



2018年度
第5回バイオインフォマティクス実習

作図方法

先端医科学研究センター
バイオインフォマティクス解析室 中林潤

- 大容量、高次元データを可視化
- 散布図、M-A plot、volcano plot、heatmap
- Rの作図機能を使用

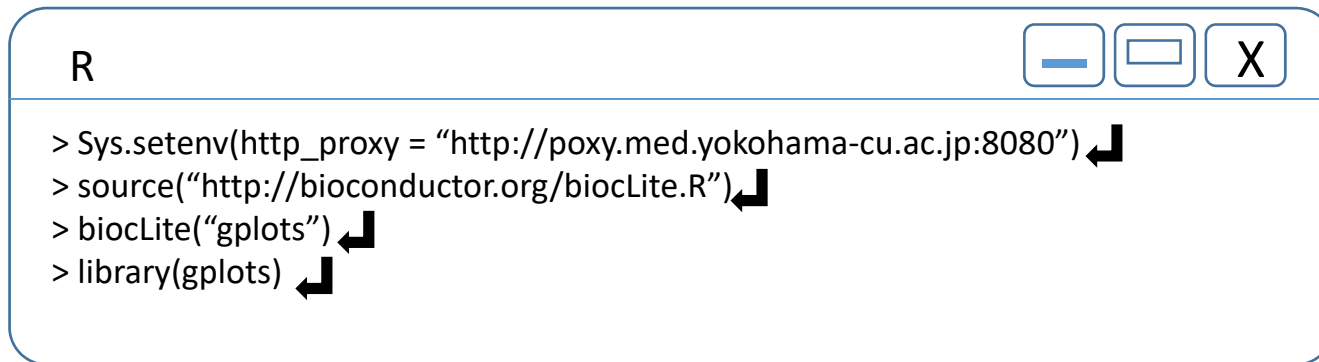
ファイルの取得

- 遺伝子発現プロファイル

GEO データベース accession number GSE52452

Expression.rev1.txtファイルを各自のデスクトップへ
ドラッグ アンド ドロップでコピー

Rの設定



```
R
```

```
> Sys.setenv(http_proxy = "http://poxy.med.yokohama-cu.ac.jp:8080")  
> source("http://bioconductor.org/biocLite.R")  
> biocLite("gplots")  
> library(gplots)
```

データの読み込み

```
R
```

```
> x <- read.table("Z:/デスクトップ/Expression.rev1.txt", header = T, sep = "¥t")
```

1～3列: 正常コントロール 3サンプル

4～6列: TAK1 siRNA 3サンプル

7列: コントロールの平均値

8列: TAK1 siRNAの平均値

9列: コントロールとTAK1 siRNAで二群間比較したP値

scatter plot (散布図)

R



```
> plot(x[,7], x[,8], pch = 20, col = "gray", xlim = c(1,15), ylim = c(1,15), xlab = "Ctrl",  
+ ylab = "TAK", main = "scatter plot")
```

散布図

データの分布を表示

Rでの行列の扱い

x[列,行]

plot(x軸, y軸, オプション)

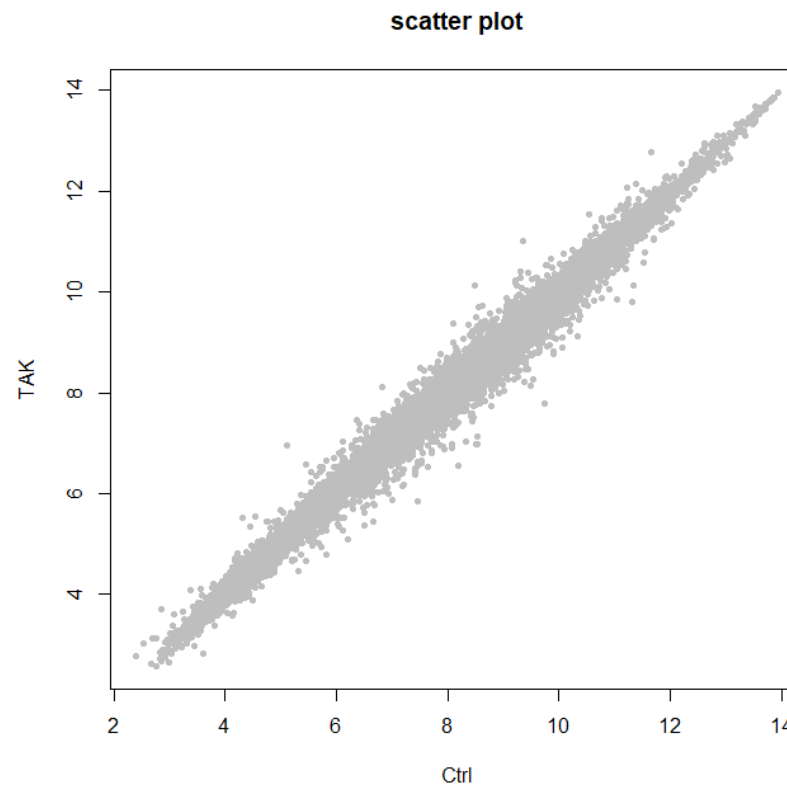
pch: 点のタイプ

col: 色

xlim, ylim: x軸、y軸の範囲

xlab, ylab: x軸、y軸のラベル

main: タイトル



M-A plot

R



```
> plot((x[,7] + x[,8])/2, x[,7] - x[,8]), pch = 20, col = "gray", xlab = "A",  
+ ylab = "M", main = "M-A plot")
```

2群間比較

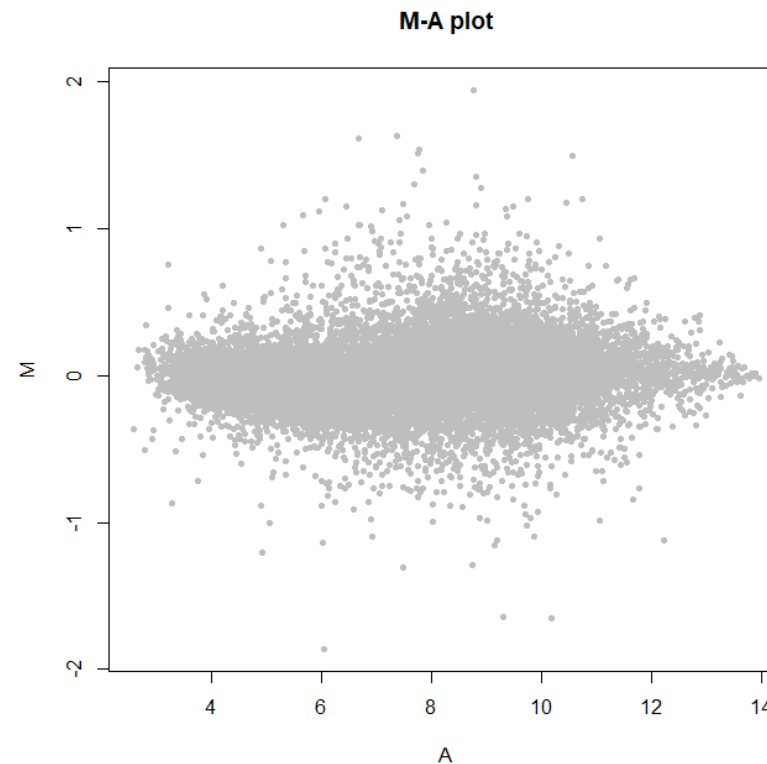
x軸:平均発現量

\log_2 変換した値を使う

$(\log_2(A) + \log_2(B)) / 2$

y軸:発現比

$\log_2(A/B)$



volcano plot

R

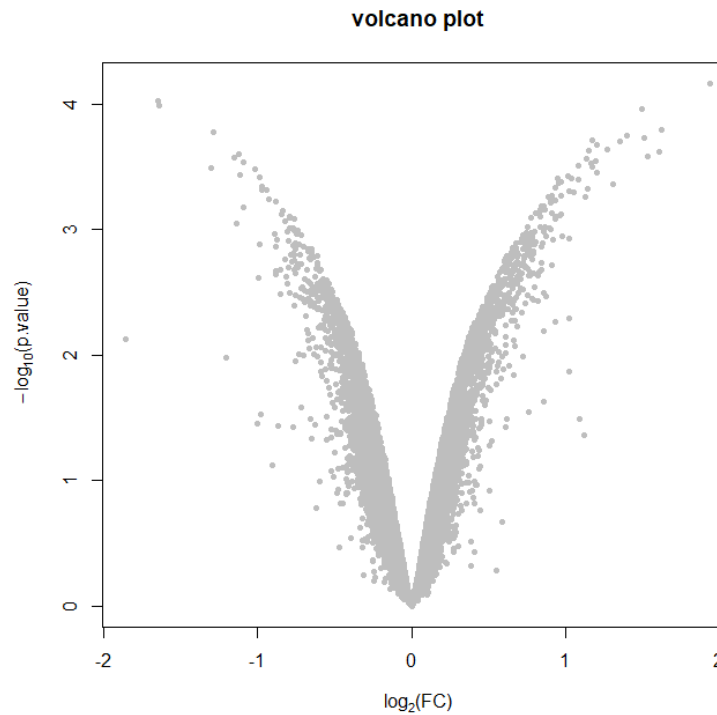


```
> plot(x[,7] - x[,8], -log10(x[,9]), xlab = expression(paste(log[2], "(FC)")),  
+ ylab = expression(paste(-log[10], "(p-value)")), col = "gray", main = "volcano plot")
```

2群間比較

x軸: 発現比

y軸: p値



heatmap

R



```
> heatmap.2(x[sample(1:21272, 1000),1:6], col = greenred(75), cexRow = 0.1, cexCol = 0.8,  
+ trace = "none", density.info = "none", main = "heatmap")
```

散布図

データの分布を表示

gplots libraryのheatmap.2関数を使用
21272遺伝子中1000遺伝子をsample
関数でランダムに抽出
cexRow, cexCol:x軸、y軸のラベル文字
の大きさ
col:greenred(階調)

