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.

```
> 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>
```

11	11	ALL	BM	T-cell	<NA>	6/25/1985	M	NA	<NA>
12	12	ALL	BM	B-cell	<NA>	9/17/1985	F	NA	<NA>
13	13	ALL	BM	B-cell	<NA>	7/27/1988	F	NA	<NA>
14	14	ALL	BM	T-cell	<NA>	11/27/1987	M	NA	<NA>
15	15	ALL	BM	B-cell	<NA>	3/25/1989	F	NA	<NA>
16	16	ALL	BM	B-cell	<NA>	2/12/1990	M	NA	<NA>
17	17	ALL	BM	B-cell	<NA>	9/26/1990	M	NA	<NA>
18	18	ALL	BM	B-cell	<NA>	<NA>	F	NA	<NA>
19	19	ALL	BM	B-cell	<NA>	<NA>	<NA>	NA	<NA>
20	20	ALL	BM	B-cell	<NA>	<NA>	<NA>	NA	<NA>
21	21	ALL	BM	B-cell	<NA>	1/24/1984	M	NA	<NA>
22	22	ALL	BM	B-cell	<NA>	5/27/1988	M	NA	<NA>
23	23	ALL	BM	T-cell	<NA>	7/9/1991	M	NA	<NA>
24	24	ALL	BM	B-cell	<NA>	5/19/1981	M	NA	<NA>
25	25	ALL	BM	B-cell	<NA>	2/18/1982	M	NA	<NA>
26	26	ALL	BM	B-cell	<NA>	<NA>	F	NA	<NA>
27	27	ALL	BM	B-cell	<NA>	<NA>	F	NA	<NA>
34	34	AML	BM	<NA>	M2	<NA>	<NA>	77	Success
35	35	AML	BM	<NA>	M1	<NA>	<NA>	67	Success
36	36	AML	BM	<NA>	M5	<NA>	<NA>	76	Success
37	37	AML	BM	<NA>	M2	<NA>	<NA>	44	Success
38	38	AML	BM	<NA>	M1	<NA>	<NA>	80	Success
28	28	AML	BM	<NA>	M2	<NA>	<NA>	79	Failure
29	29	AML	BM	<NA>	M2	<NA>	<NA>	34	Failure
30	30	AML	BM	<NA>	M5	<NA>	<NA>	93	Failure
31	31	AML	BM	<NA>	M4	<NA>	<NA>	77	Failure
32	32	AML	BM	<NA>	M1	<NA>	<NA>	86	Failure
33	33	AML	BM	<NA>	M2	<NA>	<NA>	70	Failure

	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

```
14 0.73   DFCI
15 0.98   DFCI
16 0.95   DFCI
17 0.49   DFCI
18 0.59   DFCI
19 0.80   DFCI
20 0.90   DFCI
21 0.76   DFCI
22 0.37   DFCI
23 0.77   DFCI
24 0.92   DFCI
25 0.43   DFCI
26 0.89   DFCI
27 0.82   DFCI
34 0.64  CALGB
35 0.21  CALGB
36 0.94  CALGB
37 0.95  CALGB
38 0.73  CALGB
28 0.44  CALGB
29 0.74  CALGB
30 0.80  CALGB
31 0.61  CALGB
32 0.47  CALGB
33 0.89  CALGB
```

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`.

```
> 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.