

# 5-GSEA

BAI Qiang\*

2021-09-23 15:07:14 +0200

## Contents

<b>1</b>	<b>Description</b>	<b>2</b>
<b>2</b>	<b>Preparation of data for GSEA</b>	<b>2</b>
<b>3</b>	<b>Compare cluster 1 to cluster 0 with gene_sets in Molecular Signatures Database (MSigDB)</b>	<b>2</b>
3.1	Cluster1vsCluster0_GO_MF . . . . .	2
3.2	Cluster1vsCluster0_GO_CC . . . . .	2
3.3	Cluster1vsCluster0_GO_BP . . . . .	3
3.4	Cluster1vsCluster0_Tft_GTRD . . . . .	3
3.5	Cluster1vsCluster0_Tft . . . . .	3
3.6	Cluster1vsCluster0_WikiPathways . . . . .	3
3.7	Cluster1vsCluster0_REACTOME . . . . .	4
3.8	Cluster1vsCluster0_KEGG . . . . .	4
3.9	Cluster1vsCluster0_BioCarta . . . . .	4
<b>4</b>	<b>Results</b>	<b>4</b>
<b>5</b>	<b>Session information</b>	<b>5</b>
	<b>References</b>	<b>5</b>

---

\*University Liege, mail qiang.bai@uliege.be

# 1 Description

For Gene Set Enrichment Analysis (GSEA)[1], the normalized data shown in Figure 4A were used as expression dataset and clustering information was used as phenotype data to test the AFlo and AFhi AM signatures obtained above from bulk RNA-seq data and genesets available in MSigDB database. Normalized Enrichment Score (NES), FDR and nominal p-value were calculated with 300 permutations (100 permutations for testing MSigDB genesets) between samples from different phenotypes.

NOTICE: the intermediate files used as input for GSEA were not synchronized on the platform due to the size limit. However, these files could be generated in the following steps.

## 2 Preparation of data for GSEA

The following files were used as input for GSEA analyses:

AssayData\_RNA\_data\_all\_monocytes\_all\_sample.txt - Exported expression matrix from GetAssayData(object, slot = "data", assay = "RNA") of Seurat package[2].

Class\_all\_monocytes\_all\_sample\_with\_header.cls - Class file generated from metadata of SeuratObject@meta.data\$seurat\_clusters

geneRef.chip - Gene reference file generated from CellRanger output outs/raw\_feature\_bc\_matrix/features.tsv.gz

<geneset>.gmt - Geneset files download from ftp.broadinstitute.org

## 3 Compare cluster 1 to cluster 0 with gene\_sets in Molecular Signatures Database (MSigDB)

The following codes were used to launch each of GSEA analyses:

### 3.1 Cluster1vsCluster0\_GO\_MF

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt - 1
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmx ftp
.broadinstitute.org://pub/gsea/gene_sets/c5.go.mf.v7.4.symbols.gmt -
collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -permute
phenotype -rnd_type no_balance -scoring_scheme weighted -rpt_label
Cluster1vsCluster0_GO_MF -metric Signal2Noise -sort real -order
descending -chip /geneRef.chip -create_gcts false -create_svgs false -
include_only_symbols true -make_sets true -median false -num 100 -
plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500 -
set_min 15 -zip_report false -out /GSEA_results
```

### 3.2 Cluster1vsCluster0\_GO\_CC

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt - 1
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmx ftp
.broadinstitute.org://pub/gsea/gene_sets/c5.go.cc.v7.4.symbols.gmt -
collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -permute
phenotype -rnd_type no_balance -scoring_scheme weighted -rpt_label
Cluster1vsCluster0_GO_CC -metric Signal2Noise -sort real -order
descending -chip /geneRef.chip -create_gcts false -create_svgs false -
include_only_symbols true -make_sets true -median false -num 100 -
```

```
plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500 -
set_min 15 -zip_report false -out /GSEA_results
```

### 3.3 Cluster1vsCluster0\_GO\_BP

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt -
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmxf ftp
.broadinstitute.org://pub/gsea/gene_sets/c5.go.bp.v7.4.symbols.gmt -
collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -permute
phenotype -rnd_type no_balance -scoring_scheme weighted -rpt_label
Cluster1vsCluster0_GO_BP -metric Signal2Noise -sort real -order
descending -chip /geneRef.chip -create_gcts false -create_svgs false -
include_only_symbols true -make_sets true -median false -num 100 -
plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500 -
set_min 15 -zip_report false -out /GSEA_results
```

### 3.4 Cluster1vsCluster0\_Tft\_GTRD

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt -
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmxf ftp
.broadinstitute.org://pub/gsea/gene_sets/c3.tft.gtrd.v7.4.symbols.gmt -
collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -permute
phenotype -rnd_type no_balance -scoring_scheme weighted -rpt_label
Cluster1vsCluster0_Tft_GTRD -metric Signal2Noise -sort real -order
descending -chip /geneRef.chip -create_gcts false -create_svgs false -
include_only_symbols true -make_sets true -median false -num 100 -
plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500 -
set_min 15 -zip_report false -out /GSEA_results
```

### 3.5 Cluster1vsCluster0\_Tft

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt -
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmxf ftp
.broadinstitute.org://pub/gsea/gene_sets/c3.tft.v7.4.symbols.gmt -
collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -permute
phenotype -rnd_type no_balance -scoring_scheme weighted -rpt_label
Cluster1vsCluster0_Tft -metric Signal2Noise -sort real -order
descending -chip /geneRef.chip -create_gcts false -create_svgs false -
include_only_symbols true -make_sets true -median false -num 100 -
plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500 -
set_min 15 -zip_report false -out /GSEA_results
```

### 3.6 Cluster1vsCluster0\_WikiPathways

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt -
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmxf ftp
.broadinstitute.org://pub/gsea/gene_sets/c2.cp.wikipathways.v7.4.
symbols.gmt -collapse Collapse -mode Max_probe -norm meandiv -nperm 100
-permute phenotype -rnd_type no_balance -scoring_scheme weighted -
rpt_label Cluster1vsCluster0_WikiPathways -metric Signal2Noise -sort
real -order descending -chip /geneRef.chip -create_gcts false -
```

```
create_svgs false -include_only_symbols true -make_sets true -median
false -num 100 -plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false
-set_max 500 -set_min 15 -zip_report false -out /GSEA_results
```

### 3.7 Cluster1vsCluster0\_REACTOME

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt -
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmxf ftp
.broadinstitute.org://pub/gsea/gene_sets/c2.cp.reactome.v7.4.symbols.
gmt -collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -
permute phenotype -rnd_type no_balance -scoring_scheme weighted -
rpt_label Cluster1vsCluster0_REACTOME -metric Signal2Noise -sort real -
order descending -chip /geneRef.chip -create_gcts false -create_svgs
false -include_only_symbols true -make_sets true -median false -num 100
-plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500
-set_min 15 -zip_report false -out /GSEA_results
```

### 3.8 Cluster1vsCluster0\_KEGG

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt -
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmxf ftp
.broadinstitute.org://pub/gsea/gene_sets/c2.cp.kegg.v7.4.symbols.gmt -
collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -permute
phenotype -rnd_type no_balance -scoring_scheme weighted -rpt_label
Cluster1vsCluster0_KEGG -metric Signal2Noise -sort real -order
descending -chip /geneRef.chip -create_gcts false -create_svgs false -
include_only_symbols true -make_sets true -median false -num 100 -
plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500 -
set_min 15 -zip_report false -out /GSEA_results
```

### 3.9 Cluster1vsCluster0\_BioCarta

```
gsea-cli.sh GSEA -res /AssayData_RNA_data_all_monocytes_all_sample.txt -
cls /Class_all_monocytes_all_sample_with_header.cls#1_versus_0 -gmxf ftp
.broadinstitute.org://pub/gsea/gene_sets/c2.cp.biocarta.v7.4.symbols.
gmt -collapse Collapse -mode Max_probe -norm meandiv -nperm 100 -
permute phenotype -rnd_type no_balance -scoring_scheme weighted -
rpt_label Cluster1vsCluster0_BioCarta -metric Signal2Noise -sort real -
order descending -chip /geneRef.chip -create_gcts false -create_svgs
false -include_only_symbols true -make_sets true -median false -num 100
-plot_top_x 20 -rnd_seed timestamp -save_rnd_lists false -set_max 500
-set_min 15 -zip_report false -out /GSEA_results
```

## 4 Results

All GSEA results were not presented in the manuscript but can be found via the link:

[https://github.com/BlanQwall/Human\\_monocytes\\_and\\_AM/tree/main/scRNAseq\\_analysis/5-GSEA/GESA\\_results](https://github.com/BlanQwall/Human_monocytes_and_AM/tree/main/scRNAseq_analysis/5-GSEA/GESA_results)

## 5 Session information

Java version

```
java -version
```

```
## openjdk version "11.0.11" 2021-04-20
## OpenJDK Runtime Environment (build 11.0.11+9-Ubuntu-0ubuntu2.20.04)
## OpenJDK 64-Bit Server VM (build 11.0.11+9-Ubuntu-0ubuntu2.20.04, mixed
mode, sharing)
```

GSEA version

```
GSEA v4.1.0 [build: 27]
```

```
uname -a
```

```
## Linux uliege-TUF-X299-MARK-2 5.4.0-84-generic #94-Ubuntu SMP Thu Aug 26
20:27:37 UTC 2021 x86_64 x86_64 x86_64 GNU/Linux
```

## References

1. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 2005;
2. Hao Y, Hao S, Andersen-Nissen E, Mauck III WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zagar M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LB, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. *Cell* [Internet] 2021; Available from: <https://doi.org/10.1016/j.cell.2021.04.048>.