



第2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

- マイクロアレイデータの可視化
- 樹形図
- 散布図、MA-Plot
- ヒートマップ

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
- f) コマンドの途中でenter キーを押しても、そのまま入力を続行できる

答え合わせ

- 課題 GSE26910からデータを取得し、正規化、ファイルへの出力を行う
- 足りないライブラリ

pd.hg.u133.plus.db

R console



```
> biocLite("pd.hg.u133.plus.2") ↩  
> library(pd.hg.u133.plus.2) ↩
```

/geo/query/acc.cgi?acc=GSE26910

Submission date Jan 27, 2011
 Last update date Dec 03, 2014
 Contact name Paolo Provero
 E-mail paolo.provero@unito.it
 Organization name University of Turin
 Department Molecular Biotechnology and Health Sciences
 Street address Via Nizza 52
 City Torino
 ZIP/Postal code I-10100
 Country Italy

Platforms (1) [GPL570 \[HG-U133_Plus_2\] Affymetrix Human Genome U133 Plus 2.0 Array](#)

Samples (24) [GSM662756 prostate normal 1](#)

[More...](#) [GSM662757 prostate tumor 1](#)

[GSM662758 prostate normal 2](#)

Relations

BioProject [PRJNA136171](#)

Analyze with GEO2R

Download family

[SOFT formatted family file\(s\)](#)

Format

SOFT [?](#)

[MINiML formatted family file\(s\)](#)

MINiML [?](#)

[Series Matrix File\(s\)](#)

TXT [?](#)

Supplementary file	Size	Download	File type/resource
GSE26910_RAW.tar	104.1 Mb	(http)(custom)	TAR (of CEL)

Raw data provided as supplementary file

Processed data included within Sample table

platform
 Human
 Genome
 U133
 Plus 2.0

- 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

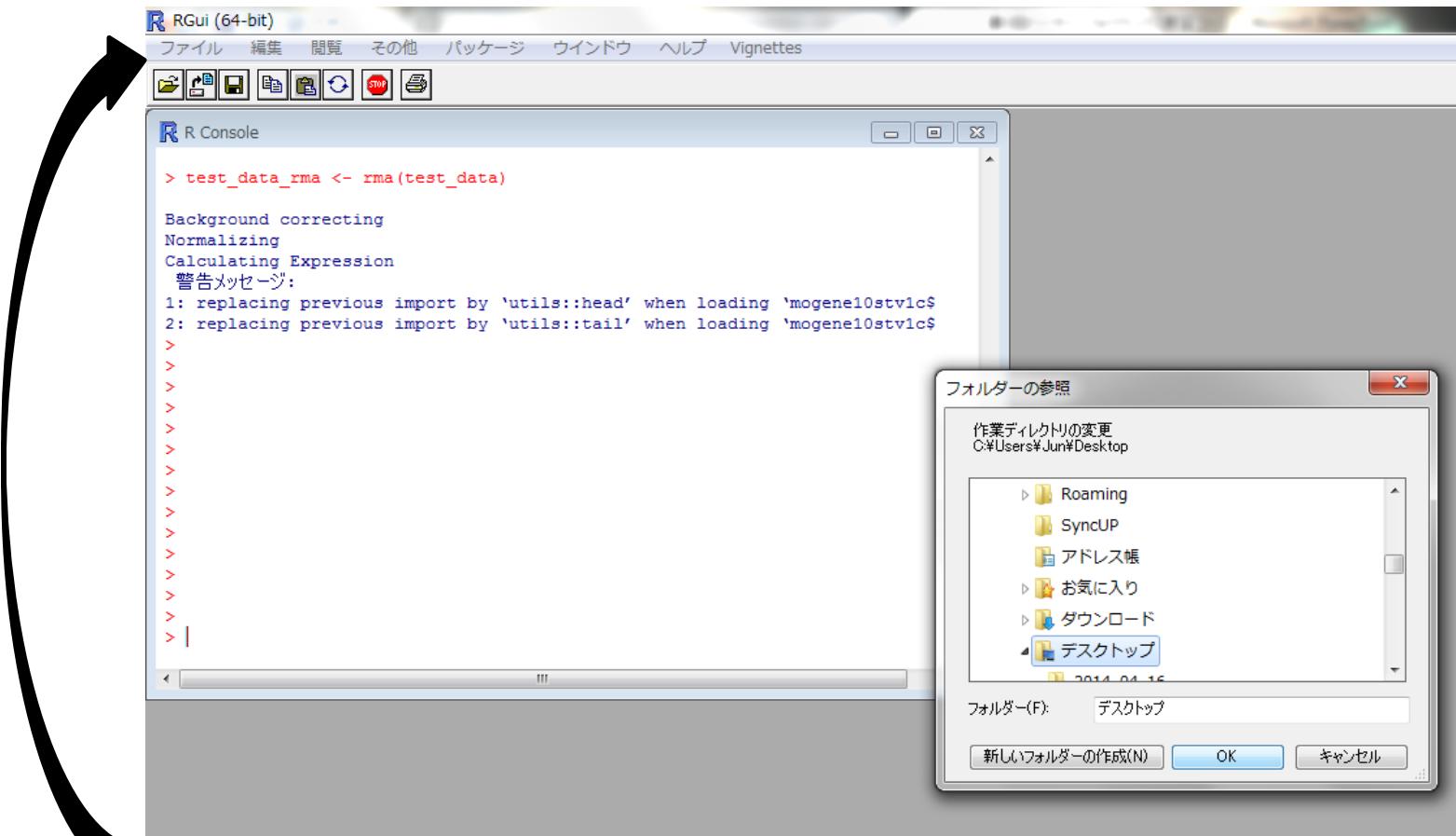
R console — □ X

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

正規化後の遺伝子発現ファイル

- 課題配布フォルダ
→GSE40493→GSE40493_Normalized.txt
- 各自のデスクトップフォルダ→GSE40493フォルダにコピー

作業ディレクトリを変更



ファイルメニューから“ディレクトリの変更”を選択
各自のデスクトップ→GSE40493フォルダを選択

ファイルの読み込み

R console



```
> x <- read.table("GSE40493_Normalized.txt", header=T, sep="\t") ↵  
> head(x)
```

GEOデータベース

<http://ncbi.nlm.nih.gov>

NCBI Resources How To

Sign in to NCBI

NCBI National Center for Biotechnology Information

All Databases

Recent

All Databases

All

- All Databases
- Assembly
- BioProject
- BioSample
- BioSystems
- Books
- ClinVar
- Clone
- Conserved Domains
- dbGaP
- dbVar
- Epigenomics
- EST
- Gene
- Genome
- GEO DataSets**
- GEO Profiles

Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI Announcements

- BLAST+ 2.2.30 released
- A new version (2.2.30) of the stand-alone BLAST executables is now available, bringing several improvements to
- New Genome BLAST selector on the BLAST homepage
- You can now easily find Genome-specific BLAST pages using the search box on
- Next NCBI webinar on November 5th
- On November 5th, NCBI will have a webinar entitled "Exploring and

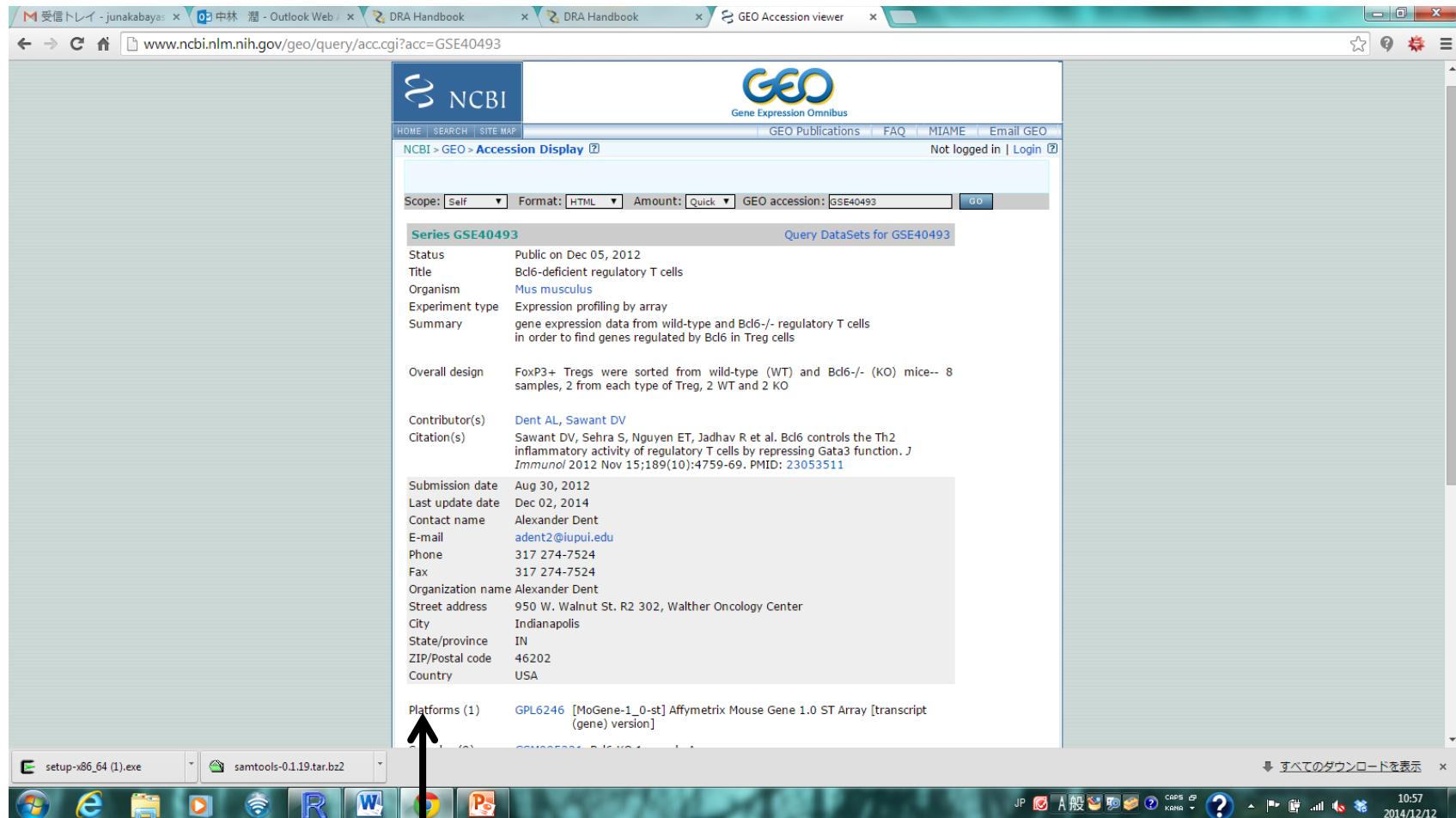
GEO Datasets を選択

アクセスション番号を入力

GSE40493

More...

Platform情報の取得



www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40493

Scope: Self Format: HTML Amount: Quick GEO accession: GSE40493 00

Series GSE40493

Status: Public on Dec 05, 2012
Title: Bcl6-deficient regulatory T cells
Organism: *Mus musculus*
Experiment type: Expression profiling by array
Summary: gene expression data from wild-type and Bcl6-/- regulatory T cells in order to find genes regulated by Bcl6 in Treg cells
Overall design: FoxP3+ Tregs were sorted from wild-type (WT) and Bcl6-/- (KO) mice-- 8 samples, 2 from each type of Treg, 2 WT and 2 KO

Contributor(s): Dent AL, Sawant DV
Citation(s): Sawant DV, Sehra S, Nguyen ET, Jadhav R et al. Bcl6 controls the Th2 inflammatory activity of regulatory T cells by repressing Gata3 function. *J Immunol* 2012 Nov 15;189(10):4759-69. PMID: 23053511

Submission date: Aug 30, 2012
Last update date: Dec 02, 2014
Contact name: Alexander Dent
E-mail: adent2@iupui.edu
Phone: 317 274-7524
Fax: 317 274-7524
Organization name: Alexander Dent
Street address: 950 W. Walnut St. R2 302, Walther Oncology Center
City: Indianapolis
State/province: IN
ZIP/Postal code: 46202
Country: USA

Platforms (1): **GPL6246** [MoGene-1_0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]

setup-x86_64 (1).exe samtools-0.1.19.tar.bz2

JP A般 CAPS KANA ? 10:57 2014/12/12

Pltaform

Platform情報の取得

受信トレイ - junakabayashi 中林 潤 - Outlook Web / DRA Handbook DRA Handbook GEO Accession viewer

www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6246

known genes (multipart).

mRNA_assignment Description of the public mRNAs that should be detected by the sets within this transcript cluster based on sequence alignment (multipart).

category Array design category of the transcript cluster

Data table

ID	GB_LIST	SPOT_ID	segname	RANGE_
10344614		chr1:3044314-3044814	chr1	NC_000014
10344616		chr1:3092097-3092206	chr1	NC_000016
10344618		chr1:3266404-3267429	chr1	NC_000018
10344620		chr1:3670652-3670993	chr1	NC_000020
10344622	NM_024221, BC094468	chr1:4761212-4762280	chr1	NC_000022
10344624	NM_008866, BC013536	chr1:4797943-4836817	chr1	NC_000024
10344633	NM_011541, NM_001159750, NM_001159751, BC083127	chr1:4847895-4887990	chr1	NC_000033
10344637	NM_133826, BC009154	chr1:5073253-5152630	chr1	NC_000037
10344653	NM_011011, L11065	chr1:5578574-5592947	chr1	NC_000053
10344658	NM_009826, AB070619, AB050017, BC150774, AK165119, AK076550, AK020027, AK033709	chr1:6204743-6265656	chr1	NC_000058
10344674	NM_001195732	chr1:6349422-6381175	chr1	NC_000074
10344679	NM_173868, BC118528	chr1:6720132-6851021	chr1	NC_000079
10344705		chr1:6864053-6864139	chr1	NC_000005
10344707	NM_183028, BC110360	chr1:7079231-7163709	chr1	NC_000007
10344713	NM_016661, L32836	chr1:7167820-7169118	chr1	NC_000013
10344715	AK036865	chr1:7488080-7488381	chr1	NC_000015
10344717		chr1:8806219-8806328	chr1	NC_000017
10344719		chr1:8846844-8847185	chr1	NC_000019
10344721		chr1:0448751-0448853	chr1	NC_000001

Total number of rows: 35557

Table truncated, full table size 32596 Kbytes.

[Download full table...](#)

[Annotation SOFT table...](#)

[Download family](#)

Format

SOFT [?](#)

MTNIMI [?](#)

www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6246

setup-x86_64 (1).exe samtools-0.1.19.tar.bz2

すべてのダウンロードを表示

JP A般 CAPS KANA ? 10:58 2014/12/12

Download full tableをクリック

array ID \longleftrightarrow その他の ID

Data table

ID	GB_LIST	SPOT_ID	seqname	RANGE_
10344614		chr1:3044314-3044814	chr1	NC_000001
10344616		chr1:3092097-3092206	chr1	NC_000001
10344618		chr1:3266404-3267429	chr1	NC_000001
10344620		chr1:3670652-3670993	chr1	NC_000001
10344622	NM_024221, BC094468	chr1:4761212-4762280	chr1	NC_000001
10344624	NM_008866, BC013536	chr1:4797943-4836817	chr1	NC_000001
10344633	NM_011541, NM_001159750, NM_001159751, BC083127	chr1:4847895-4887990	chr1	NC_000001
10344637	NM_133826, BC009154	chr1:5073253-5152630	chr1	NC_000001
10344653	NM_011011, L11065	chr1:5578574-5592947	chr1	NC_000001
10344658	NM_009826, AB070619, AB050017, BC150774, AK165119, AK076550, AK020027, AK033709	chr1:6204743-6265656	chr1	NC_000001
10344674	NM_001195732	chr1:6349422-6381175	chr1	NC_000001
10344679	NM_173868, BC118528	chr1:6720132-6851021	chr1	NC_000001
10344705		chr1:6864053-6864139	chr1	NC_000001
10344707	NM_183028, BC110360	chr1:7079231-7163709	chr1	NC_000001
10344713	NM_016661, L32836	chr1:7167820-7169118	chr1	NC_000001
10344715	AK036865	chr1:7488080-7488381	chr1	NC_000001
10344717		chr1:8806219-8806328	chr1	NC_000001
10344719		chr1:8846844-8847185	chr1	NC_000001
10344721		chr1:9448751-9448853	chr1	NC_000001

Total number of rows: 35557

Table truncated. Full table size 32506 Kbytes

Platform情報の取得

- 課題配布フォルダ
→BioInfoJishu→GPL_6246-
21513_ID_NAME
- 各自のデスクトップのGSE40493フォルダに
コピー

IDファイルの読み込み アレイIDと遺伝子名の関連付け

R console



```
> id <- read.table("GPL6246-21513_ID_NAME.txt", header=T, sep="\t") ↴  
> head(id) ↴  
> x_id <- merge(id, x, by.x="ID", by.y="X") ↴  
> head(x_id) ↴
```

merge(A, B, by.x="Aの項目", by.y="Bの項目")
AとBを項目Aと項目Bとで関連付け

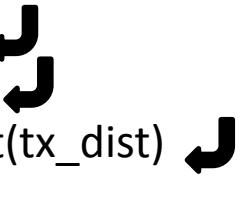
階層的クラスタリングと樹形図

距離行列
階層的クラスタリング
樹形図

R console

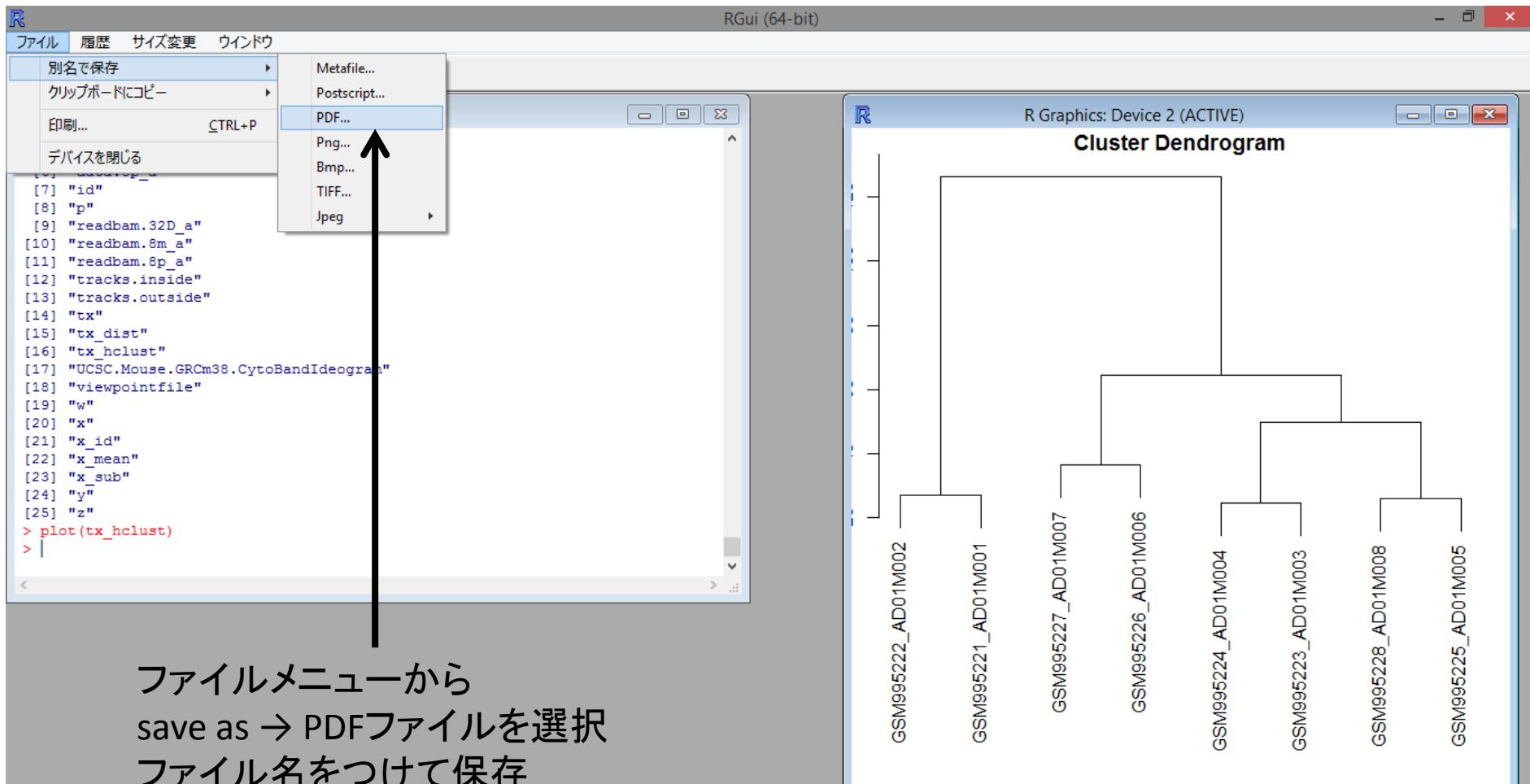


```
> tx <- t(x_id[,3:10])  
> tx_dist <- dist(tx)  
> tx_cluster <- hclust(tx_dist)  
> plot(tx_cluster)
```

Four handwritten black arrows point from the bottom right towards the lines of R code: one to the first line, one to the second, one to the third, and one to the fourth.

転置行列 $t(\text{行列})$
距離行列 $\text{dist}(\text{行列})$
 $\text{hclust}(\text{距離行列})$

図のPDFファイルを保存



散布図

plot(x値, y値, option)

option : type = “p” or “l” or “b”

 xlab = “x軸名”

 ylab = “y軸名”

 pch = •の種類 (数字)

 col = “色”

R console



X

```
> plot(x_id[,3], x_id[,4], xlab="KO1_1", ylab="KO1_2", main = "Scatter Plot", pch=20) ↴  
> plot(x_id[,3], x_id[,7], xlab="KO1_1", ylab="WT1_1", main = "Scatter Plot", pch=20) ↴
```

MA-Plot

R console



```
> plot((x_id[,4] + x_id[,3]) / 2, x_id[,4] - x_id[,3], xlab="A", ylab="M", ↵
+ main = "MA-Plot", ylim = c(-7, 7), pch=20) ↵
> plot((x_id[,7] + x_id[,3]) / 2, x_id[,7] - x_id[,3], xlab="A", ylab="M", ↵
+ main = "MA-Plot", ylim = c(-7, 7), pch=20) ↵
```

WT1の平均値とKO1の平均値の行列を生成
WT/KO > 2の遺伝子を選別

R console



```
> x_mean <- cbind((x_id[,3] + x_id[,4]) / 2, ↴  
+ (x_id[,7] + x_id[,8]) / 2) ↴  
> x_mean <- data.frame(x_mean) ↴  
> rownames(x_mean) <- x_id$GeneName ↴  
> colnames(x_mean) <- c("KO", "WT") ↴  
> head(x_mean) ↴  
> x_sub <- subset(x_mean, x_mean$WT / x_mean$KO > 1.1) ↴  
> nrow(x_sub) ↴  
> head(x_sub) ↴
```

cbind(ベクトル1, ベクトル2, ...)

Bioconductor, biocLiteの設定

R console



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

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 "gplots"  
heatmap.2
```

R console



```
> biocLite("gplots") ↴  
> library (gplots) ↴  
> heatmap.2(as.matrix(x_sub), trace = "none", density.info = "none", ↴  
+ col = greenred(75), cexCol = 2.0, cexRow = 0.25) ↴
```

課題

- GSE26910のデータについても樹形図、散布図、MA-plot、ヒートマップを描いてください。

第3回

日時：平成27年2月9日（月）17:00～

「マイクロアレイデータ解析 3」

- 発現変動遺伝子の抽出
- オントロジー解析