



第4回バイオインフォマティクス実習コース  
横浜市大 先端医科学研究センター  
バイオインフォマティクス研究室

室長 田村智彦  
准教授 中林潤  
免疫学 藩龍馬

- 発現変動遺伝子の抽出

- R package “limma” を使って発現変動遺伝子を抽出
- データ GSE40493 (Affimetrics 正規化済み)  
GSE40493\_Normalized\_ID.txt  
課題配布フォルダ

# コマンド入力時の注意事項

- a) 大文字、小文字は区別する
- b) スペースは入力する必要はない
- c) 配布資料中の **↵** はenterキー
- d) 配布資料中の“¥”はバックスラッシュ
- e) ↑キーで入力のやり直しができる

- PC起動 各自のアカウントでログイン
- R起動 スタートメニュー 4.統計解析ツール
- Proxyの設定

R console



```
> Sys.setenv(http_proxy="http://proxy.yokohama-cu.ac.jp:8080")  
> Sys.getenv("http_proxy")
```

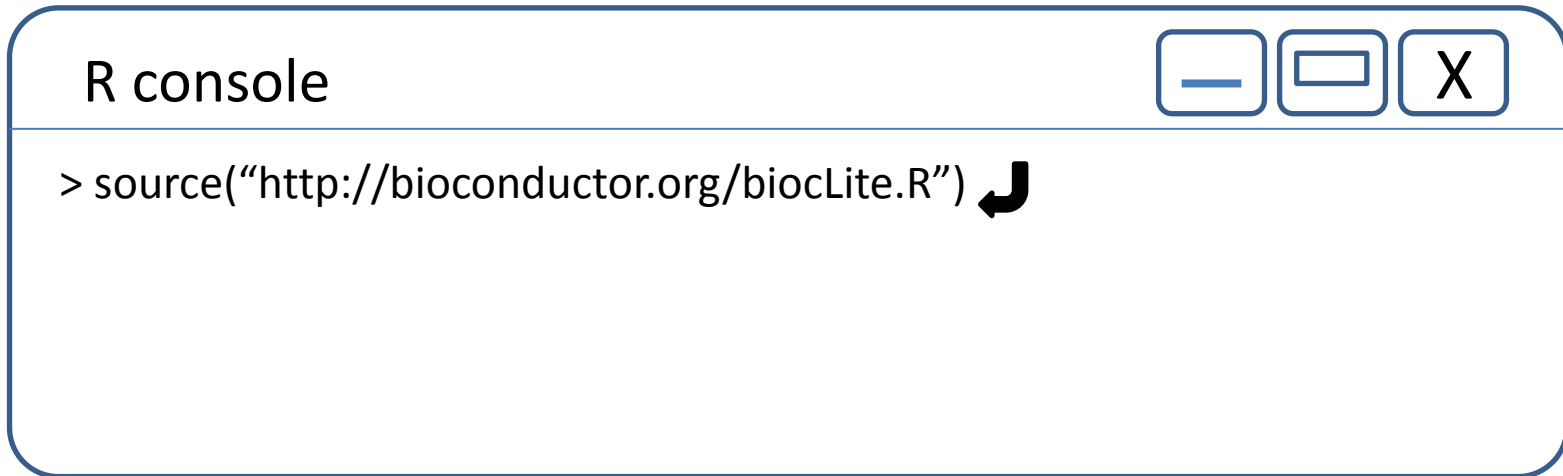
# ファイルの読み込み

R console



```
> x <- read.table("GSE40493_Normalized_ID.txt", header=T, sep="¥t") ↵  
> head(x) ↵  
> xx <- x[,3:10] ↵  
> rownames(xx) <- x$GeneName ↵  
> head(xx) ↵
```

# Bioconductor, biocLiteの設定



R console

```
> source("http://bioconductor.org/biocLite.R") ↵
```

The image shows a simulated R console window with a blue border and rounded corners. The title bar contains the text 'R console' and three window control buttons (minimize, maximize, close). The main area contains a single line of R code: '> source("http://bioconductor.org/biocLite.R")' followed by a black arrow pointing downwards, representing the Enter key.

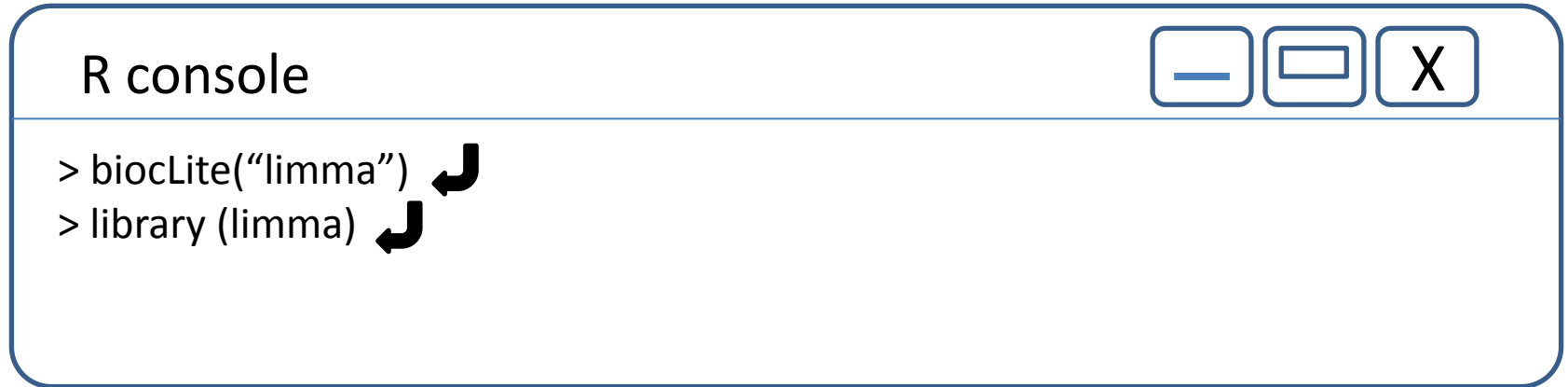
Bioconductor

バイオインフォマティクス関連のパッケージを配布しているサイト

biocLite.R

バイオインフォマティクス関連のパッケージをインストールするインストーラ  
パッケージ間の依存関係やバージョンの整合性を調整してくれる。

# Package “limma”



R console

```
> biocLite("limma") ↵  
> library (limma) ↵
```

The image shows a simulated R console window with a blue border and rounded corners. The title bar contains the text "R console" on the left and three window control buttons (minimize, maximize, close) on the right. The main area of the window contains two lines of R code, each followed by a right-pointing arrowhead symbol (↵) representing a carriage return.

# Design Matrix

R console



```
> groups <- factor(c(1,1,1,1,2,2,2,2)) ↵  
> groups ↵  
> design <- model.matrix(~ 0 + groups) ↵  
> colnames(design) <- c("KO", "WT") ↵  
> design ↵
```

design matrix を定義する  
KO 4サンプル、WT 4サンプル

	KO	WT
1	1	0
2	1	0
3	1	0
4	1	0
5	0	1
6	0	1
7	0	1
8	0	1



# Contrast Matrix

R console



```
> contrast.matrix <- makeContrasts(KOvsWT = KO - WT, levels = design)↵  
> contrast.matrix ↵
```

contrast matrix を定義する  
KO vs WT

KO	1
WT	-1

# 発現変動遺伝子の抽出

R console



```
> fit <- lmFit(xx, design) ↵  
> fit2 <- contrasts.fit(fit, contrast.matrix) ↵  
> fit2 <- eBayes(fit2) ↵
```

$$E(y_j) = X\alpha_j$$

$$\beta_j = C^T \alpha_j$$

$y_j$ : expression data of gene  $j$

$X$ : Design Matrix

$\alpha_j$ : vector of coefficients

$C$ : contrasts matrix

# 結果の表示

R console



```
> topTable(fit2) ↵  
> topTable(fit2, number=20, sort="logFC", adjust="fdr") ↵  
> summary(decideTests(fit2)) ↵
```

topTableのオプション

number = : 表示する遺伝子の数

sort = : 並べる順番, "B", "t", "p", "P", "logFC"

adjust = : 補正方法, "fdr", "BH"

# 作図

R console



```
> plotMA(fit2) ↵  
> plotSA(fit2) ↵  
> volcanoplot(fit2) ↵
```

# 3group の場合の design matrix

R console



```
> groups <- factor(c(1,1,2,2,3,3)) ↵  
> groups ↵  
> design <- model.matrix(~ 0 + groups) ↵  
> colnames(design) <- c("sample1", "sample2", "sample3") ↵  
> design ↵
```

	sample1	sample2	sample3
1	1	0	0
2	1	0	0
3	0	1	0
4	0	1	0
5	0	0	1
6	0	0	1

# 3group の場合のcontrast matrix

R console



```
> contrast.matrix <- makeContrasts(sample1 - sample2, sample2-sample3,
+ sample3-sample1, levels = design)
> contrast.matrix
```

3goups の場合のcontrast matrix

	Sample1- sample2	Sample2- sample3	Sample3- sample1
sample1	1	0	-1
sample2	-1	1	0
sample3	0	-1	1

# 課題

- GSE40493のデータをduplicateの4サンプルとして、design matrix、contrast matrixを定義し、発現変動遺伝子を抽出しなさい。

# 第5回

日時：平成27年3月30日（月）17:00～

「RNA-seq データ解析2」