

CellEnrich Manual

2024.05.16 @hainct

Introduction

Testing environment:

- 1.Ubuntu 22.04.2 LTS
- 2.R version 4.3.0
- 3.RStudio 2023.03.0+386
- 4.Windows 10 + Rtools 4.3

NOTE: on a fresh installation, users may need to install some required interpreter compilers for the system to install other R packages:

- C++ compiler
- gfortran compiler (FYI: tips for [MAC](#) users or [other OS](#))
- Seurat >= version 5.0.0 is REQUIRED

Install Dependent Packages:

- Also, some package needs **install from sources which needs compilation**, and it needs proper version of [Rtools](#) installed.
- It's recommended to install **devtools**, **BiocManager** packages first before installing the following packages in Github/Bioconductor (not in CRAN).

```
library(devtools)
```

```
library(BiocManager)
```

Bioconductor Packages:

- 1.SingleCellExperiment
- 2.Scran

```
BiocManager::install('SingleCellExperiment')
```

```
BiocManager::install('scrn')
```

Other Packages:

1. Seurat
2. farver
3. ggbiplot
4. waiter

```
install.packages('Seurat') # version >= 5.0.0 is REQUIRED!  
install.packages('farver')  
devtools::install_github('JohnCoene/waiter')  
devtools::install_github('vqv/ggbiplot')
```

After installing these packages, we can install `CellEnrich` now.

Running this script will install most of the other dependent packages.

```
install_github('noobCoding/CellEnrich')  
library(CellEnrich)
```

Download Data

`CellEnrich` provides example data and genesets.

Using this function, users can download them in the current working directory.

```
## For downloading human PBMC data & genesets  
CellEnrich::DownloadData(type='test')
```

```
## Use type='all' for all available data
```

Here is a list of data.

- Human Gene-sets
 - Human_Reactome.RData
 - Human_WikiPathways.RData
 - Human_GOBP.RData
 - Human_GOCC.RData
 - Human_GOMF.RData
- Mouse Gene-sets
 - Mouse_Reactome.RData
 - Mouse_WikiPathways.RData
 - Mouse_GOBP.RData
 - Mouse_GOCC.RData
 - Mouse_GOMF.RData
- Human Data
 - pbmcClustInfo.RData ([Seurat PBMC guide](#))
 - pbmcData.RData
 - [DuoClustering\(link\)](#)
 - koh.RData
 - kohInfo.RData
- Mouse Data
 - [Seurat MCA guide](#)

Tutorial with PBMC 3k data

This tutorial will use PBMC analysis data from [Seurat](#).

To get a complete description of the experiment, refer [here](#)

We assume that all dependent packages are already installed. Please refer to the **Install CellEnrich** session for details.

Required Data

Pre-processed data is included in GitHub [repository](#).

Filename	Parameter	Type
pbmcData.RData	CountData	dgCMatrx
pbmcClustInfo.RData	GroupInfo	Character

```
# Download data, if not downloaded
CellEnrich::DownloadData()
```

```
load("pbmcData.RData")
load("pbmcClustInfo.RData")
```

```
CountData <- pbmcData
GroupInfo <- pbmcClustInfo
```

```
# CellEnrich use normalized count as input for user flexible usage
CountData <- Normalized(CountData)
```

```
# This will open CellEnrich browser rShiny interface
CellEnrich(CountData, GroupInfo)
```

Result

Set options before starting CellEnrich.

Possible options are :

- Strategy for Differential Expressed Genes (Median, FGSEA)
- Plot Options
- Gene-sets and their size cutoffs to be used in the analysis.

After a few minutes, the analysis result will appear.

Test environment used:

- OS : Ubuntu 22.04 LTS
- CPU: iAMD® Ryzen 9 5950x 16-core processor × 32
- RAM: 64GB

CellEnrich consists of **five** modules.

1. Parameter Selection

CellEnrich

Options

MEDIAN

N-th Quartiles
3Q

Fold Change (max=4) from N-th Quartiles
1.5

N(%) of top-depth cells
100

FGSEA

Scatter Plot

PCA

TSNE

UMAP

N principal components
50

Genesets

Human-Reactome

Human-WikiPathway

Human-KEGG

Human-GOBP

Human-GOCC

Human-GOMF

Mouse-Reactome

Mouse-WikiPathway

Mouse-KEGG

Mouse-GOBP

Mouse-GOCC

Mouse-GOMF

User defined geneset
BROWSE...

RData format is required!

Geneset Options

Minimum Geneset Size
15

Maximum Geneset Size
500

Pathway Frequency
0.3

Q-value threshold
0.05

RUN

2. Scatter Area

- The top-left plot is a scatter plot created with PCA, TSNE or UMAP and ggplot2 that the user selected in the option.
- The top-right plot is a histogram plot created with high charter to see a distribution of Group / Cell labels.
- The bottom-left plot shows the significance level (darker is more significant) of enriched pathways dominant in each cell of a group.
- The bottom-right plot shows the significance level of enriched pathways in the whole data.



- Users can use the designed buttons to change the enriched criteria:
 1. The ODDS RATIO button will colorize the enriched cell in each group with the **highest** odds ratio.
 2. The FREQUENCY button will colorize frequently enriched cells in each group.
- Users also can use the designed buttons to download analyzed figures and results.

3. Pathways Module

This pathway module will show significant pathways for each group based on the Frequency (of pathways in a group), Size, OddsRatio and Q-value (of pathways) in table format.

For each group, **only 1 pathway** can be selected to show in the scatter plot at a time.

The corresponding **gene activity map** of the selected pathway will be shown accordingly.

B

CD14+ MONO

CD8 T

DC

FCGR3A+ MONO

MEMORY CD4 T

NAIVE CD4 T

NK

PLATELET

B

Pathway	Frequency	Size	-log10 Qvalue	OddsRatio
Generation of second messenger molecules	56 /58	30	15.3525	27.5789
Translocation of ZAP-70 to Immunological synapse	56 /58	16	11.5385	20
PD-1 signaling	56 /58	19	11.3728	19.6903
Phosphorylation of CD3 and TCR zeta chains	56 /58	19	11.2084	19.3846
Downstream TCR signaling	57 /58	91	3.4017	10.4798
Antigen activates B Cell Receptor (BCR) leading to generation of second messengers	9 /58	28	6.6602	7.898
Antigen Presentation	56 /58	28	3.8793	7.1164
TCR signaling	56 /58	112	3.3989	6.3814
Costimulation by the CD28 family	39 /58	60	10.7308	5.4218
MHC class II antigen presentation	48 /58	99	7.8847	5.1

Previous

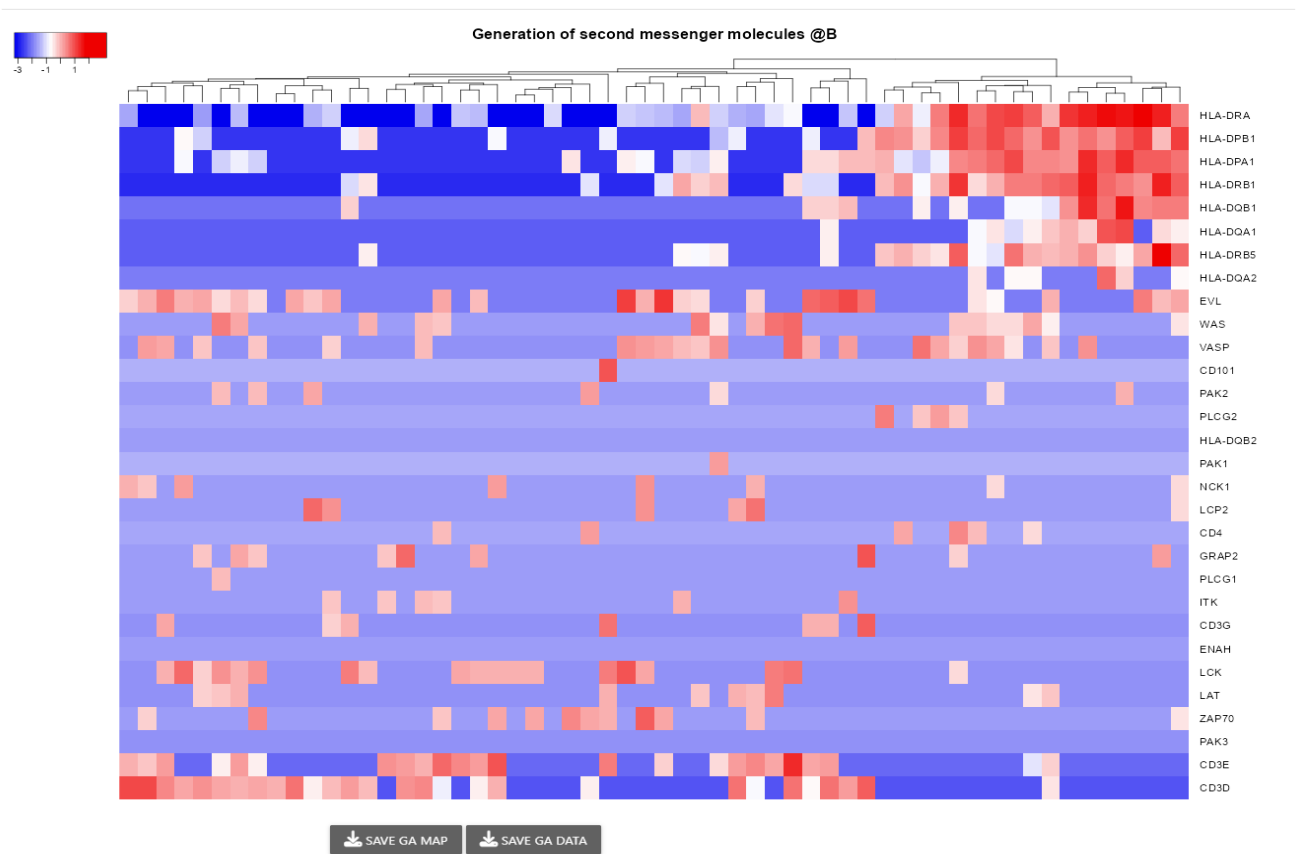
1

2

3

Next

Gene Activity (GA) Map



User chosen Pathways

Pathway	Group
Cell-Cell Communication R-HSA-1500931	Platelet
Respiratory Electron Transport R-HSA-611105	FCGR3A+ Mono
rRNA Modification In Nucleus And Cytosol R-HSA-6790901	Naive CD4 T
Activation Of BAD And Translocation To Mitochondria R-HSA-111447	DC

Previous1Next



















PLOT THE SELECTED PATHWAYS

CLEAR LIST

NOTE: please switch any element’s position to activate the plot button

After clicking the ‘PLOT THE SELECTED PATHWAYS’ button, the top cell enriched by the selected pathways will be shown in the Scatter & Bar session.

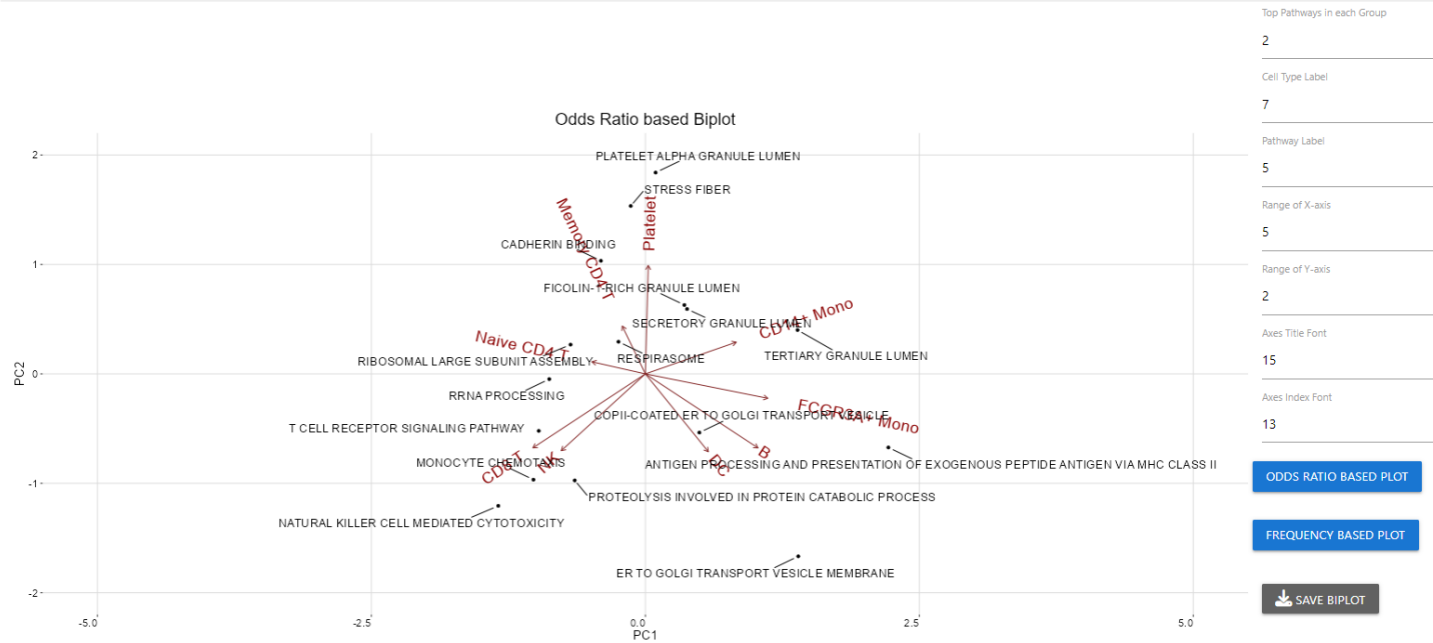
The corresponding legend will also be displayed. And it can be downloaded separately.

Scale	Pathway	Group
High  Low	 B CELL ACTIVATION	B
High  Low	 POSITIVE REGULATION OF INFLAMMATORY RESPONSE	CD14+ Mono
High  Low	 NATURAL KILLER CELL MEDIATED CYTOTOXICITY	CD8 T
High  Low	 COPII-COATED ER TO GOLGI TRANSPORT VESICLE	DC
High  Low	 SPECIFIC GRANULE MEMBRANE	FCGR3A+ Mono
High  Low	 T CELL ACTIVATION	Memory CD4 T
High  Low	 CELLULAR RESPONSE TO INTERLEUKIN-4	Naive CD4 T
High  Low	 MONOCYTE CHEMOTAXIS	NK
High  Low	 PLATELET FORMATION	Platelet

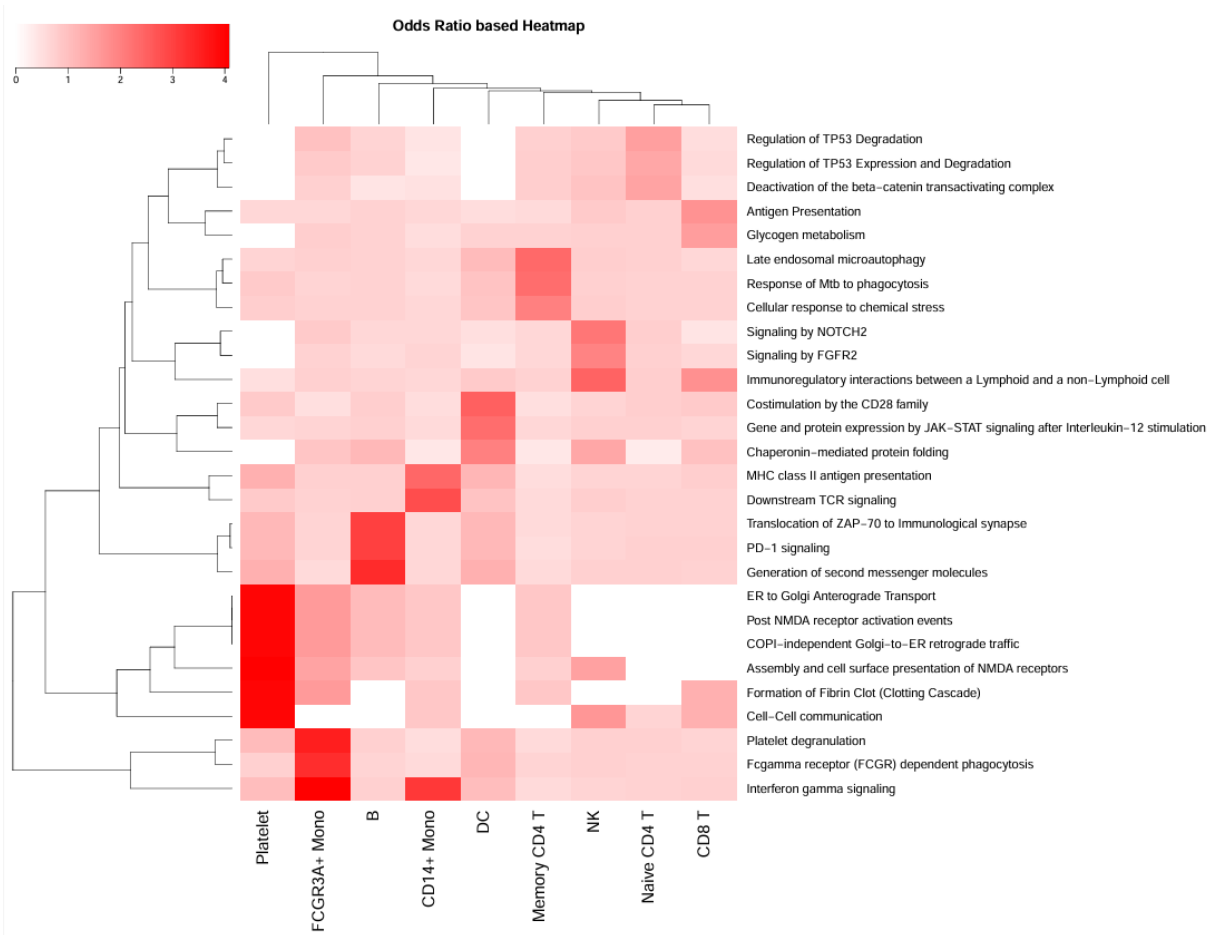
4. BiPlot & HeatMap Module

This section shows the BiPlot presentation of the correlation between categories and top significant pathways.

Biplot between top pathways and groups



Users can check the BiPlot under the perspective of **FREQUENCY** or **ODDS RATIO**. This will also update the corresponding HeatMap.



5. Marker Module

MarkerGenes

DE from each Cell specific

Group	Top	genes	FDR					
All	All	All	All					
1	1	CD3D	0					
1	1	RPS27A	0					
1	1	RPS27	0					
1	3	LTB	0					
1	5	RPL9	0					
1	2	RPL21	0					
1	3	RPS3A	0					
1	2	LDHB	0					
1	4	RPS25	0					
1	2	CD3E	0					
Previous	1	2	3	4	5	...	36	Next

DE - Pathway from each Cell specific

genes	Count	Group	Top	FDR				
All	All	All	All	All				
ACTN1	1	0	4	0				
CD3D	2	0	1	0				
CD3E	2	0	2	0				
CD3G	1	0	7	0				
CD8B	3	0	3	0				
IL7R	1	0	5	0				
LCK	2	0	5	0				
LDHB	1	0	2	0				
LEF1	1	0	8	0				
RPL10	1	0	9	0				
Previous	1	2	3	4	5	...	28	Next

This marker module will show Differentially Expressed genes in the following:

- 1. Each group (using findMarker in scrn)
- 2. Each group and pathway-specific (using Fisher’s exact test)

Developers

- Hai Nguyen (hainct@unist.ac.kr)
Author, maintainer
- Jinhwan Kim
Author
- Dougu Nam (dougnam@unist.ac.kr)
Author, project manager

License

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