the R function `example` to run them. Try `example(exprSet)`.

To see what packages are currently loaded into your R session you can use `search`.
You can list the functions in any package that is attached by using `objects("package:ts")`,
for example. This will list all the objects in the time series package `ts`. Another useful
command is `find` which will tell you which package contains the definition of a function.