

2012.08.24

R/Bioconductor によるNGS解析: Part 1

# R/Bioconductor の基礎

Itoshi NIKAIDO, Ph.D.

RIKEN CDB@Kobe

# NGS現場の会

---



[ホーム](#) [NGS現場の会とは？](#) [NGS現場の会の活動](#) [バナー募集](#) [入会方法](#) [研究会](#)

## HOME

[NGS現場の会 第二回研究会 \(2012/05/23-25, 大阪\) のページはこちら](#)

### NGS現場の会とは？

NGS (Next Generation Sequencing) 分野の技術革新は、もはや個々人では把握できない程、急速に進んでいます。激動のNGS分野の動向をフォローし、知識やノウハウをアップデートしてゆくことを目的として、NGS研究や技術開発、関連企業などの現場にいる人たちで、現場の人のためのコミュニティを立ち上げました。それが、NGS現場の会です。

<http://ngs-field.org/>

# はじめに

講義で使用するレジュメ、ソースコードとデータ  
はすべて以下からダウンロードできます。

<http://cat.HackingIsBelieving.org/lecture/>



この作品は [クリエイティブ・コモンズ 表示 - 非営利 2.1 日本 ライセンス](#)の下に提供されています。

# 今日のメニュー

RとBioCを理解し、次世代DNAシーケンサー解析の基礎を作る

1. R/Bioconductorの基礎
2. RNA-seqデータ解析の基礎
3. ChIP-seqデータ解析の基礎

# 今日のメニュー

RとBioCを理解し、次世代DNAシーケンサー解析の基礎を作る

基本的なR/BioCの使いかたを覚えれば、exomeやCNV, SNP, microarray, Proteomics解析などを自習することも可能

# What's R

統計解析環境Rとそのバイオインフォマティクスパッケージ集


- 本格的な統計解析環境
- オープンソース
- 豊富なパッケージ集
  - CRAN, Bioconductor

# Bioconductor

レビューされたバイオインフォマティクスのソフトウェア集

- バイオインフォの解析ツール
  - Software package
- たくさんデータベース
  - annotation package

<http://www.bioconductor.org>



**Bioconductor**  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

[Home](#)[Install](#)[Help](#)[Developers](#)[About](#)

Search:

[Home](#) » [Bioconductor 2.10](#) » [Software Packages](#) » BrainStars

## BrainStars

### query gene expression data and plots from BrainStars (B\*)

Bioconductor version: Release (2.10)

This package can search and get gene expression data and plots from BrainStars (B\*). BrainStars is a quantitative expression database of the adult mouse brain. The database has genome-wide expression profile at 51 adult mouse CNS regions.

Author: Itoshi NIKAIDO <dritoshi at gmail.com>

Maintainer: Itoshi NIKAIDO <dritoshi at gmail.com>

To install this package, start R and enter:

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

To cite this package in a publication, start R and enter:

```
citation("BrainStars")
```

#### Workflows »

Common Bioconductor workflows include:

- [Oligonucleotide Arrays](#)
- [High-throughput Sequencing](#)
- [Annotation](#)
- [Variants](#)
- [Flow Cytometry](#) and other assays

#### Mailing Lists »

Post questions about Bioconductor packages to our mailing lists. Read the [posting guide](#) before posting!

- [bioconductor](#)
- [bioc-devel](#)

BrainStars for R: 世界最大級のマウス脳トランスクリプトームデータの解析ツール  
2.10時点で日本人唯一のパッケージ採録者



# Incentive to study R/BioC

R/Bioconductor を学ぶたったひとつの理由

- 実験生物学と統計解析は切り離せない
- ほかのプログラミング言語よりも身近に使う一生モノの技術になる

# Rの理解への最短距離

プログラミング言語の学びかた

- データの型と構造とその変換を知る
- 構文を理解する
  - 3つの基本的な制御
- 自分が利用する機能(関数, メソッド)から覚える

# データ構造

Rで書くようになる

vector

3	2	4	1	...
---	---	---	---	-----

`c(3, 2, 4, 1)`

list

3	A	4	G	...
---	---	---	---	-----

1D

`list(3, "A", 4, "G")`

matrix

3	2	4	1	...
2	0	5	3	
1	1	7	3	
⋮				

```
matrix(  
  c(3, 2, 1, 2, 0, 1, 4, 5, 7, 1, 3, 3),  
  nrow = 3  
)
```

same types

data frame

3	A	4	C	...
2	T	5	G	
1	G	7	C	
⋮				

2D

```
data.frame(  
  first      = c(3, 2, 1),  
  first.base = c("A", "T", "G"),  
  second     = c(4, 5, 7),  
  second.base = c("C", "G", "C")  
)
```

differential types

# データにアクセスする

vector と list の場合

vector

3	2	4	1
---	---	---	---

 ...

list

3	A	4	G
---	---	---	---

 ...

1D

```
> x <- c(3,2,4,1)
> x[1]
4
> x[1] <- 4
> x
4 2 4 1
> x[1:2]
4 2
> x[1:2] <- c(3,8)
> x
3 8 4 1
```

```
> x <- list(3,2,4,1)
> x[[1]]
3
> x[[2]] <- 5
x
[[1]]
3
[[2]]
2
[[3]]
4
[[4]]
1
```

# データにアクセスする

matrix と data frame の場合

matrix

3	2	4	1	...
2	0	5	3	
1	1	7	3	

⋮

```
> x <- matrix(
  c(3,2,1,2,0,1,4,5,7,1,3,3),
  nrow = 3
)
> x[1,]      # 1行目にアクセス
3 2 4 1
> x[,2]      # 2列目にアクセス
2 0 1
> x[3,2]     # 3行2列目にアクセス
1
> x[3,2] <- 9 # 3行2列目に9を
> x[3,]
1 9 7 3
```

same types

data frame

3	A	4	C	...
2	T	5	G	
1	G	7	C	

⋮

```
> x <- data.frame(
  first      = c(3,2,1),
  first.base = c("A", "T", "G"),
  second     = c(4,5,7),
  second.base = c("C", "G", "C")
)
> x$first
[1] 3 2 1
> x$first.base
[1] A T G
Levels: A G T
> x$first.base <- c("T", "C", "A")
> x$first.base
[1] T C A
Levels: T C A
```

differential types

2D