

# 5-GSEA

BAI Qiang\*

2021-10-13 14:04:25 +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 2 to REST with gene_sets in Molecular Signatures Database (MSigDB)</b>	<b>2</b>
<b>4</b>	<b>Results</b>	<b>2</b>
<b>5</b>	<b>Session information</b>	<b>2</b>
	<b>References</b>	<b>3</b>

---

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

# 1 Description

For Gene Set Enrichment Analysis (GSEA)[1], the integrated data were used as expression dataset and clustering information was used as phenotype data to test the cluster 2 and rest of VMT or cluster 0. Normalized Enrichment Score (NES), FDR and nominal p-value were calculated with 100 permutations for testing MSigDB geneset) 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\_integrated\_data\_Zoom\_memory\_T.txt - Exported expression matrix from GetAssayData(object, slot = "data", assay = "RNA") of Seurat package[2].

Class\_Zoom\_memory\_T\_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 2 to REST with gene\_sets in Molecular Signatures Database (MSigDB)

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

For a GSEA analysis with \${geneset}.gmt (downloaded from \${databasename} on the Broad Institut website. we use the following chunk:

```
gsea-cli.sh GSEA -res AssayData_integrated_data_Zoom_memory_T.txt -cls
Class_Zoom_memory_T_with_header.cls#2_versus_REST -gmx ftp.
broadinstitute.org://pub/gsea/gene_sets/${geneset}.gmt -collapse
Collapse -mode Max_probe -norm meandiv -nperm 100 -permute phenotype -
rnd_type no_balance -scoring_scheme weighted -rpt_label
ZoomVT_Cluster2_vs_REST_${databasename} -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/VM\\_T/tree/main/4-GSEA%20analysis/GSEA\\_results](https://github.com/BlanQwall/VM_T/tree/main/4-GSEA%20analysis/GSEA_results)

## 5 Session information

Java version

```
java -version
```

```
## openjdk version "11.0.11" 2021-04-20 1
## OpenJDK Runtime Environment (build 11.0.11+9-Ubuntu-0ubuntu2.20.04) 2
## OpenJDK 64-Bit Server VM (build 11.0.11+9-Ubuntu-0ubuntu2.20.04, mixed 3
mode, sharing)
```

GSEA version

```
GSEA v4.1.0 [build: 27] 1
```

```
uname -a 1
```

```
## Linux uliege-TUF-X299-MARK-2 5.4.0-88-generic #99-Ubuntu SMP Thu Sep 23 1
17:29:00 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>.