



WPI Osaka University  
**iFReC**



Yoshiaki Yasumizu  
@yyoshiaki

# ブラウザで完結するRNA-seqデータ解析

安水 良明

大阪大学 免疫学フロンティア研究センター 実験免疫学  
大阪大学先導的学際研究機構生命医科学融合フロンティア研究部門

2022/12/22 AJACS オンライン

# 今日の話の対象となる人

- RNAseqってなに？
- バルクのRNAseqに興味はあるけど解析難しそう
- 解析は得意だけど、毎回のバルクRNAseqの解析に時間がかかってしまう

近年の生命科学研究ではなくてはならない技術です。  
息を吸うようにRNAseqをしましょう。

# 息をするようにRNAseqをするには？

解析に1ヶ月もかけてられない

-> ブラウザ上で素早く解析

こういう実験・検体のデータはすでに誰かがやっているかも？

-> 公共データ解析

※ no code ですべてができるわけではありません

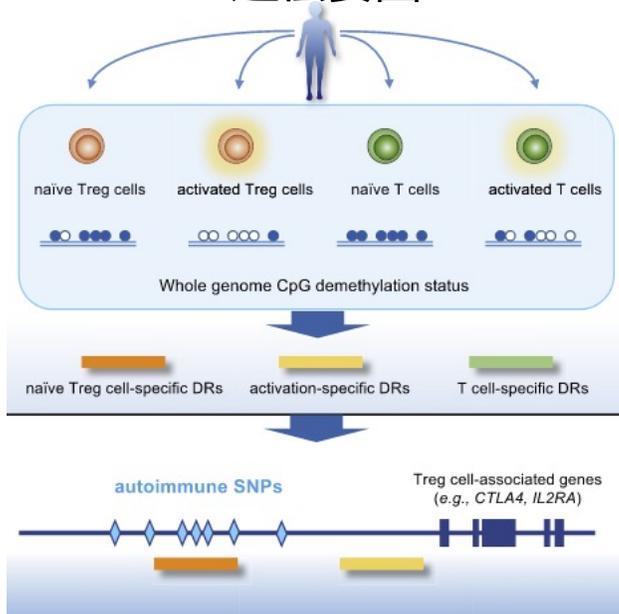
# 自己紹介

## 経歴

2017年- 大阪大学 MD研究者育成プログラム (坂口志文先生)  
 Univ. Bonn Genomics and Immunoregulation (J. Schultze Lab), 大阪大学 遺伝統計学教室 (岡田随象先生) インターン  
 2019年 大阪大学 医学部医学科卒業  
 2019年 - 2021年 大阪大学医学部附属病院 初期研修医  
 2019年 - 2022年 大阪大学大学院医学系研究科 神経内科学 招聘教員  
 2020年 - 大阪大学先導的学際研究機構生命医科学融合フロンティア研究部門 若手兼任教員  
 2021年 - 大阪大学免疫学フロンティア研究センター 実験免疫学 博士課程 (卓越大学院プログラム)

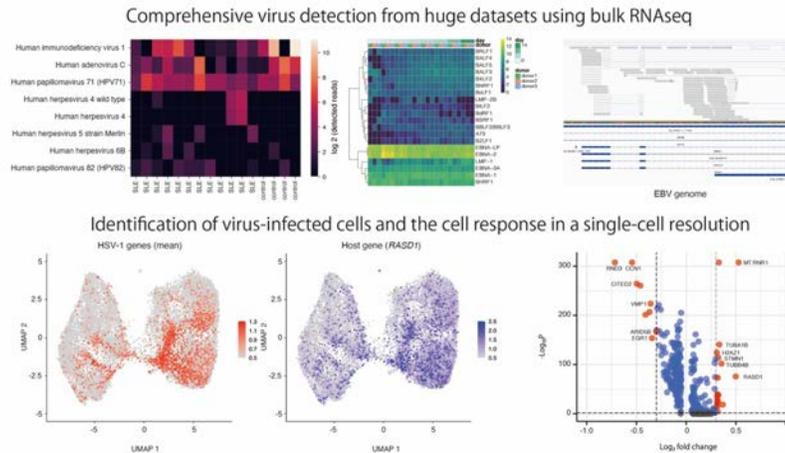


## 遺伝要因



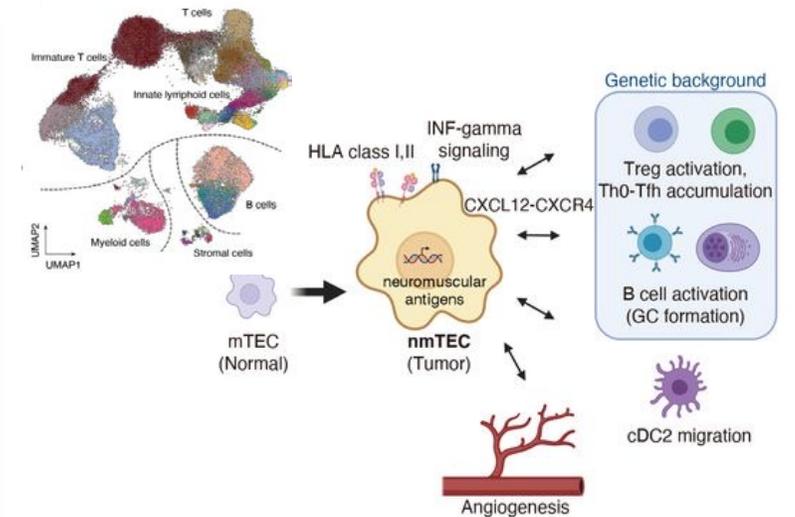
Ohkura and Yasumizu *et al.*, 2020, *Immunity*

## 環境要因



Yasumizu *et al.*, 2020, *Bioinformatics*

## 疾患解析 (重症筋無力症)



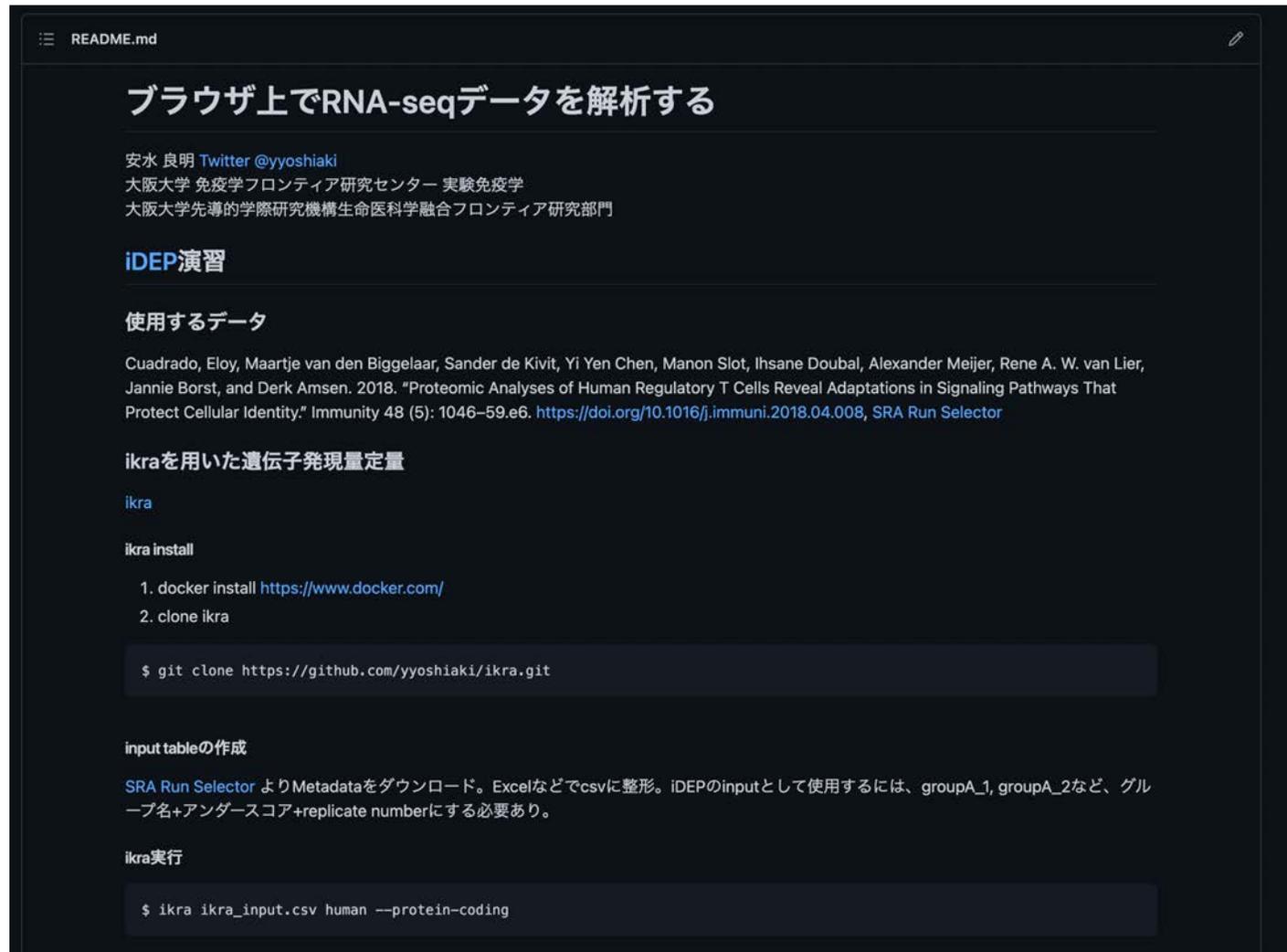
Yasumizu *et al.*, 2022, *Nat. Commun.*

# Agenda

- バルクRNAseqとは？
- バルクRNAseq解析の実際
- ブラウザを用いたバルクRNAseq解析
  - iDEP(本日のメイン)
  - BioJupies
  - RaNA-seq
- 実例紹介（胸腺腫合併重症筋無力症の解析）

**各サイトへのアクセスは、本会終了後にお願いいたします。**

# GitHub上に本日の資料があります



The image shows a screenshot of a GitHub repository's README file. The title is "ブラウザ上でRNA-seqデータを解析する". The author is 安水 良明 (@yyoshiaki) from Osaka University. The main heading is "iDEP演習". Under "使用するデータ", it lists a paper by Cuadrado et al. (2018) on proteomic analyses of human regulatory T cells. The "ikraを用いた遺伝子発現量定量" section includes a link to the "ikra" repository and a list of steps: 1. docker install, 2. clone ikra. A terminal snippet shows the command to clone the repository. The "input tableの作成" section explains how to download metadata from SRA Run Selector and format it as a CSV file. The "ikra実行" section shows a terminal command to run the ikra tool on a specific CSV file.

```
README.md
```

## ブラウザ上でRNA-seqデータを解析する

安水 良明 [Twitter @yyoshiaki](#)  
大阪大学 免疫学フロンティア研究センター 実験免疫学  
大阪大学先導的学際研究機構生命医科学融合フロンティア研究部門

### iDEP演習

#### 使用するデータ

Cuadrado, Eloy, Maartje van den Biggelaar, Sander de Kivit, Yi Yen Chen, Manon Slot, Ihsane Doubal, Alexander Meijer, Rene A. W. van Lier, Jannie Borst, and Derk Amsen. 2018. "Proteomic Analyses of Human Regulatory T Cells Reveal Adaptations in Signaling Pathways That Protect Cellular Identity." *Immunity* 48 (5): 1046–59.e6. <https://doi.org/10.1016/j.immuni.2018.04.008>, SRA Run Selector

#### ikraを用いた遺伝子発現量定量

[ikra](#)

#### ikra install

1. docker install <https://www.docker.com/>
2. clone ikra

```
$ git clone https://github.com/yyoshiaki/ikra.git
```

#### input tableの作成

SRA Run Selector よりMetadataをダウンロード。Excelなどでcsvに整形。iDEPのinputとして使用するには、groupA\_1, groupA\_2など、グループ名+アンダースコア+replicate numberに必要あり。

#### ikra実行

```
$ ikra ikra_input.csv human --protein-coding
```

<https://github.com/AJACS-training/AJACS95>

こちらも参考に <https://github.com/AJACS-training/AJACS84>

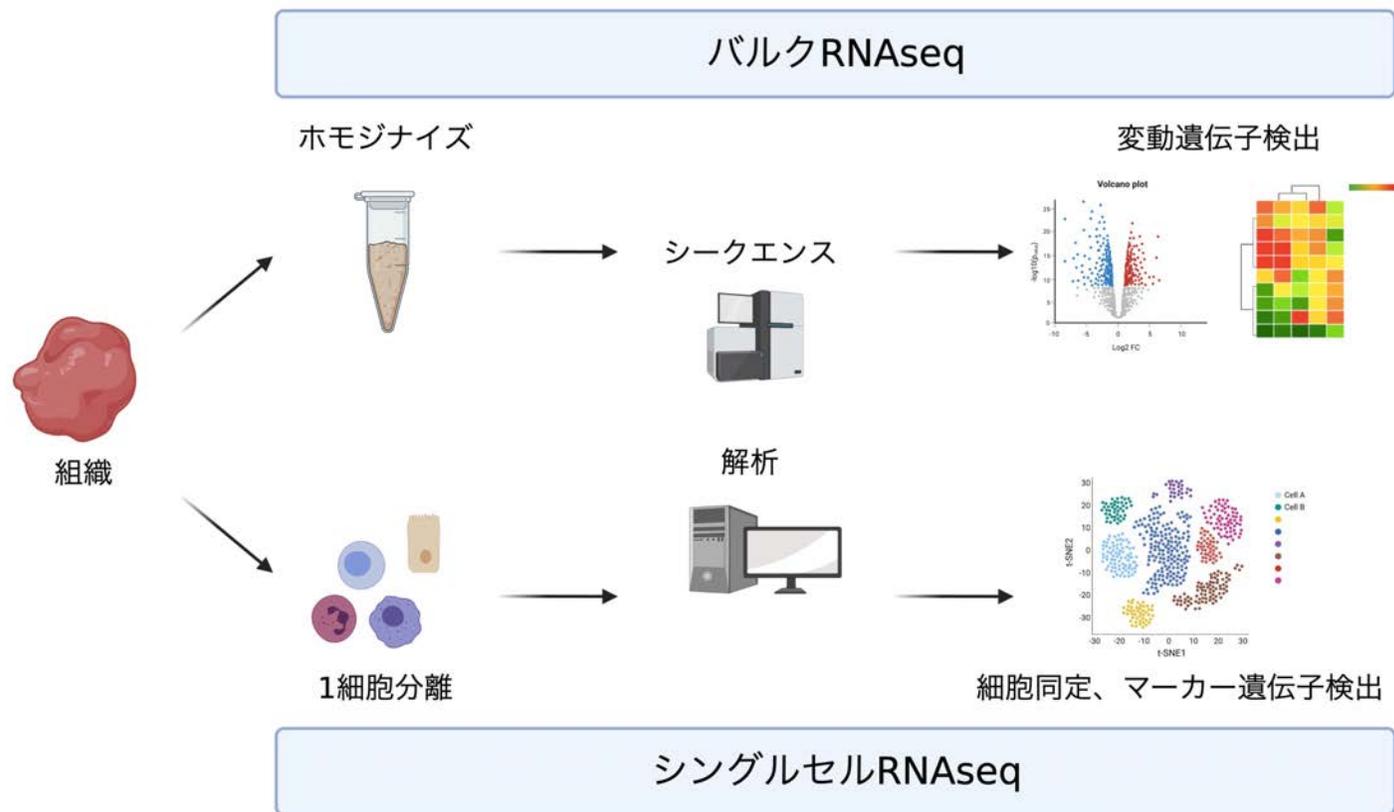
# Agenda

- **バルクRNAseqとは？**
- バルクRNAseq解析の例
- ブラウザを用いたバルクRNAseq解析
  - iDEP(本日のメイン)
  - BioJupies
  - RaNA-seq
- 実例紹介（胸腺腫合併重症筋無力症の解析）

**各サイトへのアクセスは、本会終了後にお願いいたします。**

# バルクRNAseqとは？ シングルセル解析とは？

多サンプルのプロファイリングが得意、解析も比較的簡単



バルクRNAseqはコスト・解析の容易さよりまだまだ強力なツール

**実験計画は慎重に！ (sc)RNAseqがすべてを解決してくれるわけではないが、使いこなせると非常に強力。データ特性をよく理解して研究に活かそう！**

# よくある風景：WetとDryの分離

研究室A



研究室B



Single-cellやったらすべてわかるはず  
高いからn=1でいいか

Single-cellのデータとったんだけど  
やっぱり解析してくれない？

有意差できるようにいい感じに検定し直し  
てくれない？

XXX細胞？XXX遺伝子？  
わからないけどこんな結果出ました  
解釈はそちらでお願いします

最新の機械学習手法でとりあえずツール作ってみた  
何に使うかはユーザーよろしく

# 良いサイエンスへの理想



こういう現象を解きたい、実験計画はこれでよい？

ここは検出力低そうだからこういうサンプリング、  
検体数はどうでしょう？  
こういう解析ができそうです。

釣れてきた遺伝子をこの実験に落とし込もうと思う

今日の話を通じてWet研究者がDryリテラシーを持つのも一つの解決策



GSEA



WGCNA



DESeq2



scVelo



Sckit-learn



Scanpy



Seurat

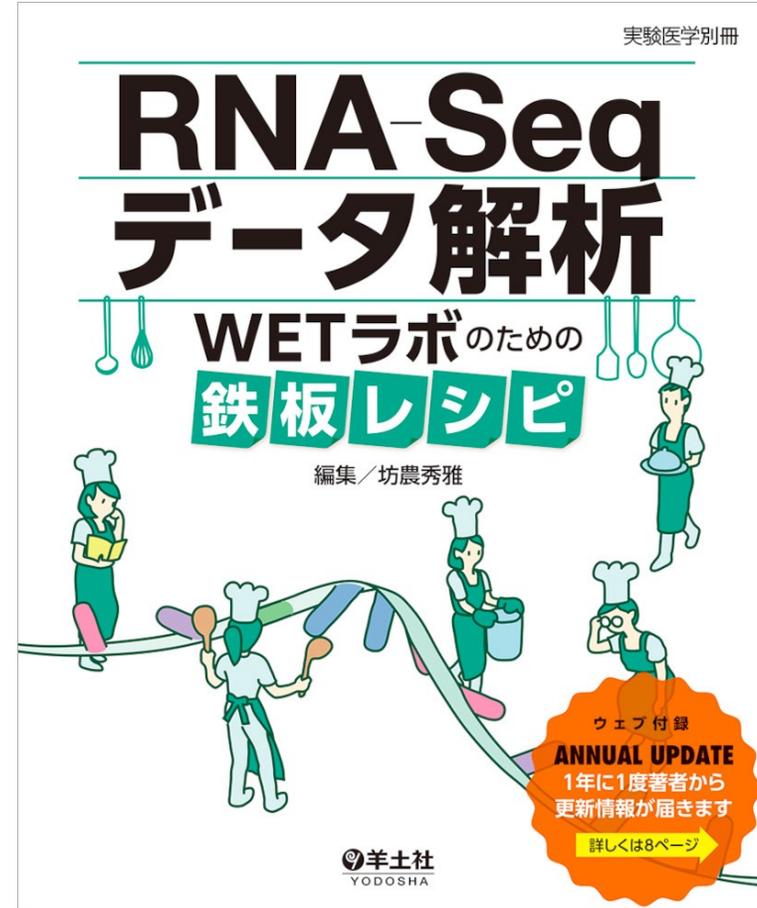


IGV



Pytorch

# 参考図書



本日は左の書籍をベースにお話します

# 関連する統合TV

TOGO TV

🏠 トップページ 📌 TogoTVについて 🎥 動画を探す 📄 イラストを探す 📁 講習会資料を探す 🗨️ お問い合わせ 🏠 マイページ 🔍 キーワードから動画を探す



## 見どころダイジェスト

- 00:36 1. アップロードするデータの説明
- 02:54 2. デモデータのアップロード
- 03:21 3. Pre-processの説明
- 04:42 4. Heatmapの説明
- 05:17 5. DEGの説明
- 05:58 6. Pathwayの説明
- 06:37 7. Genomeの説明
- 07:01 8. Biclusteringの説明

## この動画のタグ

- 発現解析
- 遺伝子
- NGS
- RNA-seq
- 可視化
- パスウェイ解析

## 動画ファイルのダウンロード

200118\_iDEP.mov

## DOI

<https://doi.org/10.7875/togotv.2020.018>

## 再利用時のライセンス

クリエイティブ・コモンズ CC-BY-4.0

この動画が再生されない場合は、YouTubeでご覧ください。 <https://youtu.be/Hl4zJAMi8YQ>

📅 2020.01.18 🕒 08:45 👤 作者: TogoTV 📝 編集: 小野 浩雅, Shuya Ikeda

再生リストに保存

## iDEPを使ってウェブブラウザ上でRNA-seqデータ解析を行う

iDEP (Integrated Differential Expression and Pathway analysis)は、ウェブブラウザ上でRNA-seqデータ解析を行うことができるウェブツールです(原著論文: "iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data"). RNA-seqやマイクロアレイ、ChIP-seq実験等で得られた遺伝子発現データ(リードカウントまたはFPKM)を入力として与えると、ヒートマップやPCA、発現差解析、パスウェイ解析、エンリッチメント解析、バークラスターリング法および共発現ネットワーク解析などの一連のデータ解析をインタラクティブに実行することができます。

- iDEP : <https://togotv.dbcls.jp/20200118.html>
- BioJupies : <https://togotv.dbcls.jp/20190730.html>
- RaNA-seq : <https://togotv.dbcls.jp/en/20210531.html>

# Agenda

- バルクRNAseqとは？
- **バルクRNAseq解析の実際**
- ブラウザを用いたバルクRNAseq解析
  - iDEP(本日のメイン)
  - BioJupies
  - RaNA-seq
- 実例紹介（胸腺腫合併重症筋無力症の解析）

各サイトへのアクセスは、本会終了後にお願いいたします。

# 自前データと公共データ



目的の遺伝子変動を調べるために、RNAseqを行うシーケンスコスト・3' kitの普及などにより安価にできるようになりつつある。



SRA



ERA

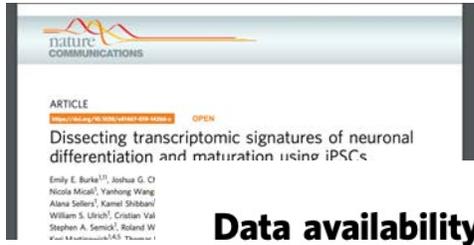


DRA

論文で使用されたシーケンスデータは、SRA/ERA/DRAなどのデータベースに登録され、他の研究者が利用できるようになる。

# 公共データへのアクセス

## 論文から辿る



### Data availability

Links for downloading all sequencing reads, including both the time-course data and all reprocessed public data, are available at <http://stemcell.umd.org/scb/>. Transcriptional data are deposited under accession code [PRJNA596331](https://www.ncbi.nlm.nih.gov/acc/record/PRJNA596331). Source data are available as a Source Data file.

### Code availability

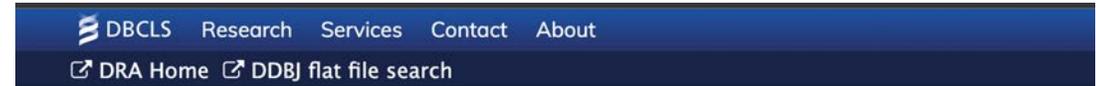
## NCBI SRA Run Selector

Sample	AGE	Bases	batch_number	Bytes		
<input type="checkbox"/> 1	SRR10738231	SAMN13618515	6	12.03 G	Batch7	4.28 Gb
<input type="checkbox"/> 2	SRR10738232	SAMN13618514	6	11.24 G	Batch7	3.97 Gb
<input type="checkbox"/> 3	SRR10738233	SAMN13618513	9	9.40 G	Batch4	3.74 Gb
<input type="checkbox"/> 4	SRR10738234	SAMN13618512	2	4.25 G	Batch4	1.73 Gb
<input type="checkbox"/> 5	SRR10738235	SAMN13618511	9	7.73 G	Batch4	3.16 Gb
<input type="checkbox"/> 6	SRR10738236	SAMN13618510	9	4.81 G	Batch4	2.15 Gb

Accession Numberからデータにアクセスできる。  
Run Selectorなどを使用すると便利に閲覧できる。

Run IDを用いてfasterq-dumpやikraからダウンロード、再解析を行うことができる。

## データベースからの検索



## DBCLS SRA > SRA

Keyword :

Accession :

<https://sra.dbcls.jp/>



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



登録データランキング

データリスト

使い方

API

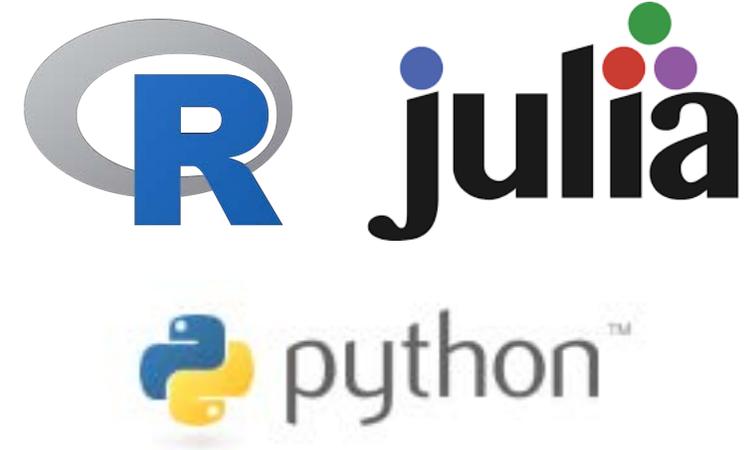
<http://aoe.dbcls.jp/>

# バイオインフォマティクス環境



```
yasumizuyoshiaki -- yyasumizu@x86_64-conda_cos6-linux-gnu:~/media32TB/bioinformatics/AJACS95 -- ssh UVI -- 118
.aliases/AJACS95
yyasumizu@x86_64-conda_cos6-linux-gnu [15時 06分 43秒] [~/media32TB/bioinformatics/AJACS95]
-> % cowsay AJACS
-----
< AJACS >
-----
  \   ^__^
   (oo)\_______
      (__)\       )\/\
         ||----w |
         ||     ||

yyasumizu@x86_64-conda_cos6-linux-gnu [16時 06分 58秒] [~/media32TB/bioinformatics/AJACS95]
-> %
```



UbuntuなどLinux OS

ターミナルを用いたCLIによる操作  
Shellの操作

プログラミング  
ライブラリの使用

サーバーやワークステーション内にこれらの環境を構築・管理する必要がある。



# 遺伝子発現量はようやく出来たが・・・

	SRR10738313	SRR10738318	SRR10738253	SRR10738254	
A1BG	69.3320313413642	25.3199528769756	141.133131572226		153.464342355173
A1CF	2.02494778903422	0	0.183603173980123	0	
A2M	0	6.97554762260232	513.88286861354	570.500486767718	
A2ML1	26.7325289046729	58.8921470633404	11.1761016694784		5.90924194893585
A3GALT2	0	0	10.5985107595544		
A4GALT	179.575161274335	98.9594279306123	59.4725598916669		104.442933609104
A4GNT	0	0	1.07927878038009	0	
AAAS	2213.34593710363	1451.044352623	1748.6158281363	1203.88492090296	
AACS	1294.71091797107	1042.0248759752	394.365087151363	305.660483357486	
AADAC	2.09932238285757	0	0	0	
AADACL2	0	0	0	0	
AADACL3	4.49089625608276	10.4786781173513	0	0	
AADACL4	1.4707018447045	0.916776830076207	0	0	
AADAT	283.513197040042	190.425302464233	479.740900972072		172.745593692753
AAGAB	2032.97049470599	1433.84901685345	2480.37818032573		1948.84902959379
AAK1	168.637767302238	212.199047091052	215.371028678045		233.980092222122
AAMDC	2587.1268484756	1292.24913201453	2360.26321921485	2078.9834905674	
AAMP	4960.63923854854	4461.07871447357	3905.84198428516	2169.07203929686	
AANAT	11.4239551008837	15.7800367229662	34.1834942489164	52.480764012642	
AAR2	1159.69536416959	758.32767960797	922.180095017134	608.891525283508	
AARD	70.9248107086759	52.9779890380827	42.0690065180639	44.0459377806264	
AARS1	5058.39709973226	4570.26148743001	3877.29446918419	2739.82040905284	
AARS2	593.004114864139	516.985627649763	491.940900096445	324.129035971875	
AARSD1	1681.65823120374	1388.47447747828	1856.25155560238	1237.39986115285	
AASDH	550.074758261778	304.44870529366	532.860077657461	453.859891386728	
AASDHPTT	1752.92247164307	1214.42874343383	4898.34817749582	2930.35816876203	
AASS	7287.75753528241	3502.80828668857	424.743602739513	454.861363745511	
AATF	2432.23478285391	1743.20500115542	1411.77973549871	819.275728358163	
AATK	86.0551215971106	25.4478311417159	411.339407092462	181.456085876676	
ABAT	184.524766445057	65.2521005757459	4493.66860945954	1563.97266904646	
ABCA1	701.909901277603	366.793565504989	819.07031720298	352.522534945071	
ABCA10	6.1400796363469	8.98097804322397	25.7341813130029	12.7337538107674	
ABCA12	0	0.881145048459206	2.64976096031239		
ABCA13	14.0699591689696	7.6345864677388	38.1151177984012	49.0493798561929	

DEG (Differentially expressed genes) は？

そもそもサンプルはきれいに  
分かれる？

交絡因子の影響は？

特徴的なパスウェイはどう  
やって求める？

発現量テーブルにしたあと、下流解析が必要。

# 標準的な下流解析の流れ (例)

```
1 #!/usr/bin/Rscript
2
3 library(tximport)
4 library(readr)
5 library(stringr)
6
7 # Rscript tximport.R.R gencode.vM19.metadata.MGI.gz Illumina_PE_SRR.csv output.
8
9 SRRs <- c("SRR10738313", "SRR10738318", "SRR10738253", "SRR10738254")
10 ids <- c("day2_1", "day2_2", "day77_1", "day77_2")
11 tx2knownGene <- read_delim("gencode.v37.metadata.HGNC.gz", "\t", col_names = c(
12
13 # print(paste(c("salmon_output_"), split_vec, c("/quant.sf"), sep=""))
14
15 # files <- paste(c("salmon_output_"), exp.table[,2], c("/quant.sf"), sep="")
16 files <- paste(c("salmon_output_"), SRRs, c("/quant.sf"), sep="")
17 names(files) <- ids
18
19 print(files)
20
21 # txi.salmon <- tximport(files, type = "salmon", tx2gene = tx2knownGene)
22 txi.salmon <- tximport(files, type = "salmon", tx2gene = tx2knownGene,
23 countsFromAbundance="scaledTPM")
24
25 write_table(txi.salmon$counts, file="output_salmon_scaledTPM.tsv")
```

Environment: Global Environment

Files: 2022\_shinkei\_handson

Name	Size	Modified
..		
.gitignore	40 B	Dec 30, 2021, 7:09 PM
2022_shinkei_handson.Rproj	205 B	Dec 30, 2021, 7:09 PM
design.csv	105 B	Dec 29, 2021, 2:43 PM
Dockerfile	595 B	Dec 29, 2021, 2:43 PM
output_salmon_scaledTPM_sample.tsv	1.2 MB	Dec 29, 2021, 2:44 PM
README.md	4 KB	Dec 29, 2021, 2:43 PM
scripts.sh	3 KB	Dec 29, 2021, 2:43 PM
tximport.R	970 B	Dec 29, 2021, 2:43 PM

Console: R 4.1.1 -- "Kick Things"  
Copyright (C) 2021 The R Foundation for Statistical Computing  
Platform: x86\_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.



Rなどのプログラミング言語 (+ライブラリ)を使用し、統計解析、可視化などを行う。

# Bioconductor



Search:

[Home](#)

[Install](#)

[Help](#)

[Developers](#)

[About](#)

## About *Bioconductor*

The mission of the *Bioconductor* project is to develop, support, and disseminate free open source software that facilitates rigorous and reproducible analysis of data from current and emerging biological assays. We are dedicated to building a diverse, collaborative, and welcoming community of developers and data scientists.

*Bioconductor* uses the R statistical programming language, and is open

### Install »

- Discover [2083 software packages](#) available in *Bioconductor* release 3.14.

Get started with *Bioconductor*

- [Install \*Bioconductor\*](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

### Learn »

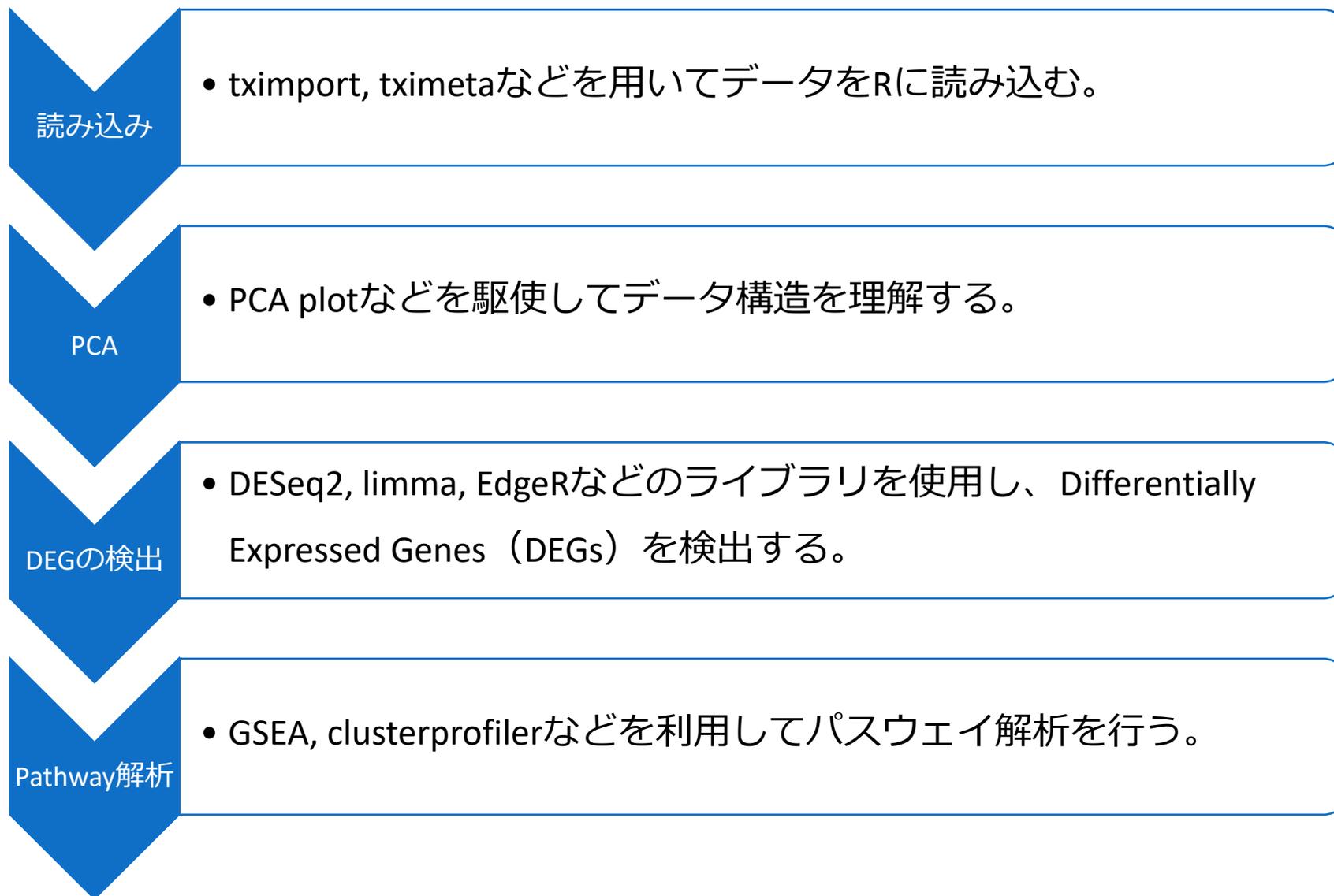
Master *Bioconductor* tools

- [Courses](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)

<http://bioconductor.org/>

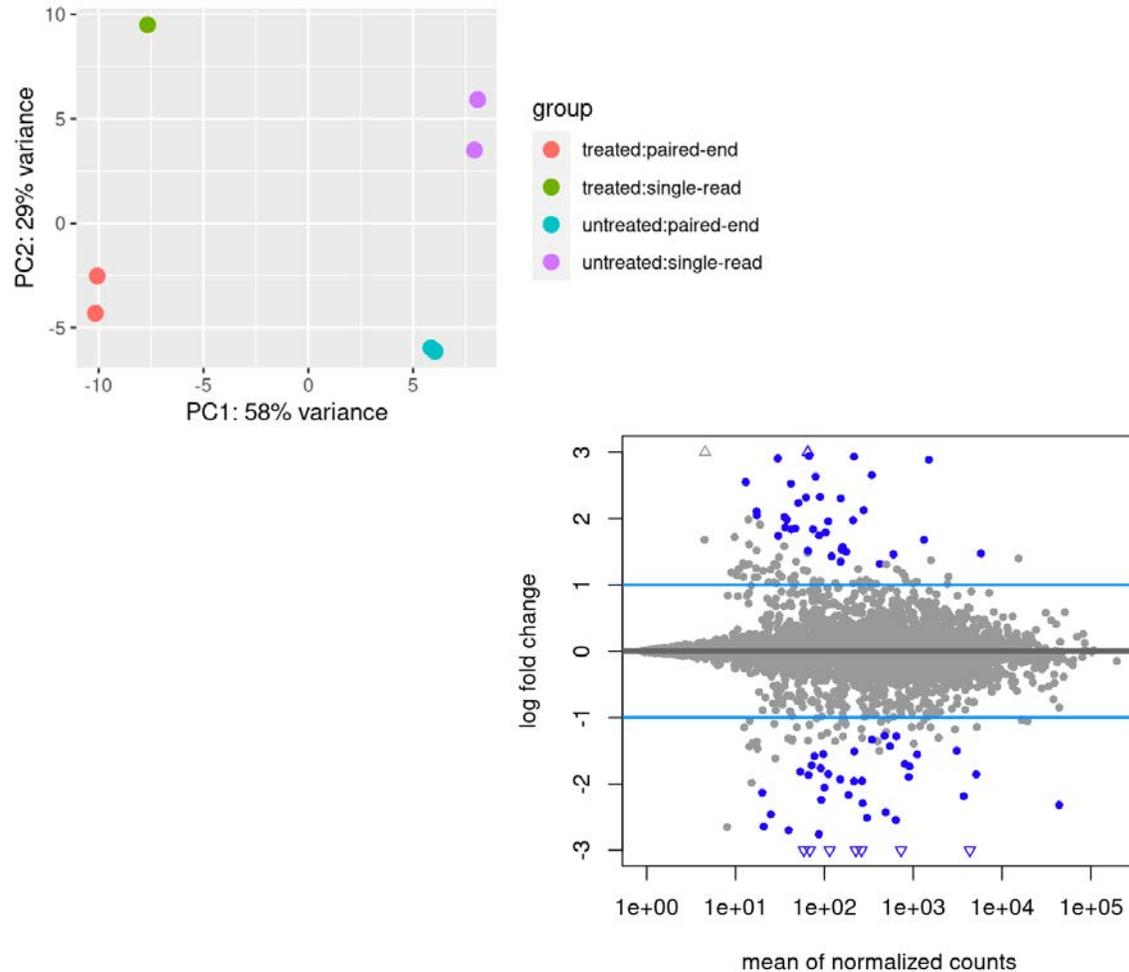
Bioinformatics用のRパッケージは、Bioconductorプロジェクトによりまとめて管理されている。

# 標準的な下流解析の流れ (例)



(加えて、パッケージのインストールなど、環境構築はもちろん必要)

# DESeq2を用いたDEG検出



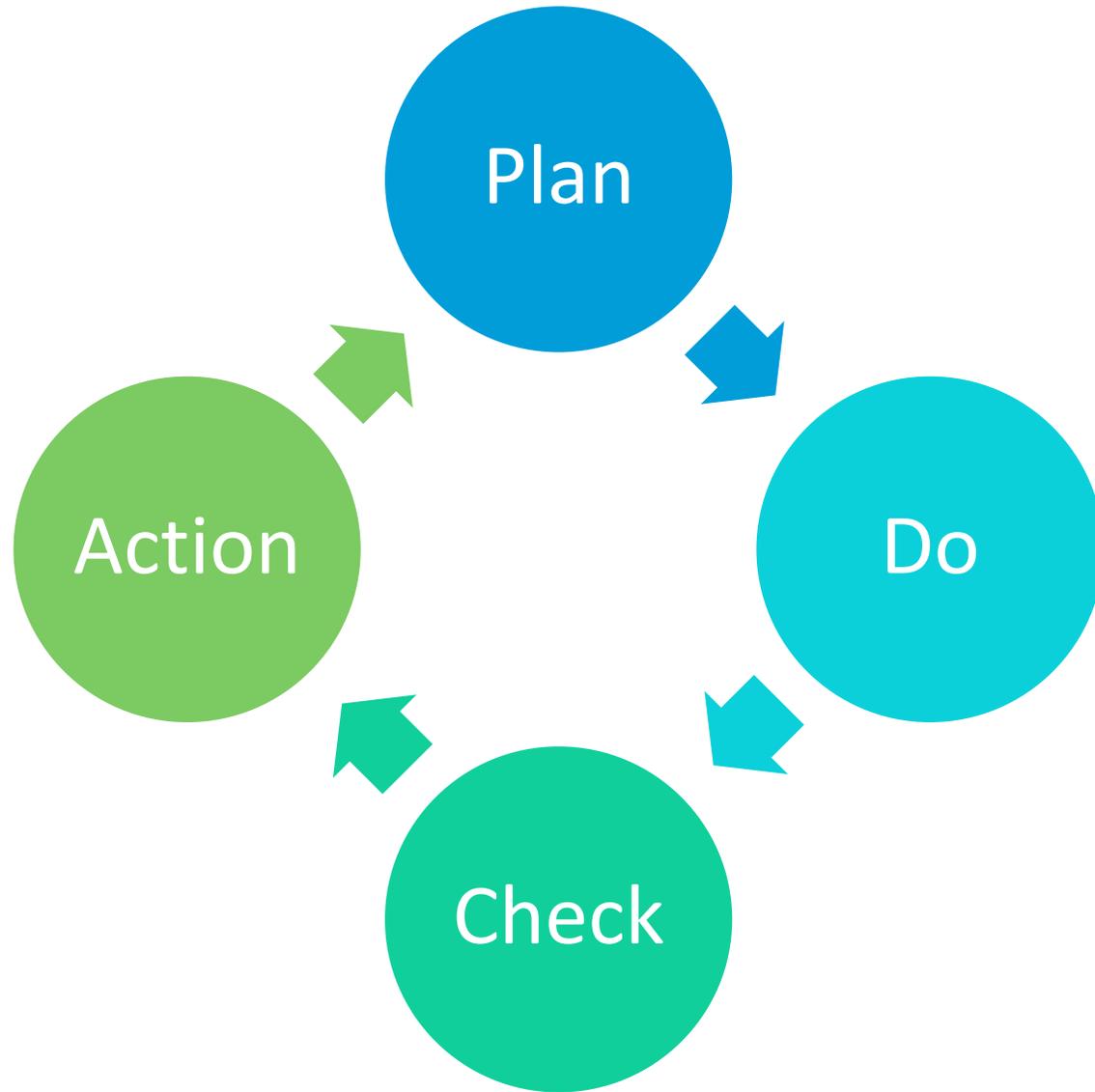
- DESeq2は遺伝子ごとにグループ間の差の統計量を計算する。
- その際、サイズファクター(サンプルごとのリード数)や多重検定補正などを考慮し、正確な検定を手軽に行ってくれる。

Love MI, 2014, *Genome Biology*.

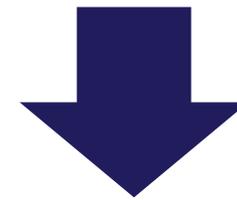
詳しくは、vignettes(**公式チュートリアル**)を参照。

<https://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html>

# 通常のRNA-seq解析は定型解析が多い割に時間がかかる



自動化・省力化による  
PDCAサイクルの高速化



Webツールを用いた解析

# Agenda

- バルクRNAseqとは？
- バルクRNAseq解析の実際
- **ブラウザを用いたバルクRNAseq解析**
  - **iDEP(本日のメイン)**
  - BioJupies
  - RaNA-seq
- 実例紹介（胸腺腫合併重症筋無力症の解析）

各サイトへのアクセスは、本会終了後にお願いいたします。

# iDEP を用いた下流解析

## 論文執筆用、再現性担保のためにバージョンを記録しておく

iDEP.94

Load Data Pre-Process Heatmap k-Means PCA DEG1 DEG2 Pathway Genome Bicluster Network R

Click here to load demo data  
and just click the tabs for some magic!

1. Optional: Select or search for your species.  
Best matching species [dropdown] Info

2. Choose data type  
 Read counts data (recommended)  
 Normalized expression values (RNA-seq FPKM, microarray, etc.)  
 Fold-changes and corrected P values from CuffDiff or any other program

3. Upload expression data (CSV or text)  
Browse... No file selected

Analyze public RNA-seq datasets for 9 species  
Optional: Upload an experiment design file (CSV or text)  
Browse... No file selected

Example gene IDs  
Try ShinyGO for GO enrichment analysis

Ready to load data files.

Nov. 15, 2021: iDEP v0.95 released in testing mode. It includes Ensembl database update, new species fr species) update from v11 to 11.5.

10/26/2021: The Genome view is now much improved! Automatically detects chromosomal regions enrich iDEP v.0.94 based on Ensembl Release 104 and STRING-db V11. 9/3/2021

10/15/21: For GO enrichment analysis, we now recommend using background genes, instead of all genes that all genes passed the filter in Pre-Process tab are used as a customized background.

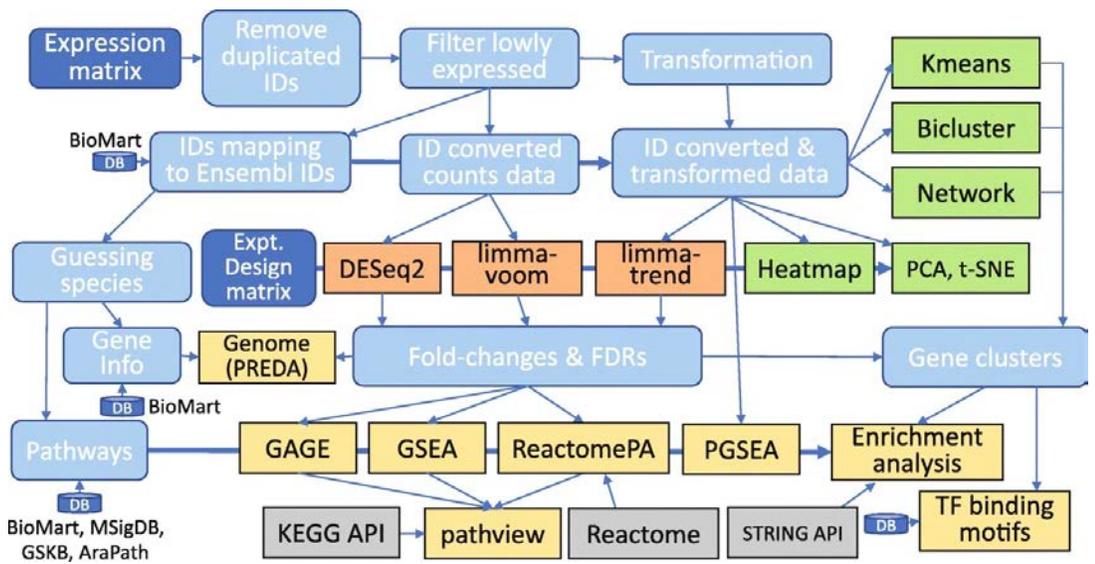
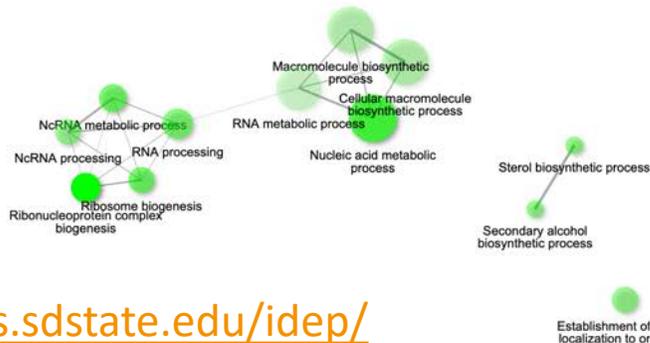
We updated instruction for local installation here. The most recent database file is now publically available

Check out the 50,000+ datasets of uniformly processed public RNA-seq data here!

Email Jenny for questions. Dr. Ge is notoristly slow in responding to emails.

iDEP has not been thoroughly tested. Please let us know if you find any issue/bug.

10/18/20: Interactive network enables users to easily visualize the relatedness of pathways, similar to Enri you can generate and export interactive networks like this one below. You can move the nodes by draggin an empty point and drag.



Steven Xijin Ge et al., 2018, BMC Bioinformatics

<http://bioinformatics.sdstate.edu/idep/>

iDEPには基本から応用まで、大体の解析が揃う。

ドキュメント: <https://idepsite.wordpress.com/>

アクセス集中を避けるため、本セッションは会が終了した後お手元でお試しく下さい。

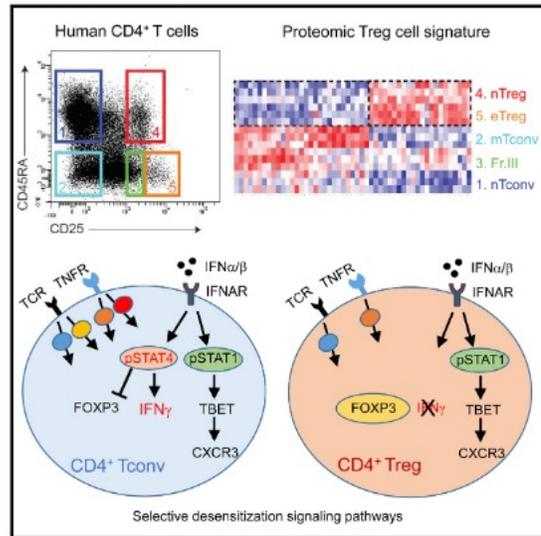
# Human PBMCから取得したCD4+T細胞のデータを用いる

Resource

## Immunity

### Proteomic Analyses of Human Regulatory T Cells Reveal Adaptations in Signaling Pathways that Protect Cellular Identity

#### Graphical Abstract



#### Authors

Eloy Cuadrado,  
Maartje van den Biggelaar,  
Sander de Kivit, ..., Rene A.W. van Lier,  
Jannie Borst, Derk Amsen

#### Correspondence

eloycua@gmail.com (E.C.),  
d.amsen@sanquin.nl (D.A.)

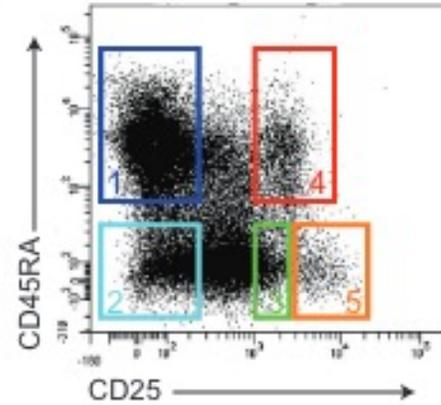
#### In Brief

Using high-resolution mass spectrometry and transcriptomics, Cuadrado et al. provide a molecular characterization of regulatory and conventional CD4<sup>+</sup> T cell subsets, yielding markers to distinguish cells with different properties and insights into mechanisms that prevent regulatory T cells from exhibiting undesirable functional activities of the related but functionally antithetical conventional T cells.

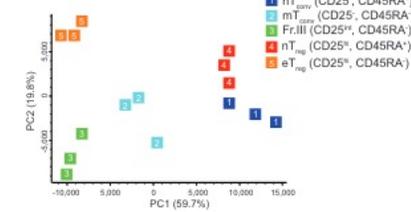
#### DATA AND SOFTWARE AVAILABILITY

The mass spectrometry data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE (Vizcaino et al., 2016) repository with the dataset identifier PDX007745, PDX007744, and PDX005477. RNASeq data can be accessed with accession number GSE90600.

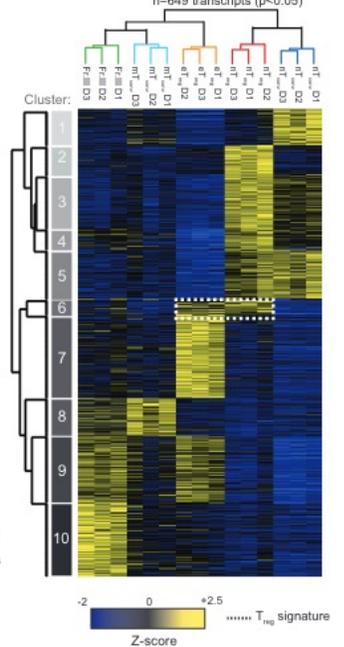
A



A



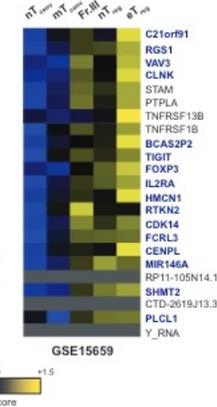
B



C



D



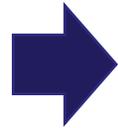
<https://doi.org/10.1016/j.immuni.2018.04.008>

# ikra v2.0.1 -RNAseq pipeline centered on Salmon-

Ikraを使用すると、簡単なcsvファイルにRun IDやfastq fileをまとめるだけで、今回の実習と同じ発現量テーブルを取得できる。

input.csv

```
name, SRR, Layout  
PID_1, SRR7501484, PE  
PID_2, SRR7501485, PE  
PID_3, SRR7501486, PE  
WT_1, SRR7501481, PE  
WT_2, SRR7501483, PE  
WT_3, SRR7501482, PE
```



ikra.sh input.csv human

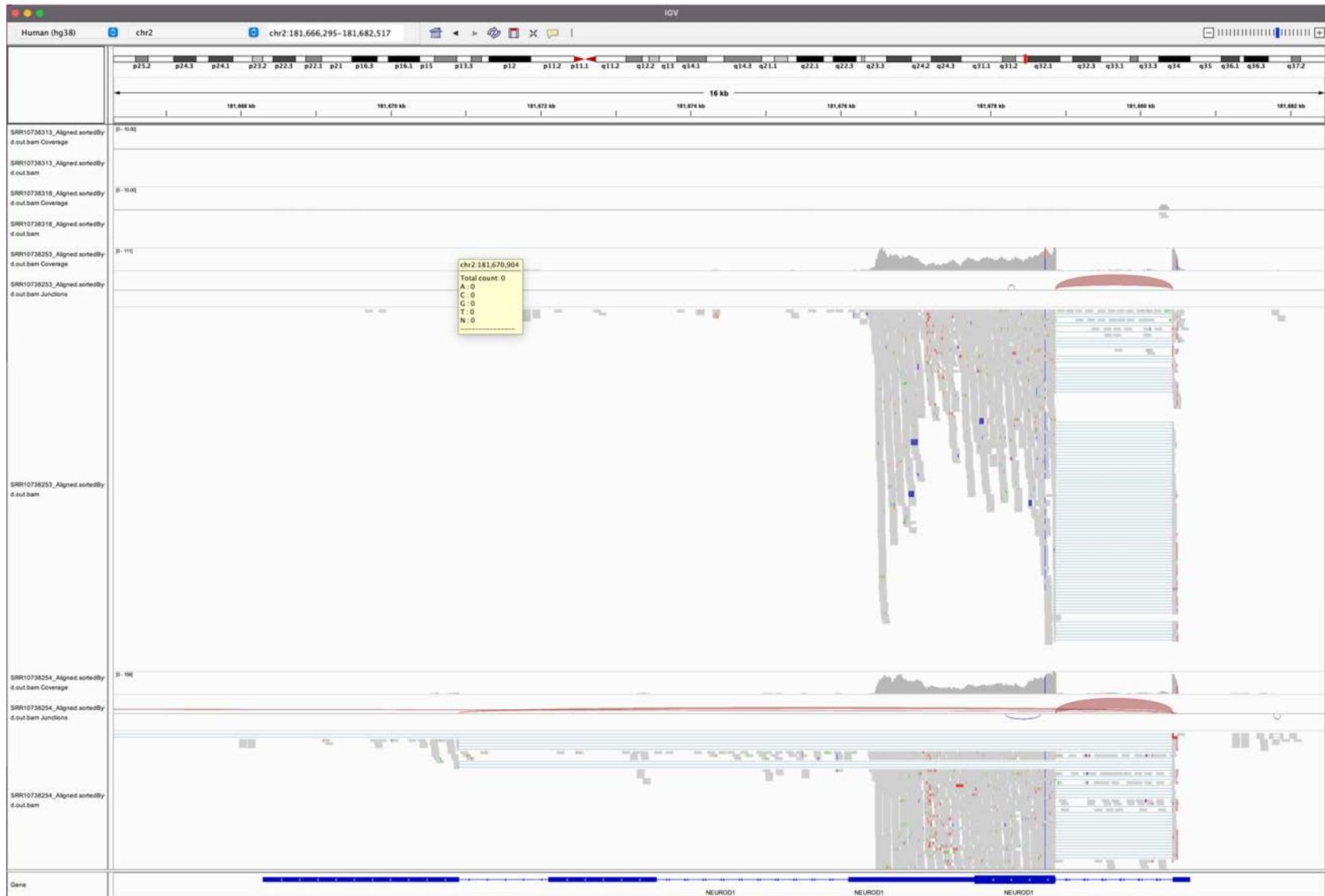
	Treg_LN_1	Treg_LN_2
0610005C13Rik	0	0
0610006L08Rik	0	1
0610009B22Rik	4	10
...		

<https://github.com/yyoshiaki/ikra>

Hiraoka Yu, Yamada Kohki, Ryuichiro Yamsasaki, YusukeKawasaki, Kitabatake Ryoko, Matsumoto Yasunari, Ishikawa Kaito, Umezu Yuto, Hirose Haruka, & Yoshiaki Yasumizu. (2021). yyoshiaki/ikra: ikra v2.0.1 (v2.0.1). Zenodo. <https://doi.org/10.5281/zenodo.5541399>

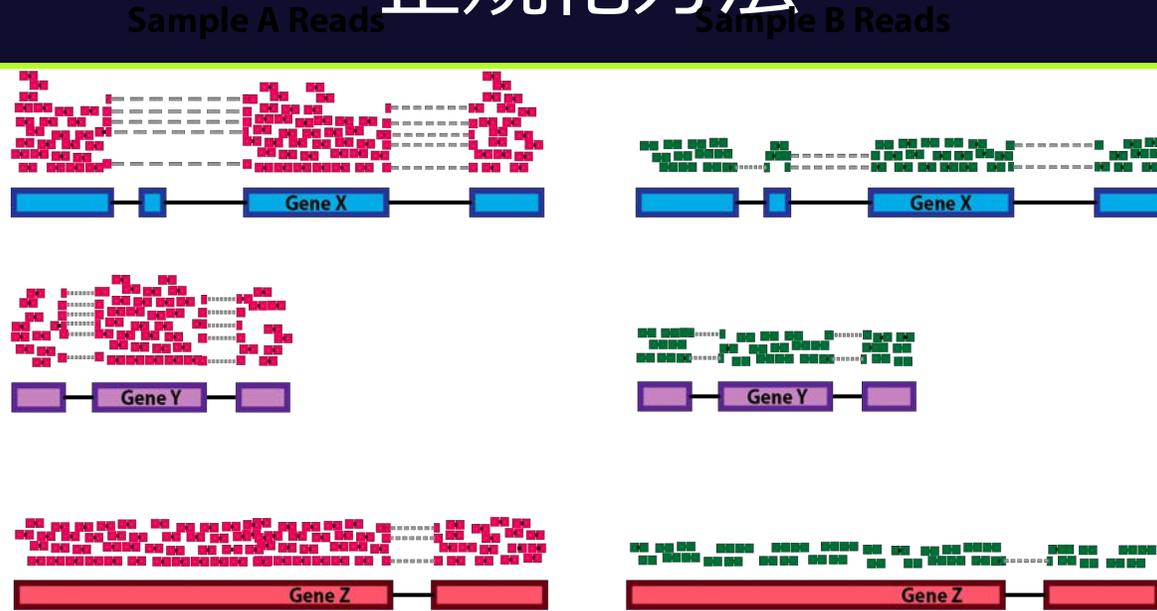


# The Integrative Genomics Viewer (IGV)



STARでマッピングを行い、IGVで確認することも重要なQC

# 正規化方法



[https://hbctraining.github.io/DGE\\_workshop/lessons/02\\_DGE\\_count\\_normalization.html](https://hbctraining.github.io/DGE_workshop/lessons/02_DGE_count_normalization.html)

RNA-seqの遺伝子発現量は生のカウントデータだと遺伝子長やサンプルごとのリード量によってばらついてしまうため、以下のような補正を行う。

- RPKM/FPKM (reads/fragments per kilobase of transcript per million reads/fragments) : **古い!**
- **TPM (Transcripts per kilobase million) : サンプル間比較に使用される**

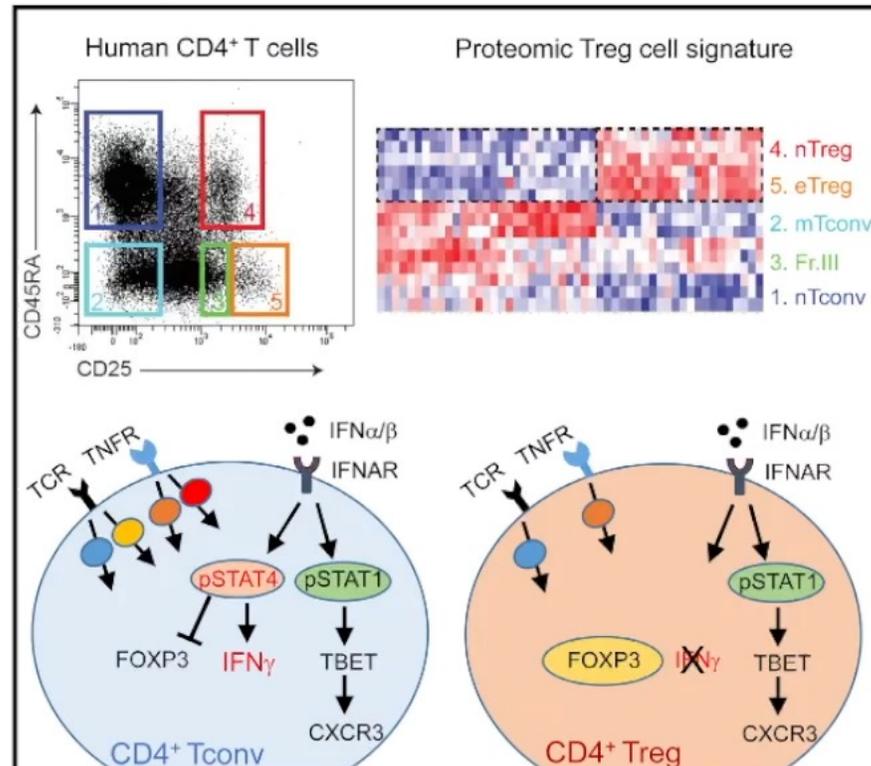
DESeq2は(負の二項分布を仮定するため)、カウントデータに相当するデータを使う必要がある。前項で用意したscaledTPM (TPMをリード量で再度補正)はカウントデータより正確なDEG検出が行えるとされている。(Soneson et al., 2015.)

Resource

# Immunity

## Proteomic Analyses of Human Regulatory T Cells Reveal Adaptations in Signaling Pathways that Protect Cellular Identity

### Graphical Abstract



### Authors

Eloy Cuadrado,  
Maartje van den Biggelaar,  
Sander de Kivit, ..., Rene A.W. van Lier,  
Jannie Borst, Derk Amsen

### Correspondence

eloycua@gmail.com (E.C.),  
d.amsen@sanquin.nl (D.A.)

### In Brief

Using high-resolution mass spectrometry and transcriptomics, Cuadrado et al. provide a molecular characterization of regulatory and conventional CD4<sup>+</sup> T cell subsets, yielding markers to distinguish cells with different properties and insights into mechanisms that prevent regulatory T cells from exhibiting undesirable functional activities of the related but

yyasumizu@x86\_64-conda\_cos6-linux-gnu [16時 20分 42秒] [~/media32TB/bioinformatics/AJACS95]

-> % █

# データのロード

iDEP.96 Load Data Pre-Process Heatmap k-Means PCA DEG1

[Click here to load demo data](#)

and just click the tabs for some magic!

Reset

1. Optional: Select or search for your species.

Human

2. Choose data type

Read counts data (recommended)

Normalized expression values (RNA-seq FPKM, microarray, etc.)

Fold-changes and corrected P values from CuffDiff or any other program

3. Upload expression data (CSV or text)

Browse... output.rmTCR.tsv

Upload complete

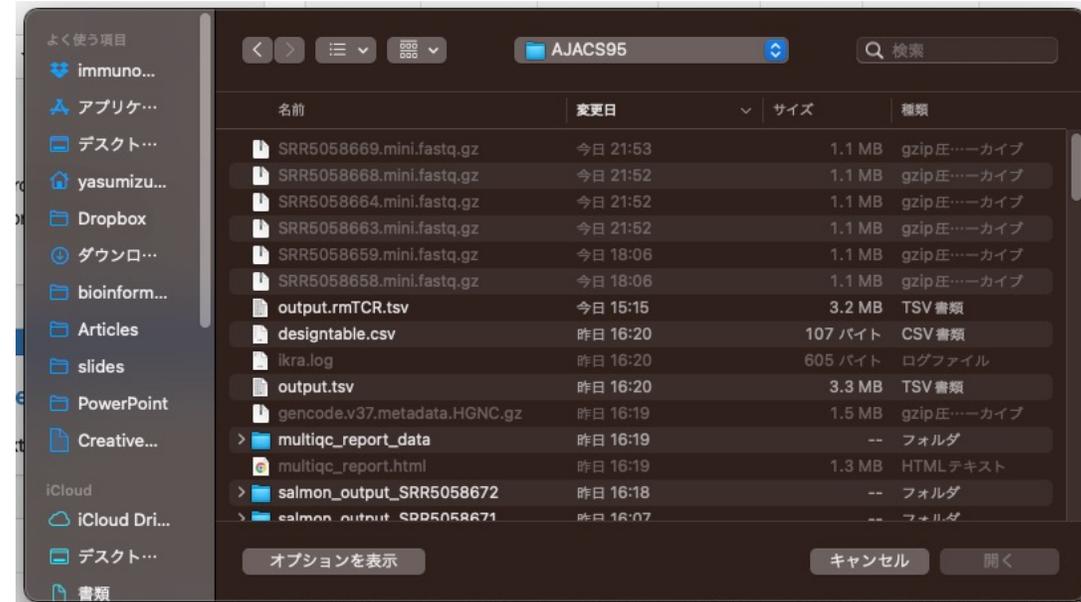
Analyze public RNA-seq datasets for 9 species

Optional: Upload an experiment design file(CSV or text)

Browse... No file selected

Matched Species (%genes)

Human (100) **Used in mapping** To change, select from above and resubmit query.
Homo sapiens STRINGdb (97)
Chimpanzee (90)
Mouse (87)
Mus musculus STRINGdb (86)



Browse > ファイル(今回はoutput.rmTCR.tsv)を選択。  
data typeはscaledTPMの場合Read countを選択。  
uploadが完了すればPre-Processのタブへ。

```
nTconv_1    mTconv_1    nTreg_1    eTreg_1    nTconv_2
A1BG      30.8536431683328    26.5128520434148    23.12628328664
A1CF      0.21936245696525    0.283808877938921    0.237596635009
```

Inputのサンプル名はアンダーバー区切りで前がグループ、後ろがreplicate number。

# Pre-Process

iDEP.94

Load Data

Pre-Process

Heatmap

基本はデフォルトでOK.

Keep genes with minimal counts per million (CPM) in at least n libraries:

Min. CPM:  n libraries:

Transform counts data for clustering & PCA.

VST: variance stabilizing transform  
 rlog: regularized log (slow)  
 EdgeR: log2(CPM+c)

Pseudo count c:

Missing values imputation:

Do not convert gene IDs to Ensembl.

[?](#)

## 低発現遺伝子の足切り

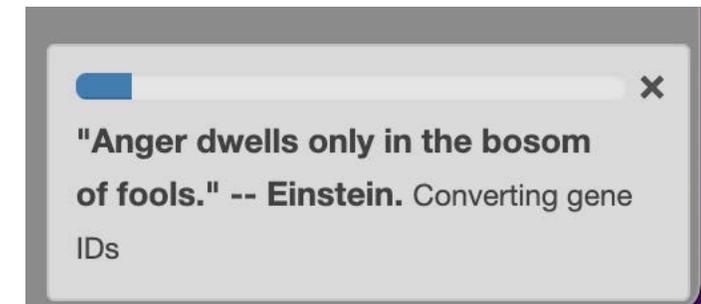
Transform方法の選択。VST (低発現遺伝子の分散を抑える。機械学習やWGCNAにおすすめ。)

なお、DEGの計算はraw countを使用している。

## 欠損値の補完方法

## 興味のある遺伝子の可視化

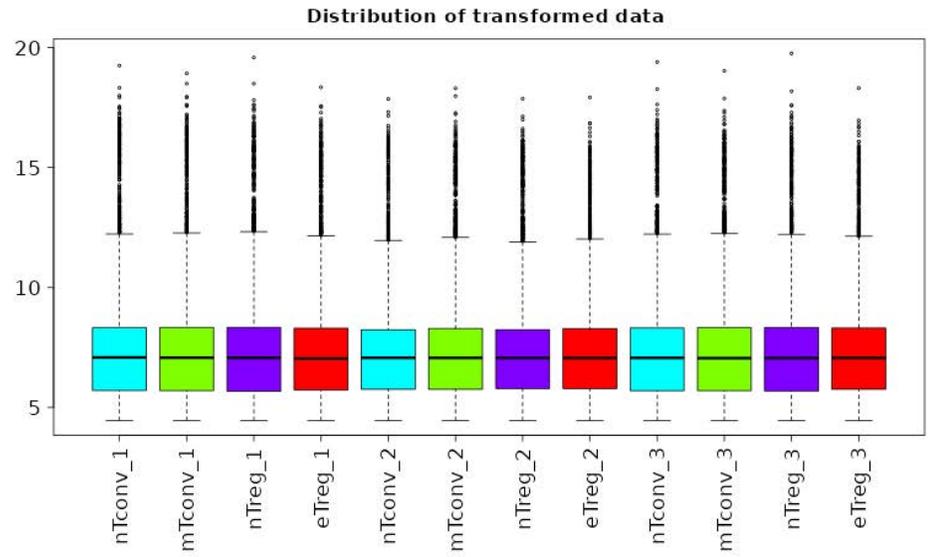
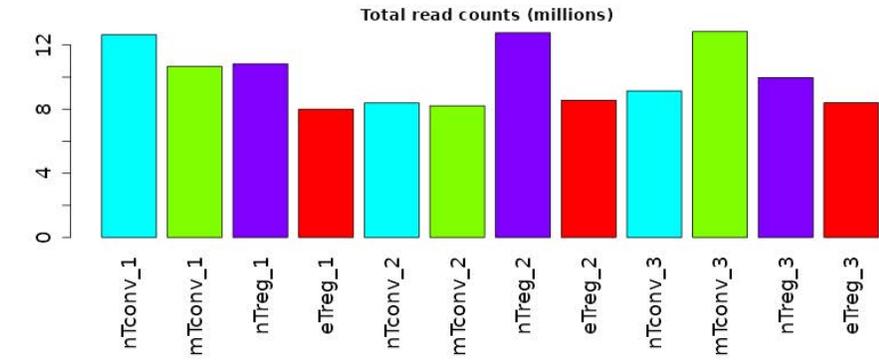
transformされた遺伝子発現データのダウンロード



右下にprogress barが出ている時は待ち時間。

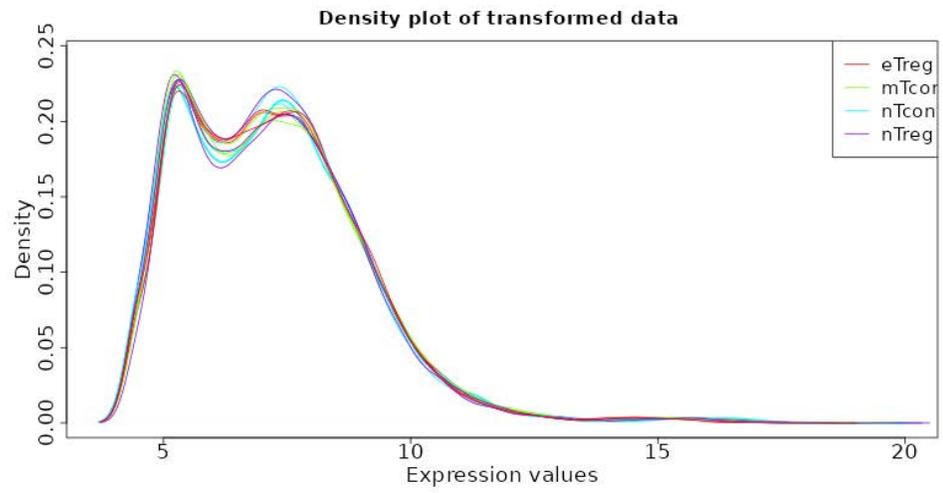
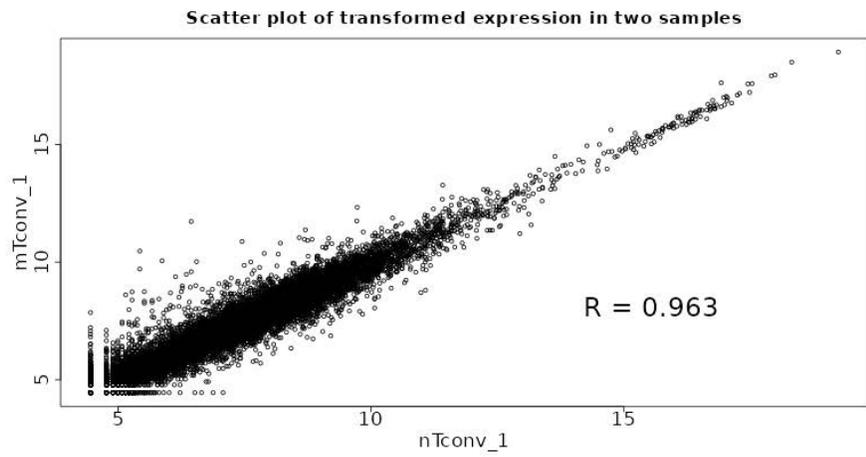
# Pre-Process

Aspect ratios of figures can be adjusted by changing the width of browser window.



Select a sample for x-axis:

Select a sample for y-axis:



サンプル内に大きな外れ値がないかをよく確認。

# 個別遺伝子発現の確認

IDEP.96

Load Data

Pre-Process

Heatmap

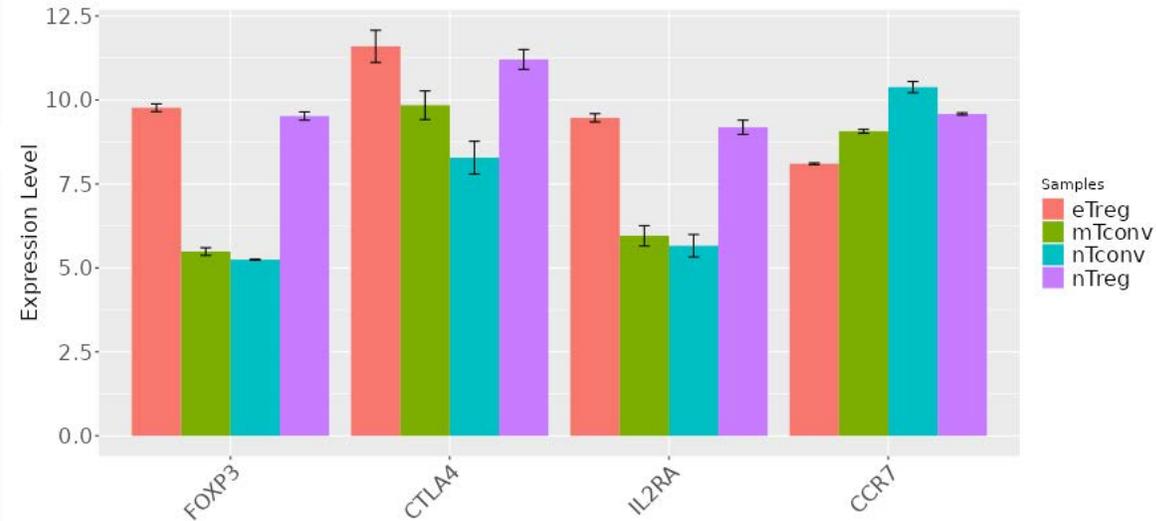
k

Search for genes

Enter full or partial gene ID, or list of genes separated by semicolon:

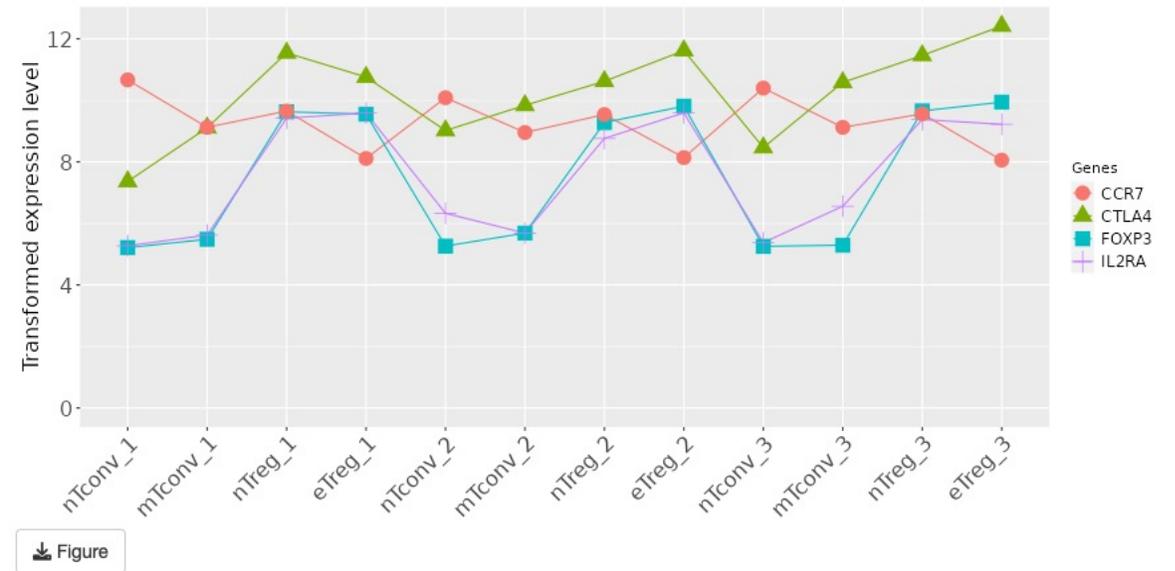
FOXP3,CTLA4,IL2RA,CCR7

Show individual samples



Use standard deviation instead of standard error

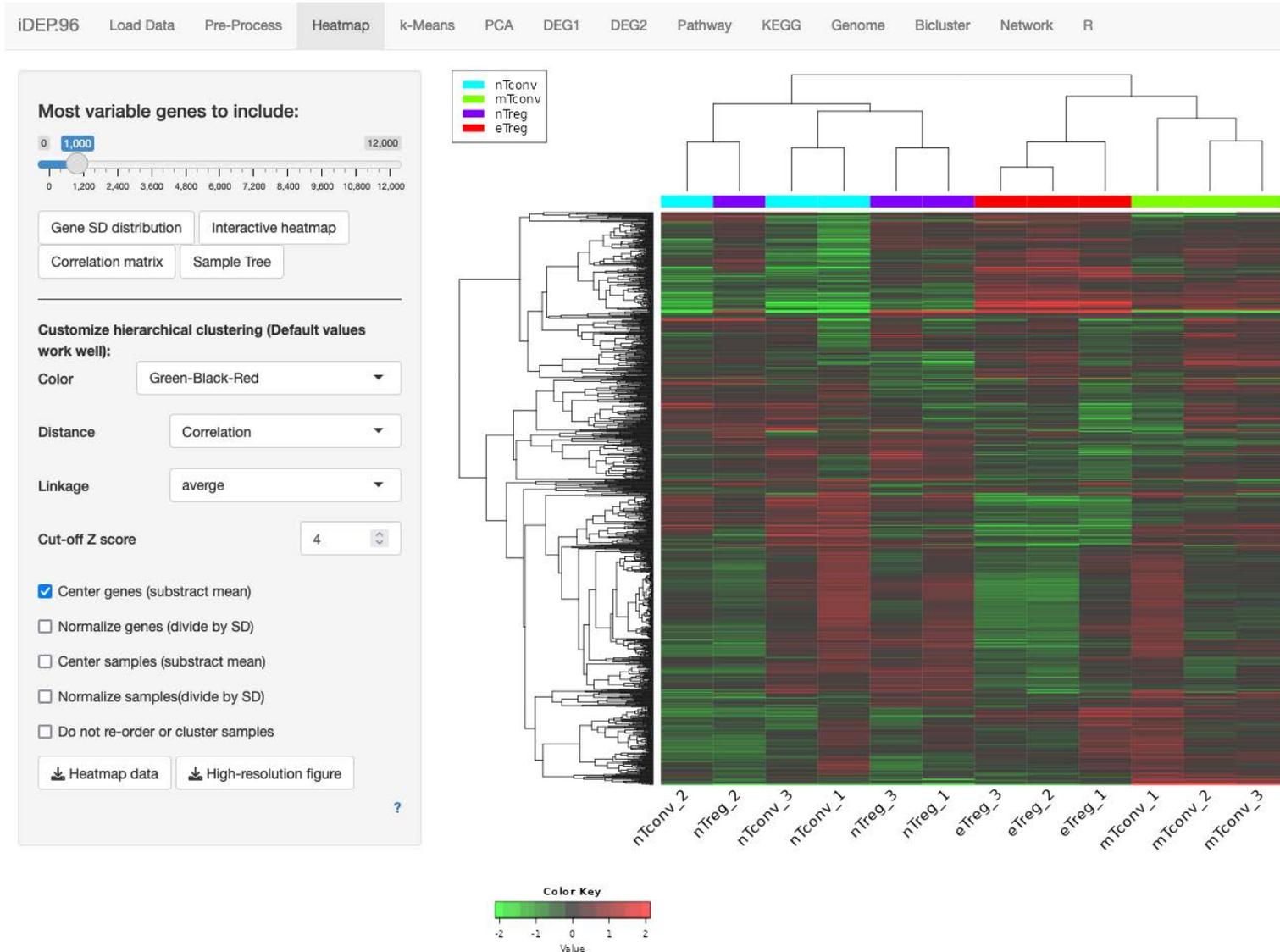
Figure



Figure

複数遺伝子も可能。画像はダウンロードボタンから適宜ダウンロード可能。

# ヒートマップによる可視化



Most variable genesについて可視化。

# K-Means法によるクラスタリング

iDEP.94

Load Data

Pre-Process

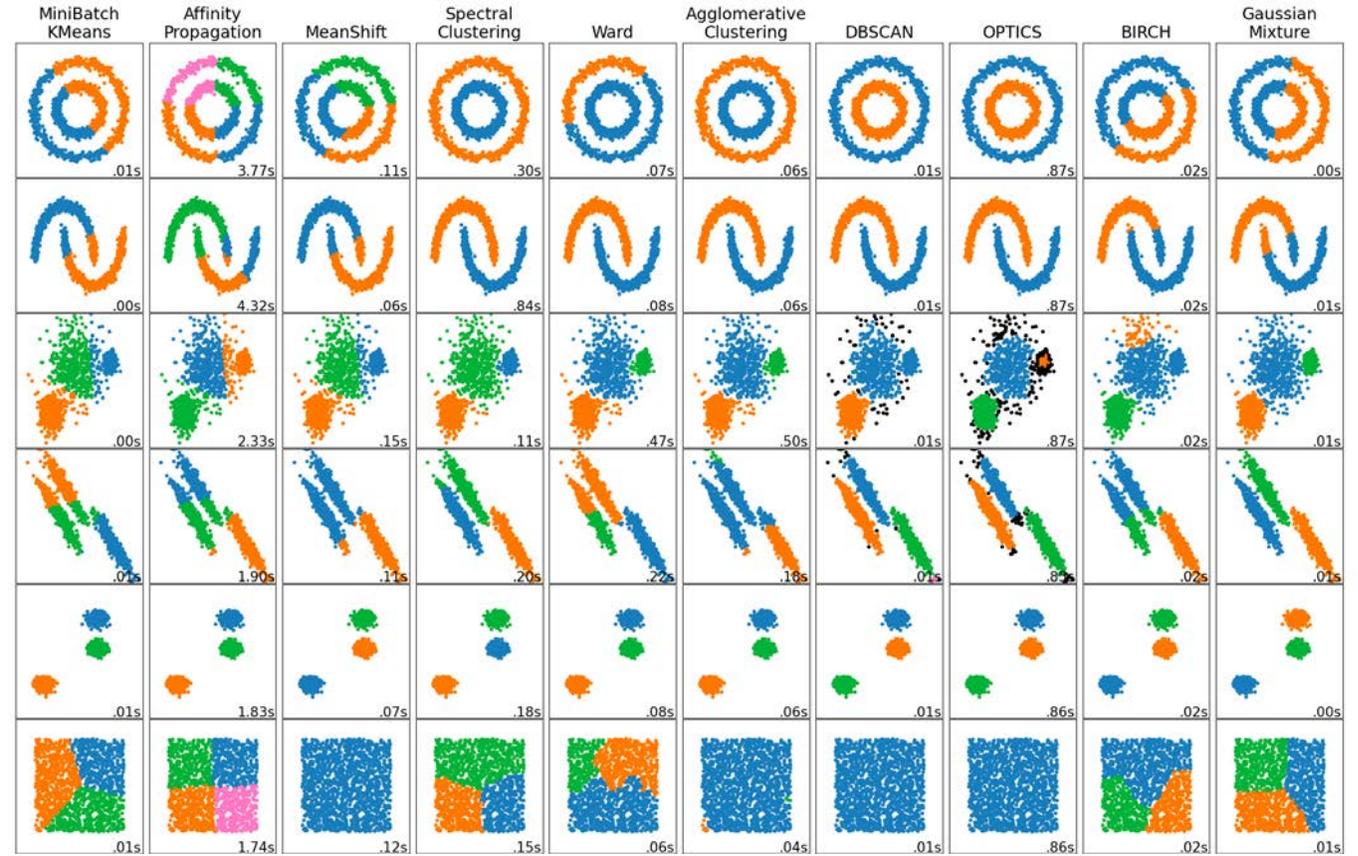
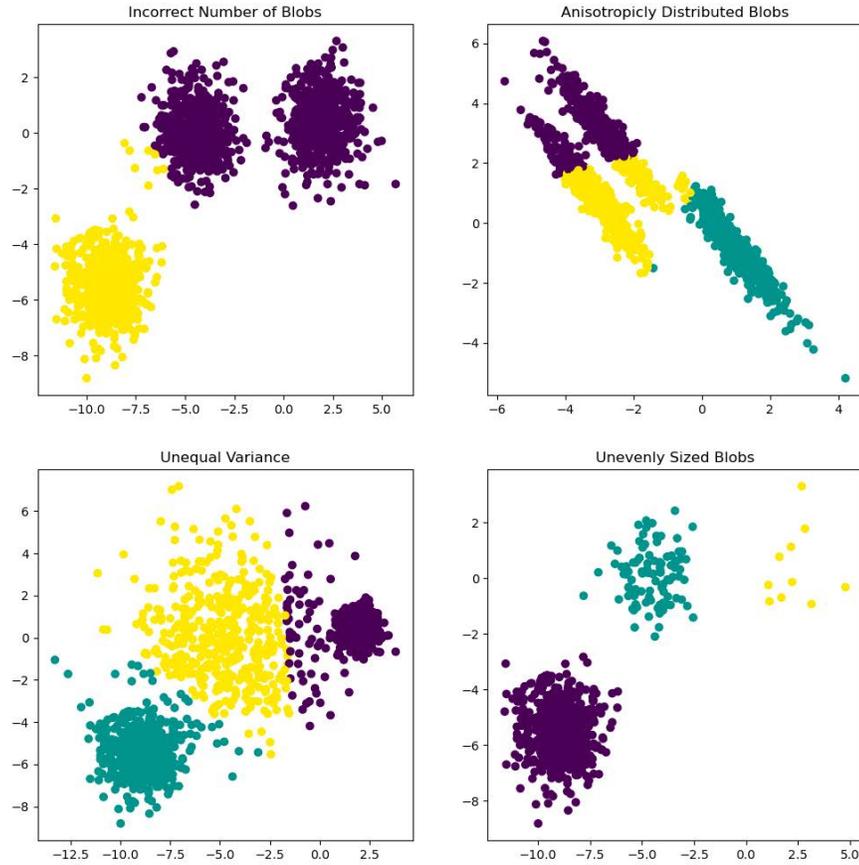
Heatmap

**k-Means**

PCA

DEG1

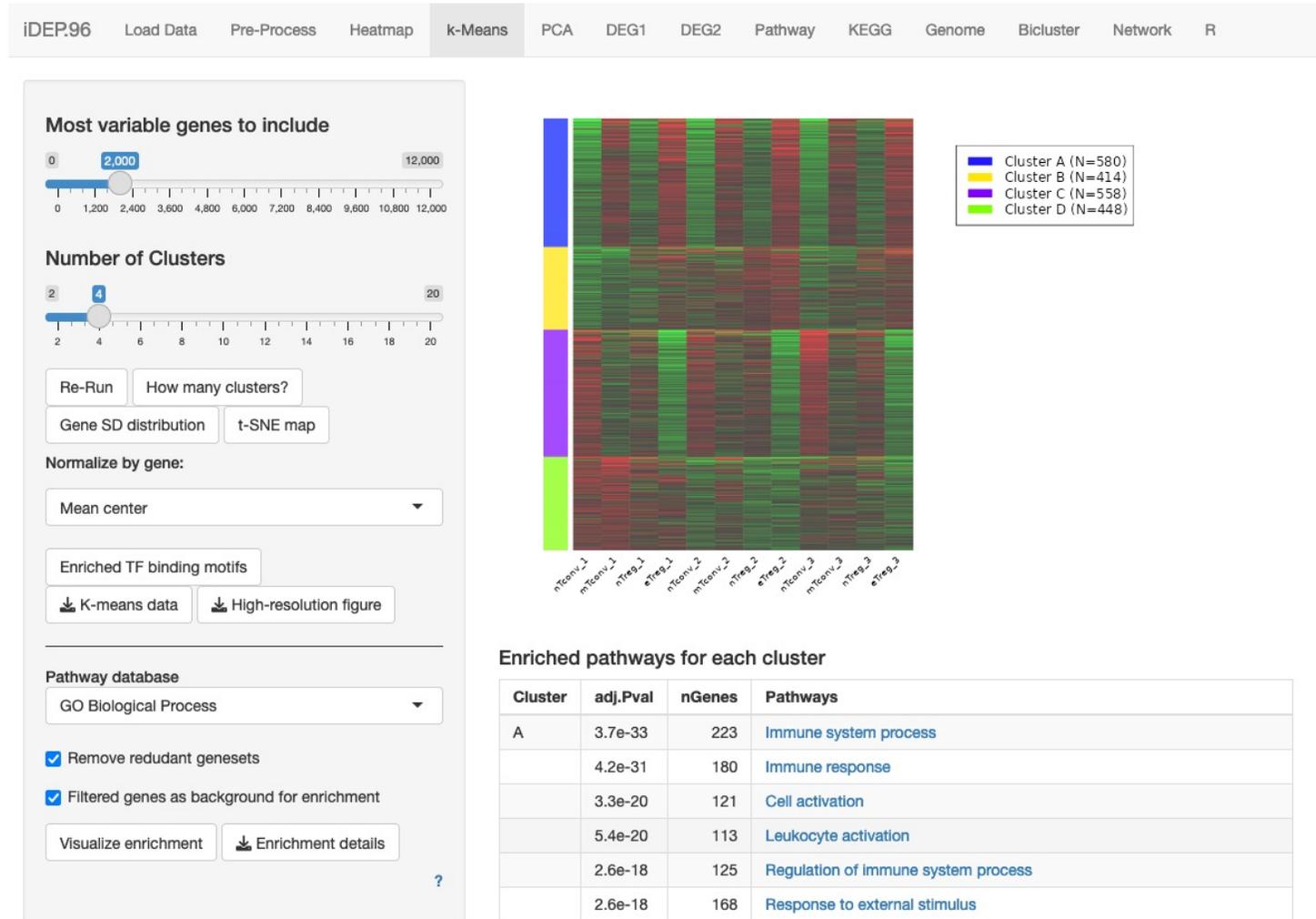
DEG2



<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

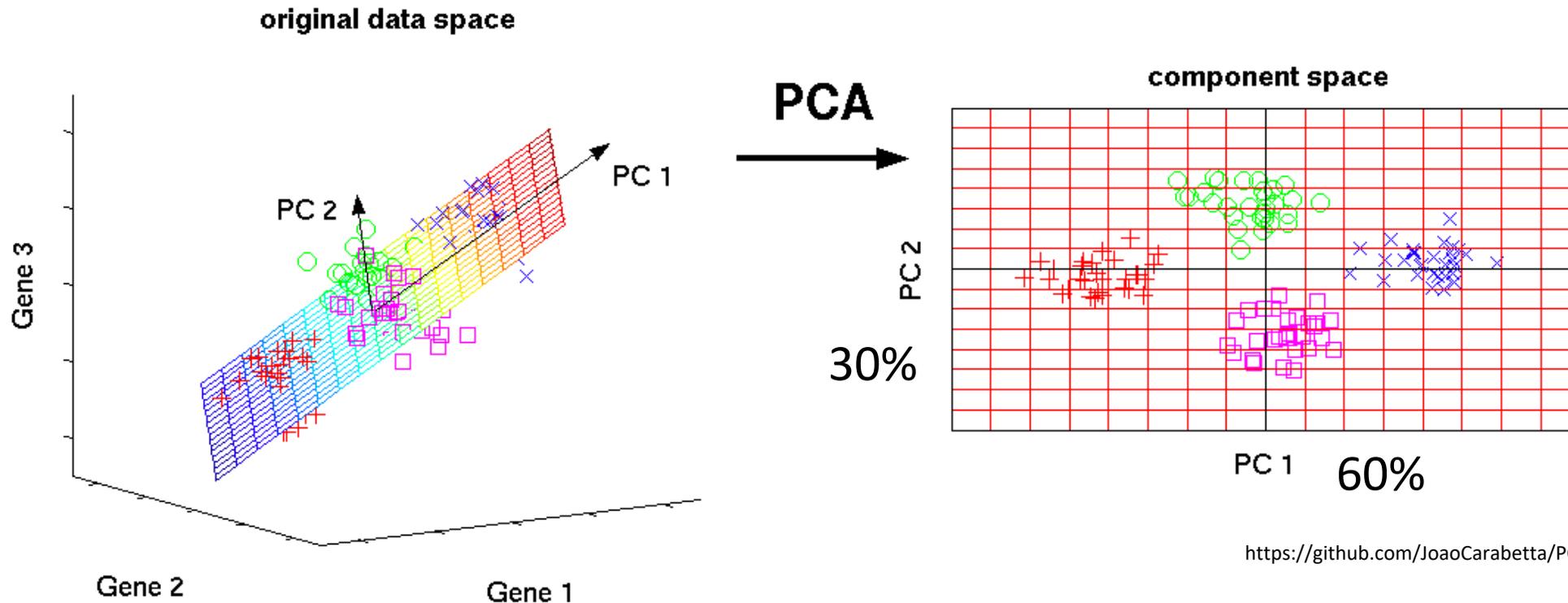
KmeansはK個のクラスターにデータ点を分割する古典的なアルゴリズム

# K-Means法によるクラスタリング



クラスターごとに集積したGOパスウェイ

# PCA (主成分分析)



<https://github.com/JoaoCarabetta/PCA-Explorer>

PCAとは、データ点のばらつきを最も表現する主成分（PC1）を抽出する。続いて、PC1に直交しつつデータ点のばらつきを最も表現する主成分PC2、そしてPC3と主成分を取っていく。

PC1, PC2の横の数字は**寄与率**といい、データ点全体の分散における各PC成分が表現している分散の割合。

各PCにおける各変数（遺伝子）の相関を**負荷量**という。

# PCA (主成分分析)

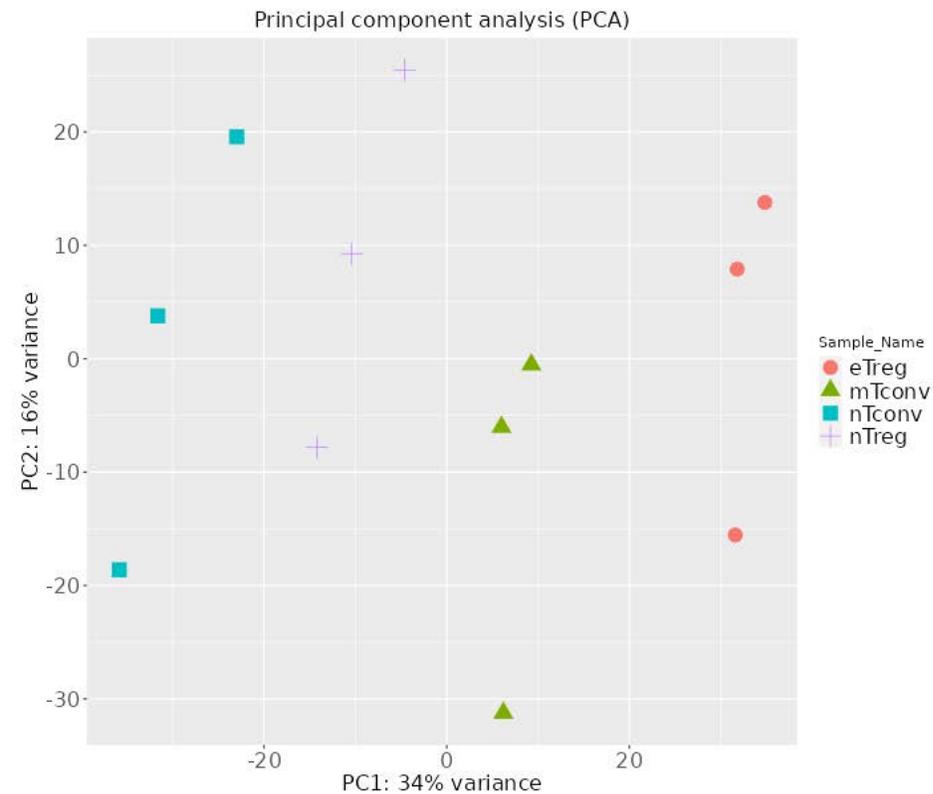
iDEP.96 Load Data Pre-Process Heatmap k-Means **PCA** DEG1 DEG2 Pathway KEGG Genome Bicluster Network R

**Methods**

- Principal Component Analysis
- Multidimensional Scaling
- t-SNE
- Pathway Analysis of PCA rotation

Upload a sample info file to customize this plot.

?

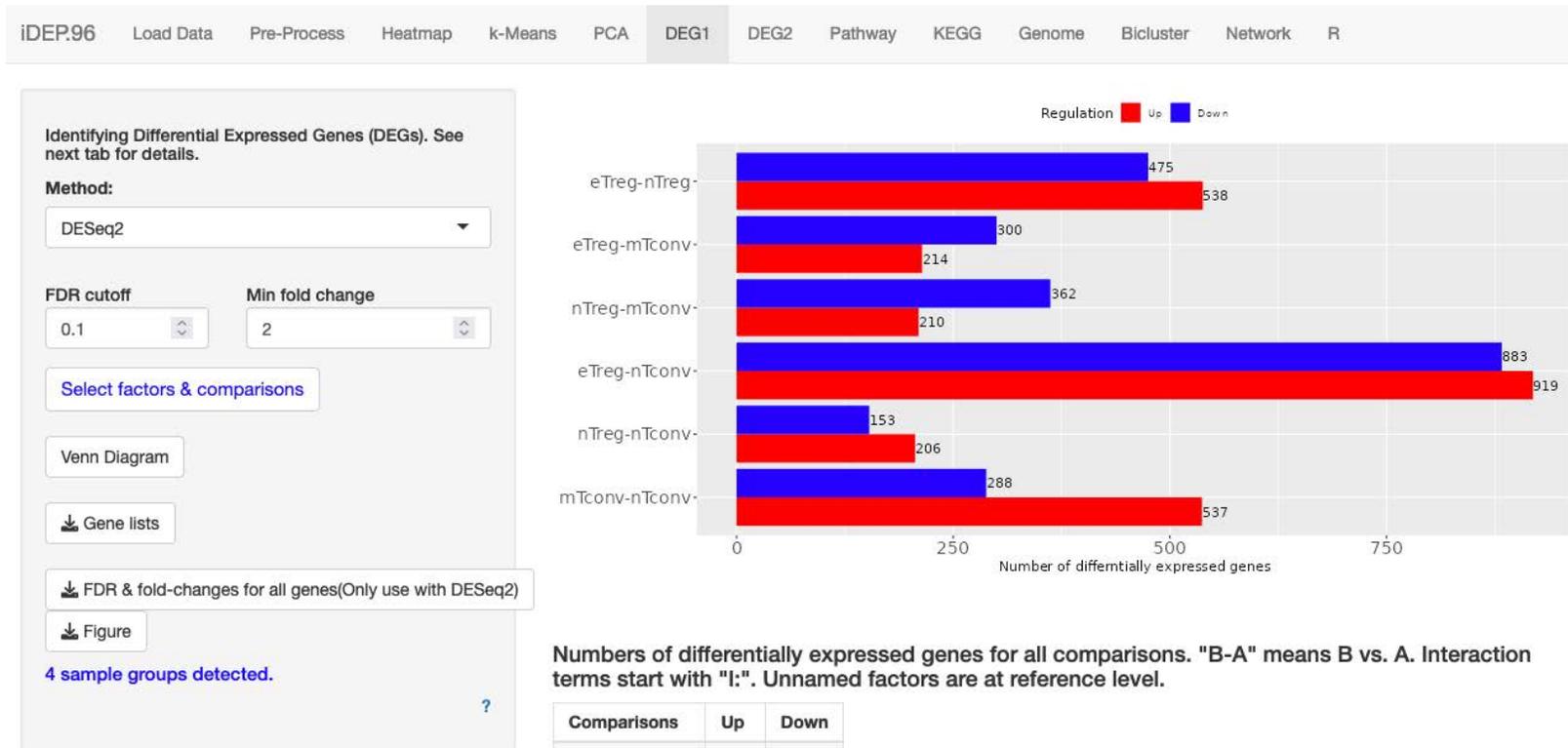


Principal component for x-axis

Principal component for y-axis

- Pathway Analysis of PCA rotationより各PCの負荷量の大きな遺伝子がどのような特徴を持つか調べられる。

# DEG1 (変動遺伝子検出)



FDR : 検定を複数回行う場合、**多重検定補正**を行う必要がある。今回False Discovery Rate (FDR)を使用する。  
Fold change : Group間の平均遺伝子発現の差の比率

# 変動遺伝子の下流解析

PCA DEG1 **DEG2** Pathway

今回は2群だが、複数の群を扱うこともできる。

Volcano plotや、TFモチーフエンリッチメント解析

Pathwayを選択できる

Examine the results of DEGs for each comparison  
Select a comparison to examine. "A-B" means A vs. B (See heatmap). Interaction terms start with "I:"

nTreg-nTconv

Volcano Plot MA Plot Scatter Plot

TF binding motifs in promoters

Gene list & data High-resolution figure

Enrichment analysis for DEGs:  
GO Biological Process

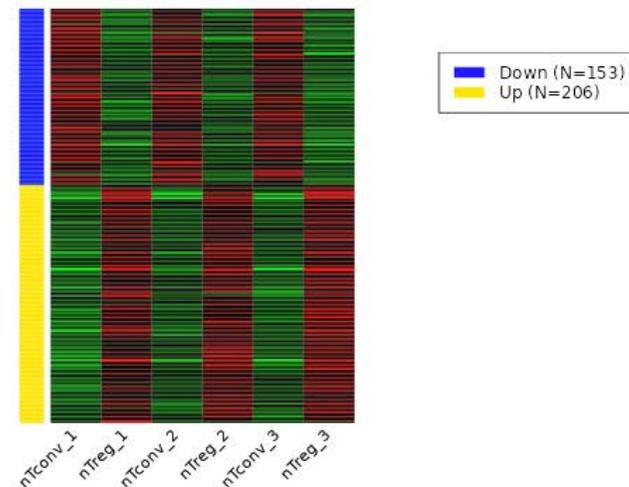
Filtered genes as background for enrichment

Enrichment tree Network (New!)

Enrichment details

Enrichment using STRING API

Also try [ShinyGO](#)



Enriched pathways in DEGs for the selected comparison:

Direction	adj.Pval	nGenes	Pathways
Down regulated	7.7e-04	32	<a href="#">Defense response</a>
	7.7e-04	11	<a href="#">Negative regulation of inflammatory response</a>
	7.7e-04	3	<a href="#">Negative regulation of neutrophil activation</a>
	7.8e-04	27	<a href="#">Cell adhesion</a>
	7.8e-04	30	<a href="#">Positive regulation of signal transduction</a>

# Pathway解析

DEG2

Pathway

Genome

Select a comparison to analyze:

mTconv-nTconv

Select method:

GAGE

Select genesets (Choose KEGG to show pathway diagrams):

GO Biological Process

Geneset size: Min.

5

Max.

2000

Pathway significance cutoff (FDR)

0.2

Number of top pathways to show

30

Use absolute values of fold changes for GSEA and GAGE

Remove genes with big FDR before pathway analysis:

1

Pathway tree

Network(New!)

↓ Pathway list w/ genes

\* Warning! The many combinations can lead to false positives in pathway analyses.

?

Direction	GAGE analysis: mTconv vs nTconv	statistic	Genes	adj.Pval
Up	T cell activation	5.4505	371	1.9e-04
	Regulation of leukocyte activation	5.0952	395	4.0e-04
	Leukocyte differentiation	5.0642	399	4.0e-04
	Regulation of cell activation	5.036	422	4.0e-04
	Regulation of T cell activation	4.8388	253	9.7e-04
	Regulation of lymphocyte activation	4.7193	339	1.3e-03
	Regulation of leukocyte cell-cell adhesion	4.6678	245	1.6e-03
	Positive regulation of cytokine production	4.5744	360	1.8e-03
	Leukocyte cell-cell adhesion	4.5384	274	1.8e-03
	Regulation of cell adhesion	4.5368	487	1.8e-03
	Positive regulation of cell activation	4.5286	263	1.8e-03
	Regulation of cell-cell adhesion	4.5121	308	1.8e-03
	Positive regulation of cell adhesion	4.4455	299	2.1e-03
	Mononuclear cell differentiation	4.4413	311	2.1e-03
	Positive regulation of leukocyte activation	4.4215	256	2.2e-03
	Positive regulation of cell-cell adhesion	4.4099	206	2.3e-03
	Lymphocyte differentiation	4.3069	285	3.1e-03
	Adaptive immune response	4.2744	325	3.4e-03
	T cell differentiation	4.1206	203	6.7e-03
	Positive regulation of leukocyte cell-cell adhesion	4.0965	182	7.0e-03
	Positive regulation of lymphocyte activation	4.0771	228	7.0e-03
	Regulation of hemopoiesis	4.0377	277	7.6e-03
	Lymphocyte mediated immunity	3.9872	187	9.6e-03
	Regulation of interferon-gamma production	3.9672	75	1.2e-02
	Interferon-gamma production	3.9536	76	1.2e-02

GAGE, PGSEA : 各gene setに含まれる全遺伝子のfold changeを用いた検定, PGSEAは各サンプルごとに統計量を出してくれる。

GSEA : 全遺伝子の中で各gene setに含まれる遺伝子の順位に偏りがあるか

# Genome

Pathway

Genome

Bicluster

Select a comparison to visualize:  
day77-day2

Genes: FDR: 0.1      Fold change: 2

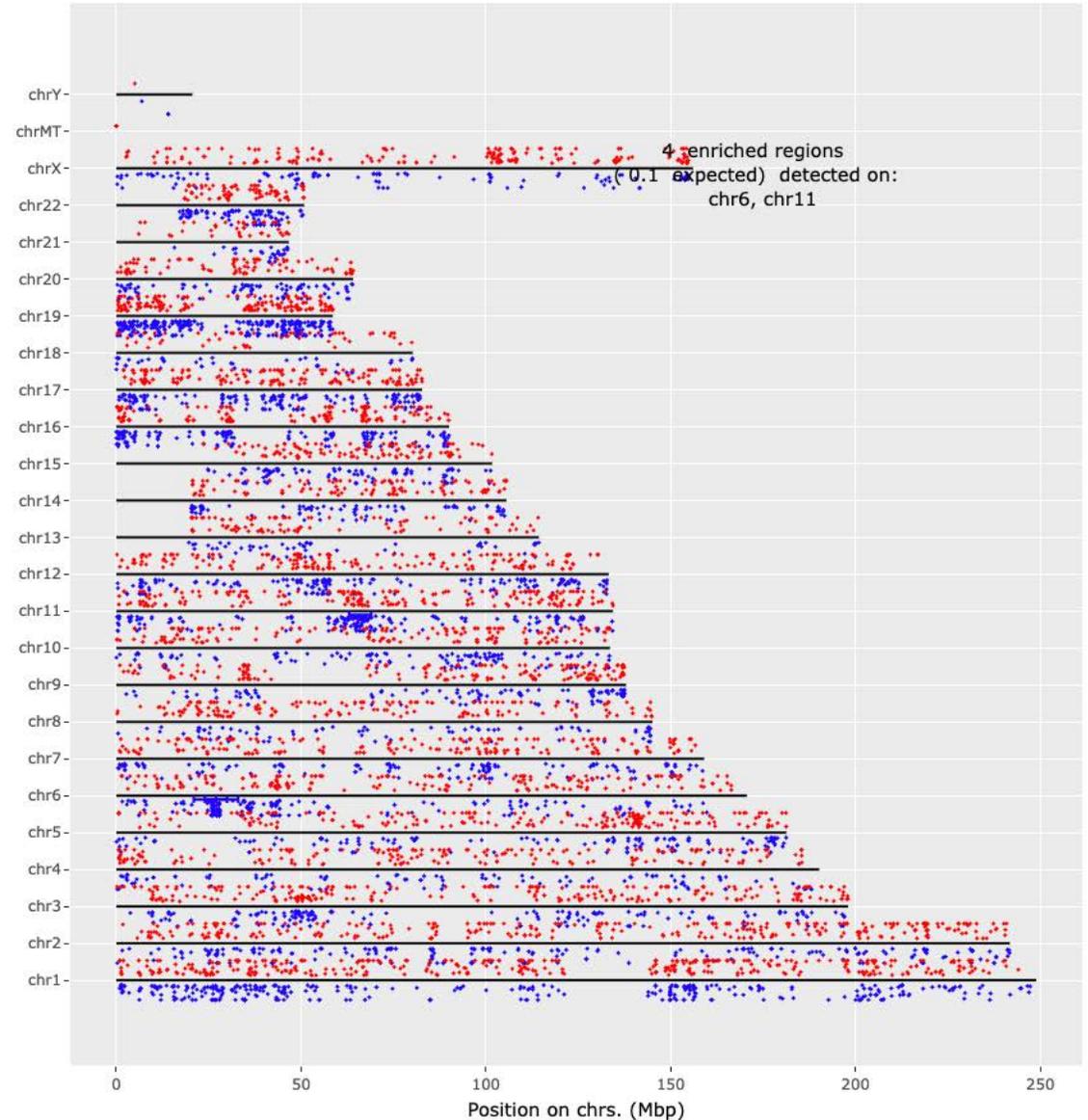
Label Genes       Coding genes only

Window Size (Mb): 6      Steps: 2

FDR cutoff for window: 1e-04

Run PREDA (5 mins)

Info



変動遺伝子のゲノム上での偏りを調べる。  
腫瘍サンプルなどは領域特異性があることも。

# Biclustering

Home Bicluster Network

Biclustering can discover genes correlated on subset of samples. Only useful when sample size is large (>10). Uses methods implemented in the biclust R package.

Most variable genes to include

3000

Method:

BCCC

Select a cluster

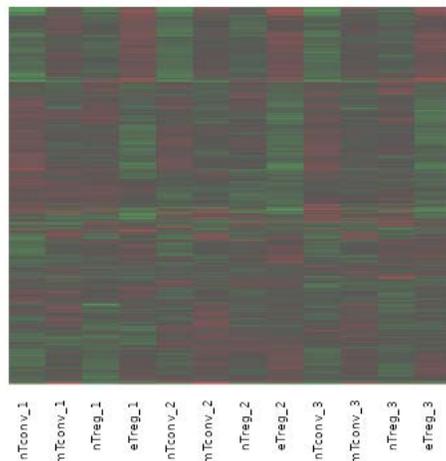
1

Enrichment database

GO Biological Process

Download all biclusters

1 clusters found. Cluster 1 has 2000 genes correlated across 12 samples.



Enriched gene sets in selected bicluster

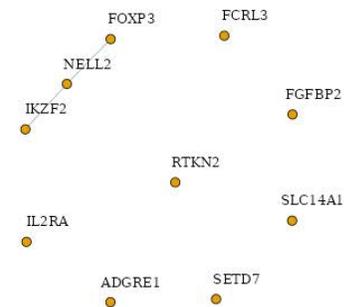
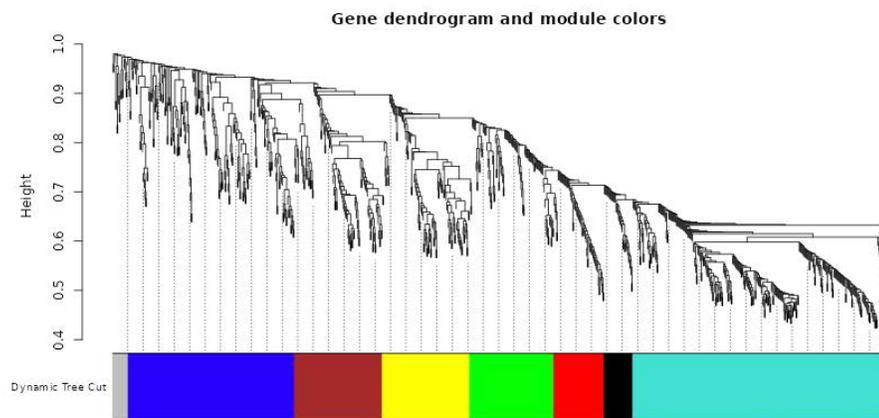
adj.Pval	Genes	Pathways
5.8e-25	274	Cell activation
9.0e-25	508	Immune system process
8.7e-24	248	Leukocyte activation
1.7e-19	157	Lymphocyte activation
2.0e-18	391	Immune response
1.8e-17	461	Cellular response to chemical stimulus
2.1e-17	95	Lymphocyte differentiation
2.3e-17	124	Leukocyte differentiation
1.9e-16	102	Mononuclear cell differentiation
7.4e-16	115	T cell activation

大きなサンプルの場合に有用。  
全サンプルの遺伝子発現を用いて相関する遺伝子群を定義、その上で各クラスターに関連する遺伝子セットを検索。

# Network (WGCNA)

大きなサンプル(n>12程度)の場合に有用。特に臨床検体など、ヘテロなサンプルの際に有用。

全サンプルの遺伝子発現のパターンより、似た遺伝子発現パターンを呈する遺伝子群（モジュール）を定義。その上で各クラスターに関連する遺伝子セットを検索。



## Enriched pathways among all genes in selected module

adj.Pval	Genes	Pathways
1.1e-03	3	RUNX1 and FOXP3 control the development of regulatory T lymphocytes Tregs

Bicluster **Network** R

Identify co-expression networks and sub-modules using [WGCNA](#). Only useful when sample size is large(>15).

Most variable genes to include (<3001)

Soft Threshold:  Min. Module Size:

*A network of 980 genes was divided into 7 modules.*

Select a module

3. brown (114 genes)

Edge Threshold:  Top genes:

Enrichment database:

Curated.Reactome

The network file can be imported to [VisANT](#) or [Cytoscape](#).

?

休憩

[Click here to load demo data](#)

and just click the tabs for some magic!

[Reset](#)

### 1. Optional: Select or search for your species.

Best match

Info

### 2. Choose data type

- Read counts data (recommended)
- Normalized expression values (RNA-seq FPKM, microarray, etc.)
- Fold-changes and corrected P values from CuffDiff or any other program

### 3. Upload expression data (CSV or text)

Browse...

output.rmTCR.tsv

## Analyze public RNA-seq datasets for 9 species

Optional: Upload an experiment design file(CSV or text)

Browse...

No file selected

Example gene IDs

Try [ShinyGO](#) for GO enrichment analysis

?

Ready to load data files.

All new **iDEP 1.0** released in testing mode!

Please [send us a brief email](#) to show your support.

If you state your general research area and how iDEP makes you more productive, we can use it as a support letter when we apply for the next round of funding. Hundreds of strong, enthusiastic letters sent to us in 2019 were essential when we applied for the current grant from NIH/NHGRI (R01HG010805), which expires in 20 months. Your letters will help sustain and improve this service.

If this server is busy, please use a mirror sever <http://ge-lab.org/idep/> hosted by NSF-funded JetStream2.

July 30, 2022: iDEP updated to v0.96. Fixed a bug in the DEG1 tab regarding the different comparisons. iDEP now works even when factors have more than two levels. The downside is that some comparisons for non-reference levels are difficult to make. Users have to change the reference levels and rerun.

April 25, 2022: Gene ID conversion is much faster now, even when species has to be guessed. So is the DEG2 tab.

April 24, 2022: Add a tab for visualizing the fold-change of all genes in all KEGG diagrams across all comparisons!

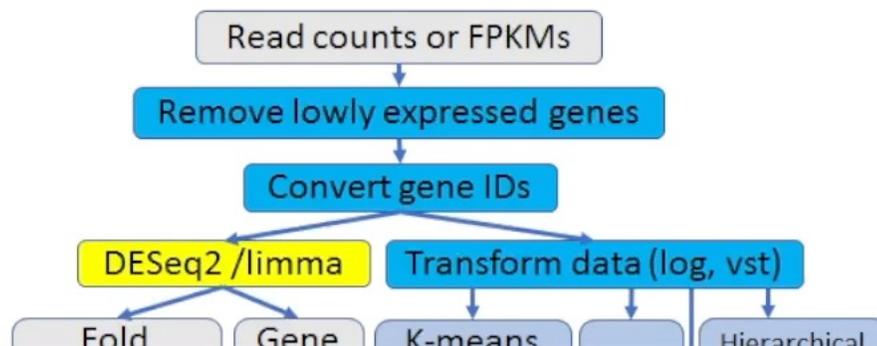
Feb. 11, 2022: Like iDEP but your genome is not covered? [Customized iDEP](#) is now available. Its database includes several custom genomes requested by users. To request to add new species/genome, fill in this [Form](#).

Email Jenny for questions. Follow [Dr Ge](#) on Twitter for updates.

If it is slow, restart from a new browser window (not a new tab). You will be assigned to a new worker computer.

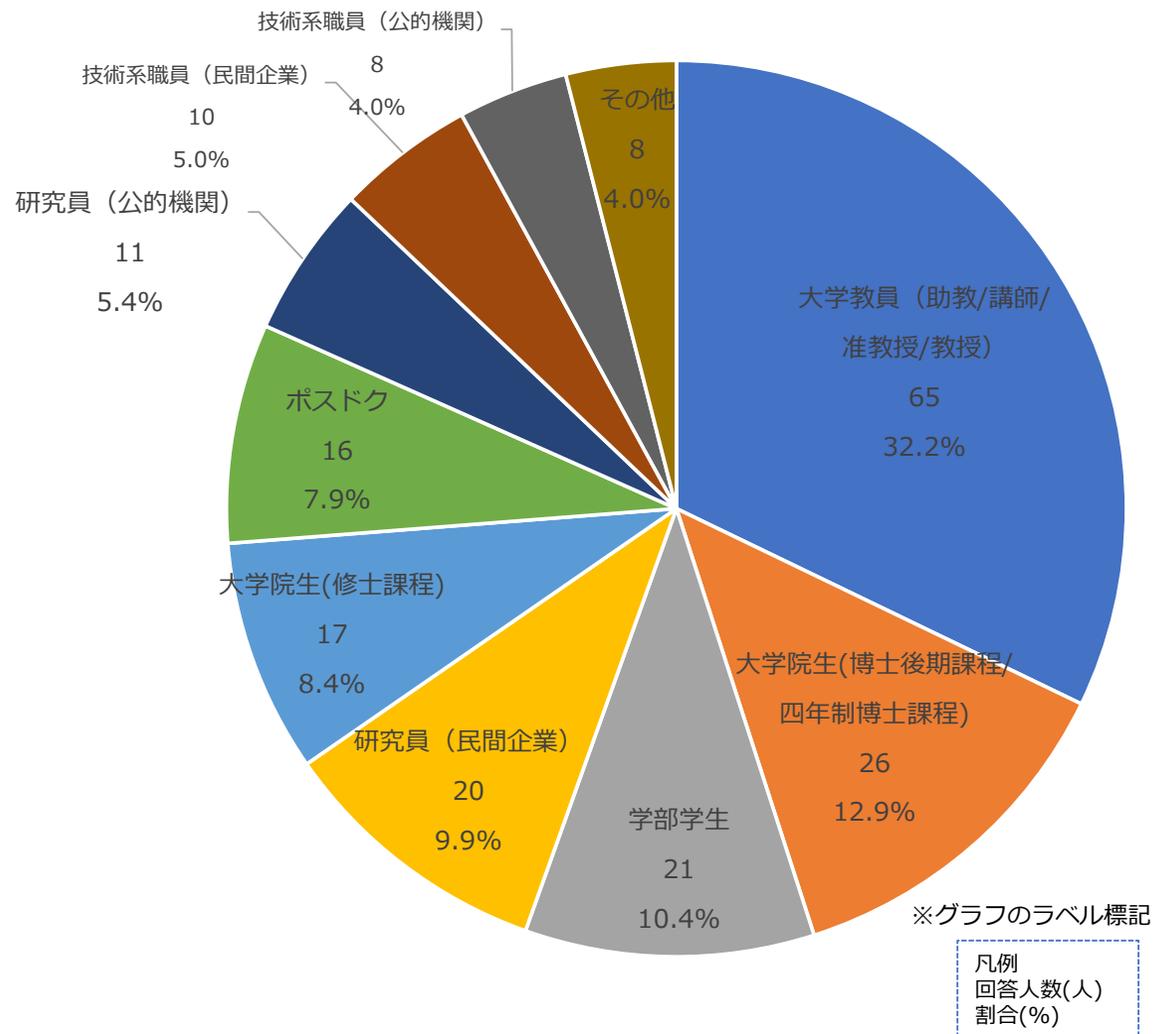
iDEP has not been thoroughly tested. Please let us know if you find any issue/bug.

## iDEP: Integrated Differential Expression and Pathway analysis



# I. 受講申込者の属性

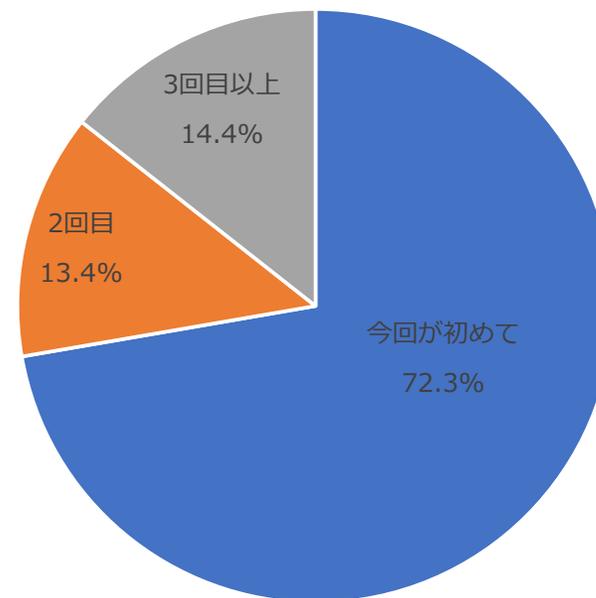
## 1. 職位



## 2. 講習会を知ったきっかけ (複数回答)

講習会を知ったきっかけ (複数回答)	回答数
SNS (Twitterなど)	58
先生・上司・知人(等)に勧められて	41
NBDCウェブサイト	38
NBDCメールマガジン	19
ポスター/チラシ	12
その他ウェブサイト	10
その他メールマガジン	6
DBCLSウェブサイト	5

## 3. AJACS参加は今回が何回目か？



# I. 受講申込者の属性

## 4. 専門分野（複数回答）

専門分野（複数回答）	回答数
生物科学	78
基礎医学	65
基礎生物学	40
臨床医学	36
農学	24
バイオインフォマティクス	21
ゲノム科学	21
情報学	6
複合化学	2
基礎化学	1

## 5. 研究対象（記述式、抜粋）

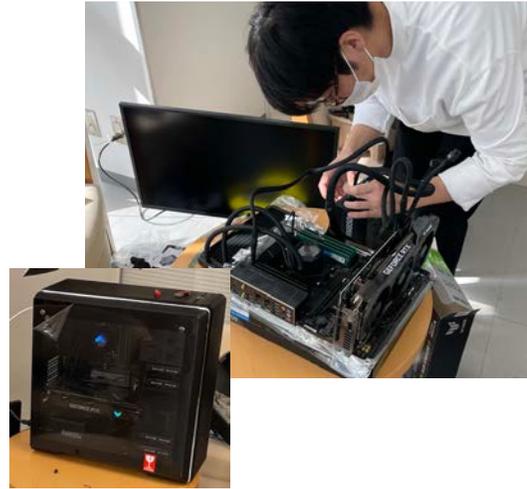
研究対象（記述式）	回答数
RNA	60
DNA	40
タンパク質	28
細胞	19
アミノ酸	6
植物	6
がん	5
微生物	4
神経	4
遺伝子	4
細菌	3
免疫	3
幹細胞	2
ショウジョウバエ	2
ゲノム	2
シロイヌナズナ	2
ウイルス	2
ウイルス	2
iPS	2
脂質	1
オミクス	1
核酸	1

# 解析に必要なコンピューターは？

## ワークステーション



- 高価
- サポートもあり
- カスタム可



## 自作PC

- 最も安価
- サポートなし
- カスタム可
- パーツの相性などで無駄な出費もありえる。
- コンピューターの理解には最適

## HPC, クラウド



- 遺伝研サーバー、大学クラスターなど、もしくはGCP, AWS, Azureなどのクラウドサービス
- 配列からの処理はHPCでおこない、可視化を手元のlaptopなどで行うのも一つ

## BTO



- 比較的安価
- サポートあり
- カスタム可

RAM >64Gb, ubuntuが望ましい。Memory, HDDは詰めるだけ積む。データ保護も忘れずに。

# その他のQ&A

## 非モデル生物の場合は？

Ikraはマウス・ヒトのみ対応

ゲノム配列 or トランスクリプト配列 と遺伝子アノテーションを用意し、自力で解析ツールを回すのがよいか。

iDEPは様々な種に対応。

## どのように勉強すれば良い？

本講習会、日本語書籍など参考書が出ています。ただし、解析にあたっては**必ず各ツールの公式サイトも見る**ようにしてください。また、参考文献のmethodも参考にしてください。

## セキュリティの問題は？

iDEPにupするデータはサマリーデータなので、個人情報にはあたりませんが、機密情報であることには代わりありません。不安であればsecureな環境で解析してください。

# その他のQ&A

## きれいなグラフを描きたい

iDEPはepsで書き出せるため、illustratorで文字やレイアウトの修正が可能です。私は有効な解析をiDEPで確認した後、論文で使用する場合はRで解析をし直して製図したりします。

## scRNAseqに興味がある

本講習会はバルクRNAseqについてののみです。ただし、single cell解析にあたってはバルクRNAseqの仕組みは理解しておく必要があります。また、実例紹介ではシングルセルとの統合についても簡単にご紹介いたします。

## Macを買えば問題ない？

Intel macはスムーズに解析出来ましたが、Apple silicon Macはツールによっては動きません。製図用や端末としての使用にとどめ、リモートのコンピューターを別途用意するのがよいです。

# その他のQ&A

## 感染症のRNAseq

ヒトに感染するウイルスならVIRTUSというツールをおすすめしています。

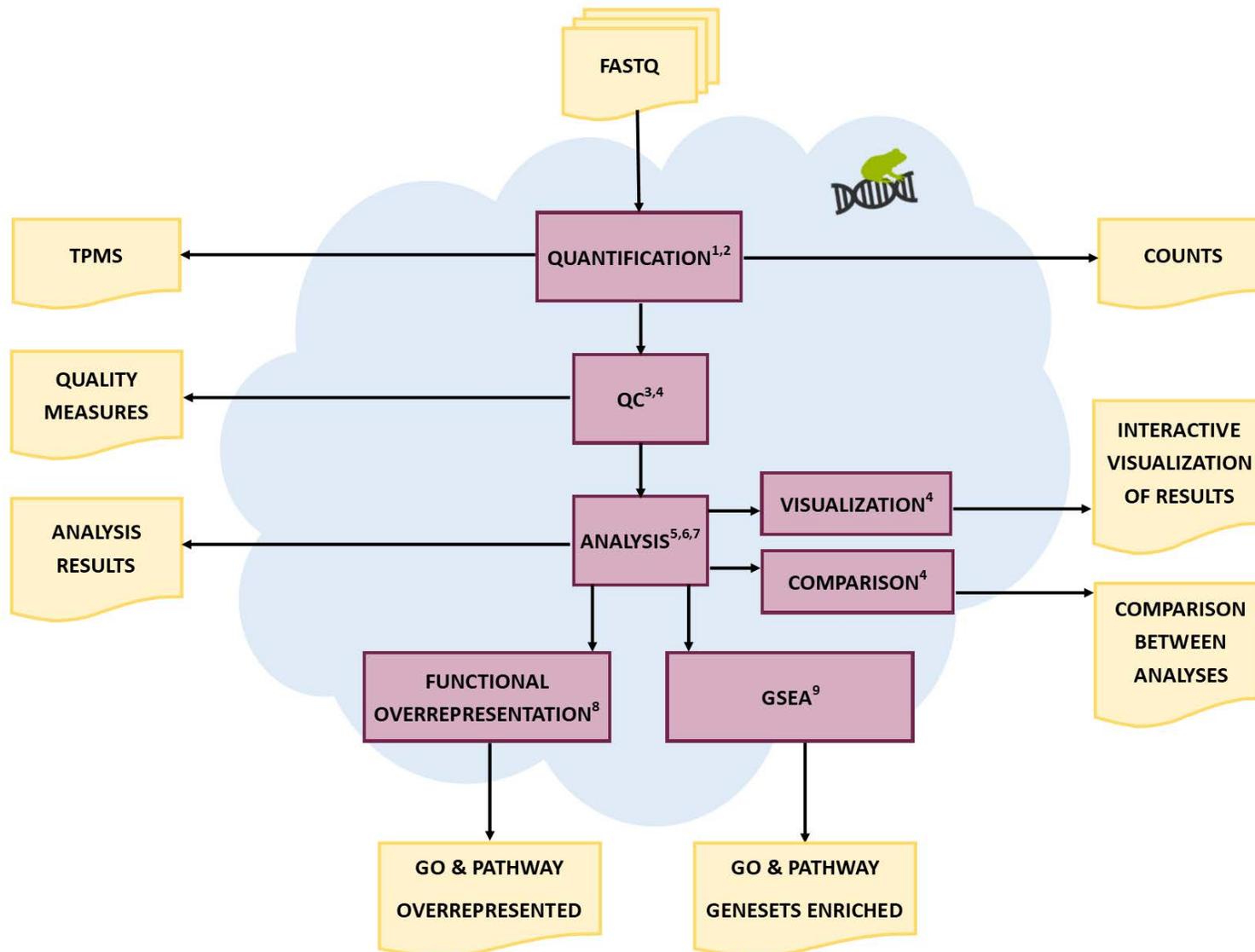
<https://github.com/yyoshiaki/VIRTUS2>

## ゲノムデータとの統合 (vcf)

eQTL, sQTL解析などが出来ますが、これらはno codeでは出来ません。

## Bioinformatics解析はすべて無料で出来て安上がり？

そんなことはありません。リソースの維持や人件費などがかかります。共同研究先に解析を依頼される場合はこのあたりにご注意ください。



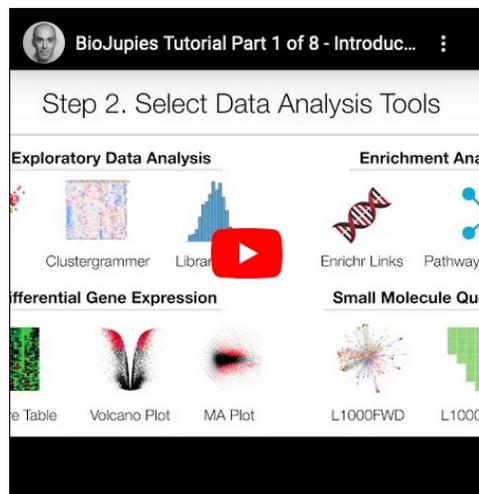
## 様々な生物種に対応

定量ステップはfastp, salmon  
DEGはDESeq2, edgeR, Voom  
GO enrichmentなども

mTconv	nTconv		
ENSG00000284949.1	0	0	
ENSG00000285378.1	0	0	
ENSG00000283860.1	0	0	
ENSG00000284249.1	0	0	
ENSG00000283628.1	0	0	
ENSG00000282068.1	0	0	
ENSG00000284812.1	0	0	
ENSG00000285419.1	0	0	
ENSG00000284854.1	0	0	

iDEPに入力するときなど、1行目の最初にタブを挿入、Ensembl IDの枝番号を削除する必要があった。

# BioJupies <https://maayanlab.cloud/biojupies/>

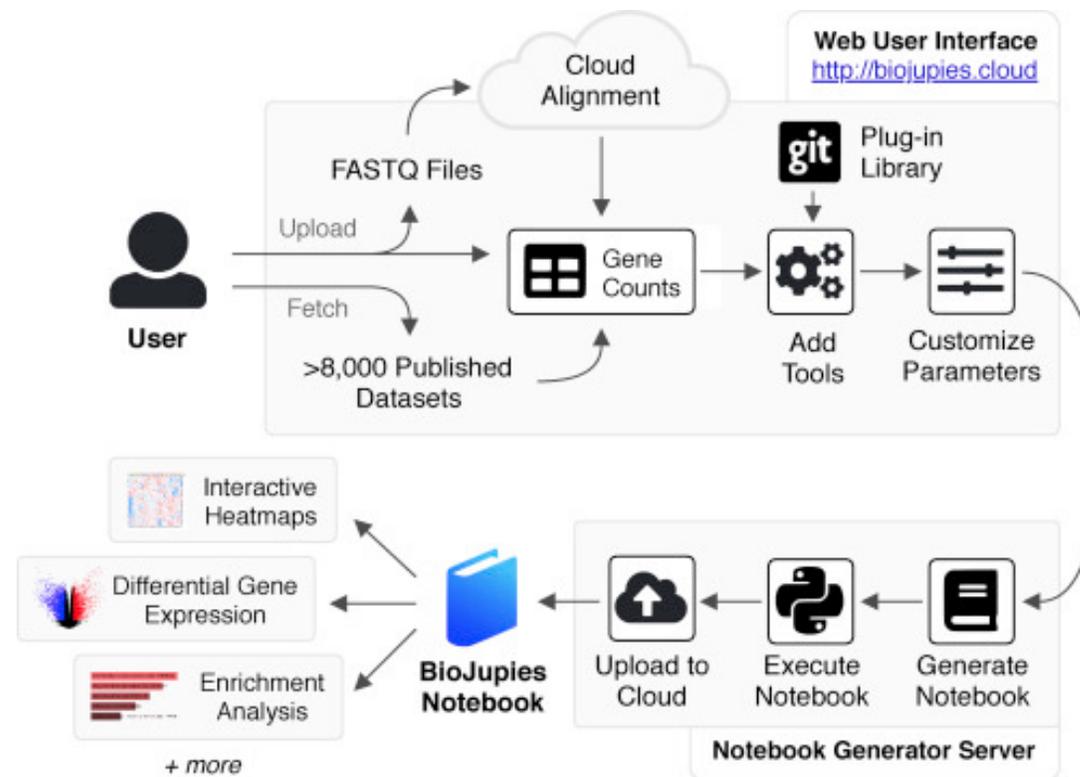


## BioJupies Automatically Generates RNA-seq Data Analysis Notebooks

With BioJupies you can produce in seconds a customized, reusable, and interactive report from your own raw or processed RNA-seq data through a simple user interface

[Get Started](#)

Torre, Denis, Alexander Lachmann, and Avi Ma'ayan. 2018. "BioJupies: Automated Generation of Interactive Notebooks for RNA-Seq Data Analysis in the Cloud." *Cell Systems* 7 (5): 556–61.e3.

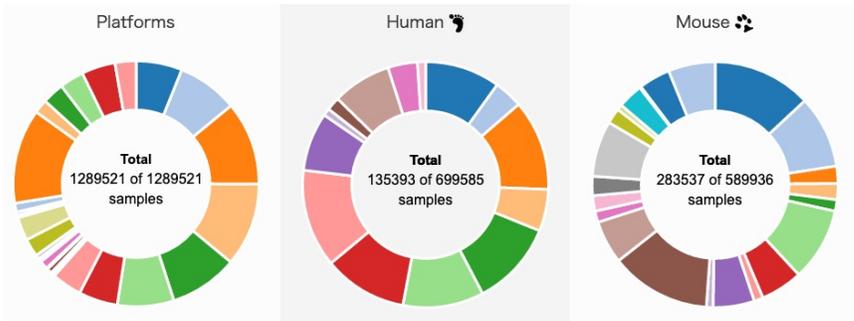
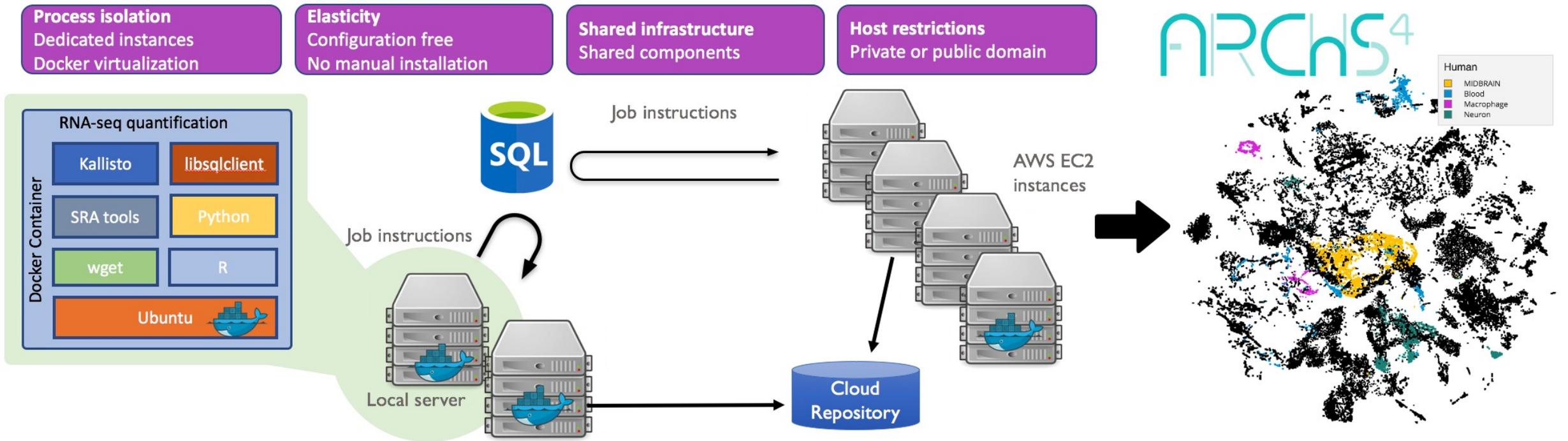


Elysiumでalignment

下流解析まで一つのプラットフォームで解決

Ma'ayan Lab得意のgene set解析のデータセットの多さやノートブックの拡張性などが特徴  
Datasetsからカウントデータをダウンロードすることもできる

# ARCHS4による処理済み公共データ



<https://maayanlab.cloud/archs4/>

GEOやSRAに登録されたデータセットをElysiumにより解析、公開している。  
iDEP, BioJupiesもARCHS4のデータセットを使用

**疾患応用：重症筋無力症合併胸腺腫**

# バイオインフォマティクスで病態を解く

nature communications

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature communications](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 22 July 2022](#)

## Myasthenia gravis-specific aberrant neuromuscular gene expression by medullary thymic epithelial cells in thymoma

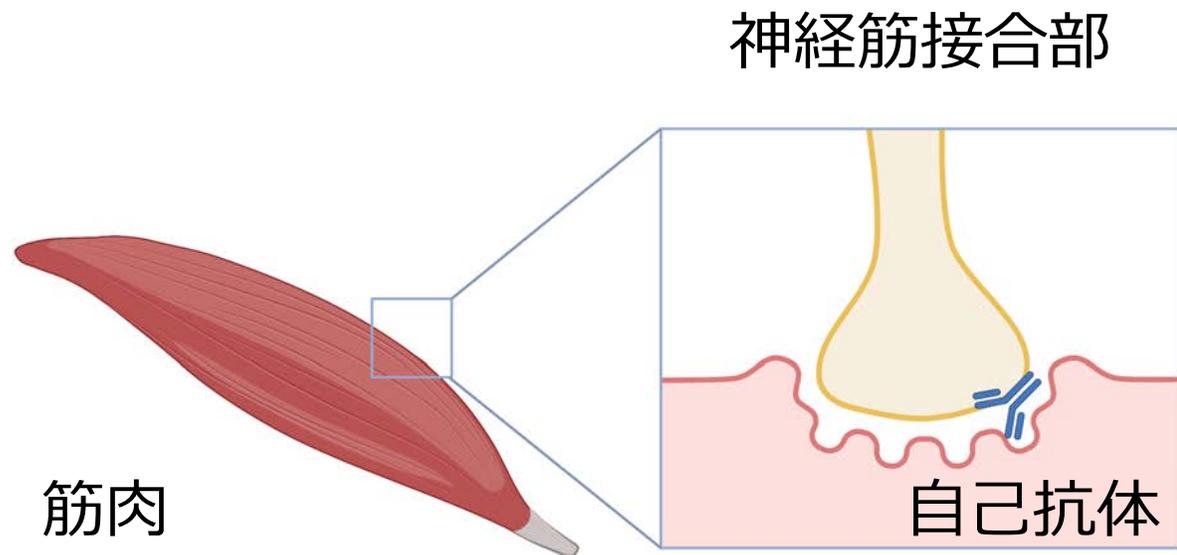
[Yoshiaki Yasumizu](#), [Naganari Ohkura](#) ✉, [Hisashi Murata](#), [Makoto Kinoshita](#), [Soichiro Funaki](#), [Satoshi Nojima](#), [Kansuke Kido](#), [Masaharu Kohara](#), [Daisuke Motooka](#), [Daisuke Okuzaki](#), [Shuji Suganami](#), [Eriko Takeuchi](#), [Yamami Nakamura](#), [Yusuke Takeshima](#), [Masaya Arai](#), [Satoru Tada](#), [Meinoshin Okumura](#), [Eiichi Morii](#), [Yasushi Shintani](#), [Shimon Sakaguchi](#), [Tatsusada Okuno](#) ✉ & [Hideki Mochizuki](#)

[Nature Communications](#) **13**, Article number: 4230 (2022) | [Cite this article](#)

**2039** Accesses | **29** Altmetric | [Metrics](#)

<https://www.nature.com/articles/s41467-022-31951-8>

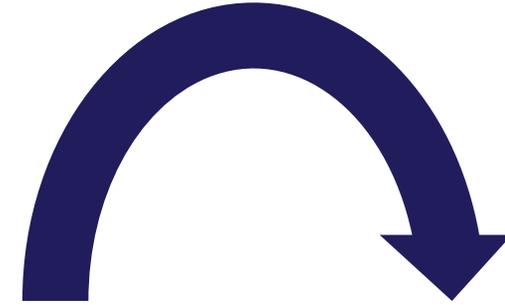
# 重症筋無力症



- 神経筋関連たんぱく質に対する自己抗体が産生されることにより引き起こされる自己免疫疾患。
- 全身の筋力低下を引き起こし、重症化すると呼吸筋麻痺による呼吸困難をきたすこともある。
- 日本では10万人あたり23人が罹患する。

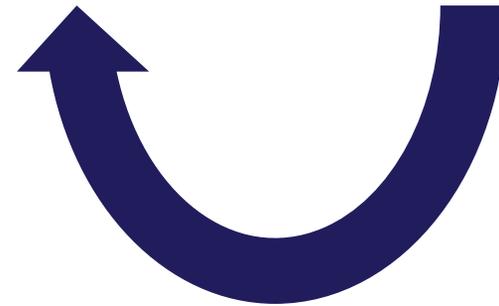
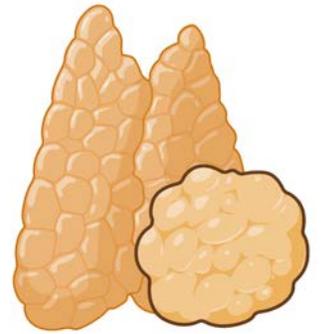
# 胸腺腫合併重症筋無力症の例

21% (Zhi-Feng Mao *et al*, 2012)

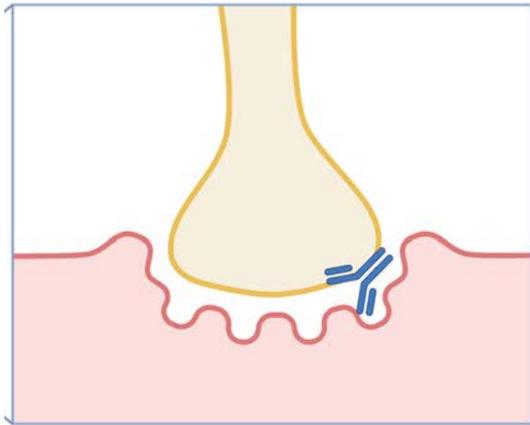


重症筋無力症 (MG)

胸腺腫



25% (Kazuya Kond *et al*, 2005)

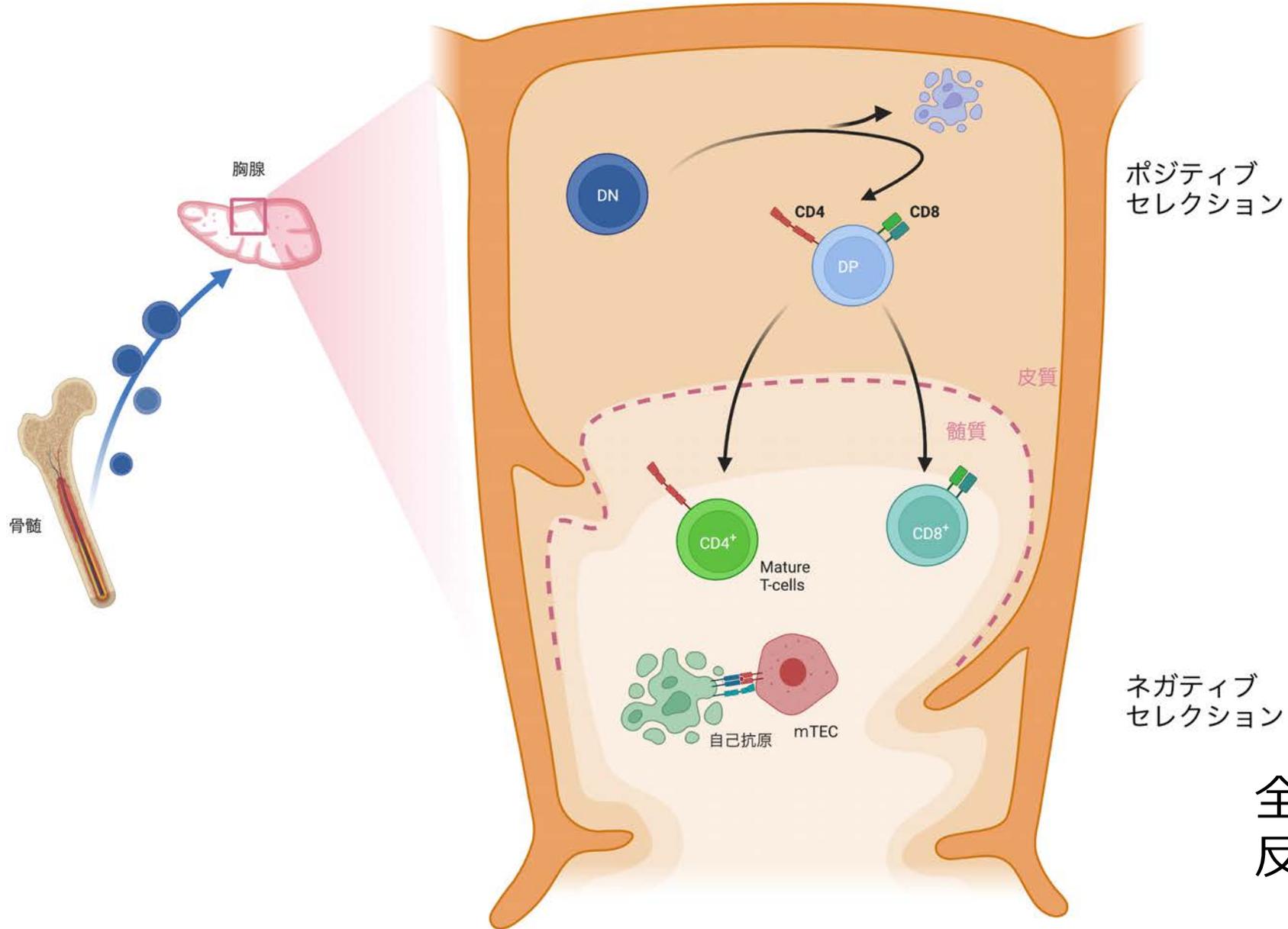


転載不可



**MG胸腺腫の中で何が起きている？**

# 胸腺でのT細胞の教育



全身の自己抗原を提示し、  
反応するT細胞を除去する

# TCGA – The Cancer Genome Atlas

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis Repository

Quick Search Manage Sets Login Cart 0 GDC Apps

## Harmonized Cancer Datasets

### Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

#### Data Portal Summary [Data Release 35.0 - September 28, 2022](#)

PROJECTS 72	PRIMARY SITES 67	CASES 86,394
FILES 843,011	GENES 21,446	MUTATIONS 2,610,557

#### Cases by Major Primary Site

Primary Site	Cases (in 1000s)
Adrenal Gland	0.1
Bile Duct	0.1
Bladder	0.2
Bone	0.1
Bone Marrow	9.0
Brain	0.2
Breast	9.0
Cervix	0.2
Colorectal	8.0
Esophagus	0.2
Eye	0.1
Head and Neck	0.3
Kidney	0.3
Liver	0.2
Lung	11.0
Lymph Nodes	0.1
Nervous System	4.0
Ovary	0.2
Pancreas	0.3
Pleura	0.1
Prostate	0.2
Skin	0.3
Soft Tissue	0.2
Stomach	0.2
Testis	0.1
Thymus	0.1
Thyroid	0.2
Uterus	0.2

#### GDC Applications

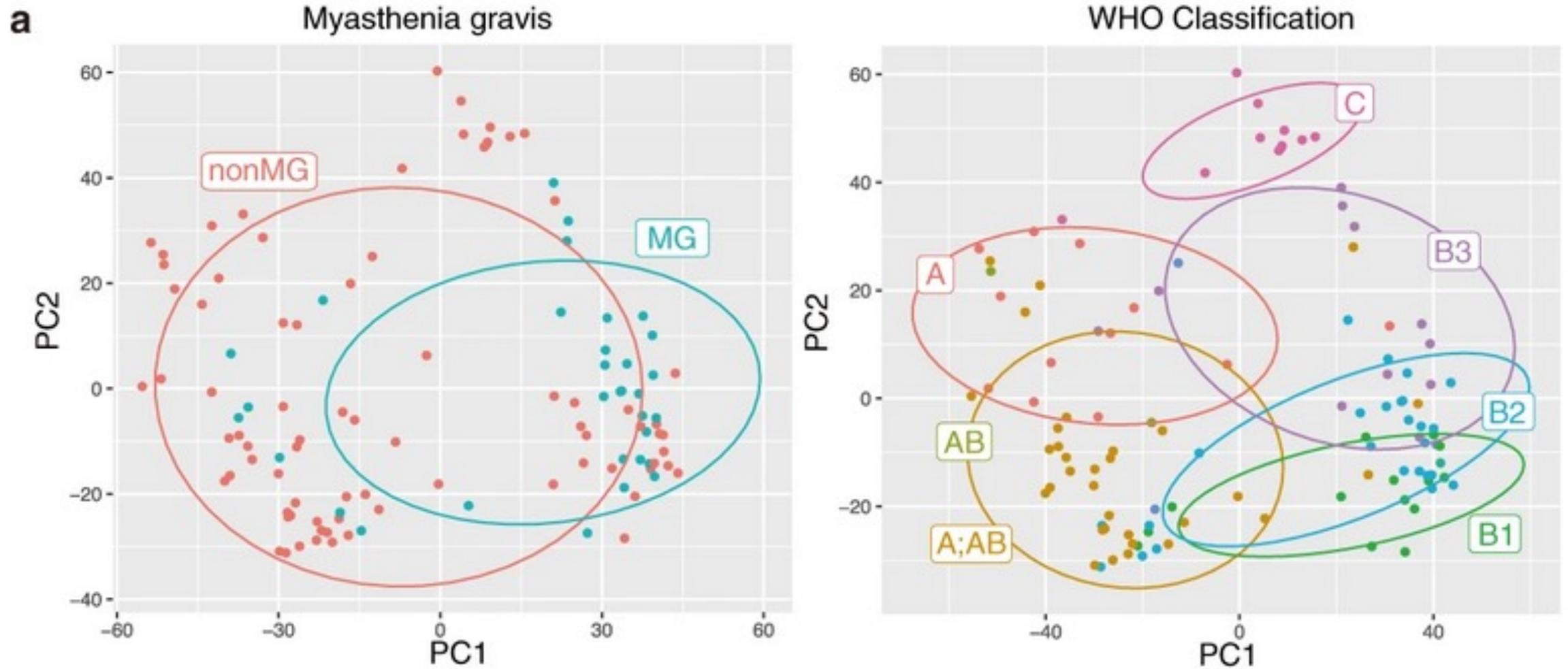
The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

- Data Portal
- Website
- API
- Data Transfer Tool
- Documentation
- Data Submission Portal
- Legacy Archive
- Publications

Gibbs, R. A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

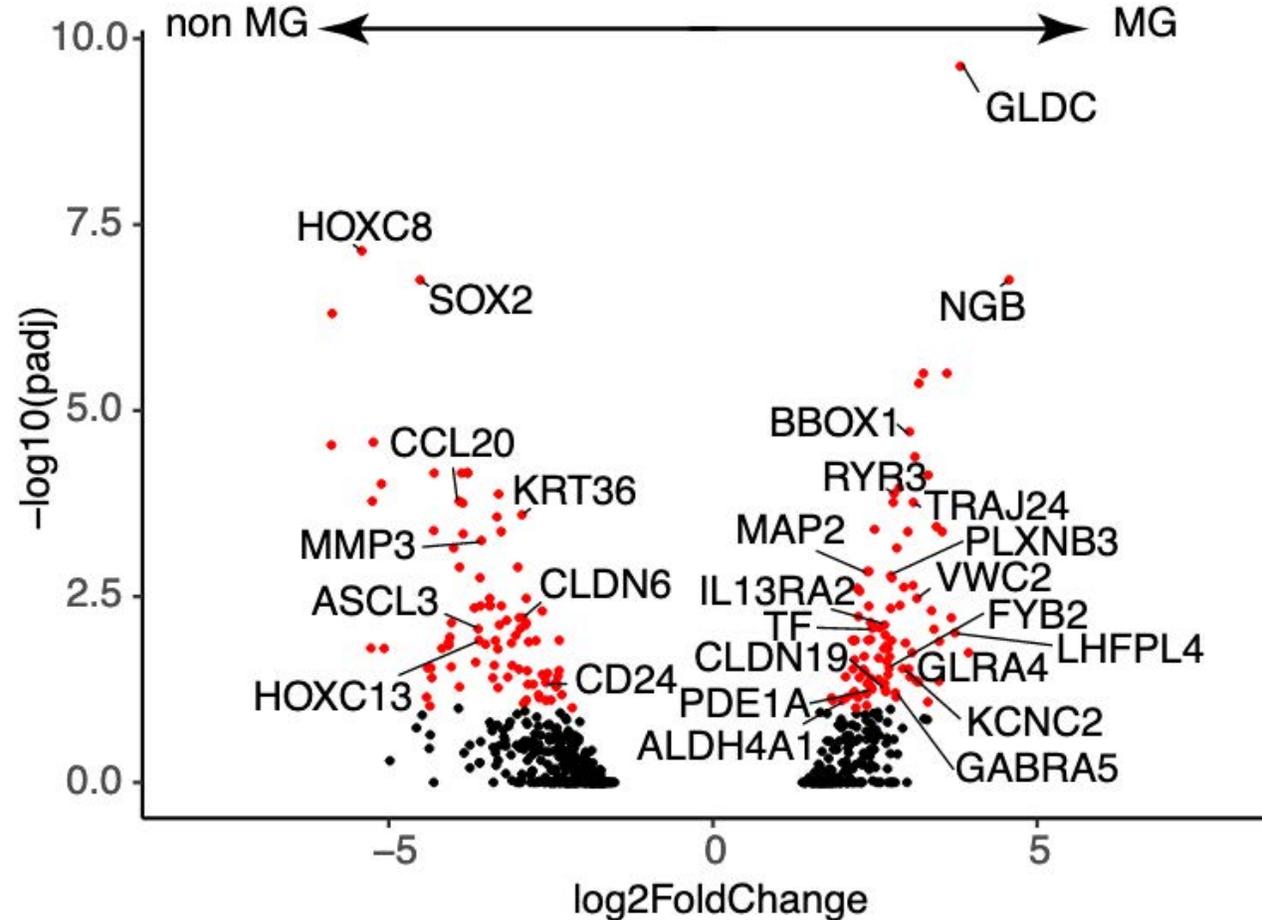
114 thymoma (~30 with MG)のbulk RNAseqが登録されていた。Thymoma検体を再解析した。

# MG/non-MGやWHO分類は特徴的なトランスクリプトームを有する



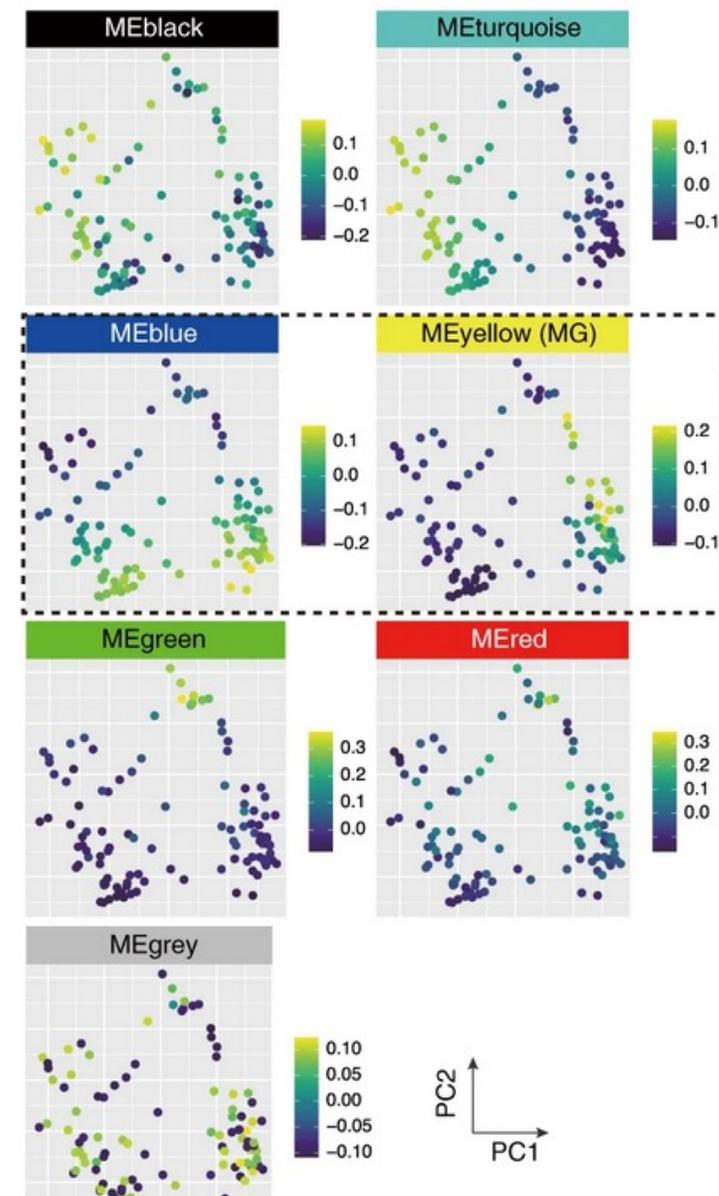
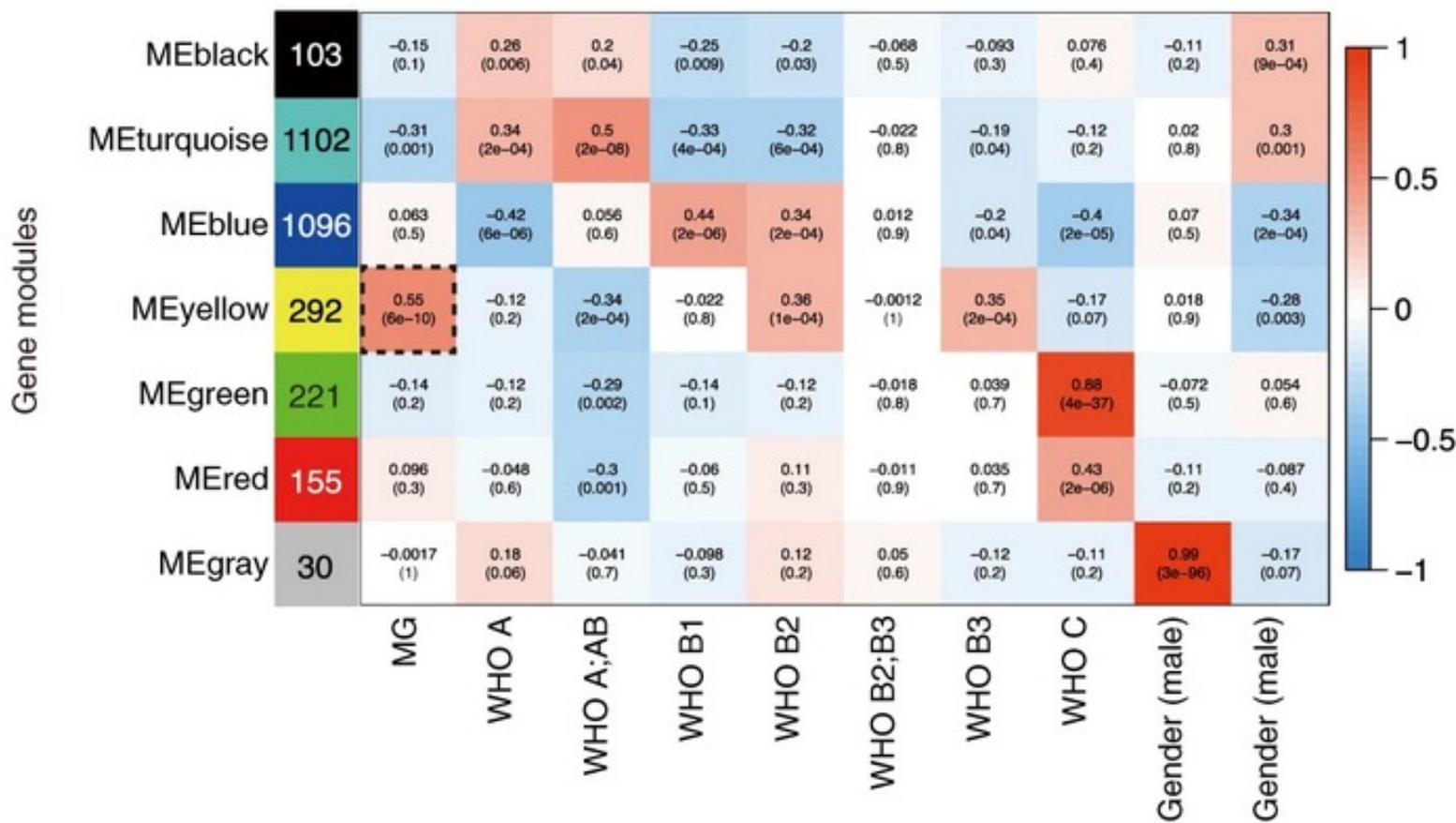
DESeq2を用いてPCA解析を行った

# DESeq2におけるMG-nonMGの変動遺伝子の検出



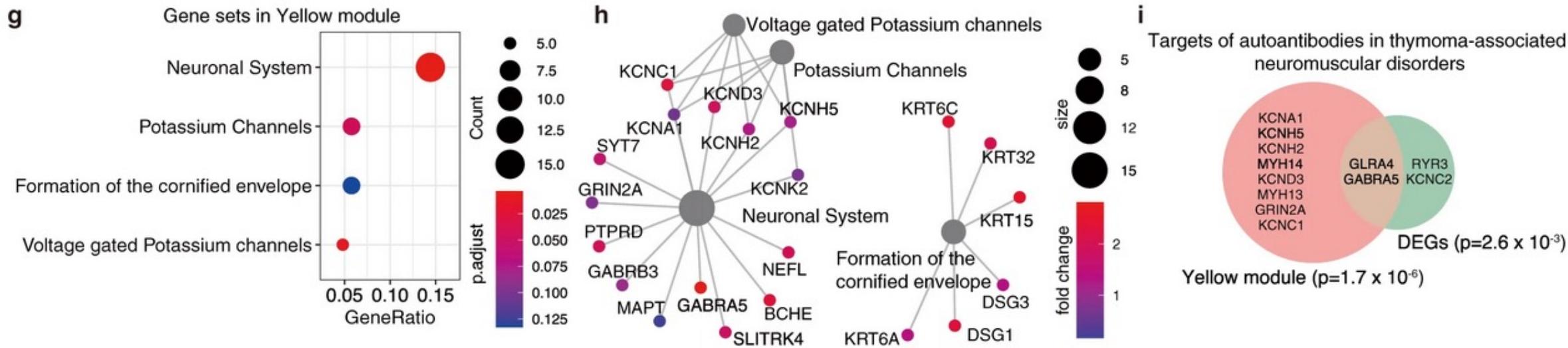
いくつかの遺伝子変動は見つかったが、WHO分類など交絡要因などの影響は？  
-> より全体の遺伝子変動を調べたい

# MGに相関するyellow moduleを同定した

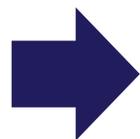


WGCNAを用いた遺伝子ネットワーク解析

# MG合併胸腺腫で神経関連分子の高発現



Yellow module のReactome 遺伝子セット集積探索  
(clusterprofiler)



どの細胞が神経関連分子を出している？

# どの細胞が神経関連抗原を発現しているかを調べるためにシングルセル解析をおこなった

MG21



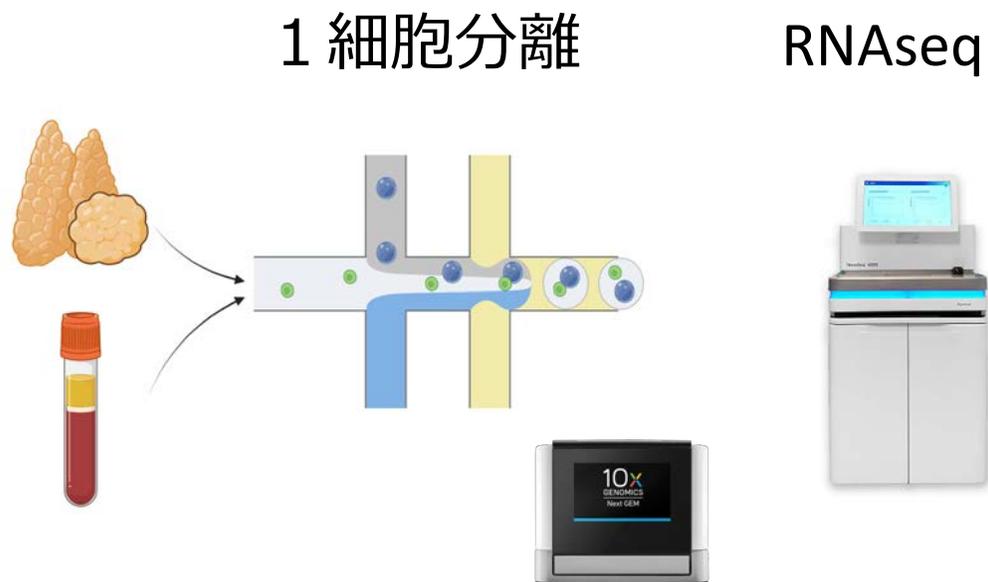
MG22



MG03



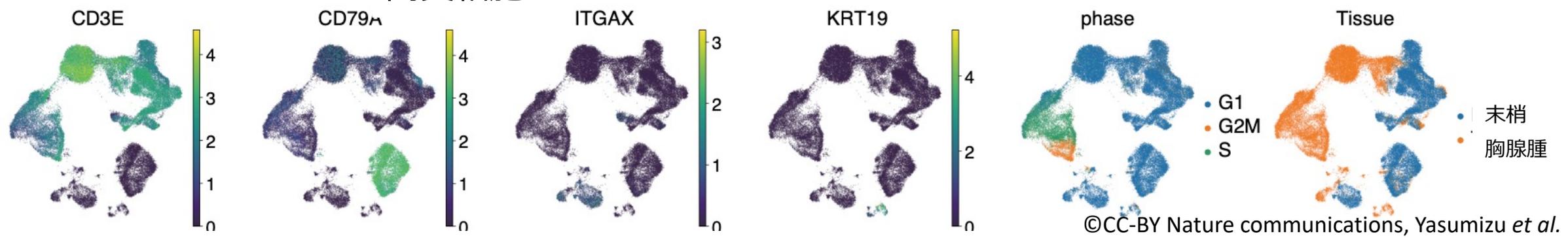
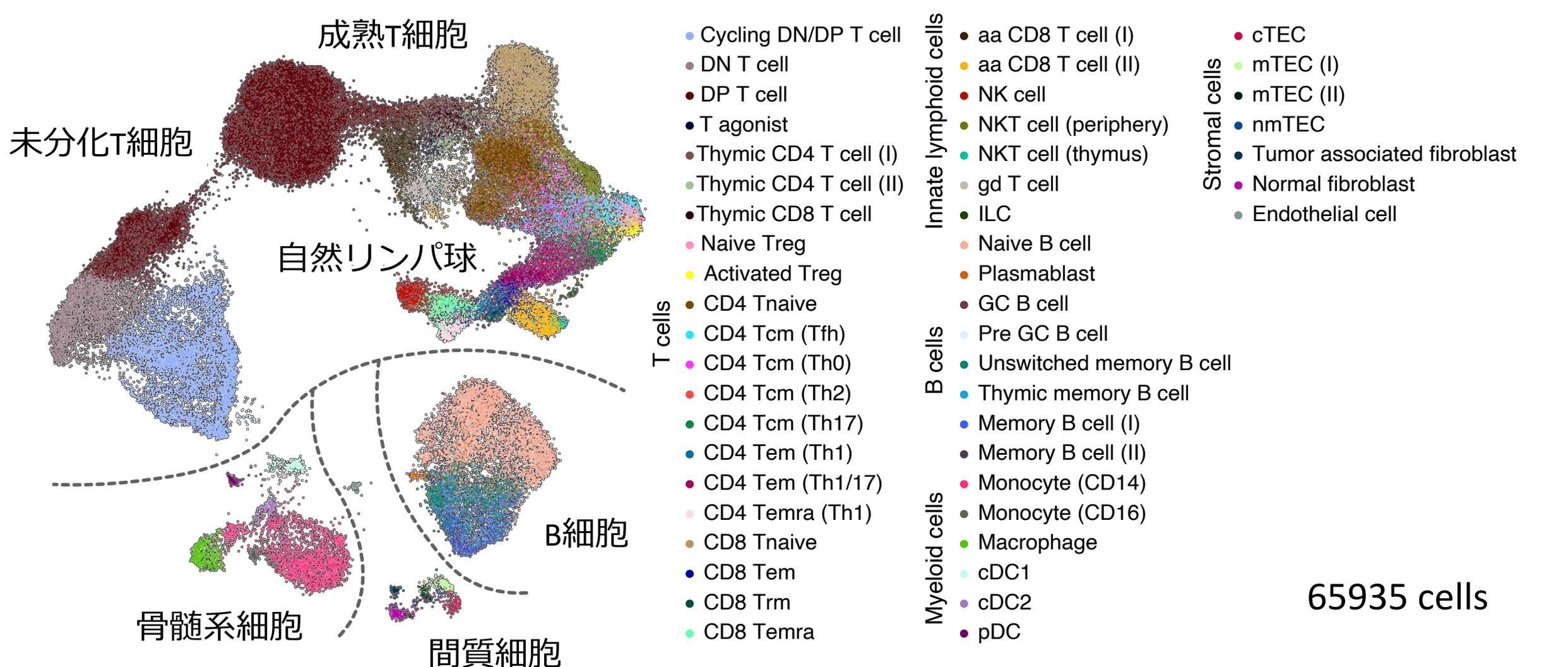
MG23



胸腺腫合併MG患者さんの胸腺腫、末梢血を  
1細胞ごとに単離、RNAseqを行った。

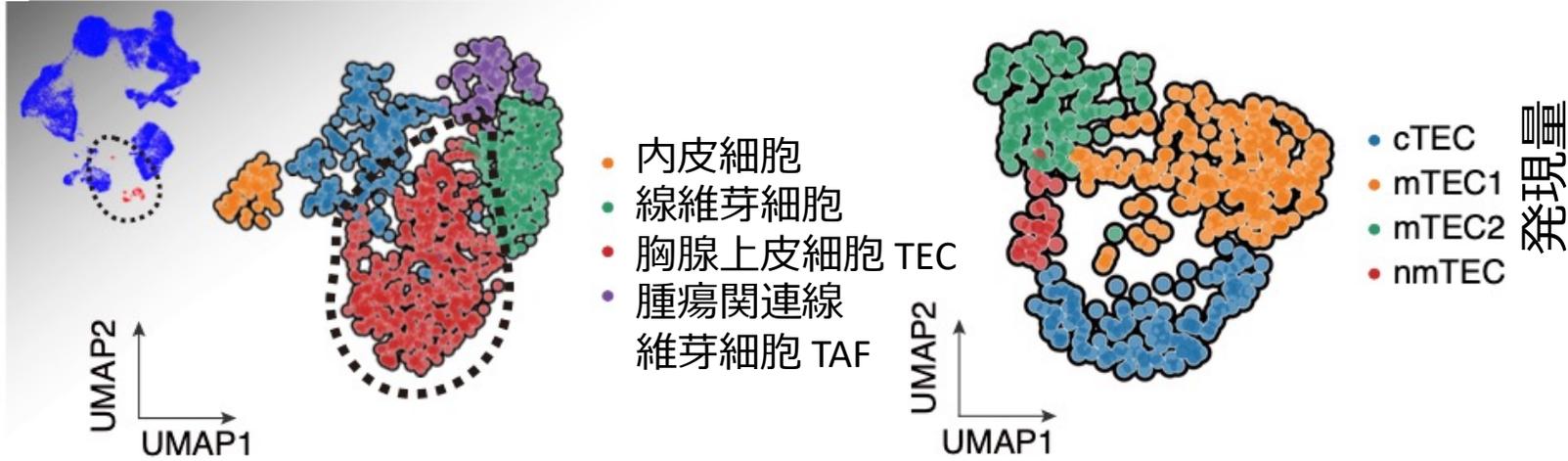


木下先生、村田先生  
(神経内科学)

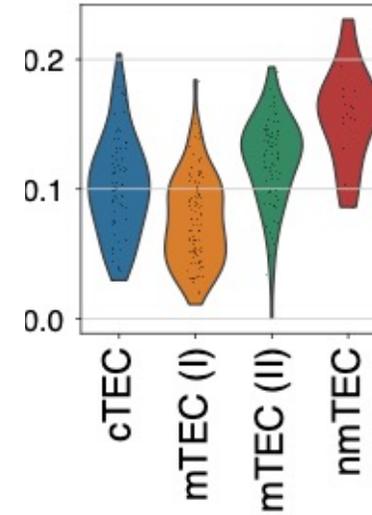


# 神経抗原を発現するnmTECを同定した

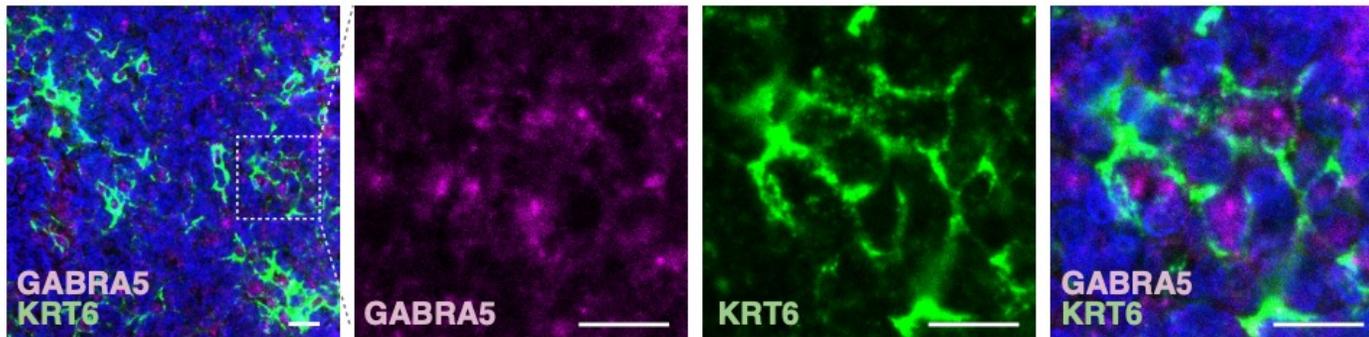
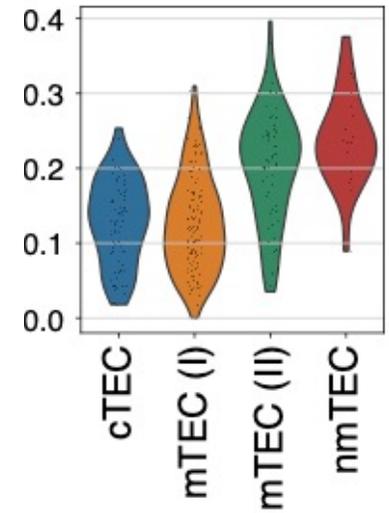
間質細胞



神経抗原

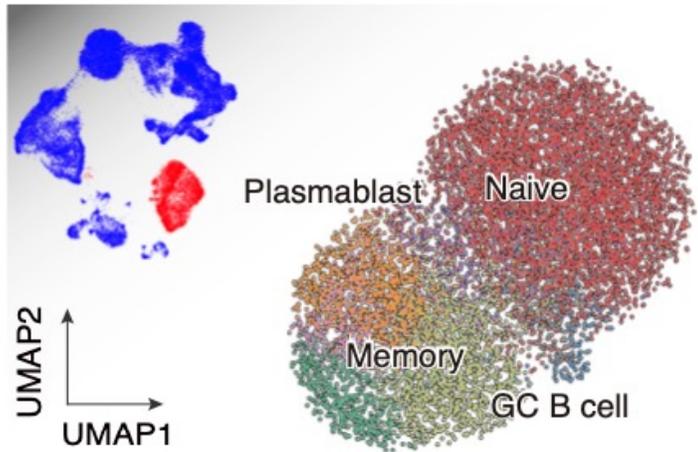


筋抗原

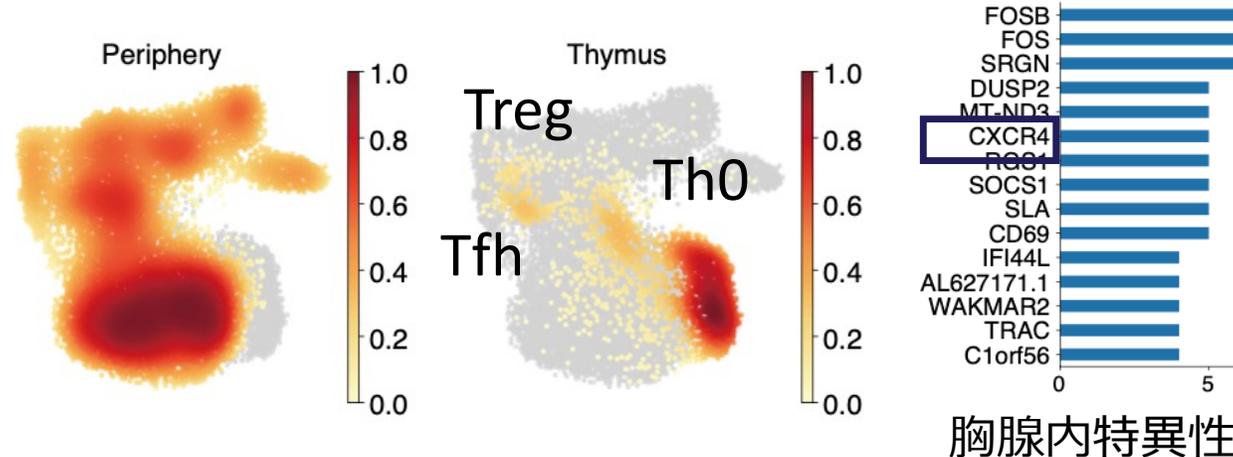
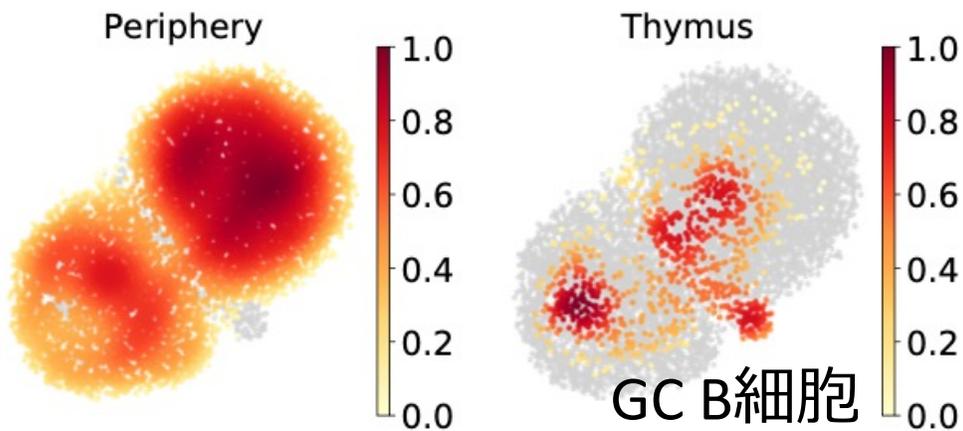
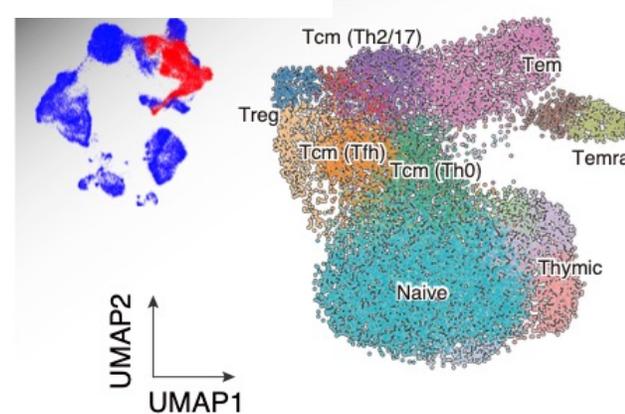


# 胸腺内での胚中心形成、Treg, Th0-Tfhの活性化

## B細胞



## CD4+T細胞



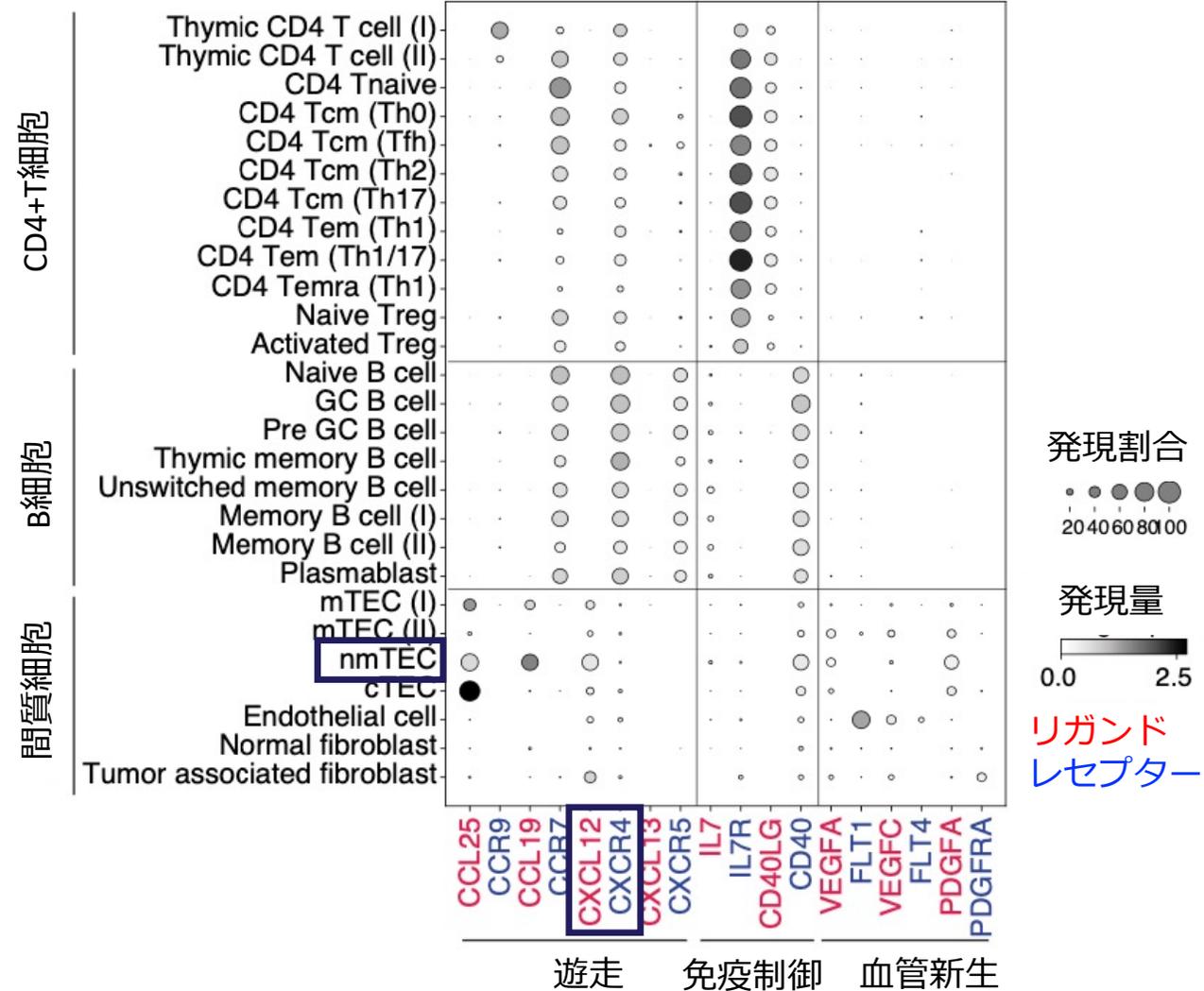
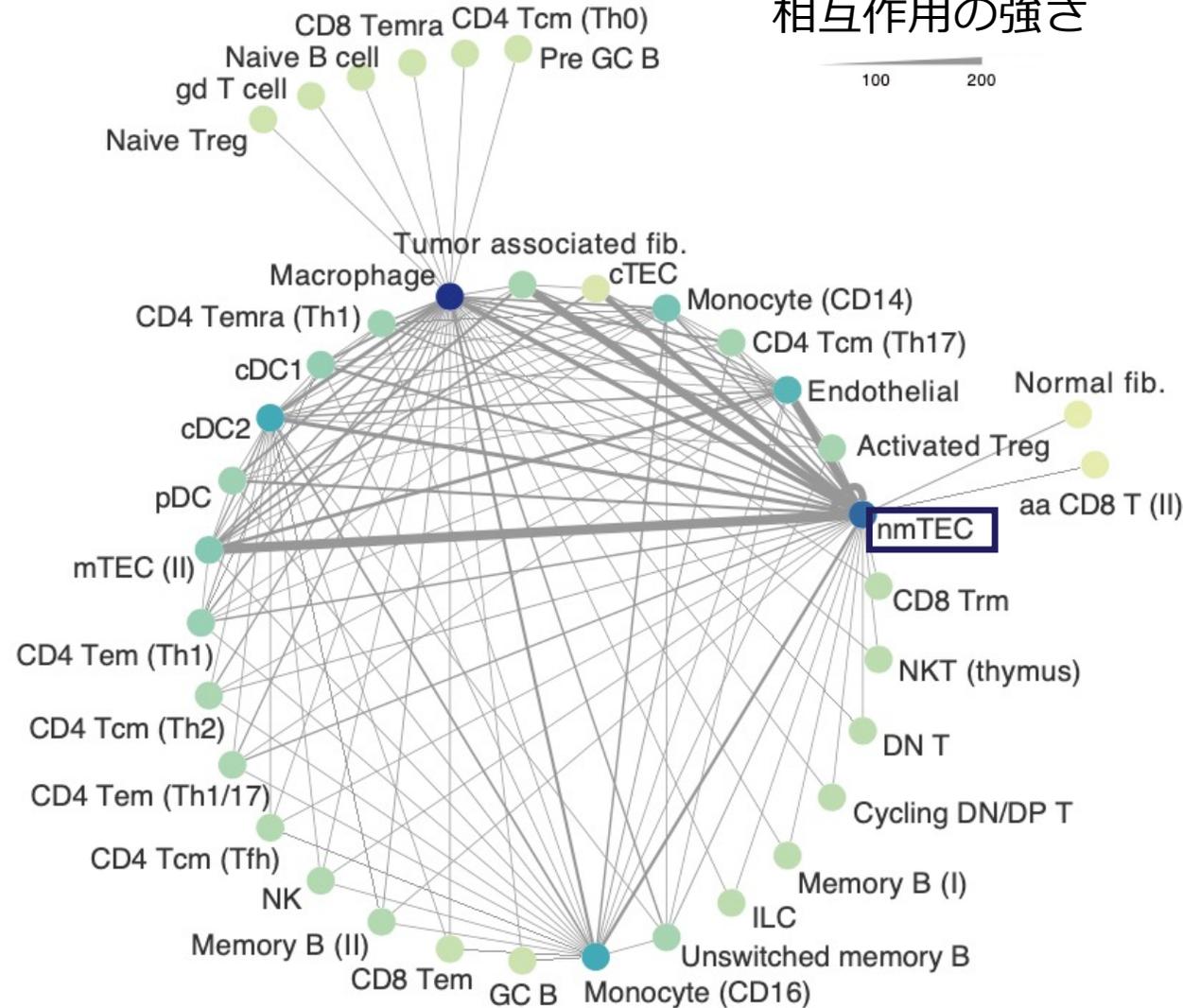
※ 胚中心 (GC) : 免疫応答の場として形成される微小構造で、T細胞や樹状細胞を介したB細胞の成熟が起きる。

※ 濾胞性T細胞(Tfh) : ヘルパーT細胞の一種で、肺中心に存在し、B細胞の抗体産生やクラススイッチを誘導する。

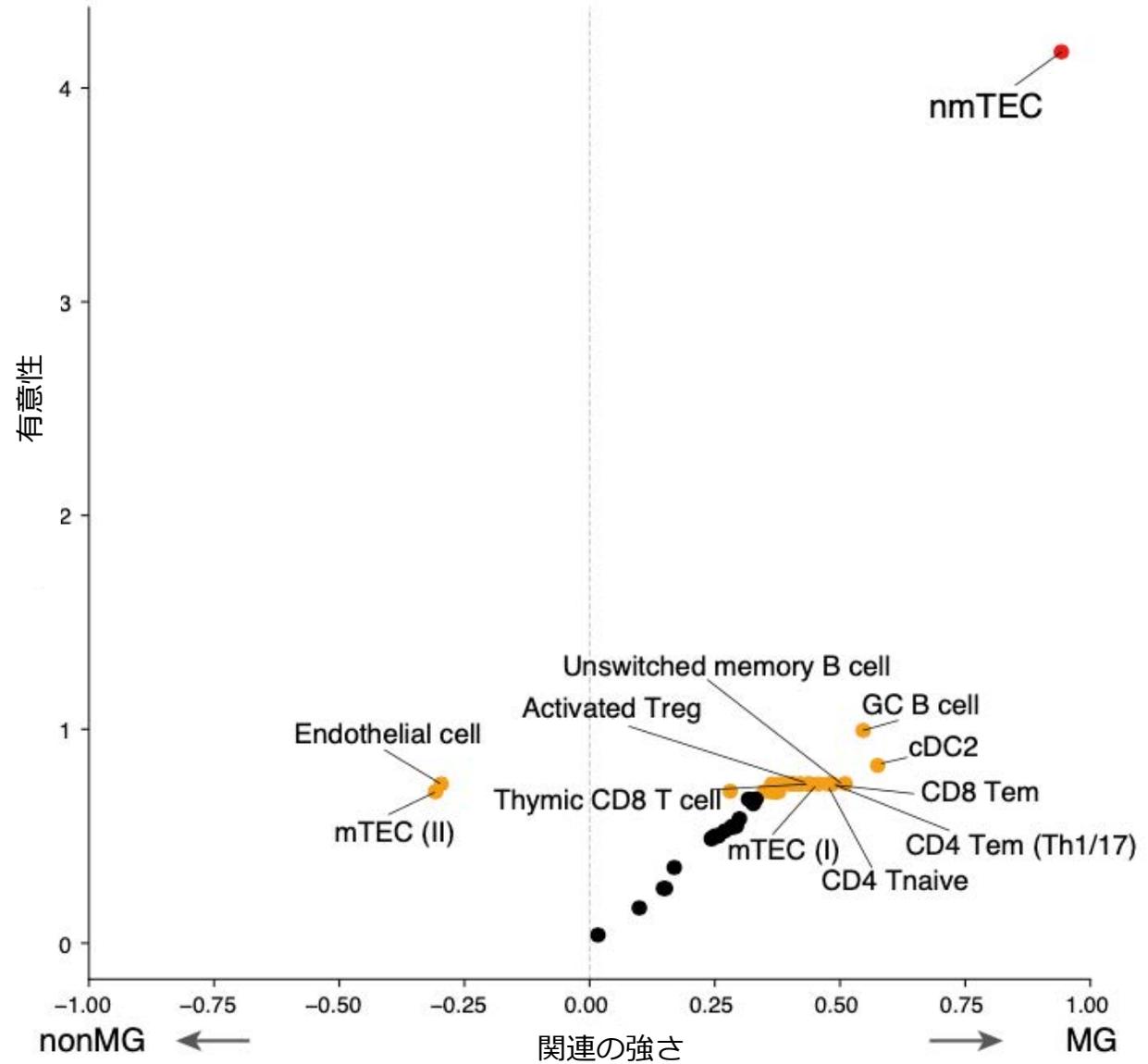
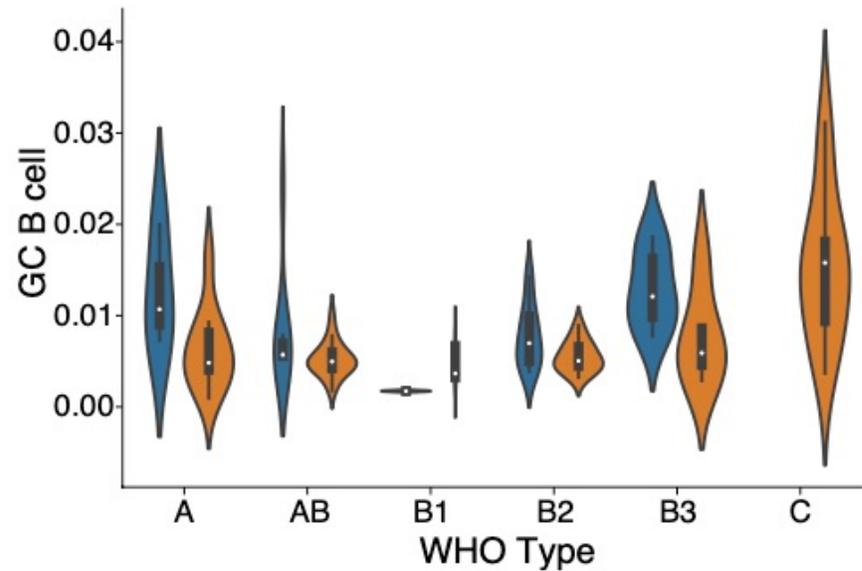
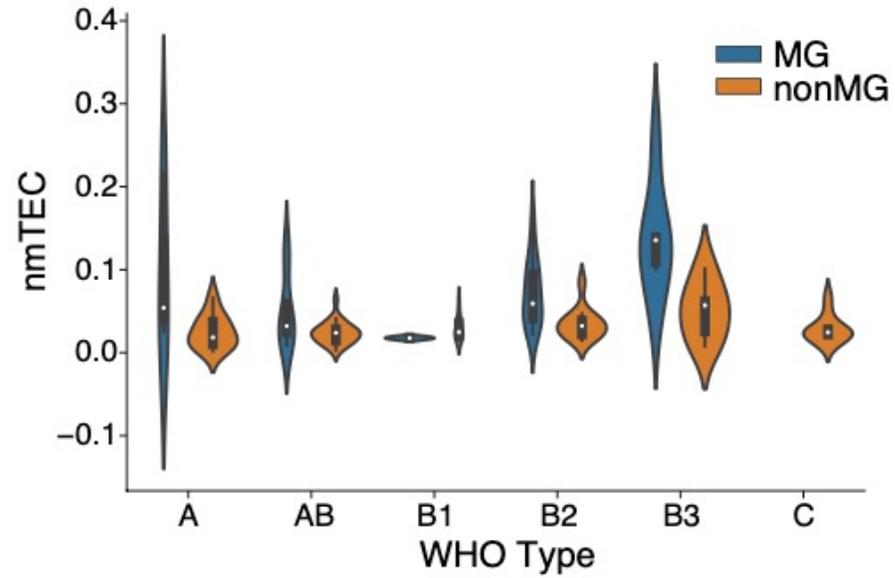
※ 制御性T細胞(Treg) : 免疫応答を抑制し、自己免疫寛容の維持や過度の免疫応答による組織障害を防ぐ。

# 細胞間コミュニケーション解析により、CXCL12-CXCR4の重要性が明らかに

相互作用の強さ



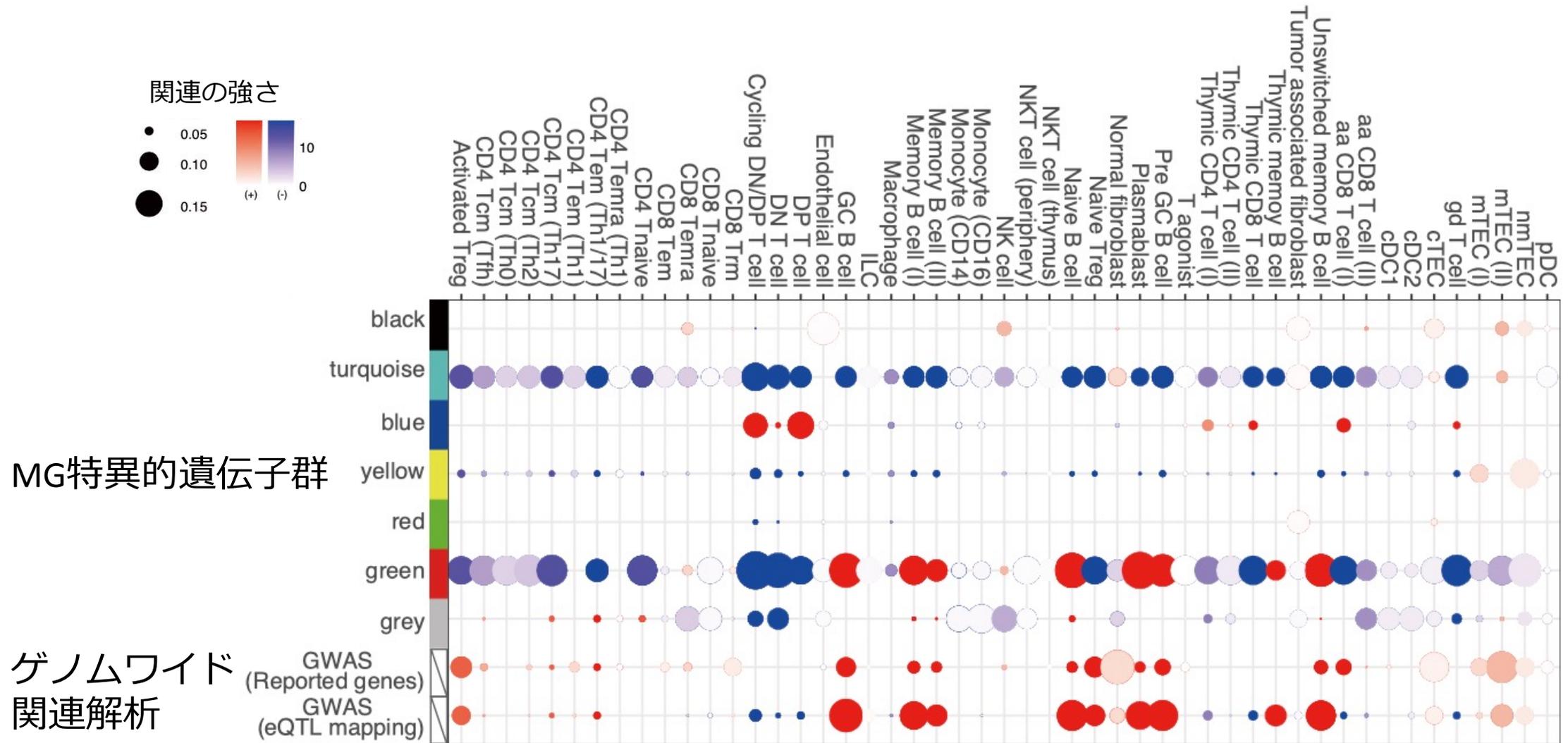
116例胸腺腫bulk RNAseqのデコンボリューションにより、MGにおけるnmTEC, GC B cell, cDC2の集積が明らかに



※ デコンボリューション:細胞頻度予測

©CC-BY Nature communications, Yasumizu *et al.*

# 全細胞種におけるMG特異的発現変化および遺伝要因集積

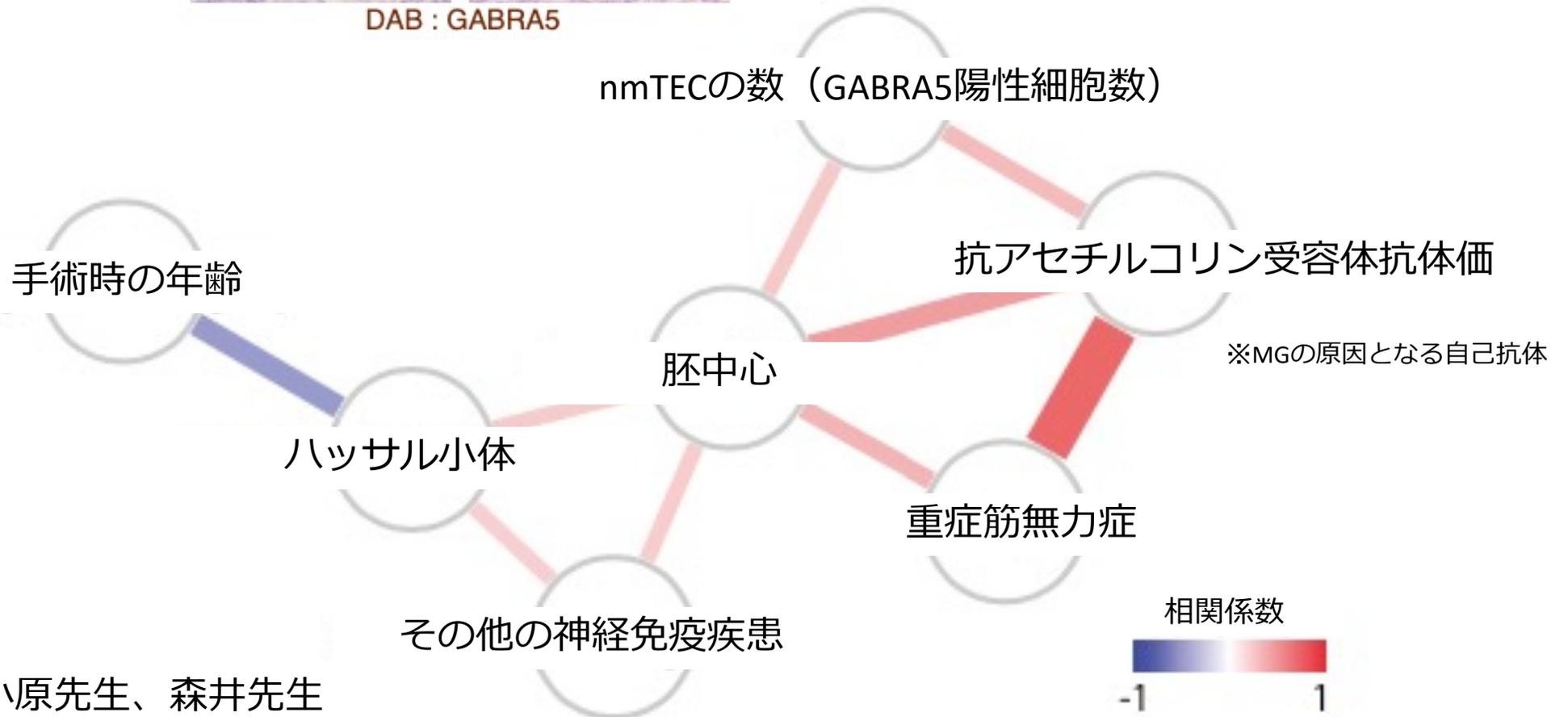
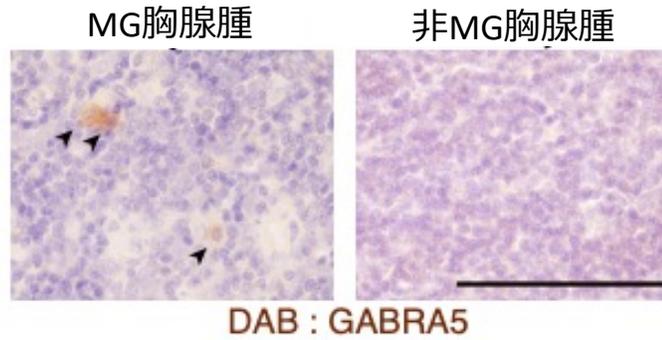


nmTECにMG特異的遺伝子群が集積していることがわかった。  
TregやB細胞に遺伝的要因が集積していた。

※ゲノムワイド関連解析  
全ゲノム上で疾患・形質と関連する  
座位を網羅的に同定する手法。

# 組織学的にもnmTECおよび胚中心とMG発症の関連があった

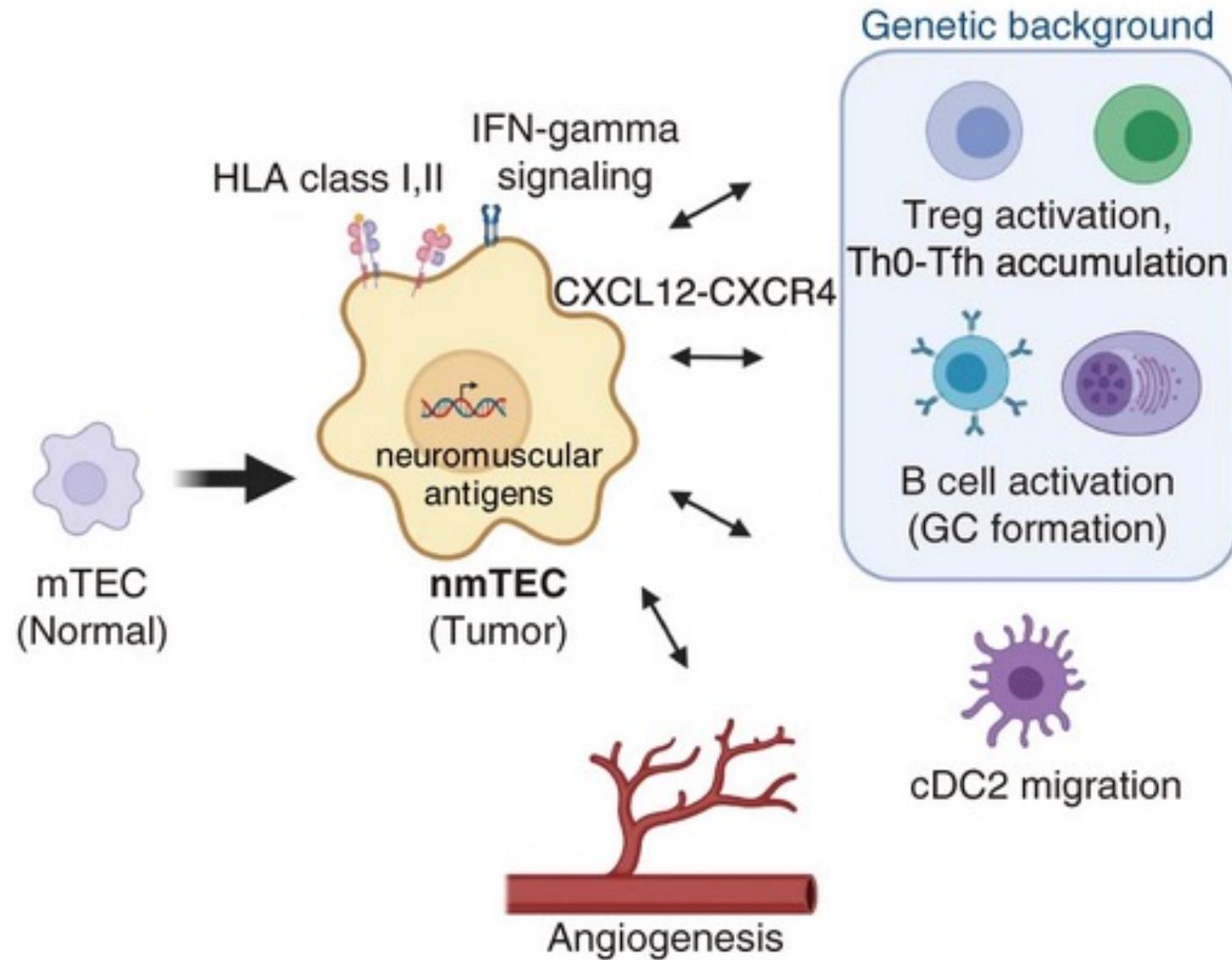
63 samples  
(WHO Type AB-B3)



共同研究  
病態病理学 野島先生、小原先生、森井先生

# 今回明らかにした病態

k



# おしまい

- バルクRNAseqとは？
- バルクRNAseq解析の実際
- ブラウザを用いたバルクRNAseq解析
  - iDEP(本日のメイン)
  - BioJupies
  - RaNA-seq
- 実例紹介（胸腺腫合併重症筋無力症の解析）

今日の話が皆様の研究の一助となれば幸いです。