

統計解析ソフトRを使った バイオインフォマティクス解析実習

東京科学大学 ILA国府台
中林 潤

JH人材育成課 バイオインフォマティクソン育成講座 ⑫



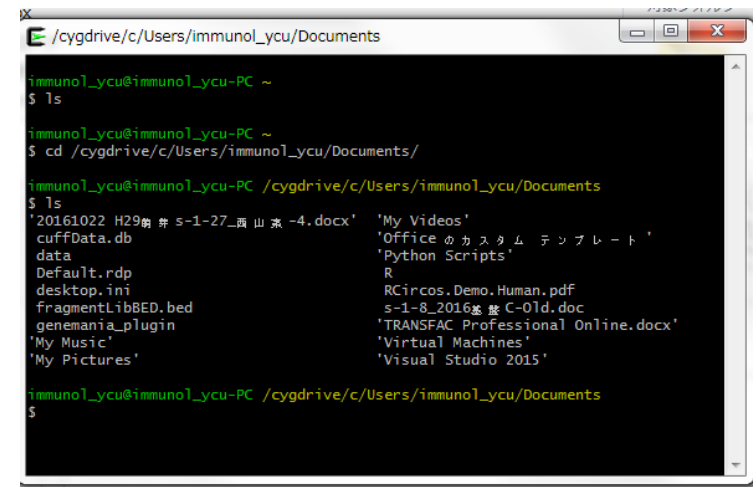
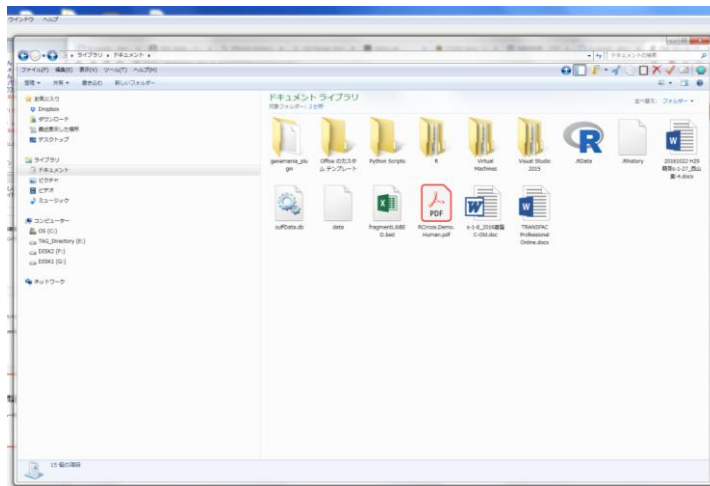
GUI と CUI

- Graphical User Interface (GUI)

コンピュータのディスプレイ上で、アイコンや画像などを多用し、マウスなどのポインティングデバイスで直感的操作を可能とするユーザーインターフェース

- Character User Interface (CUI)

コマンドや情報を文字によって表示し、コンピュータの操作を行うユーザーインターフェース



統計解析ソフトR

- オープンソースの統計解析ソフトウェア
- バイオインフォマティクス解析用のライブラリが豊富
- 配布元
<https://cran.r-project.org>



OSを選択

The screenshot shows the CRAN website with the following content:

- 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 \(Debian, Fedora/Redhat, Ubuntu\)](#)
 - [Download R for macOS](#)
 - [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 (2026-04-24, Because it was There) [R-4.6.0.tar.gz](#), read [what's new](#) in the latest version.
 - The CRAN directory [src/base-prerelease](#) contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
 - Between releases, the same directory [src/base-prerelease](#) contains snapshots of current patched and development versions.
Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
 - Alternatively, daily snapshots are [available here](#).
 - 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.

Linux, Mac, Windows
のどれかを選択

The screenshot shows the CRAN website for Windows binaries. The page title is "R for Windows". Under "Subdirectories:", there are four links: "base", "contrib", "old contrib", and "Rtools". Each link has a corresponding description: "base" is for the first-time installation; "contrib" and "old contrib" are for contributed packages; "Rtools" are for building packages. A note at the bottom states that CRAN checks for viruses but cannot guarantee safety. Two arrows point from the text "R本体" to the "base" link and from "Rtools" to the "Rtools" link. The browser's taskbar at the bottom shows the time as 18:46 on 2026/05/07.

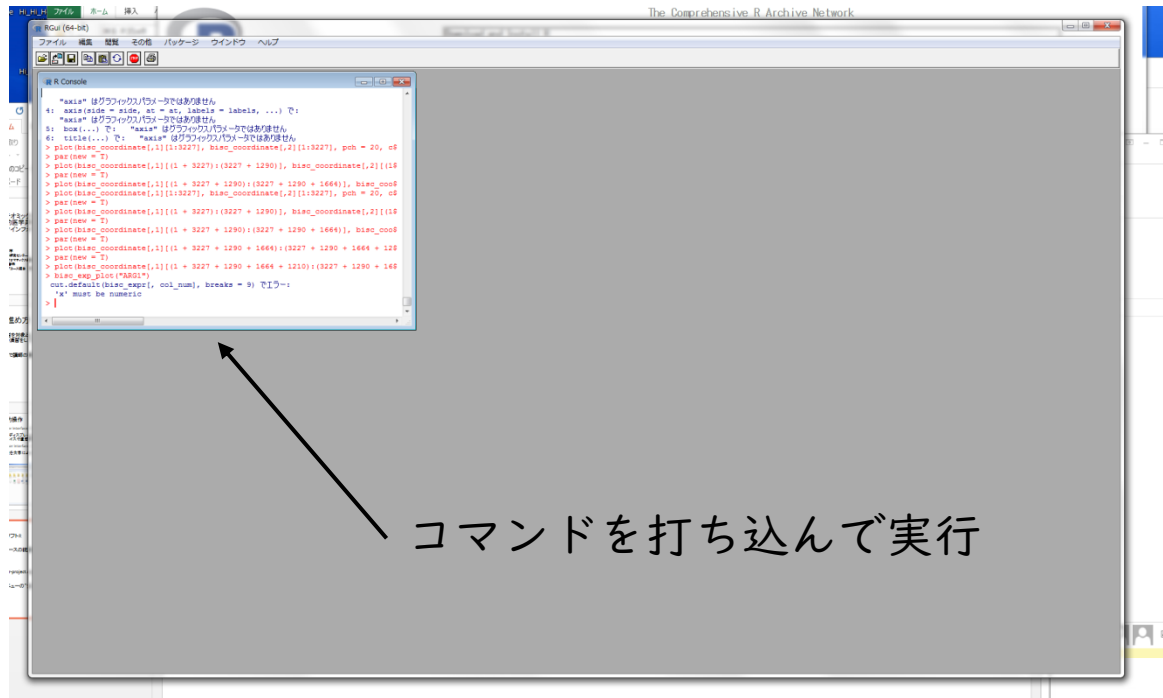
R本体とRtoolsをインストール

R本体

Rtools



Rコンソール



R console



```
> 1+5 ↵  
[1] 6  
> x <- 1 ↵  
> y <- 5 ↵  
> z <- x + y ↵  
> z ↵  
[1] 6
```

↵ : エンター
<- : 代入



Gene Expression Omnibus (GEO)

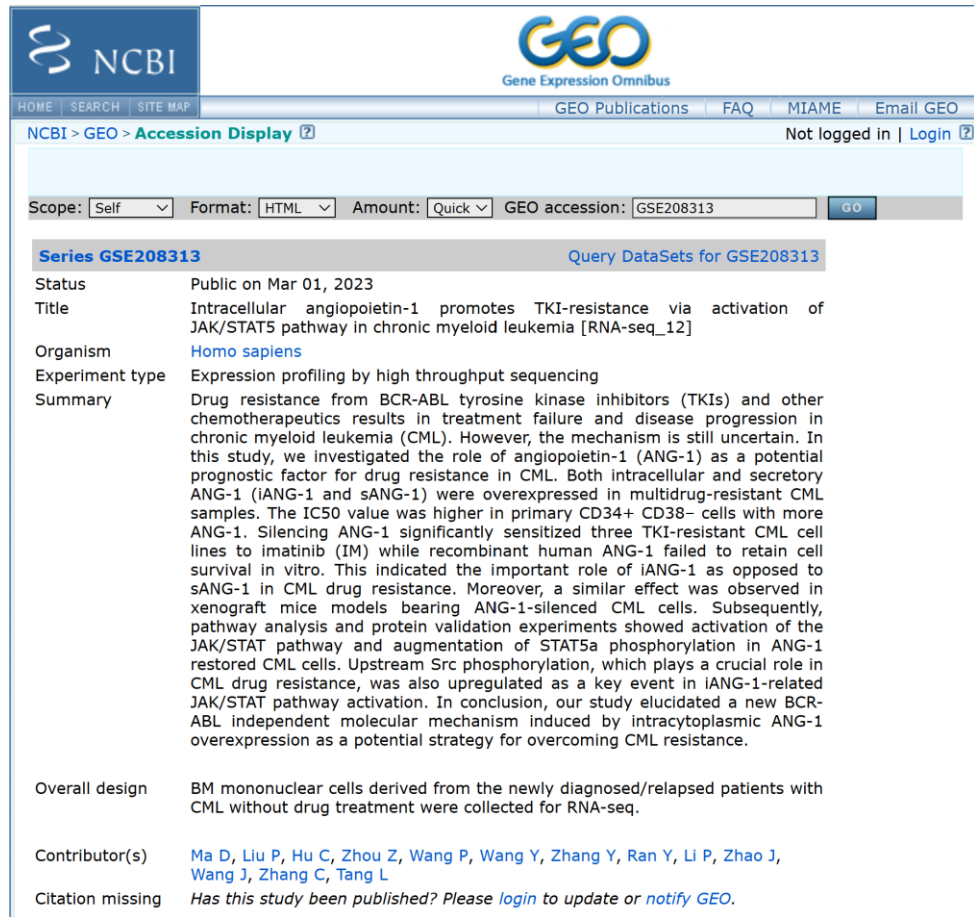
遺伝子発現データのデータベース

The screenshot shows the homepage of the Gene Expression Omnibus (GEO) website. The browser address bar displays <https://www.ncbi.nlm.nih.gov/geo/>. The page features a navigation menu with links for GEO Home, Documentation, Query & Browse, and Email GEO. A search bar is located on the right side of the page, labeled "Keyword or GEO Accession" with a "Search" button. The main content area is divided into three columns: "Getting Started" (Overview, FAQ, About GEO DataSets, About GEO Profiles, About GEO2R Analysis, How to Construct a Query, How to Download Data), "Tools" (Search for Studies at GEO DataSets, Search for Gene Expression at GEO Profiles, Search GEO Documentation, Analyze a Study with GEO2R, Studies with Genome Data Viewer Tracks, Programmatic Access, FTP Site, ENCODE Data Listings and Tracks), and "Browse Content" (Repository Browser, DataSets: 4348, Series: 283558, Platforms: 28383, Samples: 8486611). Below these columns is the "Information for Submitters" section, which includes links for Login to Submit, Submission Guidelines, Update Guidelines, MIAME Standards, Citing and Linking to GEO, Guidelines for Reviewers, and GEO Publications. The footer of the page contains links for NLM, NIH, Email GEO, Disclaimer, Accessibility, and HHS Vulnerability Disclosure. The Windows taskbar at the bottom shows the time as 14:16 on 2026/05/08.

<https://www.ncbi.nlm.nih.gov/geo/>



Gene Expression Omnibus (GEO)



The screenshot shows the NCBI GEO website interface. At the top, there is a navigation bar with links for HOME, SEARCH, SITE MAP, GEO Publications, FAQ, MIAME, and Email GEO. The main content area displays the accession display for GSE208313. The search bar shows the accession number GSE208313 and a GO button. Below the search bar, there is a section for Series GSE208313 with a link to Query DataSets for GSE208313. The main content area is divided into several sections: Status (Public on Mar 01, 2023), Title (Intracellular angiopoietin-1 promotes TKI-resistance via activation of JAK/STAT5 pathway in chronic myeloid leukemia [RNA-seq_12]), Organism (Homo sapiens), Experiment type (Expression profiling by high throughput sequencing), Summary (Drug resistance from BCR-ABL tyrosine kinase inhibitors (TKIs) and other chemotherapeutics results in treatment failure and disease progression in chronic myeloid leukemia (CML). However, the mechanism is still uncertain. In this study, we investigated the role of angiopoietin-1 (ANG-1) as a potential prognostic factor for drug resistance in CML. Both intracellular and secretory ANG-1 (iANG-1 and sANG-1) were overexpressed in multidrug-resistant CML samples. The IC50 value was higher in primary CD34+ CD38- cells with more ANG-1. Silencing ANG-1 significantly sensitized three TKI-resistant CML cell lines to imatinib (IM) while recombinant human ANG-1 failed to retain cell survival in vitro. This indicated the important role of iANG-1 as opposed to sANG-1 in CML drug resistance. Moreover, a similar effect was observed in xenograft mice models bearing ANG-1-silenced CML cells. Subsequently, pathway analysis and protein validation experiments showed activation of the JAK/STAT pathway and augmentation of STAT5a phosphorylation in ANG-1 restored CML cells. Upstream Src phosphorylation, which plays a crucial role in CML drug resistance, was also upregulated as a key event in iANG-1-related JAK/STAT pathway activation. In conclusion, our study elucidated a new BCR-ABL independent molecular mechanism induced by intracytoplasmic ANG-1 overexpression as a potential strategy for overcoming CML resistance.), Overall design (BM mononuclear cells derived from the newly diagnosed/relapsed patients with CML without drug treatment were collected for RNA-seq.), Contributor(s) (Ma D, Liu P, Hu C, Zhou Z, Wang P, Wang Y, Zhang Y, Ran Y, Li P, Zhao J, Wang J, Zhang C, Tang L), and Citation missing (Has this study been published? Please login to update or notify GEO.).

GSE208313

Intracellular angiopoietin-1 promotes TKI-resistance via activation of JAK/STAT pathway chronic myeloid leukemia.

Ma D, Liu P, Hu C, Zhou Z, Wang P, Wang Y, Zhang Y, Ran Y, Li P, Zhao J, Wang J, Zhang C, Tang L.

GSE208313_all.counts.txtファイルを使用



RNA-seqのカウントデータ

The screenshot shows an Excel spreadsheet with the following data:

AccID	LR-EV1	LR-EV2	LR-EV3	LR-SI-1	LR-SI-2	LR-SI-3	EV-1	EV-2	EV-3	iRNA-1	iRNA-2	iRNA-3
3.8-1.2	0	0	0	0	0	0	0	0	0	0	0	0
3.8-1.3	0	0	0	0	0	0	0	0	0	0	0	0
3.8-1.4	0	0	0	0	0	0	0	0	0	0	0	0
3.8-1.5	0	0	0	0	0	0	0	0	0	0	0	0
5-HT3C2	0	0	0	0	0	0	0	0	0	0	0	0
A1BG	22	26	27	38	29	27	39	39	28	66	26	30
A1BG-AS1	34	56	77	60	57	51	69	55	75	61	49	65
A1CF	0	0	0	0	0	0	0	0	0	0	0	0
A2M	0	2	0	5	5	2	48	72	59	74	90	58
A2M-AS1	0	0	1	0	4	0	0	0	0	0	0	0
A2ML1	0	0	0	0	0	0	1	2	0	0	1	3
A2MP1	0	0	0	0	0	0	0	0	0	0	0	0
A3GALT2	0	0	0	0	0	0	0	0	0	0	0	0
A4GALT	146	194	205	257	208	168	1	0	0	4	0	2
A4GNT	0	0	0	0	0	0	0	0	0	0	0	0
AA06	0	0	0	0	0	0	0	0	0	0	0	0
AAAS	1210	1187	1492	1559	1104	1234	1645	2088	2088	1548	1899	1571
AACS	116	129	162	177	173	144	525	664	656	591	596	599
AACSP1	0	0	0	0	0	0	80	106	114	174	178	103
AADAC	0	0	0	0	0	0	0	0	0	1	0	0
AADACL2	0	0	0	0	0	0	0	0	0	0	0	0
AADACL2-	0	0	0	0	0	0	0	0	0	0	0	0
AADACL3	0	0	0	0	0	0	0	0	0	0	0	0

タブ区切りテキスト
ファイル

K562+emp 1~3

K562+AGN-1 4~6



Rのライブラリ

- 一連の計算手順をひとまとめにしたもの。
- さまざまな解析用のライブラリが公開されている。
- インストール、ロードして使用する。
- 書式に従って引数を渡すと計算結果が返ってくる。



edgeRライブラリ

“edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets.”

Chen Y, Chen L, Lun ATL, Baldoni P, Smyth GK (2025). *Nucleic Acids Research*, 53(2), gkaf018.

ユーザーマニュアル

<https://bioconductor.org/packages//release/bioc/vignettes/edgeR/instant/doc/edgeRUsersGuide.pdf>



edgeRライブラリのインストール

```
R console — □ X  
>install.packages("BiocManager") ↵  
>BiocManager::install("edgeR")↵  
>library(edgeR)↵
```

1. BiocManager（パッケージ管理用ライブラリ）のインストール
2. BiocManagerでedgeRをインストール
3. edgeRをロード



ミラーサイトの選択

The screenshot shows the RGui interface with the R Console window open. The console displays the R version (4.5.1) and the command `install.packages("BiocManager")`. A dialog box titled "Secure CRAN mirrors" is open, listing various international mirrors. The "Japan (Yonezawa) [https]" option is highlighted in blue. The dialog box has "OK" and "キャンセル" (Cancel) buttons at the bottom.

R version 4.5.1 (2025-06-13 ucrt) -- "Great Square Root"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、`'license()'` あるいは `'licence()'` と入力してください。

R は多くの貢献者による共同プロジェクトです。
詳しくは `'contributors()'` と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
`'citation()'` と入力してください。

`'demo()'` と入力すればデモをみることができます。
`'help()'` とすればオンラインヘルプが出ます。
`'help.start()'` で HTML ブラウザによるヘルプがみられます。
`'q()'` と入力すれば R を終了します。

[以前にセーブされたワークスペースを復帰します]

```
> install.packages("BiocManager")
```

パッケージを 'C:/Users/Nakabayashi/AppData/Local/R/win-library/4.5' 中にインストール
(`'lib'` が指定されていないため)
--- このセッションで使うために、CRAN のミラーサイトを選んでください ---

Secure CRAN mirrors

- China (Hong Kong) [https]
- China (Jinan) [https]
- China (Lanzhou) [https]
- China (Nanjing) [https]
- China (Shanghai 2) [https]
- China (Shenzhen) [https]
- China (Wuhan) [https]
- Costa Rica [https]
- Cyprus [https]
- Denmark [https]
- East Asia [https]
- Ecuador (Cuenca) [https]
- Finland (Helsinki) [https]
- France (Lyon 1) [https]
- France (Lyon 2) [https]
- France (Paris 1) [https]
- Germany (Erlangen) [https]
- Germany (Göttingen) [https]
- Germany (Leipzig) [https]
- Germany (Münster) [https]
- Greece [https]
- Hungary [https]
- Iceland [https]
- India (Bhubaneswar) [https]
- India (New Delhi) [https]
- Indonesia (Banda Aceh) [https]
- Iran (Mashhad) [https]
- Italy (Milano) [https]
- Italy (Padua) [https]
- Japan (Yonezawa) [https]
- Mexico (Mexico City) [https]
- Mexico (Texcoco) [https]
- Morocco [https]
- Netherlands (Dronnten) [https]
- New Zealand [https]
- Norway [https]
- Poland [https]
- Saudi Arabia (Riyadh) [https]
- Spain (A Coruña) [https]
- Spain (Madrid) [https]
- Sweden (Umeå) [https]
- Switzerland (Zurich 1) [https]
- Taiwan (Taipei) [https]
- UK (Bristol) [https]

OK キャンセル

Japanを選択

ダウンロード完了

```
R Console
R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()' と入力してください。

'demo()' と入力すればデモをみることができます。
'help()' とすればオンラインヘルプが出ます。
'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

[以前にセーブされたワークスペースを復帰します]

> install.packages("BiocManager")
パッケージを 'C:/Users/Nakabayashi/AppData/Local/R/win-library/4.5' 中にイン$
('lib' が指定されていないため)
--- このセッションで使うために、CRAN のミラーサイトを選んでください ---
URL 'https://ftp.yz.yamagata-u.ac.jp/pub/cran/bin/windows/contrib/4.5/BiocManag$
Content type 'application/zip' length 664051 bytes (648 KB)
downloaded 648 KB

パッケージ 'BiocManager' は無事に展開され、MD5 サムもチェックされました

ダウンロードされたパッケージは、以下にあります
C:\Users\Nakabayashi\AppData\Local\Temp\Rtmpy8apDf\downloaded_packages
> |
```



ライブラリのインストール

```
R Console
C:\Users\Nakabayashi\AppData\Local\Temp\Rtmpy8apDf\downloaded_packages
> BiocManager::install("edgeR")
'getOption("repos")' replaces Bioconductor standard repositories, see
'help("repositories", package = "BiocManager")' for details.
Replacement repositories:
  CRAN: https://ftp.yz.yamagata-u.ac.jp/pub/cran
Bioconductor version 3.21 (BiocManager 1.30.27), R 4.5.1 (2025-06-13 ucrt)
Installation paths not writeable, unable to update packages
path: C:/Program Files/R/R-4.5.1/library
packages:
  boot, cluster, foreign, lattice, Matrix, mgcv, nlme, rpart, survival
Old packages: 'base64enc', 'BH', 'bslib', 'cli', 'cppl1', 'crosstalk', 'curl',
'data.table', 'digest', 'dplyr', 'evaluate', 'fastDummies', 'fitdistrplus',
'fs', 'future', 'future.apply', 'ggplot2', 'ggrepel', 'ggridges', 'globals',
'glue', 'gplots', 'here', 'highr', 'htmltools', 'httpuv', 'httr', 'igraph',
'irlba', 'isoband', 'knitr', 'later', 'lazyeval', 'leidenbase', 'lifecycle',
'limma', 'listenv', 'magrittr', 'openssl', 'parallely', 'patchwork',
'pillar', 'plotly', 'png', 'progressr', 'promises', 'purrr', 'rappdirs',
'Rcpp', 'RcppAnnoy', 'RcppArmadillo', 'reshape2', 'reticulate', 'rlang',
'rmarkdown', 'ROCR', 'rprojroot', 'sctransform', 'Seurat', 'SeuratObject',
'shiny', 'sourcetools', 'sp', 'spam', 'spatstat.data', 'spatstat.explore',
'spatstat.geom', 'spatstat.random', 'spatstat.univar', 'spatstat.utils',
'statmod', 'stringr', 'tibble', 'tidyr', 'tinytex', 'uwot', 'vctrs',
'viridisLite', 'xfun', 'xtable', 'yaml', 'zoo'
Update all/some/none? [a/s/n]: |
```

aを入力



インストール完了

```
R Console
パッケージ 'tibble' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'tibble' の既存のインストールを取り除くことが出来ませんでした
警告: 'tibble' を復帰させました
パッケージ 'tidyr' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'tidyr' の既存のインストールを取り除くことが出来ませんでした
警告: 'tidyr' を復帰させました
パッケージ 'tinytex' は無事に展開され、MD5 サムもチェックされました
パッケージ 'uwot' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'uwot' の既存のインストールを取り除くことが出来ませんでした
警告: 'uwot' を復帰させました
パッケージ 'vctrs' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'vctrs' の既存のインストールを取り除くことが出来ませんでした
警告: 'vctrs' を復帰させました
パッケージ 'viridisLite' は無事に展開され、MD5 サムもチェックされました
パッケージ 'xfun' は無事に展開され、MD5 サムもチェックされました
パッケージ 'xtable' は無事に展開され、MD5 サムもチェックされました
パッケージ 'yaml' は無事に展開され、MD5 サムもチェックされました
パッケージ 'zoo' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'zoo' の既存のインストールを取り除くことが出来ませんでした
警告: 'zoo' を復帰させました

ダウンロードされたパッケージは、以下にあります
      C:\Users\Nakabayashi\AppData\Local\Temp\Rtmpy8apDf\downloaded_packages
38 件の警告がありました (警告を見るには warnings() を使って下さい)
> |
```

警告は無視して大丈夫



edgeRをロード

```
R Console
警告: 'tibble' を復帰させました
パッケージ 'tidyr' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'tidyr' の既存のインストールを取り除くことが出来ませんでした
警告: 'tidyr' を復帰させました
パッケージ 'tinytex' は無事に展開され、MD5 サムもチェックされました
パッケージ 'uwot' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'uwot' の既存のインストールを取り除くことが出来ませんでした
警告: 'uwot' を復帰させました
パッケージ 'vctrs' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'vctrs' の既存のインストールを取り除くことが出来ませんでした
警告: 'vctrs' を復帰させました
パッケージ 'viridisLite' は無事に展開され、MD5 サムもチェックされました
パッケージ 'xfun' は無事に展開され、MD5 サムもチェックされました
パッケージ 'xtable' は無事に展開され、MD5 サムもチェックされました
パッケージ 'yaml' は無事に展開され、MD5 サムもチェックされました
パッケージ 'zoo' は無事に展開され、MD5 サムもチェックされました
警告: パッケージ 'zoo' の既存のインストールを取り除くことが出来ませんでした
警告: 'zoo' を復帰させました

ダウンロードされたパッケージは、以下にあります
  C:\Users\Nakabayashi\AppData\Local\Temp\Rtmpy8apDf\downloaded_packages
38 件の警告がありました (警告を見るには warnings() を使って下さい)
> library(edgeR)
要求されたパッケージ limma をロード中です
> |
```

edgeRパッケージが使用可能な状態



データの読み込み

R console



```
> x <- read.table("GSE208313_all.counts.txt", header = T, sep = "¥t",  
row.name = 1) ↵
```

ファイルからデータを読み込み変数xに代入

read.table(): ファイルを読み込み。ファイル名は各自のPC環境に合わせて変更

“/Users/ユーザ名/フォルダ名/ファイル名”

header=T: 列名ありで読み込み

row.name=1: 1列目を行名にする

sep="¥t": タブ区切り。¥はバックスラッシュ



データの成型

R console



```
> x <- x[,1:6]↵  
> head(x)↵
```

1. `x[,1:6]` : 行列`x`の1列目から6列目までを抽出
`x[行,列]`で行列の一部を抽出
`x[a:b,c:d]` : 行列`x`の`a`~`b`行、`c`~`d`列を抽出
2. `head()` : 行列の上6行を表示する関数



解析に使うデータ

```
R Console
file(file, "rt") で:
ファイル '/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/GSE208313$
> x <- read.table("/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/$
エラー: 想定外のシンボルです ( "x <- read.table("/Users/Nakabayashi/Dropbox/PC$
> x <- read.table("/Users/Nakabayashi/Dropbox/PC\ (2)/GSE208313_all.counts.txt/$
file(file, "rt") でエラー: コネクションを開くことができません
追加情報: 警告メッセージ:
file(file, "rt") で:
ファイル '/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/GSE208313$
> x <- read.table("/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/GS
file(file, "rt") でエラー: コネクションを開くことができません
追加情報: 警告メッセージ:
file(file, "rt") で:
ファイル '/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/GSE208313$
> x <- read.table("/Users/Nakabayashi/Dropbox/PC (2)/Downloads/GSE208313_all.co$
> x <- x[,1:6]
> head(x)
      LR.EV1 LR.EV2 LR.EV3 LR.SI.1 LR.SI.2 LR.SI.3
3.8-1.2      0      0      0      0      0      0
3.8-1.3      0      0      0      0      0      0
3.8-1.4      0      0      0      0      0      0
3.8-1.5      0      0      0      0      0      0
5-HT3C2      0      0      0      0      0      0
A1BG         22     26     27     38     29     27
> |
```

上6行が表示される

6列を解析に使う



edgeRでデータを読み込み

R console



```
> d <- DGEList(counts = x, group = c(rep("A", 3), rep("B", 3)))
```

カウントデータをedgeRに渡し、結果を変数dに代入

DGEList(): edgeRで解析するデータの読み込み

counts: カウントデータ

group: サンプルのグループ分け

rep(): 繰り返し行列を作る関数

rep("A", 3)で{"A" "A" "A"}を作成

c(): 行列を繋げる関数

c(rep("A", 3), rep("B", 3))で{"A" "A" "A" "B" "B" "B"}を作成

カウントデータの1~3列目をAグループ、4~6列目をBグループとする。



発現変動遺伝子の抽出

R console



```
> d <- calcNormFactors(d)␣  
> d <- estimateCommonDisp(d)␣  
> d <- estimateTagwiseDisp(d)␣  
> result <- exactTest(d)␣
```

1. 正規化
2. 全体の分散を計算
3. グループ間の分散を計算
4. 発現変動遺伝子を抽出し、結果を変数resultに代入



発現変動遺伝子の抽出

```
R Console
file(file, "rt") でエラー: コネクションを開くことができません
追加情報: 警告メッセージ:
file(file, "rt") で:
  ファイル '/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/GSE208313$
> x <- read.table("/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/G$
file(file, "rt") でエラー: コネクションを開くことができません
追加情報: 警告メッセージ:
file(file, "rt") で:
  ファイル '/Users/Nakabayashi/Dropbox/PC (2)/GSE208313_all.counts.txt/GSE208313$
> x <- read.table("/Users/Nakabayashi/Dropbox/PC (2)/Downloads/GSE208313_all.co$
> x <- x[,1:6]
> head(x)
      LR.EV1 LR.EV2 LR.EV3 LR.SI.1 LR.SI.2 LR.SI.3
3.8-1.2      0      0      0        0        0        0
3.8-1.3      0      0      0        0        0        0
3.8-1.4      0      0      0        0        0        0
3.8-1.5      0      0      0        0        0        0
5-HT3C2      0      0      0        0        0        0
AlBG         22     26     27       38       29       27
> d <- DGEList(counts = x, group = c(rep("A",3), rep("B", 3)))
> d <- calcNormFactors(d)
> d <- estimateCommonDisp(d)
> d <- estimateTagwiseDisp(d)
> result <- exactTest(d)
> |
```

エラーが表示されなければ解析完了



結果のファイル出力

R console



```
> result.table <- topTags(result, n = Inf)@.Data[[1]] ↵  
> result.sig <- subset(result.table, result.table$FDR < 0.05) ↵  
> write.table(result.sig, "result.sig.txt", quote=F, sep="¥t") ↵  
> head(result.sig) ↵
```

1. resultのロットからデータをresult.tableに代入
2. $FDR < 0.05$ を満たすものだけをresult.sigに代入
subset(条件) : 条件を満たすものだけを選別する関数
result.table\$FDR : result.tableのFDR列
3. result.sigをテキストファイルに書き出し
write.table() : ファイルを書き出す関数
quote=F : “” なし



統計学的に有意に発現変動している遺伝子

```
R Console
3.8-1.2    0    0    0    0    0    0
3.8-1.3    0    0    0    0    0    0
3.8-1.4    0    0    0    0    0    0
3.8-1.5    0    0    0    0    0    0
5-HT3C2    0    0    0    0    0    0
A1BG       22   26   27   38   29   27
> d <- DGEList(counts = x, group = c(rep("A",3), rep("B", 3)))
> d <- calcNormFactors(d)
> d <- estimateCommonDisp(d)
> d <- estimateTagwiseDisp(d)
> result <- exactTest(d)
> result.table <- topTags(result, n = Inf)@Data[[1]]
エラー: 名前 "Data" というスロットが、クラス "TopTags" のこのオブジェクトには存在
> result.table <- topTags(result, n = Inf)@.Data[[1]]
> result.sig <- subset(result.table, result.table$FDR < 0.05)
> write.table(result.sig, "/Users/Nakabayashi/Dropbox/PC (2)/Downloads/GSE20831$
> head(result.sig)
      logFC  logCPM      PValue      FDR
ANGPT1 -1.742428  7.001475  5.735647e-205  3.112220e-200
BMP6    -2.159118  5.459206  3.272902e-200  8.879546e-196
GALNT1  1.456208  8.037654  1.814305e-176  3.281533e-172
HMCN1   -2.655147  4.194349  6.988917e-134  9.480640e-130
CD84    -1.536556  5.612708  1.609246e-116  1.746386e-112
TIMP3   -1.012372  8.069077  1.126812e-113  1.019033e-109
> |
```

Fold Change (FC) : 発現比 A/B
FDR : False Discovery Rate
補正された有意水準
FDR<0.05を有意とする



出力されたファイル

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		logFC	logCPM	PValue	FDR															
2	ANGPT1	-1.74243	7.001475	5.7356473	3.11221960461e-200															
3	BMP6	-2.15912	5.459206	3.2729016	8.8795458882199e-196															
4	GALNT1	1.456208	8.037654	1.8143047	3.28153304463465e-172															
5	HMCN1	-2.65515	4.194349	6.9889167	9.48064031665043e-130															
6	CD84	-1.53656	5.612708	1.6092462	1.74638618403467e-112															
7	TIMP3	-1.01237	8.069077	1.1268121	1.01903253953213e-109															
8	HSPA5	-0.84208	9.373579	1.9718398	1.52848576727209e-94															
9	RPRD1A	1.014084	8.277033	5.7583939	3.90570269446082e-93															
10	SLC39A6	1.051719	7.087404	2.6564454	1.6015709642396e-84															
11	ELP2	0.972029	8.129984	1.1428147	6.20102703792218e-84															
12	TMED2	-0.83844	8.430051	7.0391762	3.4722976834778e-83															
13	SUPT16H	-0.84911	8.285728	1.6484395	7.45383156745589e-82															
14	CALR	-0.77933	10.36637	4.6826584	1.95450563239047e-80															
15	RYR3	2.674071	3.687303	1.0942884	4.24122764267777e-72															
16	SS18	0.868965	7.09017	1.4876305	5.3813547404698e-67															
17	ZNF24	1.119849	8.950841	3.0049726	1.01908014030726e-62															
18	TPGS2	0.850634	7.191031	2.0598813	6.574777248475e-61															
19	C18orf21	1.041796	5.833411	2.7095741	8.16801119533692e-59															
20	SEPHS2	-0.84927	7.769534	4.5422625	1.29719847288659e-57															
21	IGFBP4	-0.99378	8.096243	1.5840074	4.29749134428272e-56															
22	MYO10	-1.60999	4.084572	1.8257923	4.71758665820086e-56															
23	ACVR1	-1.65693	3.866688	2.2505067	5.55067025104458e-56															
24	PHF6	-1.10203	5.950838	1.4524363	3.42654993588808e-55															

Volcano Plot

R console



```
> x <- result.table$logFC ↵  
> y <- -log10(result.table$PValue)↵  
> pval = -log10(0.05)↵  
> plot(x, y, pch = 20, cex = 0.5, xlab = "logFC", ylab = "-log10(P-  
value)", col = ifelse(y <= pval, "black", ifelse(x >= 1, "red", ifelse(x  
<= -1, "blue", "black"))))↵
```

x軸 : logFC

y軸 : $-\log_{10}$ (p-value)

plot(x,y) : プロット

pch=20 : 小さい点で表示

cex : フォントサイズ

xlab : x軸ラベル

ylab : y軸ラベル

col : 色 $y \leq pval$ なら黒

$x \geq 1$ なら赤

$x \leq -1$ なら青



Volcano Plot

R console

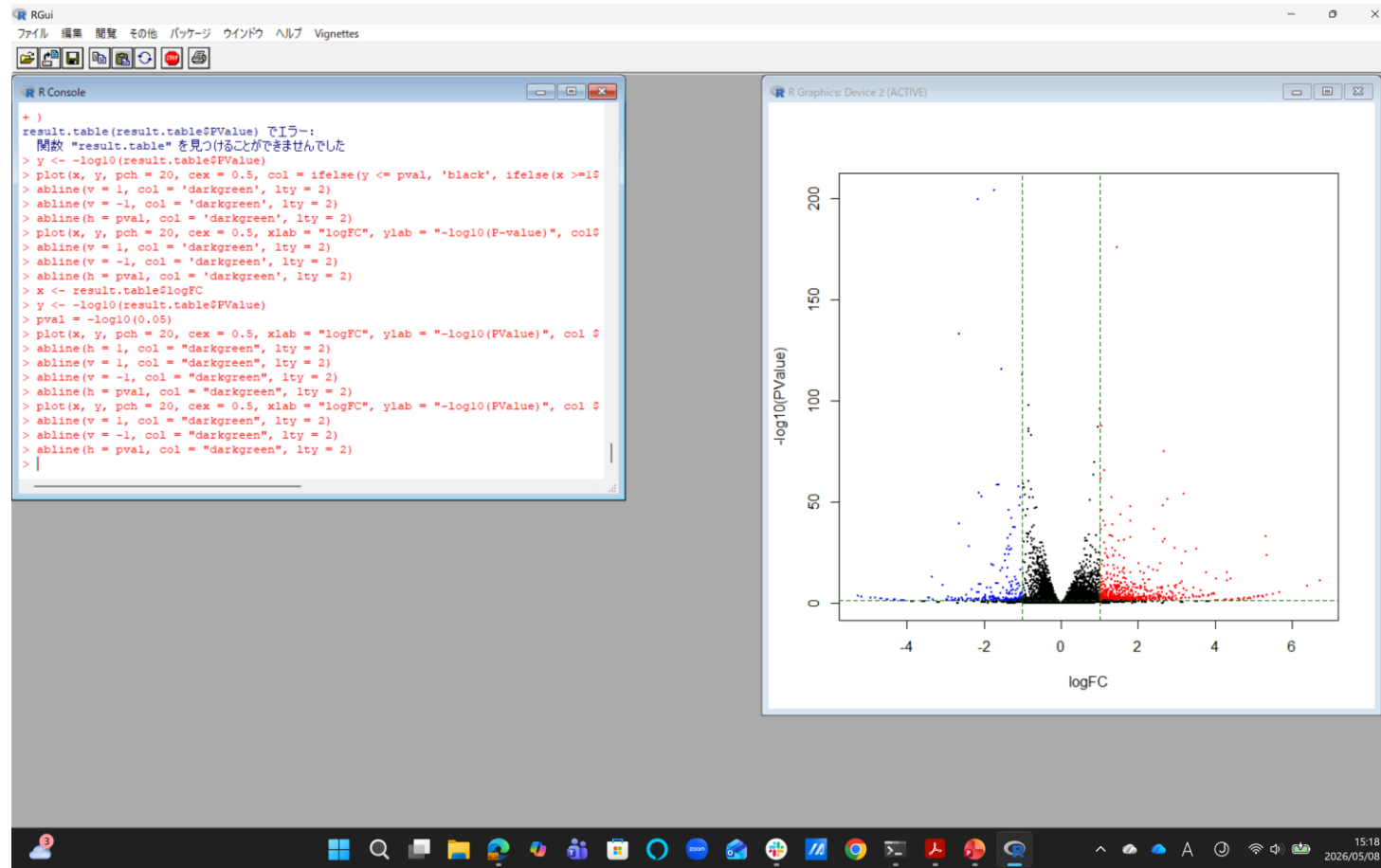


```
> abline(v = 1, col = "darkgreen", lty = 2)↵  
> abline(v = -1, col = "darkgreen", lty = 2)↵  
> abline (h = pval, col = "darkgreen", lty = 2)↵
```

x=1, -1に垂直線
y=pvalに水平線
lty=2 : 破線



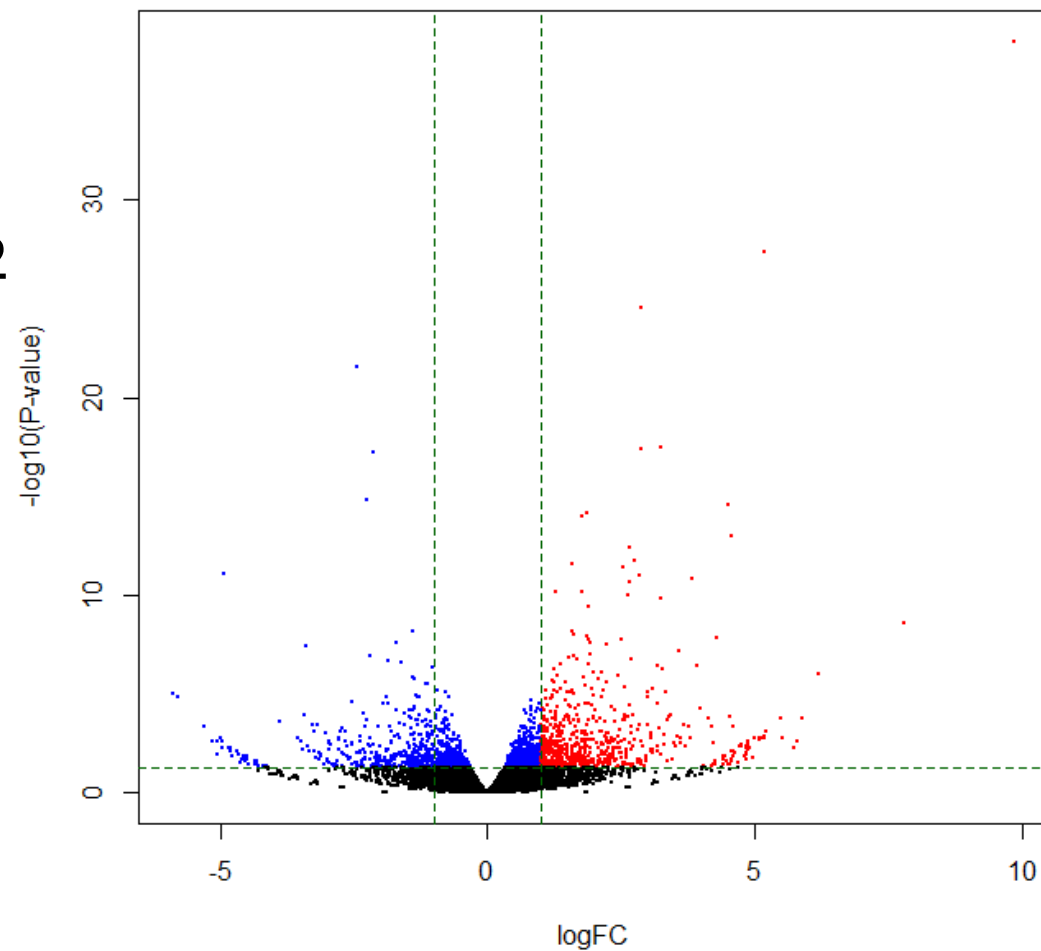
Volcano Plot



7~12列のvolcano plot

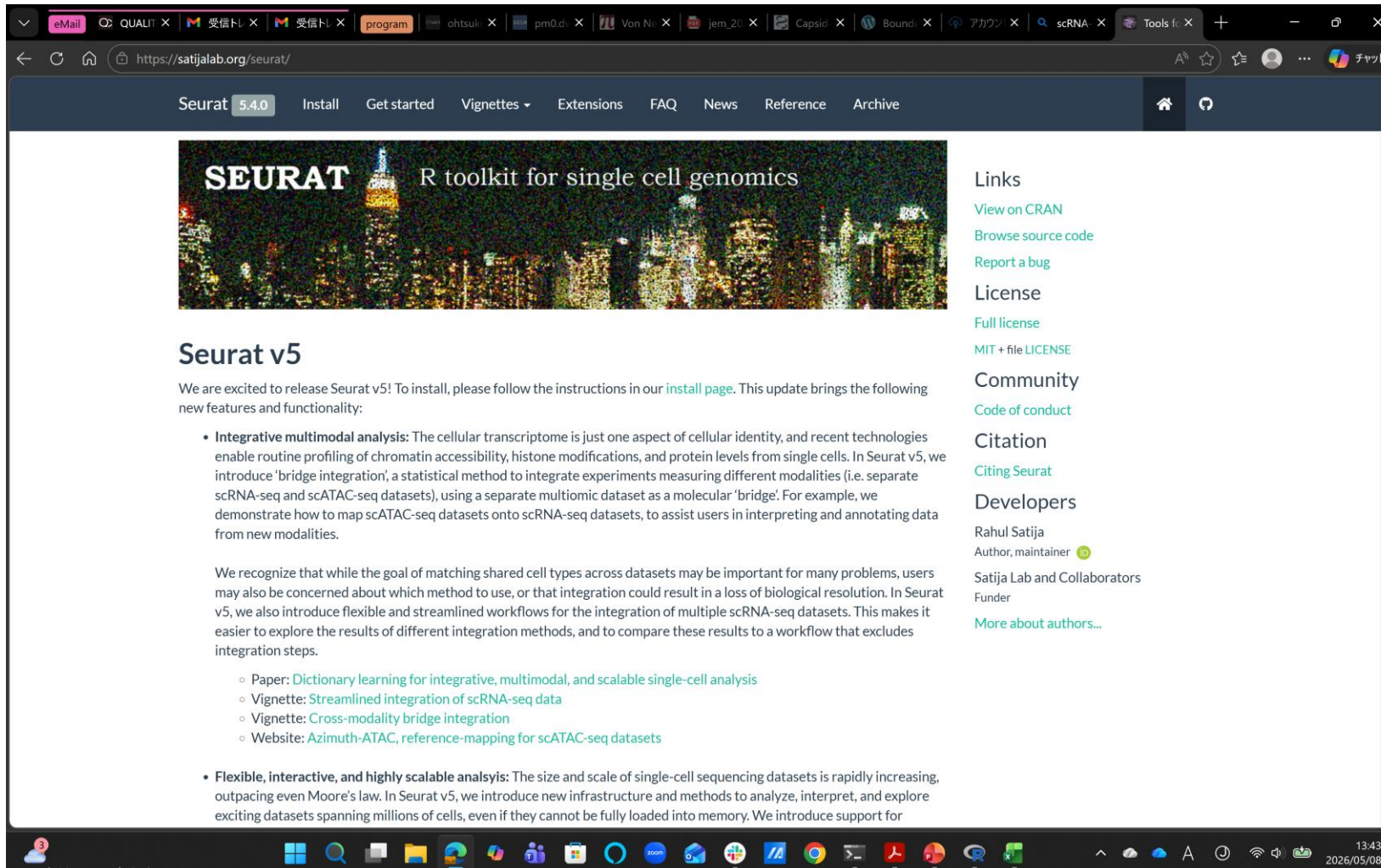
LAMA84R+emp 7~9

LAMA84R+AGN-1 10~12



Seurat

- scRNA-seq解析用のRパッケージ
<https://satijalab.org/seurat/>



The screenshot shows the Seurat website homepage. The main heading is "SEURAT R toolkit for single cell genomics". Below this, there is a section titled "Seurat v5" with a sub-heading "We are excited to release Seurat v5!". The text describes the new features and functionality, including "Integrative multimodal analysis" and "Flexible, interactive, and highly scalable analysis". There are also navigation links for "Install", "Get started", "Vignettes", "Extensions", "FAQ", "News", "Reference", and "Archive".

Seurat v5

We are excited to release Seurat v5! To install, please follow the instructions in our [install page](#). This update brings the following new features and functionality:

- **Integrative multimodal analysis:** The cellular transcriptome is just one aspect of cellular identity, and recent technologies enable routine profiling of chromatin accessibility, histone modifications, and protein levels from single cells. In Seurat v5, we introduce 'bridge integration', a statistical method to integrate experiments measuring different modalities (i.e. separate scRNA-seq and scATAC-seq datasets), using a separate multiomic dataset as a molecular 'bridge'. For example, we demonstrate how to map scATAC-seq datasets onto scRNA-seq datasets, to assist users in interpreting and annotating data from new modalities.

We recognize that while the goal of matching shared cell types across datasets may be important for many problems, users may also be concerned about which method to use, or that integration could result in a loss of biological resolution. In Seurat v5, we also introduce flexible and streamlined workflows for the integration of multiple scRNA-seq datasets. This makes it easier to explore the results of different integration methods, and to compare these results to a workflow that excludes integration steps.

- Paper: [Dictionary learning for integrative, multimodal, and scalable single-cell analysis](#)
- Vignette: [Streamlined integration of scRNA-seq data](#)
- Vignette: [Cross-modality bridge integration](#)
- Website: [Azimuth-ATAC, reference-mapping for scATAC-seq datasets](#)

- **Flexible, interactive, and highly scalable analysis:** The size and scale of single-cell sequencing datasets is rapidly increasing, outpacing even Moore's law. In Seurat v5, we introduce new infrastructure and methods to analyze, interpret, and explore exciting datasets spanning millions of cells, even if they cannot be fully loaded into memory. We introduce support for

Links

- [View on CRAN](#)
- [Browse source code](#)
- [Report a bug](#)

License

- [Full license](#)
- [MIT + file LICENSE](#)

Community

- [Code of conduct](#)

Citation

- [Citing Seurat](#)

Developers

- Rahul Satija
Author, maintainer
- Satija Lab and Collaborators
Funder
- [More about authors...](#)

Get started



Guided tutorial
2700PBMCs



- 習うより慣れよ

