



Functional enrichment data analysis

Yan Li, Ph.D.

Bioinformatics Core

CRI, University of Chicago

November 19th, 2025

➤ Prerequisite Preparation

- xQuartz installation (mac user)
- access to randi
- initiate rstudio for last hands-on section

➤ Input test data (e-mail): download & unzip

macOS

1. Install XQuartz

- Download and install from <https://www.xquartz.org/>.
- Log out and back in, or restart your computer.

2. Enable X11 Forwarding in SSH

- Use the `-Y` option when connecting to `randi` to enable trusted X11 forwarding.

Windows

1. Install an X11 Server

- Recommended: VcXsrv ([link](#)) or Xming ([link](#)).
- Start the X11 server **before** opening your SSH session.

2. Use an SSH Client with X11 Forwarding

- Recommended: [PuTTY](#)
- Settings:
 - `Connection > SSH > X11`
 - Check `Enable X11 forwarding`
 - `X display location : localhost:0` (default)

Outline

- Overview of Functional Enrichment Analysis Concepts
 - over-representation analysis (ORA)
 - gene set enrichment analysis (GSEA)

Hands-on: Using the Functional Enrichment Analysis Application

- https://biocoreapps.bsd.uchicago.edu/gsea_shiny/
Gene Sets Over Representation (GSOR) Analysis
- <https://biocoreapps.bsd.uchicago.edu/feavisapp/>
enrichment analysis results visualization application

Hands-on: Implementing GSOR via rstudio on Randi - Single gene list - Multiple gene lists

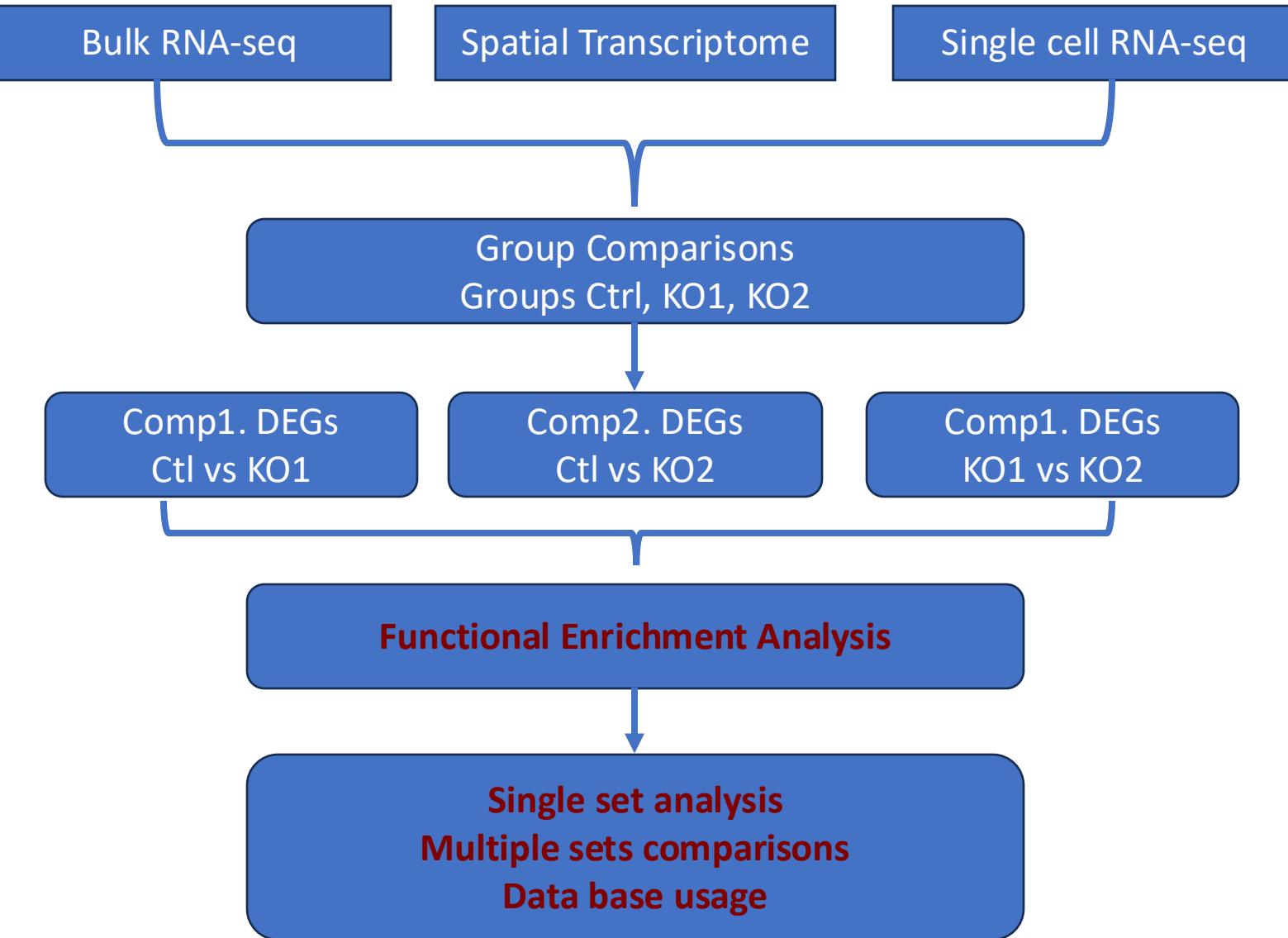
Bioinformatics Workshop Series 2025

- **Target: 1-2 workshops each quarter semester**
- Winter Quarter
 - ❖ February: Bulk RNA-seq Pipeline Hands-on Training
- Spring Quarter
 - ❖ April: Introduction to Spatial Transcriptomics
 - ❖ June: 10x Genomics scRNA-seq data analysis workshop
- Summer Break
- Autumn Quarter
 - ❖ October: Bulk RNA-seq Pipeline Hands-on Training
 - ❖ November: Functional Enrichment Analysis workshop
- workshop github: <https://github.com/CRI-Biocore/>

Bioinformatics Workshop Series 2025

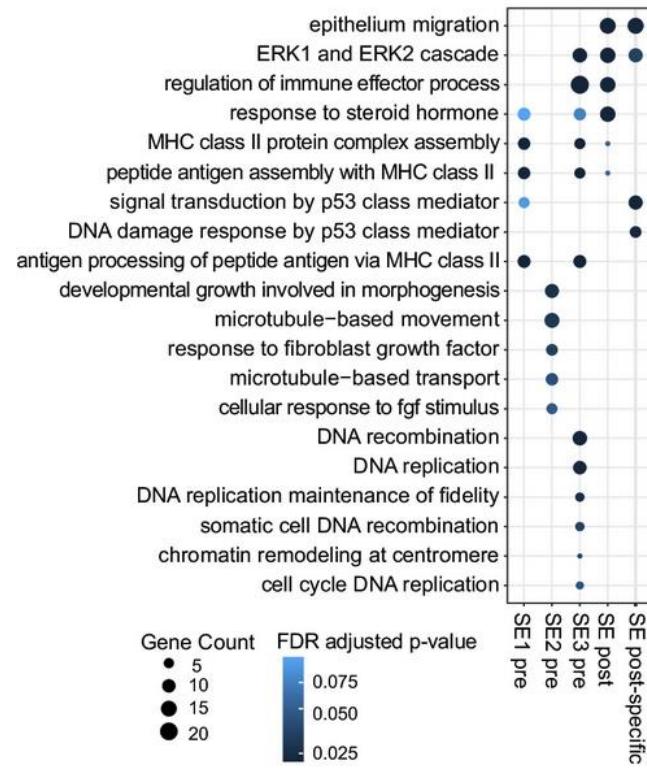
- **Target: 1-2 workshops each quarter semester**
- Winter Quarter
 - ❖ February: Bulk RNA-seq Pipeline Hands-on Training
- Spring Quarter
 - ❖ April: Introduction to Spatial Transcriptomics
 - ❖ June: 10x Genomics scRNA-seq data analysis workshop
- Summer Break
- Autumn Quarter
 - ❖ October: Bulk RNA-seq Pipeline Hands-on Training
 - ❖ **November: Functional Enrichment Analysis workshop (Today)**
- workshop github: <https://github.com/CRI-Biocore/>
- feedback: <https://mycri.cri.uchicago.edu/educations/trainings/79/survey/>

Why Functional Enrichment Analysis?

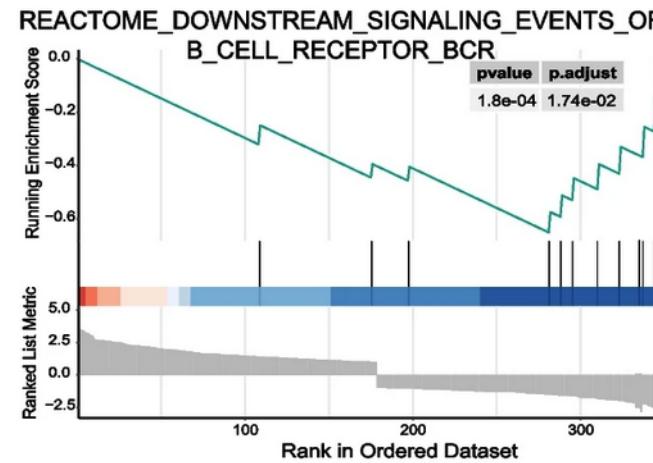
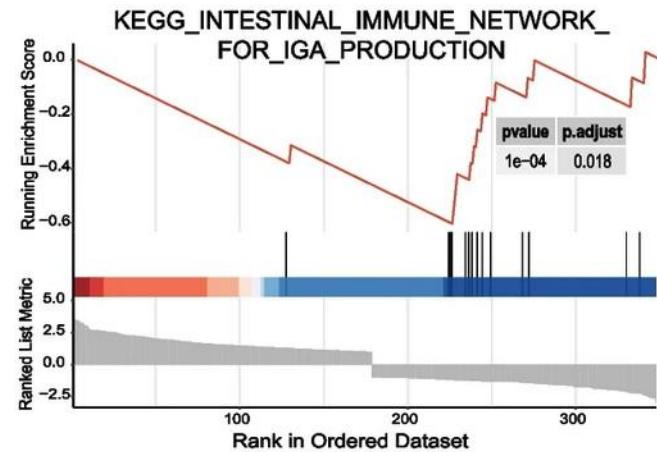


Functional Enrichment Analysis Overview

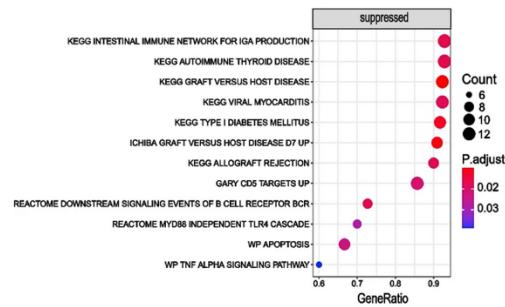
A. Gene Sets Over Representation (GSOR)



B. Gene Sets Enrichment Analysis (GSEA)



C.



Pics. source

- A). Nature communication, Weigert M., Li Y, et al. (2025) at <https://www.nature.com/articles/s41467-024-55440-2>
- B)& C). BMC, Zhang L et al. (2022) at <https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-022-01375-w>

Gene Sets Over Representation (GSOR)

- Hypergeometric Test
- P-value is calculated by:

$$p = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

N: total number of genes in the background distribution
M: the number of genes within that distribution that are annotated to the gene set of interest
n: the size of the list of genes of interest
K: the number of genes within that list which are annotated to the gene set

- For each functions in your functional database, you are checking whether your input gene list is over-represented or not?

| | Gene no interest | Genes in interest | Row sum |
|--------------------------|------------------|-----------------------------|--|
| In function category | M-k | k | M (known functional genes) |
| Not in function category | N-M-(n-k) | n-k | N-M |
| Col sum | N-n | n (input) | N |

- This small p-value indicates the observed enrichment is unlikely to have occurred by random chance

Gene Sets Over Representation (GSOR) cont.



| | Gene no interest | Genes in interest | Row sum |
|--------------------------|------------------|--------------------|-------------------------------------|
| In function category | M-k | k | M (known functional genes) |
| Not in function category | N-M-(n-k) | n-k | N-M |
| Col sum | N-n | n (input) | N |

N: total number of genes in the background distribution

M: the number of genes within that distribution that are annotated to the gene set of interest

n: the size of the list of genes of interest

K: the number of genes within that list which are annotated to the gene set

➤ As n increases, do we have higher chance to find more enriched functions to be over-represented?

- ❖ Shall we use the balanced/same number of gene sets as input for comparison?
- ❖ DEGs with larger and small difference?
- ❖ Any test does not rely on the size of input gene list?



Gene Sets Enrichment Analysis (GSEA)

- Utilize the entire gene sets
- ES (enrichment score) is calculated for each function category
- The set of genes in a function is ranked based on the ES trend
- A set of leading genes are positively (top list of genes) or negatively (bottom list of genes) correlated with the function of interest
- p-value of ES is based on permutation test

Gene Sets Enrichment Analysis (GSEA)

- Utilize the entire gene sets
- ES (enrichment score) is calculated for each function category
- The set of genes in a function is ranked based on the ES trend
- A set of leading genes are positively (top list of genes) or negatively (bottom list of genes) correlated with the function of interest
- p-value of ES is based on permutation test
- **Multiple test correction (FDR correction)**



Gene Sets Enrichment Analysis (GSEA)

Key differences

| Feature | GSEA (Gene Set Enrichment Analysis) | ORA (Over-Representation Analysis) |
|-------------------------------|--|--|
| Input Data | Uses the full ranked list of all genes based on their differential expression, without thresholds. | Requires a pre-selected list of significantly up- or down-regulated genes, often based on fold-change and p-value cutoffs. |
| Gene Selection | Includes all genes in the analysis, making it sensitive to subtle, coordinated changes across a gene set. | Relies on arbitrary thresholds, potentially missing important genes that don't meet the criteria. |
| Pathway Analysis | Determines if a gene set is enriched at the top (upregulated) or bottom (downregulated) of the ranked list, accounting for both. | Tests for the statistical over-representation of a gene set among the "significant" genes. |
| Biological Information | Captures continuous expression shifts and correlations within a gene set. | Ignores the degree of differential expression for individual genes once they are included in the "significant" list. |
| Typical Use Case | Best when there is no clear distinction between "significant" and "non-significant" genes or when the differential expression is subtle but coordinated. | Useful when only a small number of highly significant genes are identified, or for a quick analysis. |



Functional Enrichment Analysis Databases

- Gene Ontology DB (GOs)
 - <https://geneontology.org/docs/ontology-documentation/>
 - ❖ Biological Processing (BP)
 - ❖ Cellular Component (CC)
 - ❖ Molecular Function (MF)
- KEGG DB (KEGG)
 - ❖ <https://www.genome.jp/kegg/pathway.html>
- reactome
 - ❖ <https://reactome.org/>
- Broad MSigDB (molecular signature databases)
 - ❖ <https://www.gsea-msigdb.org/gsea/msigdb>



Functional Enrichment Analysis Databases



Molecular Signatures Database

Overview

The Molecular Signatures Database (MSigDB) is a resource of tens of thousands of annotated gene sets for use with GSEA software, divided into **Human** and **Mouse** collections. From this web site, you can

- ▶ **Examine** a gene set and its annotations. See, for example, the [HALLMARK_APOTOPSIS](#) human gene set page.
- ▶ **Browse** gene sets by name or collection.
- ▶ **Search** for gene sets by keyword.
- ▶ **Investigate** gene sets:
 - ▶ **Compute overlaps** between your gene set and gene sets in MSigDB.
 - ▶ **Categorize** members of a gene set by gene families.
 - ▶ **View the expression profile** of a gene set in a provided public expression compendia.
 - ▶ Investigate the gene set in the online **biological network repository NDEX**
- ▶ **Download** gene sets.

License Terms

GSEA and MSigDB are available for use under [these license terms](#).

Please [register](#) to download the GSEA software and the MSigDB gene sets, and to use our web tools. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

Current Version

Human MSigDB v2025.1.Hs updated June 2025. [Release notes](#).

Mouse MSigDB v2025.1.Mm updated June 2025. [Release notes](#).

Citing the MSigDB

To cite your use of the Molecular Signatures Database (MSigDB), a joint project of UC San Diego and Broad Institute, please reference Subramanian, Tamayo, et al. (2005, PNAS) and one or more of the following as appropriate: Liberzon, et al. (2011, Bioinformatics), Liberzon, et al. (2015, Cell Systems), and also the source for the gene set as listed on the gene set page. If you use Mouse MSigDB, please also cite Castanza, et al. (2023, Nature Methods).

Funding

GSEA and MSigDB are currently funded by a grant from NCI's [Informatics Technoloy for Cancer Research \(ITCR\)](#)

Human Collections

H **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

C1 **positional gene sets** corresponding to human chromosome cytogenetic bands.

C2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

C3 **regulatory target gene sets** based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.

C4 **computational gene sets** defined by mining large collections of cancer-oriented expression data.

C5 **ontology gene sets** consist of genes annotated by the same ontology term.

C6 **oncogenic signature gene sets** defined directly from microarray gene expression data from cancer gene perturbations.

C7 **immunologic signature gene sets** represent cell states and perturbations within the immune system.

C8 **cell type signature gene sets** curated from cluster markers identified in single-cell sequencing studies of human tissue.

Mouse Collections

MH **mouse-ortholog hallmark gene sets** are versions of gene sets in the MSigDB Hallmarks collection mapped to their mouse orthologs.

M1 **positional gene sets** corresponding to mouse chromosome cytogenetic bands.

M2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

M3 **regulatory target gene sets** based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.

M5 **ontology gene sets** consist of genes annotated by the same ontology term.

M7 **immunologic signature gene sets** represent cell states and perturbations within the immune system.

M8 **cell type signature gene sets** curated from cluster markers identified in single-cell sequencing studies of mouse tissue.

Functional Enrichment Analysis Reference Genome

- Input gene list
 - ❖ human, mouse, rat, warm et al.
 - ❖ ensemble id, entrez id or gene symbol
 - ❖ gene lists vs ranked gene lists
- Functional category
 - ❖ human specific databases?
- IPA (licensed tool)
 - ❖ functional enrichment analysis, network drawing etc.
- shiny applications
 - ❖ GSOR analysis: https://biocoreapps.bsd.uchicago.edu/gsea_shiny/
 - ❖ GSOR results visualization:
<https://biocoreapps.bsd.uchicago.edu/feavisapp/>

GSOR analysis application

➤ https://biocoreapps.bsd.uchicago.edu/gsea_shiny/

The screenshot shows the 'GSOR Functional' Shiny App interface. On the left, a sidebar menu includes 'Upload Files', 'Input Settings', 'Analysis Options', and a prominent 'Run Analysis' button. The main content area features an 'Introduction' section with a welcome message and a note about its use for biological research. Below it are sections for 'How to Use the App', 'Results key columns description', and 'Analysis Results Download'. The top of the page has a standard browser header with tabs, a search bar, and user authentication links.

GSOR visualization application

➤ <https://biocoreapps.bsd.uchicago.edu/feavisapp/>

The screenshot shows the FeaVis visualization application interface. The left sidebar has two sections: "Step 1: Select Input Type and Folder" and "Step 2: Choose Files for Analysis". The "Step 1" section includes a dropdown for "Select type of results" (set to "GSEA App results"), a text input for "Enter GSEA app token" (with placeholder "https://biocoreapps.bsd.uchicago.edu/gsea_shiny/ result token:"), and a "Load Files" button. The "Step 2" section includes a "Submit Data" button. The main area contains several search and filter components. At the top is a search bar for keywords in the description. Below it is a search panel with fields for "Select one or more keywords to search in the Description" and "Enter one or more keywords (separated by comma) to search in the Description", separated by an "OR" operator. A "QUERY SEARCH" button is located at the bottom right of this panel. Further down are sections for "Gene sets over representation summary" (green background) and "Over represented gene sets table" (green background). At the bottom is a "Gene sets over representation plot" section with a dropdown for "Select Functions to Display" (set to "Top 10 Functions") and a "DOWNLOAD DOTPLOT" button.



Hands-on: GSOR analysis application

➤ https://biocoreapps.bsd.uchicago.edu/gsea_shiny/

The screenshot shows a web browser window for the 'GSOR Functional' app at the URL https://biocoreapps.bsd.uchicago.edu/gsea_shiny/. The interface has a dark sidebar on the left with options: 'Upload Files', 'Input Settings', 'Analysis Options', and a prominent 'Run Analysis' button. The main content area is titled 'Introduction' and contains a welcome message: 'Welcome to the GSOR Functional Enrichment Analysis Shiny App! This app enables users to perform gene set over-representation (GSOR) analysis and visualize the results in an intuitive, interactive interface. Originally designed as a set of R scripts for HPC or local execution, this tool now allows you to upload your gene lists, run functional enrichment analyses across multiple databases (GO, KEGG, Reactome, Broad MSigDB, Network Cancer Genes), and generate publication-ready visualizations—all without writing any code.' Below this is a note: 'The app is ideal for researchers seeking to uncover biological functions, pathways, or gene clusters enriched in their gene sets, with customizable options for selection, clustering, and visualization.' There are three expandable sections: 'How to Use the App', 'Results key columns description', and 'Analysis Results Download', each preceded by a '+' sign.



Hands-on: GSOR analysis application

- Input gene lists download via e-mail or zoom chat
 - ❖ test_data_input.zip

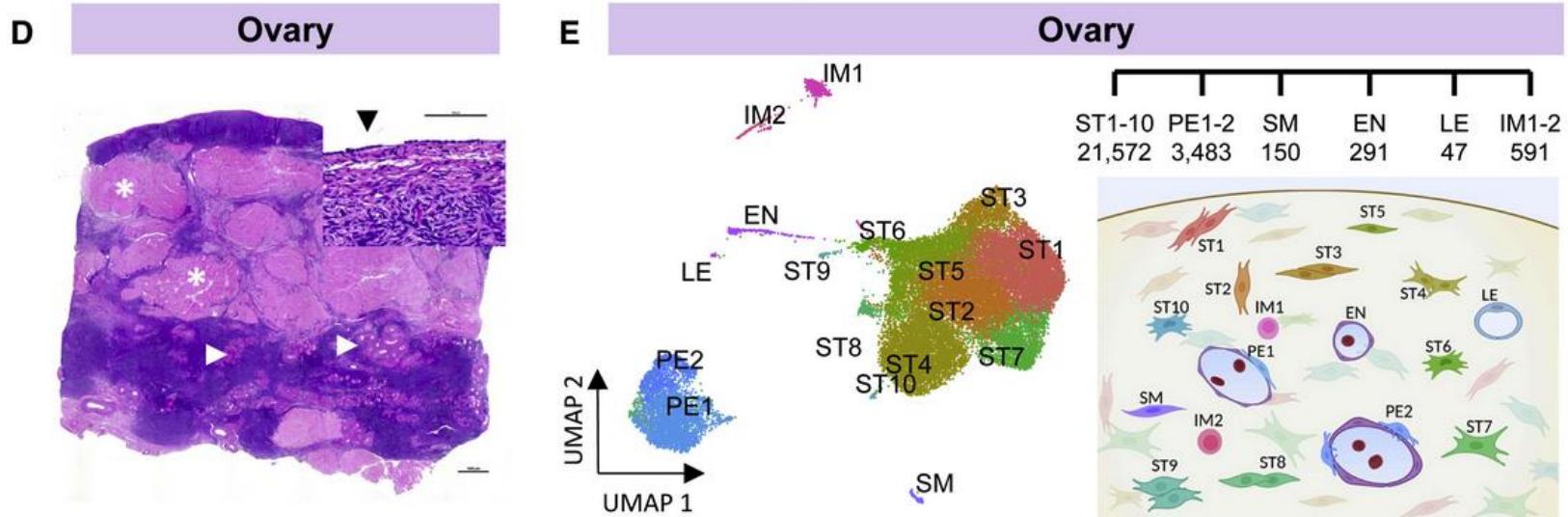
| Name | Date Modified | Size | Kind |
|--------------------|-------------------------|-----------|------------|
| gsorInput_EN.txt | Nov 14, 2025 at 7:44 PM | 5 KB | Plain Text |
| gsorInput_IM1.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_IM2.txt | Nov 14, 2025 at 7:44 PM | 3 KB | Plain Text |
| gsorInput_LE.txt | Nov 14, 2025 at 7:44 PM | 4 KB | Plain Text |
| gsorInput_PE1.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_PE2.txt | Nov 14, 2025 at 7:44 PM | 3 KB | Plain Text |
| gsorInput_SM.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_ST1.txt | Nov 14, 2025 at 7:44 PM | 576 bytes | Plain Text |
| gsorInput_ST2.txt | Nov 14, 2025 at 7:44 PM | 224 bytes | Plain Text |
| gsorInput_ST3.txt | Nov 14, 2025 at 7:44 PM | 338 bytes | Plain Text |
| gsorInput_ST4.txt | Nov 14, 2025 at 7:44 PM | 526 bytes | Plain Text |
| gsorInput_ST5.txt | Nov 14, 2025 at 7:44 PM | 139 bytes | Plain Text |
| gsorInput_ST6.txt | Nov 14, 2025 at 7:44 PM | 321 bytes | Plain Text |
| gsorInput_ST7.txt | Nov 14, 2025 at 7:44 PM | 843 bytes | Plain Text |
| gsorInput_ST8.txt | Nov 14, 2025 at 7:44 PM | 1 KB | Plain Text |
| gsorInput_ST9.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_ST10.txt | Nov 14, 2025 at 7:44 PM | 588 bytes | Plain Text |

- A molecular atlas of the human postmenopausal fallopian tube and ovary from single-cell RNA and ATAC sequencing, Lengyel E, Li Y et al., Cell Reports 2022



Hands-on: GSOR analysis application

➤ ovary scRNA-seq data



< > test_data_input

| Name | Date Modified | Size | Kind |
|--------------------|-------------------------|-----------|------------|
| gsorInput_EN.txt | Nov 14, 2025 at 7:44 PM | 5 KB | Plain Text |
| gsorInput_IM1.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_IM2.txt | Nov 14, 2025 at 7:44 PM | 3 KB | Plain Text |
| gsorInput_LE.txt | Nov 14, 2025 at 7:44 PM | 4 KB | Plain Text |
| gsorInput_PE1.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_PE2.txt | Nov 14, 2025 at 7:44 PM | 3 KB | Plain Text |
| gsorInput_SM.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_ST1.txt | Nov 14, 2025 at 7:44 PM | 576 bytes | Plain Text |
| gsorInput_ST2.txt | Nov 14, 2025 at 7:44 PM | 224 bytes | Plain Text |
| gsorInput_ST3.txt | Nov 14, 2025 at 7:44 PM | 338 bytes | Plain Text |
| gsorInput_ST4.txt | Nov 14, 2025 at 7:44 PM | 526 bytes | Plain Text |
| gsorInput_ST5.txt | Nov 14, 2025 at 7:44 PM | 139 bytes | Plain Text |
| gsorInput_ST6.txt | Nov 14, 2025 at 7:44 PM | 321 bytes | Plain Text |
| gsorInput_ST7.txt | Nov 14, 2025 at 7:44 PM | 843 bytes | Plain Text |
| gsorInput_ST8.txt | Nov 14, 2025 at 7:44 PM | 1 KB | Plain Text |
| gsorInput_ST9.txt | Nov 14, 2025 at 7:44 PM | 2 KB | Plain Text |
| gsorInput_ST10.txt | Nov 14, 2025 at 7:44 PM | 588 bytes | Plain Text |

https://biocoreapps.bsd.uchicago.edu/gsea_shiny/



Hands-on: GSOR analysis application

GSOR Functional

Introduction

Welcome to the GSOR Functional Enrichment Analysis Shiny App! This app enables users to perform gene set over-representation (GSOR) analysis and visualize the results in an intuitive, interactive interface. Originally designed as a set of R scripts for HPC or local execution, this tool now allows you to upload your gene lists, run functional enrichment analyses across multiple databases (GO, KEGG, Reactome, Broad MSigDB, Network Cancer Genes), and generate publication-ready visualizations—all without writing any code.

The app is ideal for researchers seeking to uncover biological functions, pathways, or gene clusters enriched in their gene sets, with customizable options for selection, clustering, and visualization.

Select Analysis Type:

- Run GO Analysis
- Run KEGG Analysis
- Run Reactome Analysis
- Run Broad MSigDB Analysis
- Run Human Disease Analysis (Human only)
- Run Network Cancer Genes Analysis (Human only)

https://reactome.org

Run Analysis

How to Use the App

Results key columns description

Analysis Results Download

Running functional enrichment analysis...Please do not close the page

- Do not close the application
- Take several minutes to run



Hands-on: GSOR analysis application

GSOR Functional

Upload Files

Input Settings

Analysis Options

Select Analysis Type:

- Run GO Analysis
- Run KEGG Analysis
- Run Reactome Analysis
- Run Broad MSigDB Analysis
- Run Human Disease Analysis (Human only)
- Run Network Cancer Genes Analysis (Human only)

https://reactome.org

Run Analysis

Introduction

Welcome to the GSOR Functional Enrichment Analysis Shiny App! This app enables users to perform gene set over-representation (GSOR) analysis and visualize the results in an intuitive, interactive interface. Originally designed as a set of R scripts for HPC or local execution, this tool now allows you to upload your gene lists, run functional enrichment analyses across multiple databases (GO, KEGG, Reactome, Broad MSigDB, Network Cancer Genes), and generate publication-ready visualizations—all without writing any code.

The app is ideal for researchers seeking to uncover biological functions, pathways, or gene clusters enriched in their gene sets, with customizable options for selection, clustering, and visualization.

How to Use the App

Results key columns description

Analysis Results Download

feavis_input

results_cpEnrich_reactome_clustered_res_top10_category.pdf

results_cpEnrich_reactome_clustered_res.txt

Generate results string

- Red: analysis results to download
- Green: ‘generate results string’ button click for visualization application





Hands-on: GSOR analysis application

The screenshot shows the 'GSOR Functional' Shiny App interface. On the left, a sidebar lists 'Upload Files', 'Input Settings', and 'Analysis Options'. Under 'Analysis Options', 'Run Reactome Analysis' is selected. Below the sidebar, a URL 'https://reactome.org' is shown. A large central area contains sections: 'Introduction' (with a welcome message), 'How to Use the App' (with a plus sign icon), 'Results key columns description' (with a plus sign icon), and 'Analysis Results Download' (with three download buttons: 'feavis_input', 'results_cpEnrich_reactome_clustered_res_top10_category.pdf', and 'results_cpEnrich_reactome_clustered_res.txt'). A green box highlights the 'biocorecce9a87d' token under 'Your visualization key' and the 'Go to Visualization' button below it.

- Red: analysis results to download
- Green: ‘generate results string’ button click for visualization application
 - Red color string is your token used for visualization application at <https://biocoreapps.bsd.uchicago.edu/feavisapp/>



Hands-on: GSOR visualization application

➤ <https://biocoreapps.bsd.uchicago.edu/feavisapp/>

The screenshot shows the FeaVis application interface. On the left, there's a sidebar with sections for 'INTRODUCTION' and 'DEMO DATA DOWNLOAD'. Below that, under 'Step 1: Select Input Type and Folder', it says 'Select type of results: GSEA App results' and provides a URL 'Enter GSEA app https://biocoreapps.bsd.uchicago.edu/gsea_shiny/ result token:' followed by a text input field and a 'Load Files' button. Under 'Step 2: Choose Files for Analysis', there's a 'Submit Data' button. The main area has several search and filter options: a top search bar 'Search in the description for keywords', a section for 'Select one or more keywords to search in the Description' with an 'OR' option and another search bar for 'Enter one or more keywords (separated by comma) to search in the Description', a 'QUERY SEARCH' button, a 'Filter by nominal p-value or FDR adjusted p-value' section, and two summary tables: 'Gene sets over representation summary' and 'Over represented gene sets table'. At the bottom, there's a 'Gene sets over representation plot' section with a dropdown for 'Select Functions to Display: Top 10 Functions' and a 'DOWNLOAD DOTPLOT' button.



Hands-on: GSOR visualization application



The screenshot shows the FeaVis application interface. On the left, there's a sidebar with sections for 'INTRODUCTION' and 'DEMO DATA DOWNLOAD'. Below that is 'Step 1: Select Input Type and Folder', which is highlighted with a red box. It contains a dropdown menu 'Select type of results:' with 'GSEA App results' selected, and a text input field 'Enter GSEA app https://...'. There's also a 'Load Files' button. To the right, there's a search bar 'Search in the description for keywords' and a section for searching by keyword. Below this is a 'Filter by nominal p-value or FDR adjusted p-value' section. On the far right, there are two green boxes: 'Gene sets over representation summary' and 'Over represented gene sets table'. At the bottom, there's a 'Gene sets over representation plot' section with a dropdown 'Select Functions to Display: Top 10 Functions' and a 'DOWNLOAD DOTPLOT' button.

➤ Select type of results (token)

- ❖ **GSEA App results (delete after application close)**
- ❖ upload your own (delete after application close)
- ❖ **CRI-BIO project results** (token is kept after application close)
 - MZCpXyDW_GO, yxKBqMDN_Hallmark, V6br7IZP_KEGG



Hands-on: GSOR visualization application

FeaVis

INTRODUCTION

DEM0 DATA DOWNLOAD

Step 1: Select Input Type and Folder

Select type of results:

GSEA App results

Enter GSEA app https://biocoreapps.bsd.uchicago.edu/gsea_shiny/result_token:

biocorecce9a87d

Load Files

Step 2: Choose Files for Analysis

Choose one or more files:

results_cpEnrich_reactome_clustered_res_Datafile1.txt

results_cpEnrich_reactome_clustered_res_Datafile2.txt

results_cpEnrich_reactome_clustered_res_Datafile3.txt

Submit Data

Search in the description for keywords

Select one or more keywords to search in the Description

OR

Enter one or more keywords (separated by comma) to search in the Description

QUERY SEARCH

Filter by nominal p-value or FDR adjusted p-value

Gene sets over representation summary

Over represented gene sets table

Gene sets over representation plot

Select Functions to Display:

Top 10 Functions

DOWNLOAD DOTPLOT





Hands-on: GSOR visualization application

INTRODUCTION

DEMOS DATA DOWNLOAD

Step 1: Select Input Type and Folder

Select type of results:

GSEA App results

Enter GSEA app https://biocoreapps.bsd.uchicago.edu/gsea_shiny/ result token:
biocorecce9a87d

Load Files

Step 2: Choose Files for Analysis

Submit Data

Search in the description for keywords

| | |
|--------|---|
| data 1 | results_cpEnrich_reactome_clustered_res_Datafile1.txt |
| data 2 | results_cpEnrich_reactome_clustered_res_Datafile2.txt |
| data 3 | results_cpEnrich_reactome_clustered_res_Datafile3.txt |

Select one or more keywords to search in the Description

search all

OR

Enter one or more keywords (separated by comma) to search in the Description

QUERY SEARCH

Filter by nominal p-value or FDR adjusted p-value

Results query is based on

Nominal p-value

FDR adjusted p-value

p-value or FDR adjusted p-value for input data 1
1

p-value or FDR adjusted p-value for input data 2
0.89

p-value or FDR adjusted p-value for input data 3
1

QUERY SEARCH

Gene sets over representation summary

| | | |
|---|--------|-----|
| 1 | data 1 | 358 |
| 2 | data 2 | 390 |
| 3 | data 3 | 316 |

Over represented gene sets table

Copy CSV Excel PDF Print Column visibility Search: _____

| ID | Description |
|---------------|--|
| R-HSA-1236977 | Endosomal/Vacuolar pathway |
| R-HSA-877300 | Interferon gamma signaling |
| R-HSA-983170 | Antigen Presentation: Folding, assembly and peptide loading of class I MHC |
| R-HSA-9012999 | RHO GTPase cycle |
| R-HSA-198933 | Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell |

Showing 1 to 6 of 358 entries

Copy CSV Excel PDF Print Column visibility Search: _____

Gene sets over representation plot

Select Functions to Display:

Top 10 Functions

Selected over represented GOs

- R-HSA-1236977
- R-HSA-877300
- R-HSA-983170
- R-HSA-9012999
- R-HSA-198933



Hands-on: GSOR visualization application

biocoreapps.bsd.uchicago.edu/feavisapp/

2 data 2 390
3 data 3 316

Gene sets over representation plot

Select Functions to Display: Top 10 Functions

Selected over represented GOs

adjusted p-value: 0.0000, 0.0005, 0.0010, 0.0015, 0.0020

Gene Ratio: ● 0.08 ● 0.12 ● 0.16

DOWNLOAD DOTPLOT

ID Description

R-HSA-1236977 Endosomal/Vacuolar pathway
R-HSA-877300 Interferon gamma signaling
R-HSA-983170 Antigen Presentation: Folding, assembly and peptide loading of class I MHC
R-HSA-9012999 RHO GTPase cycle
R-HSA-198933 Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell

Showing 1 to 6 of 358 entries

Copy CSV Excel PDF Print Column visibility Search: GeneRatio

ID Description GeneRatio

R-HSA-1236977 Endosomal/Vacuolar pathway 7/77 1
R-HSA-877300 Interferon gamma signaling 11/77 9
R-HSA-983170 Antigen Presentation: Folding, assembly and peptide loading of class I MHC 7/77 2
R-HSA-1236975 Antigen processing-Cross presentation 10/77 1
R-HSA-2172127 DAP12 interactions 7/77 4

Showing 1 to 6 of 390 entries

Copy CSV Excel PDF Print Column visibility Search: GeneRatio

ID Description GeneRatio

R-HSA-156827 L13a-mediated translational silencing of Ceruloplasmin expression 7/57
R-HSA-72706 GTP hydrolysis and joining of the 60S ribosomal subunit 7/57
R-HSA-927802 Nonsense-Mediated Decay (NMD) 7/57
R-HSA-975957 Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC) 7/57
R-HSA-2408522 Selenoamino acid metabolism 7/57

Showing 1 to 6 of 316 entries

DOWNLOAD ENRICHMENT RESULTS

Developed by bioinformatics core, Center for Research Informatics (CRI), University of Chicago



Hands-on: GSOR visualization application

https://biocoreapps.bsd.uchicago.edu/feavisapp/

2 data 2 390
3 data 3 316

ID Description

Download Dotplot

File type: PDF Width (inches): 12 Height (inches): 8 Y-axis labels: Terms
 Lock aspect ratio

Cancel Confirm & Download

Gene sets over represented

Select Functions to Display:
Top 10 Functions

R-HSA-1236977
R-HSA-877300
R-HSA-983170
R-HSA-9012999
R-HSA-198933
R-HSA-9013149
R-HSA-164940
R-HSA-8980692
R-HSA-76002
R-HSA-913531
R-HSA-1236975
R-HSA-2172127
R-HSA-1236974
R-HSA-909733
R-HSA-2424491
R-HSA-156827
R-HSA-72706
R-HSA-927802
R-HSA-975957
R-HSA-2408522
R-HSA-72613
R-HSA-72737
R-HSA-168273
R-HSA-156902
R-HSA-192823

Selected genes

Showing 1 to 6 of 390 entries

Copy CSV Excel PDF Print Column visibility Search: GeneRatio

ID Description GeneRatio

R-HSA-1236977 Endosomal/Vacuolar pathway 7/77 1
R-HSA-877300 Interferon gamma signaling 11/77 5
R-HSA-983170 Antigen Presentation: Folding, assembly and peptide loading of class I MHC 7/77 2
R-HSA-1236975 Antigen processing-Cross presentation 10/77 1
R-HSA-2172127 DAP12 interactions 7/77 4

Showing 1 to 6 of 390 entries

Copy CSV Excel PDF Print Column visibility Search: GeneRatio

ID Description GeneRatio

R-HSA-156827 L13a-mediated translational silencing of Ceruloplasmin expression 7/57 1
R-HSA-72706 GTP hydrolysis and joining of the 60S ribosomal subunit 7/57 1
R-HSA-927802 Nonsense-Mediated Decay (NMD) 7/57 1
R-HSA-975957 Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC) 7/57 1
R-HSA-2408522 Selenoamino acid metabolism 7/57 1

Showing 1 to 6 of 316 entries

DOWNLOAD ENRICHMENT RESULTS

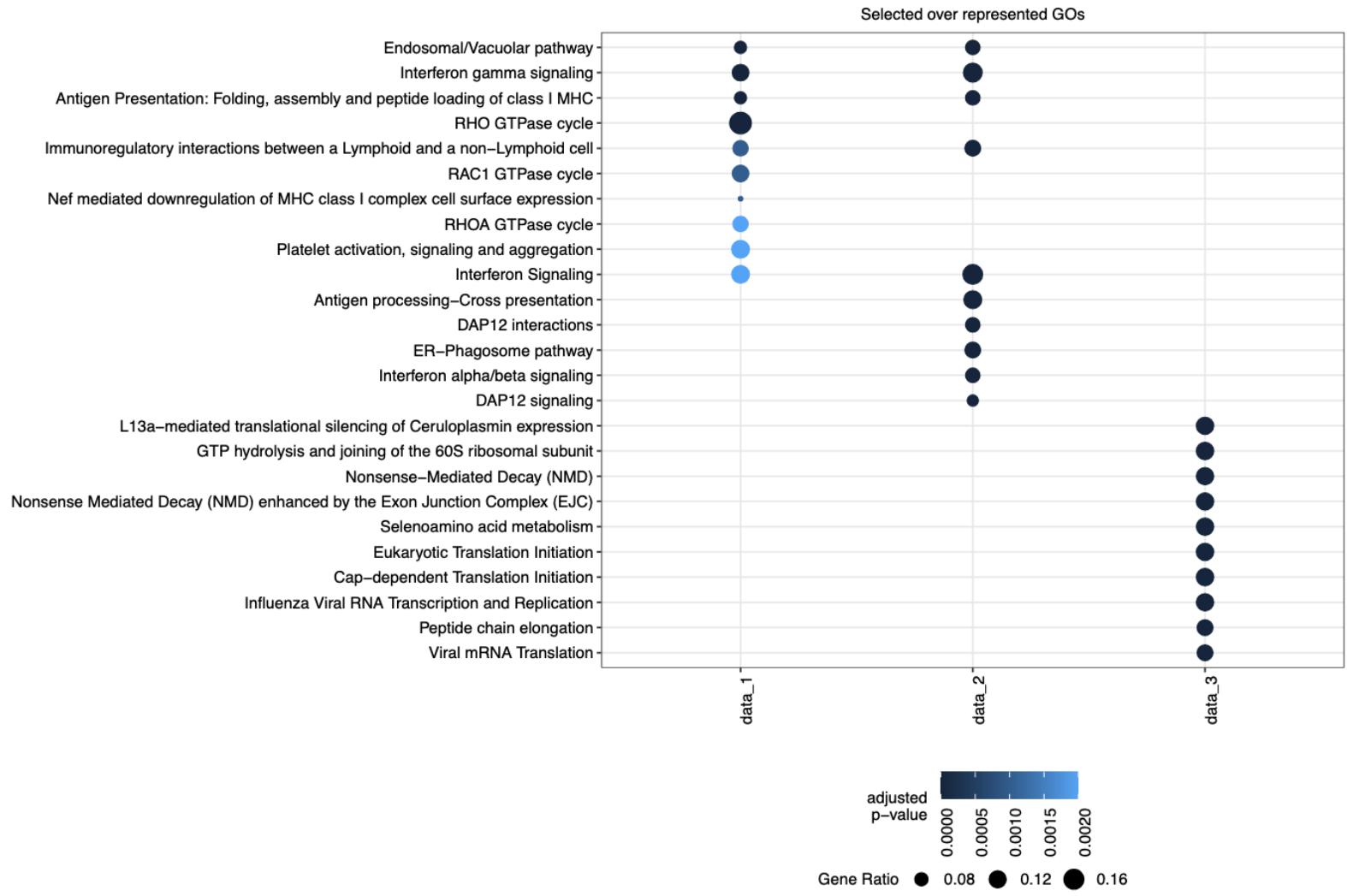
adjusted p-value
0.0000 0.0005 0.0010 0.0015 0.0020
Gene Ratio ● 0.08 ● 0.12 ● 0.16

DOWNLOAD DOTPLOT

Developed by bioinformatics core, Center for Research Informatics (CRI), University of Chicago



Hands-on: GSOR visualization application





Hands-on: GSOR visualization application

bioapps.bsd.uchicago.edu/teavisapp/

2 data 2 390
3 data 3 316

Gene sets over representation plot

Select Functions to Display:
Top 10 Functions

Selected over represented GOs

| ID | Description |
|---------------|--|
| R-HSA-1236977 | Endosomal/Vacuolar pathway |
| R-HSA-877300 | Interferon gamma signaling |
| R-HSA-983170 | Antigen Presentation: Folding, assembly and peptide loading of class I MHC |
| R-HSA-9012999 | RHO GTPase cycle |
| R-HSA-198933 | Immunoregulatory |
| R-HSA-1236975 | Antigen processing-Cross presentation |
| R-HSA-2172127 | DAP12 interactions |
| R-HSA-1236974 | |
| R-HSA-909733 | |
| R-HSA-2424491 | |
| R-HSA-156827 | |
| R-HSA-72706 | |
| R-HSA-927802 | |
| R-HSA-975957 | |
| R-HSA-2408522 | |
| R-HSA-72613 | |
| R-HSA-72737 | |
| R-HSA-168273 | |
| R-HSA-156902 | |
| R-HSA-192823 | |

adjusted p-value

Gene Ratio ● 0.08 ● 0.12 ● 0.16

data_1 data_2 data_3

DOWNLOAD DOTPLOT

resData_2025-11-18.xlsx
Completed — 83.6 KB

dotplot_2025-11-18.pdf
Completed — 8.6 KB

mmc1.pdf
Completed — 21.5 MB

gsor_analysis_workflow_md.md
File moved or missing

gsor_analysis_workflow_md.md
File moved or missing

Showing 1 to 6 of 358 entries

Copy CSV Excel PDF

ID Description

R-HSA-1236977 Endosomal/Vacuolar pathway
R-HSA-877300 Interferon gamma signaling
R-HSA-983170 Antigen Presentation: Folding, assembly and peptide loading of class I MHC
R-HSA-1236975 Antigen processing-Cross presentation
R-HSA-2172127 DAP12 interactions

neRatio 1 5 2 1 4

Showing 1 to 6 of 390 entries

Copy CSV Excel PDF Print Column visibility

Search: Gene:

ID Description

R-HSA-156827 L13a-mediated translational silencing of Ceruloplasmin expression
R-HSA-72706 GTP hydrolysis and joining of the 60S ribosomal subunit
R-HSA-927802 Nonsense-Mediated Decay (NMD)
R-HSA-975957 Nonsense Mediated Decay (NMD) enhanced by the Exon Junction Complex (EJC)
R-HSA-2408522 Selenoamino acid metabolism

Showing 1 to 6 of 316 entries

DOWNLOAD ENRICHMENT RESULTS



Hands-on: GSOR visualization application

AutoSave Home Insert Draw Page Layout Formulas Data Review View Automate

Calibri (Body) 11 A A Wrap Text General Conditional Formatting Normal Bad Good Neutral Calculation Check Cell

Paste Copy Format B I U Merge & Center Conditional Formatting as Table Insert Delete Format Clear

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|-----|--------------|-----------------|-----------|-----------|------------|-------------|--------|--------|----------|--------|---------------|-------|---------------|---------|
| 1 | ID | Description | GeneRatio | BgRatio | RichFactor | FoldEnrichn | zScore | pvalue | p.adjust | vvalue | geneID | Count | geneSymbol | Cluster |
| 2 | R-HSA-123697 | Endosomal/Va | 5/70 | 11/11091 | 0.455 | 72.019 | 18.781 | 0 | 0 | 0 | HLA-E/HLA-B/ | 5 | HLA-E/HLA-B/ | data_1 |
| 3 | R-HSA-87730C | Interferon gam | 8/70 | 96/11091 | 0.083 | 13.204 | 9.57 | 0 | 0 | 0 | HLA-E/HLA-B/ | 8 | HLA-E/HLA-B/ | data_1 |
| 4 | R-HSA-98317C | Antigen Preser | 5/70 | 29/11091 | 0.172 | 27.318 | 11.309 | 0 | 0 | 0 | HLA-E/HLA-B/ | 5 | HLA-E/HLA-B/ | data_1 |
| 5 | R-HSA-901299 | RHO GTPase c | 13/70 | 449/11091 | 0.029 | 4.587 | 6.184 | 0 | 0 | 0 | PIK3R3/SPTBN | 13 | PIK3R3/SPTBN | data_1 |
| 6 | R-HSA-198933 | Immunoregula | 7/70 | 133/11091 | 0.053 | 8.339 | 6.786 | 0 | 0.001 | 0.001 | HLA-E/HLA-B/ | 7 | HLA-E/HLA-B/ | data_1 |
| 7 | R-HSA-901314 | RAC1 GTPase | 8/70 | 185/11091 | 0.043 | 6.852 | 6.396 | 0 | 0.001 | 0.001 | PIK3R3/CAV1/ | 8 | PIK3R3/CAV1/ | data_1 |
| 8 | R-HSA-16494C | Nef mediated | 3/70 | 10/11091 | 0.3 | 47.533 | 11.732 | 0 | 0.001 | 0.001 | B2M/HLA-A/A | 3 | B2M/HLA-A/A | data_1 |
| 9 | R-HSA-898069 | RHOA GTPase | 7/70 | 147/11091 | 0.048 | 7.545 | 6.366 | 0 | 0.002 | 0.001 | CAV1/STOM/F | 7 | CAV1/STOM/F | data_1 |
| 10 | R-HSA-76002 | Platelet activa | 9/70 | 263/11091 | 0.034 | 5.422 | 5.784 | 0 | 0.002 | 0.001 | PECAM1/A2M | 9 | PECAM1/A2M | data_1 |
| 11 | R-HSA-913531 | Interferon Sig | 9/70 | 269/11091 | 0.033 | 5.301 | 5.691 | 0 | 0.002 | 0.001 | IFI27/HLA-E/H | 9 | IFI27/HLA-E/H | data_1 |
| 12 | R-HSA-909733 | Interferon alph | 5/70 | 77/11091 | 0.065 | 10.288 | 6.518 | 0 | 0.004 | 0.003 | IFI27/HLA-E/H | 5 | IFI27/HLA-E/H | data_1 |
| 13 | R-HSA-114608 | Platelet degra | 6/70 | 129/11091 | 0.047 | 7.369 | 5.799 | 0 | 0.005 | 0.004 | PECAM1/A2M | 6 | PECAM1/A2M | data_1 |
| 14 | R-HSA-217212 | DAP12 interac | 4/70 | 45/11091 | 0.089 | 14.084 | 7.009 | 0 | 0.005 | 0.004 | HLA-E/HLA-B/ | 4 | HLA-E/HLA-B/ | data_1 |
| 15 | R-HSA-76005 | Response to el | 6/70 | 134/11091 | 0.045 | 7.094 | 5.656 | 0 | 0.005 | 0.004 | PECAM1/A2M | 6 | PECAM1/A2M | data_1 |
| 371 | R-HSA-679869 | Neutrophil deg | 13/77 | 480/11091 | 0.027 | 3.901 | 5.433 | 0 | 0.001 | 0.001 | B2M/CYBA/H | 13 | B2M/CYBA/H | data_2 |
| 372 | R-HSA-970567 | SARS-CoV-2 ac | 7/77 | 126/11091 | 0.056 | 8.002 | 6.609 | 0 | 0.001 | 0.001 | B2M/HLA-A/H | 7 | B2M/HLA-A/H | data_2 |
| 373 | R-HSA-901314 | RAC1 GTPase | 8/77 | 185/11091 | 0.043 | 6.229 | 5.996 | 0 | 0.001 | 0.001 | CYBA/ARHGDI | 8 | CYBA/ARHGDI | data_2 |
| 374 | R-HSA-901342 | RAC3 GTPase | 6/77 | 94/11091 | 0.064 | 9.194 | 6.671 | 0 | 0.001 | 0.001 | CYBA/ARHGDI | 6 | CYBA/ARHGDI | data_2 |
| 375 | R-HSA-969451 | SARS-CoV-2 In | 9/77 | 299/11091 | 0.03 | 4.336 | 4.889 | 0 | 0.005 | 0.004 | B2M/HLA-A/H | 9 | B2M/HLA-A/H | data_2 |
| 376 | R-HSA-451927 | Interleukin-2 f | 4/77 | 44/11091 | 0.091 | 13.094 | 6.721 | 0 | 0.005 | 0.004 | STAT4/PIK3R1 | 4 | STAT4/PIK3R1 | data_2 |
| 377 | R-HSA-901299 | RHO GTPase c | 11/77 | 449/11091 | 0.024 | 3.529 | 4.574 | 0 | 0.005 | 0.004 | CYBA/ARHGDI | 11 | CYBA/ARHGDI | data_2 |
| 378 | R-HSA-202427 | Phosphorylati | 3/77 | 22/11091 | 0.136 | 19.642 | 7.318 | 0 | 0.009 | 0.007 | PTPRJ/HLA-DF | 3 | PTPRJ/HLA-DF | data_2 |
| 379 | R-HSA-970568 | SARS-CoV-2-h | 7/77 | 203/11091 | 0.034 | 4.967 | 4.769 | 0.001 | 0.01 | 0.008 | B2M/HLA-A/H | 7 | B2M/HLA-A/H | data_2 |
| 380 | R-HSA-164952 | The role of Ne | 3/77 | 28/11091 | 0.107 | 15.433 | 6.393 | 0.001 | 0.017 | 0.014 | B2M/HLA-A/D | 3 | B2M/HLA-A/D | data_2 |
| 381 | R-HSA-388841 | Costimulation | 4/77 | 69/11091 | 0.058 | 8.35 | 5.121 | 0.001 | 0.023 | 0.019 | PIK3R1/MAP3 | 4 | PIK3R1/MAP3 | data_2 |
| 952 | R-HSA-174178 | APC/C:Cdh1 r | 1/57 | 73/11091 | 0.014 | 2.665 | 1.026 | 0.314 | 0.485 | 0.413 | AURKA | 1 | AURKA | data_3 |
| 953 | R-HSA-70171 | Glycolysis | 1/57 | 73/11091 | 0.014 | 2.665 | 1.026 | 0.314 | 0.485 | 0.413 | RANBP2 | 1 | RANBP2 | data_3 |
| 954 | R-HSA-885451 | AURKA Activat | 1/57 | 73/11091 | 0.014 | 2.665 | 1.026 | 0.314 | 0.485 | 0.413 | AURKA | 1 | AURKA | data_3 |
| 955 | R-HSA-893921 | ESR-mediated | 2/57 | 222/11091 | 0.009 | 1.753 | 0.815 | 0.317 | 0.485 | 0.414 | CHD1/AREG | 2 | CHD1/AREG | data_3 |
| 956 | R-HSA-68075C | RNA polymer | 1/57 | 74/11091 | 0.014 | 2.629 | 1.011 | 0.318 | 0.485 | 0.414 | ELL2 | 1 | ELL2 | data_3 |
| 957 | R-HSA-983348 | PKR-mediated | 1/57 | 75/11091 | 0.013 | 2.594 | 0.996 | 0.321 | 0.488 | 0.416 | ARIH1 | 1 | ARIH1 | data_3 |



Hands-on: GSOR visualization application

The screenshot shows the FeaVis application interface. On the left, there's a sidebar with sections for 'INTRODUCTION' and 'DEMO DATA DOWNLOAD'. Below that is 'Step 1: Select Input Type and Folder', which is highlighted with a red box. It contains a dropdown menu 'Select type of results:' with 'GSEA App results' selected, and a text input field 'Enter GSEA app https://...'. Below these are fields for 'token:' and 'Load Files'. To the right, there's a search bar 'Search in the description for keywords' and a section for searching keywords in the description. Further down are sections for filtering by p-value, gene set over representation summary, and gene set over representation plot.

➤ Select type of results (token)

- ❖ GSEA App results (delete after application close)
- ❖ upload your own (delete after application close)
- ❖ **CRI-BIO project results** (token is kept after application close)
 - MZCpXyDW_GO, yxBqMDN_Hallmark, V6br7IZP_KEGG



Hands-on: GSOR analysis randi

- Instruction: https://github.com/CRI-Biocore/functional_enrichment_analysis_2025_workshop/blob/main/docs/rstudio_randi_quickstart.md

```
Mac:~ yanli$ ssh yli22@randi.cri.uchicago.edu
** Unauthorized use/access is prohibited. **
```

This computer system is owned by the University of Chicago Biological Sciences Division and is for authorized use only. Logging onto this computer verifies you have read and agree both to the statement below and to use BSD computer networks and systems in accordance with the BSD Eligibility and Acceptable Use policy and related policies.

Individuals using this computer system are subject to having all of their activities on this system monitored and recorded by system personnel. Anyone using this system expressly consents to such monitoring and is advised that if such monitoring reveals possible criminal activity or policy violation, system personnel may provide the evidence of such monitoring to law enforcement or other officials.

University of Chicago Acceptable Use Policy:
<https://itservices.uchicago.edu/policies/acceptable-use-policy>

Register this system with Red Hat Insights: insights-client --register
Create an account or view all your systems at <https://red.ht/insights-dashboard>
Last login: Mon Nov 17 22:35:23 2025 from 205.208.121.174

Home Directory (/home/yli22)



Hands-on: GSOR analysis randi

University of Chicago Acceptable Use Policy:

<https://itservices.uchicago.edu/policies/acceptable-use-policy>

Register this system with Red Hat Insights: insights-client --register

Create an account or view all your systems at <https://red.ht/insights-dashboard>

Last login: Tue Nov 18 22:13:37 2025 from 205.208.121.120

Home Directory (/home/yli22)

Used: 8.224G

Quota: 10G

Limit: 11G

Grace (days): none

Scratch Directory (/scratch/yli22)

Used: 311.3M

Quota: 10T

Limit: 20T

Grace (days): none

User: yli22

Jobs analyzed: 5

Average CPU Efficiency: 24.65%

Average Memory Efficiency: 23.84%

[yli22@cri22in002 ~]\$

[yli22@cri22in002 ~]\$

[yli22@cri22in002 ~]\$ █





Hands-on: GSOR analysis randi

```
[yli22@cri22in002 ~]$ ll /gpfs/data/biocore-workshop/functional_enrichment_workshop5/
total 34
-rwxr-x--- 1 yli22 cri-biocore_workshop 5814 Nov 17 21:51 cluster_proflier_gsor_demo.R
drwxr-s--- 5 yli22 cri-biocore_workshop 4096 Oct 15 12:59 conda_options
drwxr-s--- 22 yli22 cri-biocore_workshop 4096 Nov 17 19:51 gsor_Nov25_v2
drwxr-x--- 2 yli22 cri-biocore_workshop 4096 Nov 17 19:46 test_data_input
[yli22@cri22in002 ~]$ cp /gpfs/data/biocore-workshop/functional_enrichment_workshop5/cluster_proflier_gsor_demo.R ./
[yli22@cri22in002 ~]$
```

2. Request an Interactive Job

Without node reservation:

```
srun --mem=4Gb --pty --time=02:00:00 /bin/bash
```



With workshop node reservation:

```
srun --reservation=workshop --mem=4GB --pty --time=02:00:00 /bin/bash
```



```
[yli22@cri22in002 ~]$ srun --reservation=workshop --mem=4GB --pty --time=02:00:00 /bin/bash
[yli22@cri22cn128 ~]$
```

```
[yli22@cri22cn128 ~]$ squeue -u yli22
      JOBID PARTITION      NAME      USER ST      TIME  NODES NODELIST(REASON)
        1865412    tier1q      bash    yli22  R      1:03      1 cri22cn128
[yli22@cri22cn128 ~]$
```





Hands-on: GSOR analysis randi

3. Check Network Interfaces

```
ip a
```



- Example output:

```
[yli22@cri22cn008 ~]$ ip a
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inett6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 9000 qdisc mq state UP group default qlen 1000
    link/ether d8:5e:d3:6c:8b:db brd ff:ff:ff:ff:ff:ff
    altname eno1
    altname enp49s0f0
    altname ens8f0
    inet 10.50.46.8/23 brd 10.50.47.255 scope global noprefixroute eth0
        valid_lft forever preferred_lft forever
    inett6 fe80::da5e:d3ff:fe6c:8bdb/64 scope link
        valid_lft forever preferred_lft forever
3: eth1: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc mq state DOWN group default qlen 1000
    link/ether d8:5e:d3:6c:8b:dc brd ff:ff:ff:ff:ff:ff
    altname eno2
    altname enp49s0f1
    altname ens8f1
4: ib0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 2044 qdisc mq state UP group default qlen 1000
    link/infiniband 00:00:06:73:fe:80:00:00:00:00:08:c0:eb:03:00:75:81:6a brd 00:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff:ff
    inet 10.50.87.99/23 brd 10.50.87.255 scope global noprefixroute ib0
        valid_lft forever preferred_lft forever
    inett6 fe80::ac0:eb03:75:816a/64 scope link
        valid_lft forever preferred_lft forever
[yli22@cri22cn008 ~]$
```



Hands-on: GSOR analysis randi

```
valid_lft forever preferred_lft forever
inet6 ::1/128 scope host
    valid_lft forever preferred_lft forever
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 9000 qdisc mq state UP group default qlen 1000
    link/ether d8:5e:d3:6c:88:67 brd ff:ff:ff:ff:ff:ff
    altname eno1
    altname enp49s0f0
    altname ens8f0
    inet 10.50.46.128/23 brd 10.50.47.255 scope global noprefixroute eth0
        valid_lft forever preferred_lft forever
    inet6 fe80::da5e:d3ff:fe6c:8867/64 scope link
        valid_lft forever preferred_lft forever
3: eth1: <NO-CARRIER,BROADCAST,MULTICAST,UP> mtu 1500 qdisc mq state DOWN group default qlen 1000
    link/ether d8:5e:d3:6c:88:68 brd ff:ff:ff:ff:ff:ff
    altname eno2
    altname enp49s0f1
    altname ens8f1

Last login: Tue Nov 18 22:12:18 on ttys000

The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
BIO-ML-15:~ yanli$ ssh -Y yli22@10.50.46.128
The authenticity of host '10.50.46.128 (10.50.46.128)' can't be established.
ED25519 key fingerprint is SHA256:XswQFGmKA4QVki9GR2ZLug+5/zZTc5ZAaKRD5PZ6i90.
This key is not known by any other names.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '10.50.46.128' (ED25519) to the list of known hosts.
Register this system with Red Hat Insights: insights-client --register
Create an account or view all your systems at https://red.ht/insights-dashboard
[yli22@cri22cn128 ~]$
```



Hands-on: GSOR analysis randi

5. Load Required Modules

```
module load gcc/12.1.0
module load miniconda3/24.9.2
module list
export LD_LIBRARY_PATH=/gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2/lib:$LD_LIBRARY_PATH
```



- Example output:

```
[yli22@cri22cn008 ~]$ module load gcc/12.1.0
[yli22@cri22cn008 ~]$ module load miniconda3/24.9.2
[yli22@cri22cn008 ~]$ module list

Currently Loaded Modules:
 1) gcc/12.1.0  2) miniconda3/24.9.2
```

6. Activate Pre-set Computational Environment

```
source activate /gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2
```



- Output example:

```
[yli22@cri22cn008 ~]$ source activate /gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2
(/gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2) [yli22@cri22cn008 ~]$ █
```





Hands-on: GSOR analysis randi

Verify paths and versions:

```
which R  
which python  
python --version  
which rstudio
```



- Expected output:

```
[yli22@cri22cn008 ~]$ source activate /gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2  
(/gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2) [yli22@cri22cn008 ~]$ which R  
/gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2/bin/R  
(/gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2) [yli22@cri22cn008 ~]$ which python  
/gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2/bin/python  
(/gpfs/data/biocore-workshop/functional_enrichment_workshop5/gsor_Nov25_v2) [yli22@cri22cn008 ~]$ python --version  
Python 3.14.0
```

Note: You can skip this step, this step is more for troubleshooting.



Hands-on: GSOR analysis randi

7. Launch RStudio

rstudio

- This opens RStudio on your local machine via X11 forwarding:

The screenshot shows the RStudio interface. The R console window displays the standard R startup message, including the version (R version 4.4.3), copyright information, and license details. Below the console, a small 'Update Available' dialog box is overlaid, prompting the user to either quit and download the update or ignore it. The 'Ignore Update' button is highlighted with a red box. The main RStudio environment and history panes are visible, along with a 'Global Environment' sidebar.

R version 4.4.3 (2025-02-28) -- "Trophy Case"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-conda-linux-gnu

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.

> |

X Update Available
RStudio 2025.09.2+418 is now available (you're using 99.9.9)
Quit and Download... Remind Later Ignore Update

plot.default {graphics} R Documentation

The Default Scatterplot Function
Description



Hands-on: GSOR analysis randi

The screenshot shows the RStudio interface with the following components:

- File Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Source on Save, Run, Source.
- Environment Tab:** Global Environment. Status: Environment is empty.
- Files Tab:** Shows a directory tree:
 - /ess/home/home1/yli22/
 - .Rhistory (1.1 KB, Nov 17, 2025, 10:15 PM)
 - bash_profile_bk (2.9 KB, Feb 8, 2021, 12:58 PM)
 - cellranger_analysis_results
 - cluster_profiler_gsor_demo.R (5.7 KB, Nov 18, 2025, 10:17 PM)
 - functional_enrichment_analysis_workshops
 - logs
 - ncbi_error_report.xml (3.7 KB, Jan 24, 2022, 5:05 PM)
 - nextflow_test
 - perl5
 - R
 - R_libs_personal-setup.txt (222 B, May 10, 2017, 2:27 PM)
 - requirements.txt (43 B, Aug 24, 2016, 2:30 PM)
 - run01_cellranger_count.slurm (795 B, May 30, 2025, 4:30 PM)
 - ve_rna_seq-env.yml (4 KB, Oct 17, 2016, 4:15 PM)
 - workshop_test
 - yli2_bk
- Console Tab:** Displays the R session output:

```
R version 4.4.3 (2025-02-28) -- "Trophy Case"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-conda-linux-gnu

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.
```



Hands-on: GSOR analysis randi

The screenshot shows the RStudio interface with several panes:

- Terminal:** Displays a command-line session:

```
[yli22@cri22in002 ~]$ ll /gpfs/data/biocore-workshop/functional_enrichment_workshop5/
total 34
-rwxr-x--- 1 yli22 cri-biocore_workshop 5814 Nov 17 21:51 cluster_proflier_gsor_demo.R
drwxr-s--- 5 yli22 cri-biocore_workshop 4096 Oct 15 12:59 conda_options
drwxr-s--- 22 yli22 cri-biocore_workshop 4096 Nov 17 19:51 gsor_Nov25_v2
drwxr-x--- 2 yli22 cri-biocore_workshop 4096 Nov 17 19:46 test_data_input
[yli22@cri22in002 ~]$ cp /gpfs/data/biocore-workshop/functional_enrichment_workshop5/cluster_proflier_gsor_demo.R ./
[yli22@cri22in002 ~]$
```
- Script Editor:** Shows an R script named `cluster_proflier_gsor_demo.R` with code for performing GSOR analysis. The script includes comments, variable assignments, and function calls like `enrichGO`.
- File Browser:** Shows the directory structure under `/ess/home/home1/yli22`. A red box highlights the file `cluster_proflier_gsor_demo.R` in the list.
- Console:** Displays the R environment information and a welcome message from R version 4.4.3.

Take Home Message

- GSOR vs GESA usage
- results interpretations
- submit project request to run GSOR for visualizations
 - ❖ <https://biocore.cri.uchicago.edu/>

OUR AWESOME TEAM



Mengjie
Chen, PhD

Faculty
Director



Wenjun Kang,
MS

Technical Director



Jason Shapiro,
PhD



Yan Li, PhD

Associate
Director



Houxiang Zhu,
PhD



Diana Vera Cruz,
PhD



Qiaoshan Lin,
PhD



Yildiz Koca,
PhD



Evan Wu



Katie Aracena,
PhD



David Tieri,
PhD



Geetha Priyanka,
MS

Bioinformaticians



Genomics and proteomics data analysis

BiCF applies appropriate and state-of-the-arts statistical and bioinformatic methodologies to analyze genomics data generated from standard and emerging assays.



Consulting, grant writing and training

BiCF provides consulting services for experimental design or data analysis; grant writing assistance, including bioinformatics development, cost analysis, and documentation of tools to complete the research.



Data management system development

BiCF offers enterprise solutions for project and study management, for data production , sharing and integration .



Contact us: bioinformatics@bsd.uchicago.edu

Submit a project request: <https://biocore.cri.uchicago.edu/>

Services

- NGS sequencing data analysis
 - ❖ bulk sequencing data analysis
 - bulk RNA-seq , bulk ATAC-seq, Cut&Run, Library screen CRISPR etc.
 - standard 10 hours workflow analysis
 - 2 weeks results turn around time
 - ❖ single cell data analysis
 - scRNA-seq, scATAC-seq, OMICS etc.
 - standard 50-60 hours workflow analysis
 - 4-6 weeks results turn around time
 - ❖ spatial data analysis
 - visium, xieunm, GeoMx, spatial data deconvolution
- Grant writing assistance: Aims, power computation etc.
- Consultation: free 1hr consultation for analysis, in person office hours
- Application development Ent

Contact us: bioinformatics@bsd.uchicago.edu

Submit a project request: <https://biocore.cri.uchicago.edu/>



Acknowledgement

- HPC system team to maintain randi HPC
- HPC: Franck segbedji to reserve the computational nodes
- HPC: Mike Jarsulic to reserve the computational nodes

- My colleague: Wenjun Kang to coordinate today's zoom Q&A
- My colleague: Diana Vera Cruz to test the workshop code usage
- My colleague: Zhongyu Li
 - ❖ Developer of GSOR shiny application

- Biocore shiny application feedbacks

In Person Free Consultation

➤ HPC system team: Randi Support



- Every Tuesday afternoon 12:00 PM – 2:30 PM
- Location: Peck Pavilion N161
- Slack Group: https://join.slack.com/t/criscientific-dzi9891/shared_invite/zt-2kgihy4392-1ELPfgn8pL5BcXk4oF9D4g

➤ In-person office hours bioinformatics



- Every Tuesday afternoon 12:30 PM – 3:30 PM
- Location: Peck Pavilion N161
- Bioinformatics related questions
- Experimental design questions

➤ Biostatistical Questions



- First Tuesday of Each Month 12:30 PM – 3:30 PM
- Location: Peck Pavilion N161



THE UNIVERSITY OF
CHICAGO

Biological Sciences
Bioinformatics Core

CENTER FOR RESEARCH INFORMATICS

BIOINFORMATICS CORE

Bioinformatics Workshop Learning Series

Thanks for your attention!

Yan Li, Ph.D.

yli22@bsd.uchicago.edu

bioinformatics@bsd.uchicago.edu

Workshop feedback survey:

<https://mycri.cri.uchicago.edu/educations/trainings/79/survey/>