



第1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
小泉真一

- データベースからデータの取得
- 正規化
- ファイルへ出力

M402LL教室のPC環境

- YCUアカウントでログイン
読み書き可能フォルダ
Z:/ユーザ名
読み込み可能フォルダ
課題配布/BioInfoJishu
- インターネット接続可
proxyサーバ経由
- R ver 3.0.2がインストール済

統計解析ソフトR

- オープンソースの統計解析ソフト

<http://cran.r-project.org>
で配布

- Windows Mac Linuxで使用可能
- 様々な研究分野で広く使われている
- 参考
<http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>

http://cran.r-project.org

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages. **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2014-10-31, Pumpkin Helmet) [R-3.1.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

受信トレイ - junakabayashi × 中林 潤 - Outlook Web × www.ncbi.nlm.nih.gov/p ... ac.els-cdn.com/S002251 ... The Comprehensive R Arc ...

cran.r-project.org

CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

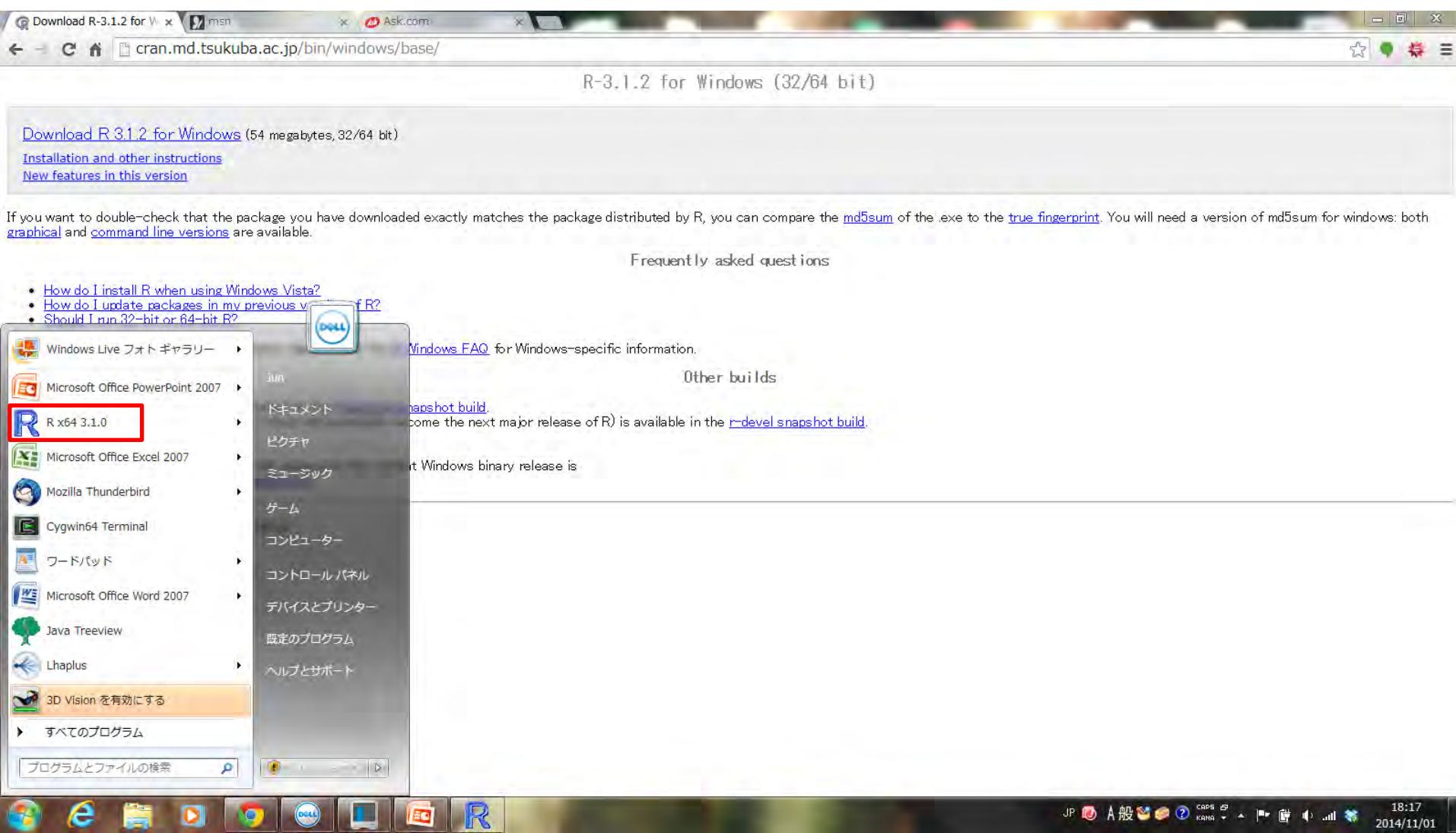
[About R](#)
[R Homepage](#)
[The R Journal](#)

[Software](#)
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

[Documentation](#)
[Manuals](#)
[FAQs](#)
[Contributed](#)

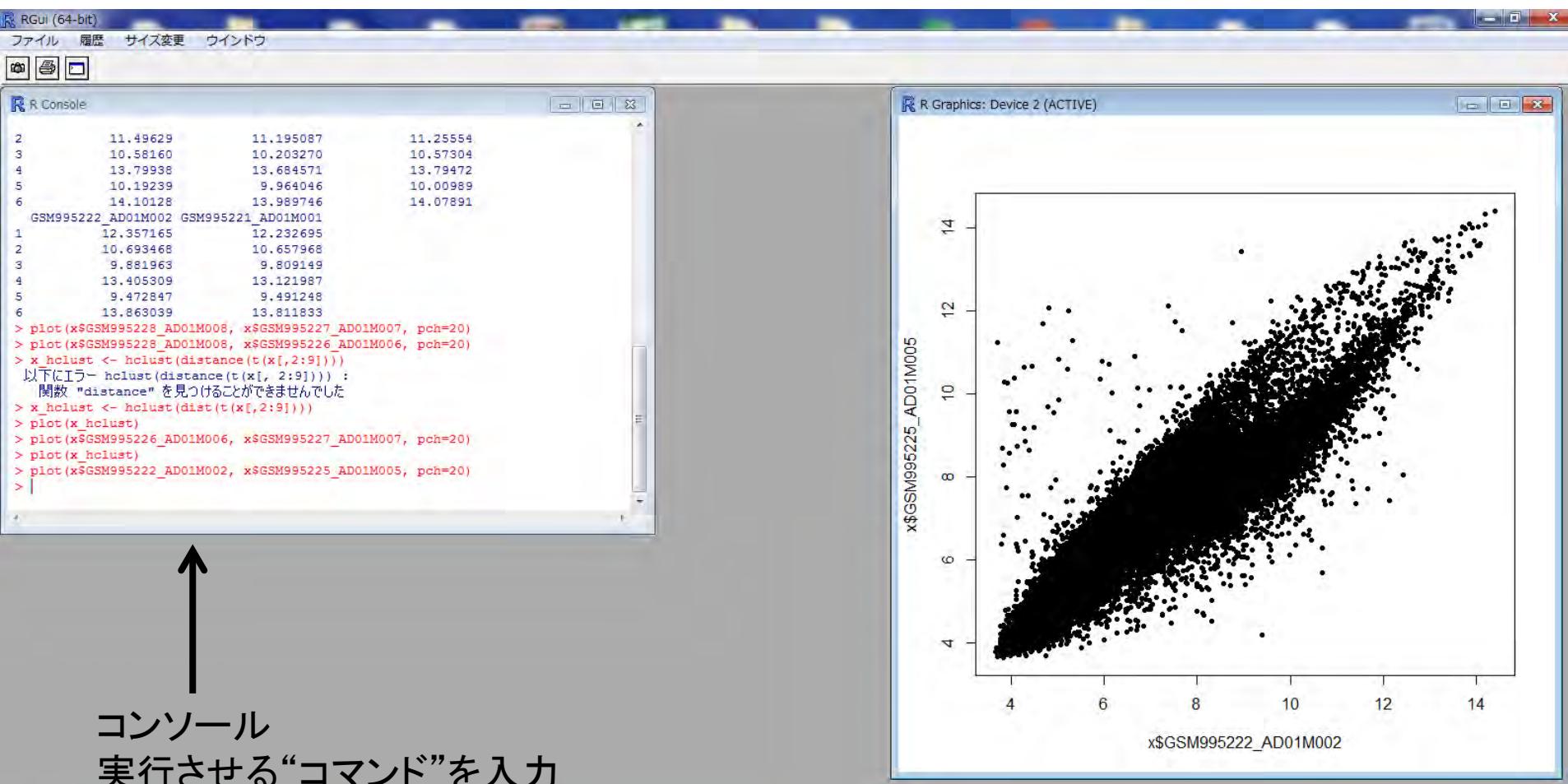
横4コマ.pptx 写真データ.zip Cancelled ni.2987.pdf nihms374495.pdf 12月マネ会議日程調...xlsx Show all downloads...

Rの起動



スタートメニューからRを選択して起動

Rのコンソール



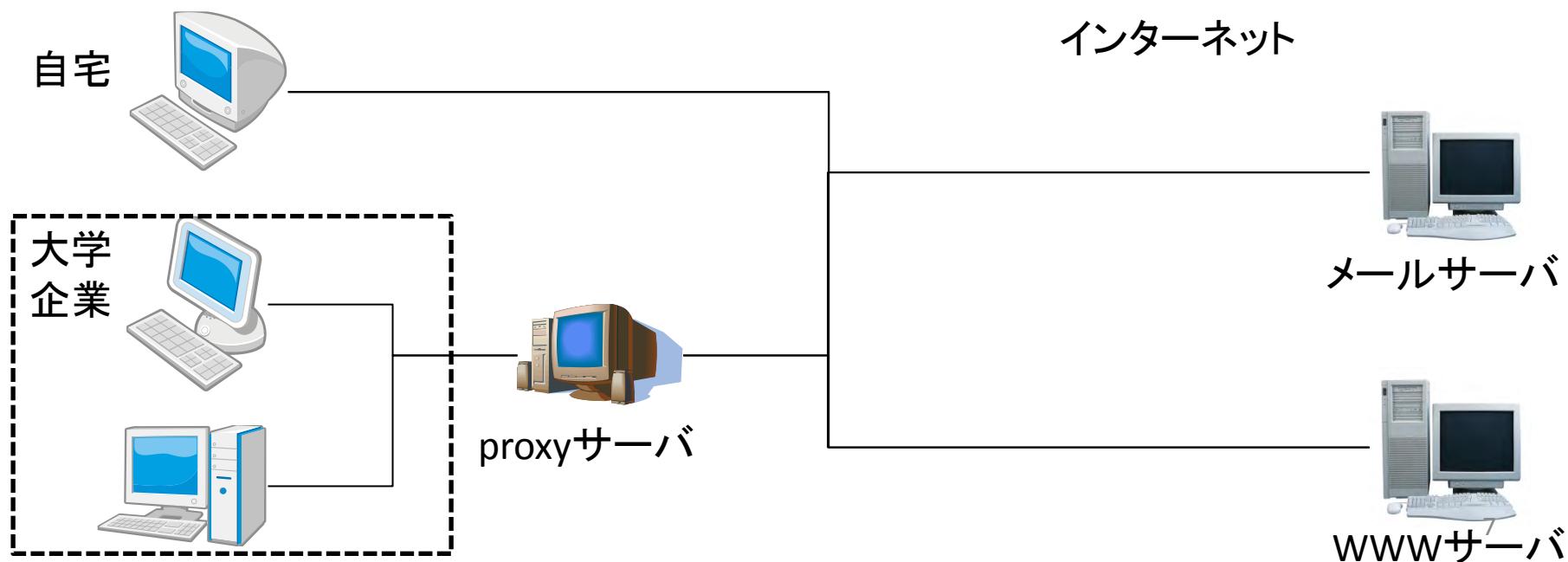
proxyの設定(横浜市大の場合)

R console



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

R起動直後に実行しないと設定されないことがあります。



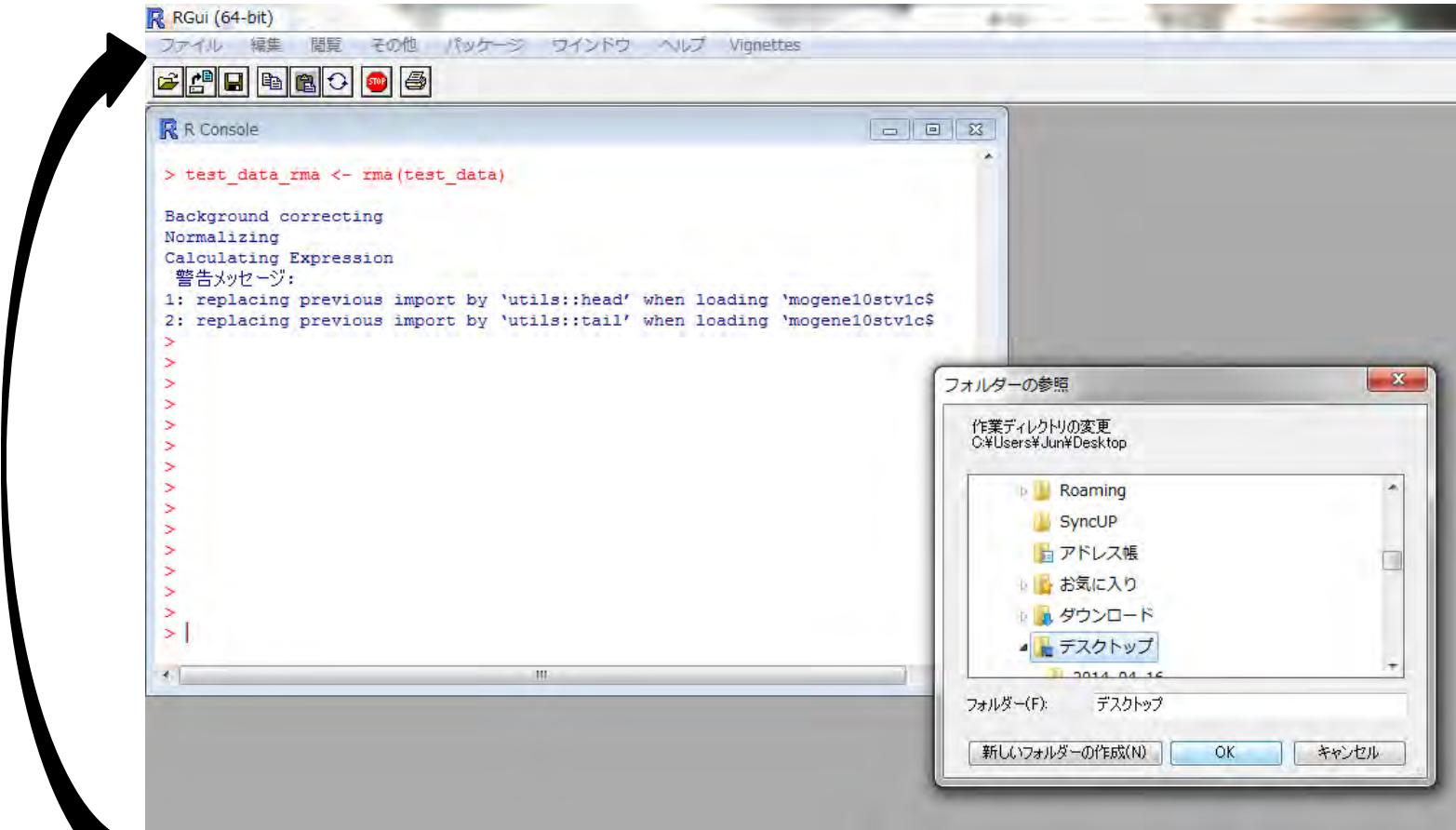
Rの基本操作

R console



```
> 34 + 58 ↴  
> 92 ↴  
> 105 / 33 ↴  
> 3.181818 ↴  
> pi ↴  
> 3.141593 ↴  
> sqrt(2) ↴  
> 1.414214 ↴  
> x <- 10 ↴  
> y <- 15 ↴  
> z <- x + y ↴  
> x <- seq(-10, 10, by=0.1) ↴  
> plot(sin(x), type="l") ↴
```

作業ディレクトリに移動



ファイルメニューから“ディレクトリの変更”を選択
作業ディレクトリを選択(読み書き可能な各自のアカウントフォルダを選択)
次回ログイン時に、今回書き込んだデータが保持されます。

ファイルの読み込み

R console



```
> p <- read.table("Kokonoe.txt", header=T, sep="\t")  
> p[1,1]  
> p[1,3]
```

Three black arrows are drawn by hand to point to specific parts of the R console code. One arrow points to the closing parenthesis at the end of the first line. Another points to the first element of the first row in the second line. A third points to the third element of the first row in the third line.

ユーザ定義関数

入力
(引数) → 関数
ユーザが定義した手順で計算を実行 → 出力
(戻り値)

R console



```
> bmi <- function(q){ ↴  
+ r <- p[q,4] / (p[q,3] / 100)^2 ↴  
> return(r)} ↴  
> bmi(3) ↴  
> p <- cbind(p, p[,4] / (p[,3] / 100)^2) ↴  
> write.table(p, "Kokonoe_rev.txt", quote=F, sep="\t") ↴
```

edit関数を使った入力

R console



```
> bmi <- edit(bmi) ↴
```

別ウインドウにテキストエディタが開くので、そこで入力の訂正を行う。

Packageのインストール

Package

複数の関数をまとめたものがパッケージとして提供されている。

Bioconductor.org

- 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

<http://bioconductor.org>

The screenshot shows the Bioconductor.org website. At the top, there's a navigation bar with links for Home, Install, Help, Developers, and About. The 'Home' link is highlighted in yellow. Below the navigation, there are several sections:

- About Bioconductor**: Describes Bioconductor as providing tools for the analysis and comprehension of high-throughput genomic data. It mentions two releases per year, software packages, and an active user community.
- News**: Lists recent updates, including the publication page, the release of Bioconductor 3.0, support site information, course material, and videos.
- Install**: Provides links to get started with Bioconductor, including install instructions, explore packages, support, latest newsletter, follow on Twitter, and using R.
- Learn**: Lists master Bioconductor tools, including courses, support site, package vignettes, literature citations, common work flows, FAQ, community resources, and videos.
- Use**: Describes creating bioinformatic solutions with Bioconductor, mentioning software, annotation, and experiment packages, Amazon Machine Image, and the latest release announcement.
- Develop**: Contribute to Bioconductor, including use Bioc 'devel', 'Dev' Software, Annotation and Experiment packages, package guidelines, and new package submission.

Packageのインストール

今回使用するpackage

- “affy”

Affymetrixデータ処理用パッケージ

- “AnnotationDbi”

アノテーション用ゲノムインデックス

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 “affy”

Package “AnnotationDbi”

R console



```
> biocLite("affy") ↴  
> library(affy) ↴  
> biocLite("AnnotationDbi") ↴  
> library(AnnotationDbi) ↴
```

GEOデータベース検索

http://ncbi.nlm.nih.gov

キーワードを入力

The screenshot shows the NCBI homepage at www.ncbi.nlm.nih.gov. A search bar at the top right contains the placeholder text "Search". Below it is a dropdown menu titled "All Databases" which is currently expanded. The "GEO DataSets" option is highlighted with a blue background and white text. To the right of the search bar, there is a "Popular Resources" sidebar listing links to PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem. At the bottom of the page, there is a "NCBI Announcements" section with news items about BLAST+ 2.2.30 and a new Genome BLAST selector.

キーワードを入力

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

Sign in to NCBI

Popular Resources

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

NCBI Announcements

- BLAST+ 2.2.30 released Oct 30, 2014
- 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 Oct 28, 2014
- You can now easily find Genome-specific BLAST servers using the search bar
- Next NCBI webinar on November 5th Oct 23, 2014
- On November 5th, NCBI will have a webinar entitled "Exploring and Downloading Databases and

More...

データの取得

- 課題配布→BioInfoJishuフォルダから
GSE40493フォルダをZ:/デスクトップに移動
- Rの作業フォルダをZ:/デスクトップ/GSE40493
に変更

R console

— □ X

```
> dir() ↩
```

コンソールにCELファイル名が表示されたら、データの取得とディレクトリの変更が完了しています。

データの読み込み

R console



```
> x <- ReadAffy()
```



作業フォルダ内のCELファイルの内容を変数xに格納する。

rma法で正規化

R console



```
>y <- rma(test_data) ↴
```

正規化したデータを変数yに格納する

RMA (Robust Multi-Array Average) 法

Exploration, normalization, and summaries of high density oligonucleotide array probe level data.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP

Biostatistics 2003 4(2):249-64

Affymetrixマイクロアレイデータの正規化法として良く用いられている手法の一つ。

以下の3ステップでデータを正規化する。

- 1) バックグラウンド補正
- 2) quantile normalization
- 3) median polish法でsammalize

write.exprsでファイルに出力

R console



```
> write.exprs(y, "GSE40493_Normalize.txt") ↴
```

yの内容をタブ区切りテキストファイルとして出力。

GSE40493フォルダにGSE40493_Normalizeと言う名前のテキストファイルができていれば作業が完了しています。

タブ区切りテキストファイルをExcelで開く

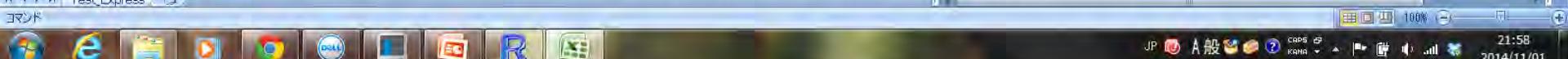
Test_Express - Microsoft Excel

ホーム挿入ページレイアウト数式データ校閲表示アドイン

貼り付け 削除 フォント 配置 数値 スタイル セル

A1

	A	B	C	D	E	F	G	H	I	J	K
1	GSM995228_AD01M008	GSM995227_AD01M007	GSM995226_AD01M006	GSM995225_AD01M005	GSM995224_AD01M004	GSM995223_AD01M003	GSM995222_AD01M002	GSM995221_AD01M001			
2	10338001	12.91442548	12.8671738	12.78268659	13.0305132	12.82905147	12.8622674	12.35716522	12.2326846		
3	10338003	11.28114413	11.30709222	11.10473015	11.49628584	11.19508745	11.25554372	10.69346842	10.65796845		
4	10338004	10.24385947	10.21488992	10.14135173	10.58159823	10.20326957	10.57304408	9.881962815	9.80914948		
5	10338017	13.63967267	13.61198487	13.66939676	13.78937742	13.684571	13.79471508	13.40530929	13.12198731		
6	10338025	9.908501846	9.938011267	9.649013557	10.192388936	9.964045748	10.00988712	9.472846696	9.491247593		
7	10338026	13.93510379	13.96977431	13.99477282	14.10127816	13.98974595	14.07891251	13.86303949	13.81183269		
8	10338029	10.67000607	10.70176467	10.60177371	10.98229195	10.59413637	10.88909846	10.20845142	10.14011654		
9	10338035	9.879418746	9.949708688	9.80272325	10.15610795	9.854045301	9.934284622	9.507837646	9.538823679		
10	10338036	10.43533886	10.4920886	10.41617879	10.98431064	10.49502855	10.79873636	10.12200472	10.01245256		
11	10338037	4.290409006	4.303970035	4.275305621	4.252870391	4.282505519	4.235840949	4.270563079	4.296882084		
12	10338041	12.13131044	12.08674964	12.06145042	12.4410089	12.10317752	12.24704861	11.55415508	11.44279685		
13	10338042	11.12622997	11.24062048	11.08221186	11.63495726	11.20597002	11.31158685	10.62185688	10.59196757		
14	10338044	12.96715264	12.90504102	12.74866645	13.13863544	12.93614621	12.98696313	12.4284398	12.26979821		
15	10338047	6.763653772	6.901370033	6.8107405	6.724367429	6.718937144	6.687530492	6.711506397	6.779435792		
16	10338056	4.159647488	4.173418903	4.131062695	4.13045013	4.165768879	4.104994121	4.154426975	4.164225438		
17	10338059	13.95348844	13.98705037	13.96089114	13.97393993	13.9276401	13.935559217	13.766423	13.67270388		
18	10338060	4.295896842	4.294548503	4.266691578	4.266635562	4.282369153	4.24607745	4.281992206	4.30901929		
19	10338063	4.225048553	4.259822836	4.212457699	4.2056999	4.214572437	4.174237693	4.19236055	4.211607208		
20	10338064	5.819037388	6.04330703	6.201893486	5.774275043	5.7686607083	5.793544556	6.022835011	6.120632578		
21	10338065	6.352138566	6.633064097	6.782333391	6.326170161	6.306145322	6.285299134	6.42007228	6.655293247		
22	10338066	4.910927396	5.06311953	5.106761426	4.914881606	4.903922725	4.948643348	5.090043501	5.236642491		
23	10338067	7.714886054	8.056897609	8.160952442	7.675949218	7.618620174	7.699587081	7.6943987	7.785893965		
24	10338068	8.638172472	8.61654494	8.40195936	8.7511163569	8.604806493	8.5239989215	7.863958602	7.707226069		
25	10338069	5.455789648	4.96948286	5.959576471	5.370045266	4.970349606	5.308601645	5.828781389	5.546517292		
26	10338070	4.774529463	4.903033703	4.623856664	4.596480622	4.715788631	5.12415018	4.730363926	4.474340236		
27	10338071	4.75308505	4.79812396	4.647918368	4.41731038	4.500078278	4.342114367	4.532176037	4.467510215		
28	10338072	4.842275309	5.01043487	4.982517004	4.991179128	5.304482953	4.932346088	4.755200176	5.175230878		
29	10338073	7.247722666	6.520938602	6.640031052	7.070276903	7.014413961	6.456815841	5.605120557	6.315544201		
30	10338074	8.104924033	8.408530411	8.401969636	8.18223739	7.404345862	8.010868736	8.534637082	8.541650866		
31	10338075	6.175731245	6.4971469	6.284237422	5.769584555	6.348672189	5.980803102	7.774690082	7.895733236		
32	10338076	6.169678012	6.581742519	6.752830992	5.887165144	6.295685281	5.750888821	6.699114267	6.485506814		
33	10338077	9.377591073	9.164535392	9.734262878	8.698280769	9.302784155	9.557002796	10.26487369	10.32188174		
34	10338078	5.784767782	5.640911444	5.436511457	5.823124266	5.709017733	5.988510278	5.395258931	5.007097466		
35	10338079	5.26842573	5.595138549	5.540079645	5.613886933	5.286217301	5.069353539	5.490831462	5.360131623		
36	10338080	6.950740114	5.600235202	6.963124094	6.156047209	5.06660116	6.0410060375	5.020054525	5.040205125		



課題

- GEOデータベース、アクセション番号 GSE62654のデータを取得し、正規化して結果をテキストファイルに出力してください。
- 今回の実習で使ったパッケージでは足りないものがあります。
- ヒント: サンプルを採取した細胞は？

第2回

- 正規化後のデータを可視化
- 散布図
- ヒートマップ
- その他