Microarray analysis of oral cancer samples

Humberto Ortiz-Zuazaga

April 27, 2011

Bioconductor

- Bioconductor `http://bioconductor.org/`
- Software suite for analysis of biological data
- emphasis on microarray and other high-throughput datasets

Loading packages

Load the `affy` and `limma` packages.

```
> library(limma)
> library(affy)
```

Targets file

A simple text file with tab separated columns can describe the microarray samples.

```
> targets <- readTargets("targets.txt")
> targets
```

```
                          FileName Target
1    OC-1_(HuGene-1_0-st-v1).CEL    pos
2    OC-5_(HuGene-1_0-st-v1).CEL    pos
3    OC-6_(HuGene-1_0-st-v1).CEL    pos
4    OC-7_(HuGene-1_0-st-v1).CEL    pos
5    OC-8_(HuGene-1_0-st-v1).CEL    pos
6   OC-10_(HuGene-1_0-st-v1).CEL    pos
7   OC-11_(HuGene-1_0-st-v1).CEL    neg
8   OC-12_(HuGene-1_0-st-v1).CEL    neg
9   OC-13_(HuGene-1_0-st-v1).CEL    neg
10  OC-14_(HuGene-1_0-st-v1).CEL    neg
11  OC-15_(HuGene-1_0-st-v1).CEL    neg
```

Reading the data

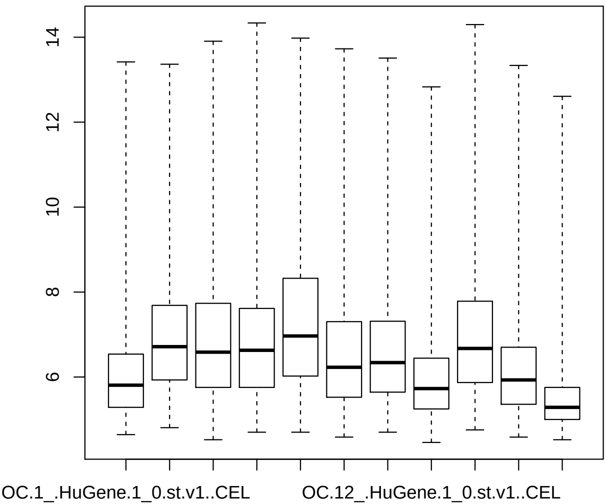
```
> ab <- ReadAffy(filenames = targets$FileName)
```

Normalization and summarization

```
> probeNames(ab)[1:10]

 [1] "7892501" "7892501" "7892501" "7892501" "7892502" "789
 [8] "7892502" "7892503" "7892503"

> eset <- rma(ab)

Background correcting
Normalizing
Calculating Expression

> featureNames(eset)[1:10]

 [1] "7892501" "7892502" "7892503" "7892504" "7892505" "789
 [8] "7892508" "7892509" "7892510"
```

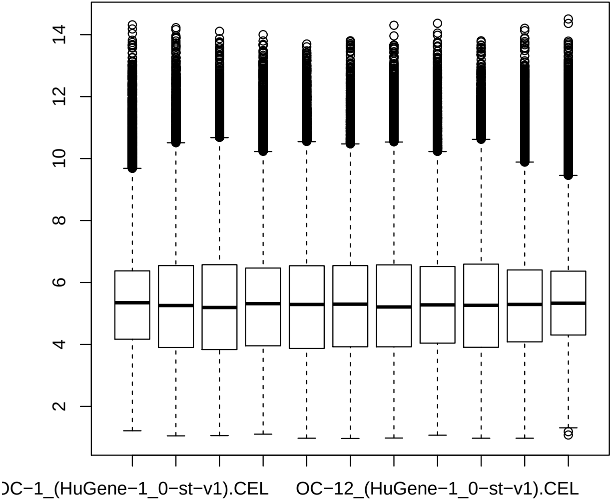
Boxplot before normalization

```
> boxplot(ab)
```



Boxplot after normalization

```
> boxplot(exprs(eset))
```



Describing the design

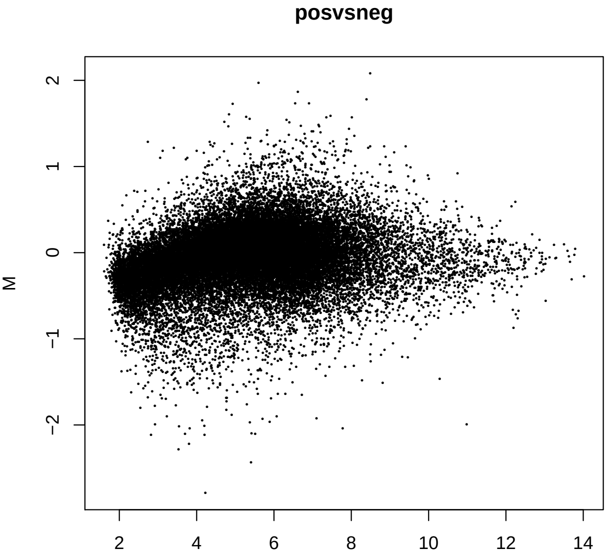
```
> f <- factor(targets$Target, levels = c("pos", "neg"))
> design <- model.matrix(~0 + f)
> colnames(design) <- c("pos", "neg")
> design
   pos neg
1    1   0
2    1   0
3    1   0
4    1   0
5    1   0
6    1   0
7    0   1
8    0   1
9    0   1
10   0   1
11   0   1
attr(,"assign")
[1] 1 1
```

Fitting a model

```
> cont.matrix <- makeContrasts(posvsneg = pos - neg, levels
> fit <- lmFit(eset, design)
> fit2 <- contrasts.fit(fit, cont.matrix)
> fit.b <- eBayes(fit2)
```

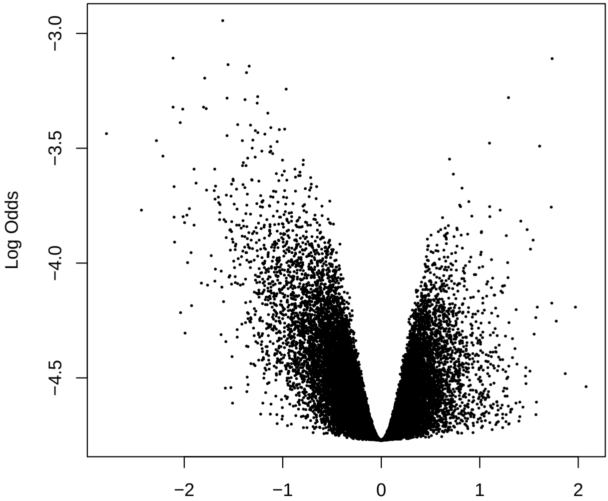
MA plot of model fit

```
> plotMA(fit.b)
```



Volcano plot of model fit

```
> volcanoplot(fit.b)
```



References

- Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 3 (Feb. 2004), 307-315.
- R. Gentleman, V. J. Carey, D. M. Bates, B.Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, and others Bioconductor: Open software development for computational biology and bioinformatics (2004). *Genome Biology*, Vol. 5, R80
- Smyth, G. K. (2005). Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor'. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds), Springer, New York, pages 397–420.