THE UNIVERSITY OF CHICAGO

Biological Sciences
Bioinformatics Core

CENTER FOR RESEARCH INFORMATICS

BIOINFORMATICS CORE

Bioinformatics Workshop Learning Series

Functional enrichment data analysis

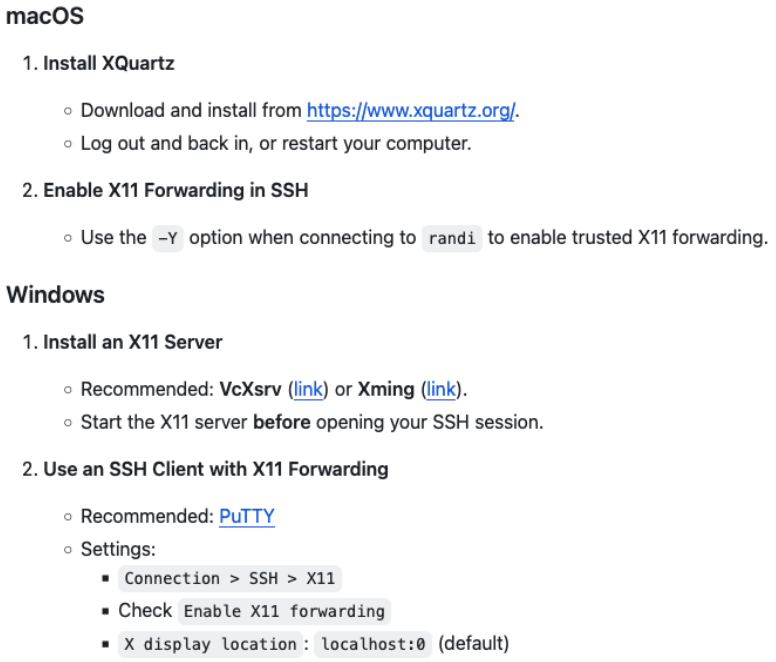
Yan Li, Ph.D.

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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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/

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Bioinformatics Workshop Series 2025

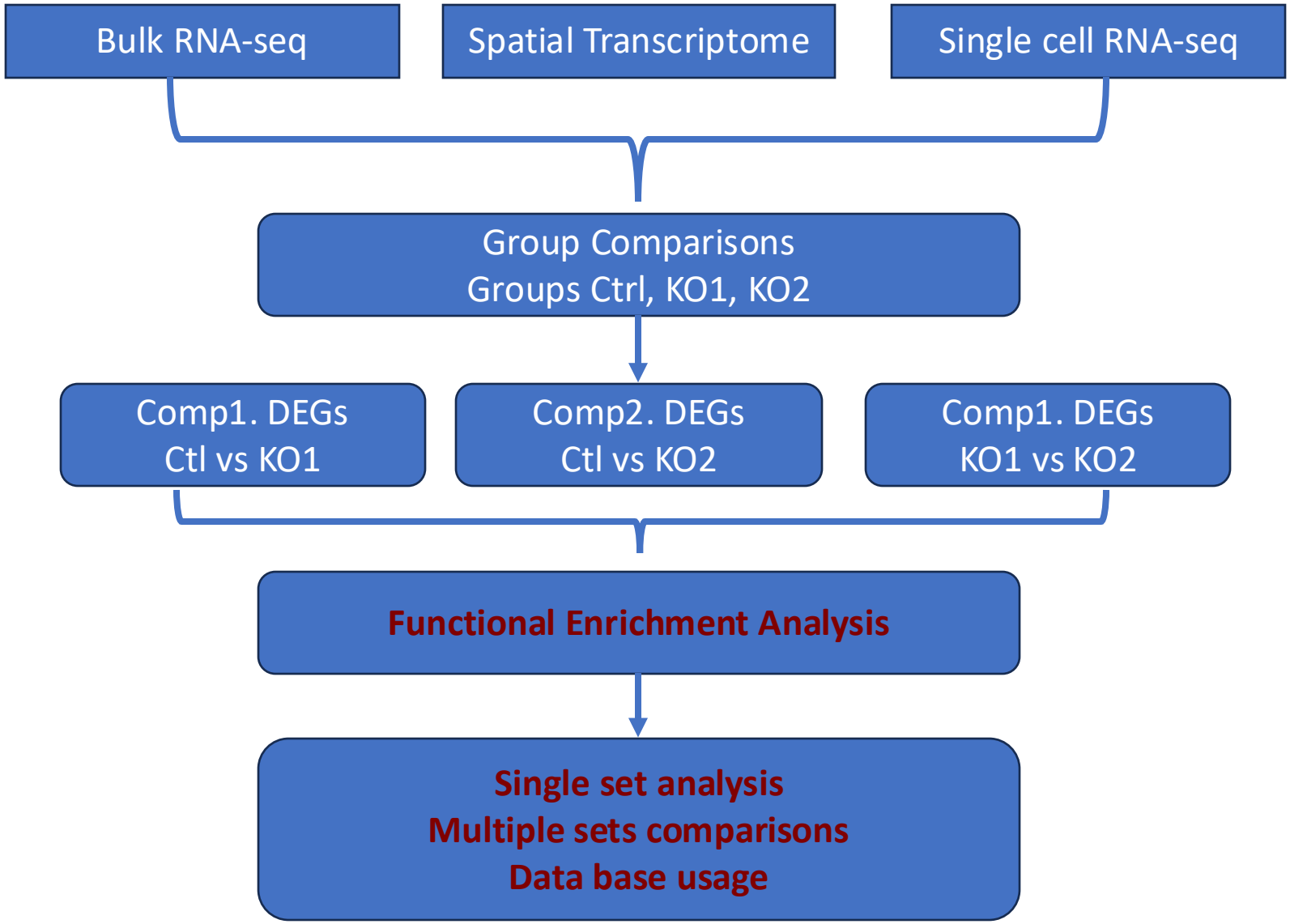
- **Target: 1-2 workshops each quarter semester**

- 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/

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

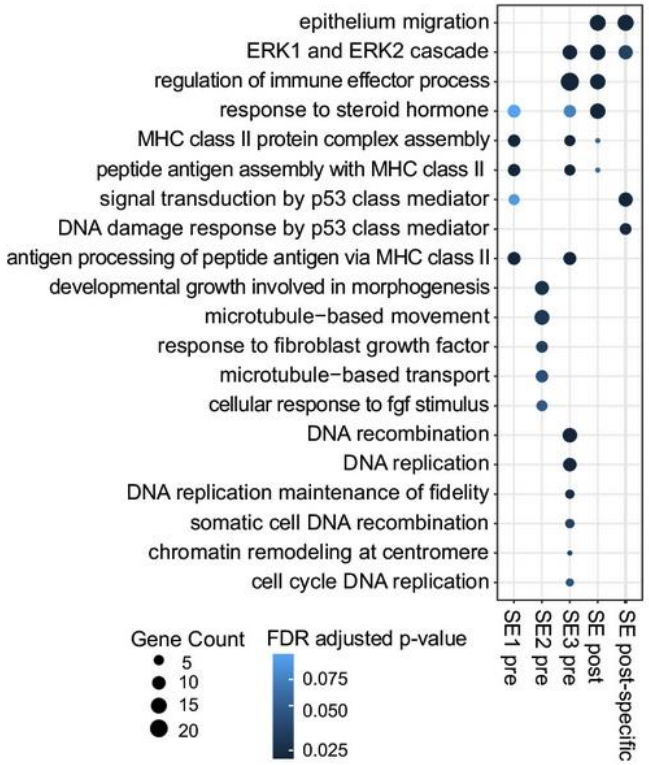
Why Functional Enrichment Analysis?



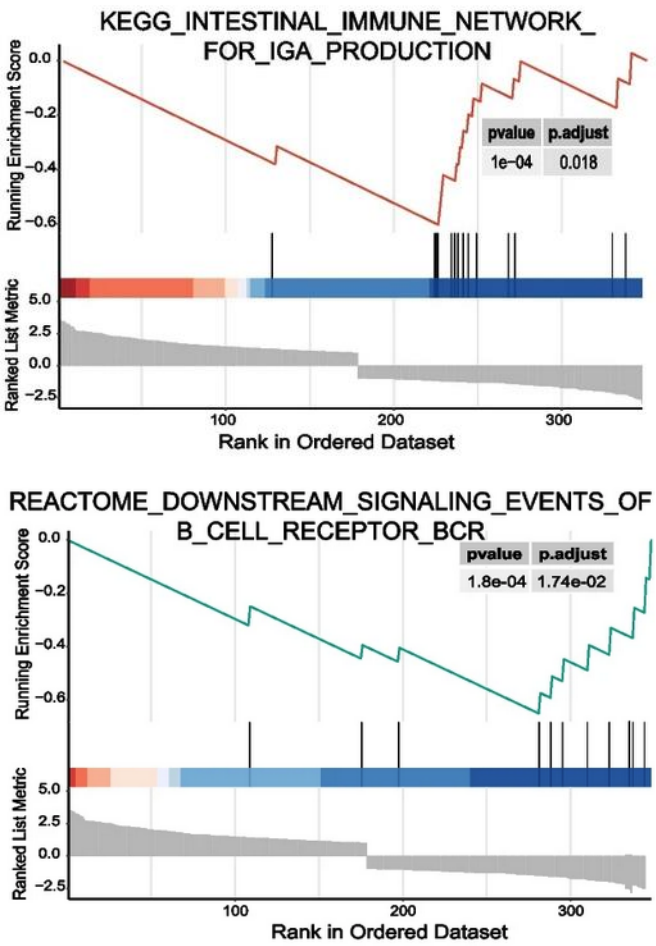
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Functional Enrichment Analysis Overview

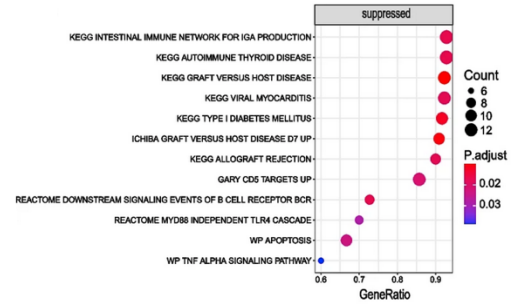
A. **Gene Sets Over Representation (GSOR)**



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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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?

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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) corelated with the function of interest
- p-value of ES is based on permutation test

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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) corelated with the function of interest
- p-value of ES is based on permutation test

- **Multiple test correction (FDR correction)**

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Gene Sets Enrichment Analysis (GSEA)

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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.

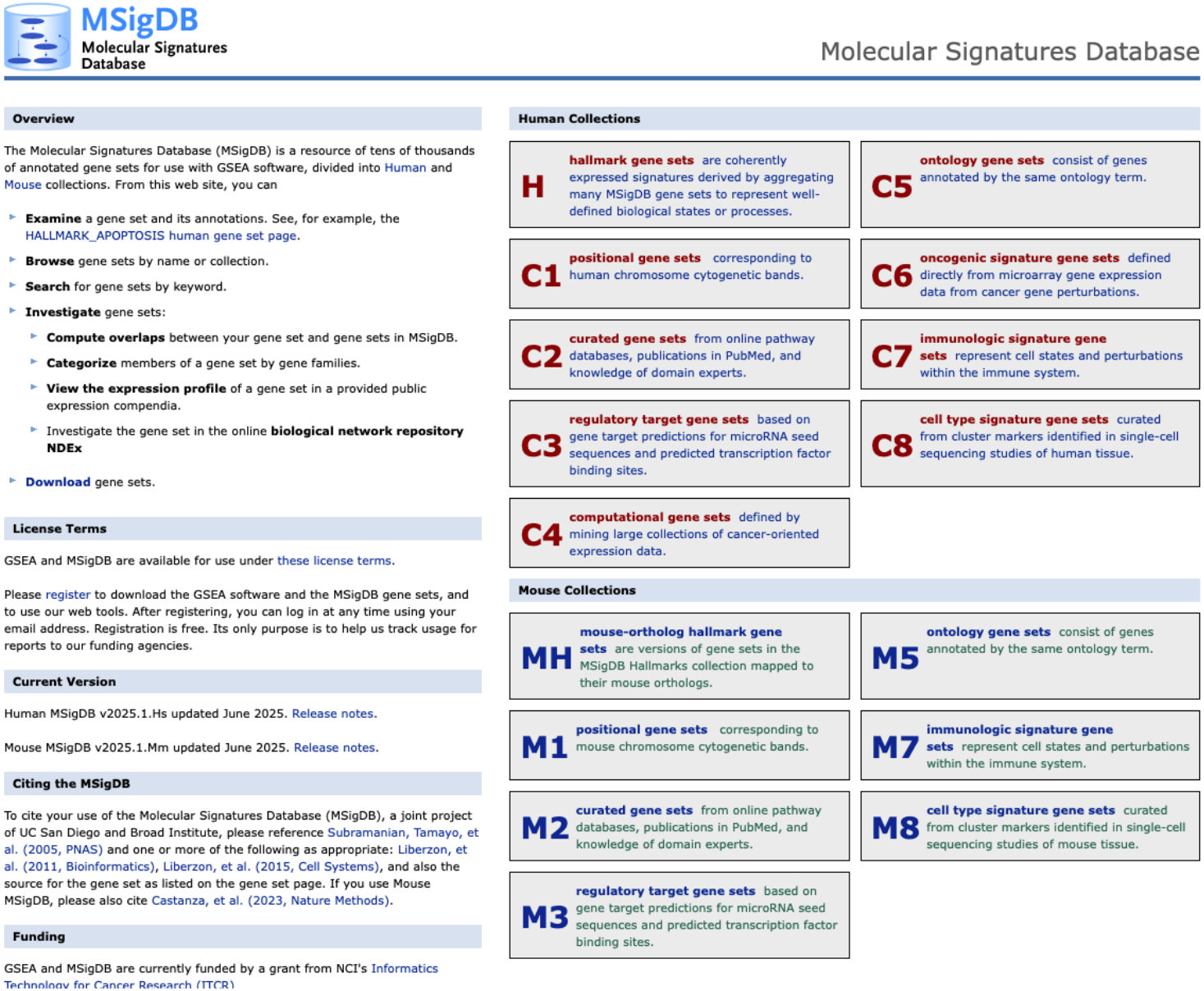
Citation: google AI comparison

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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Functional Enrichment Analysis Databases



Pic. source: https://www.gsea-msigdb.org/gsea/msigdb

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

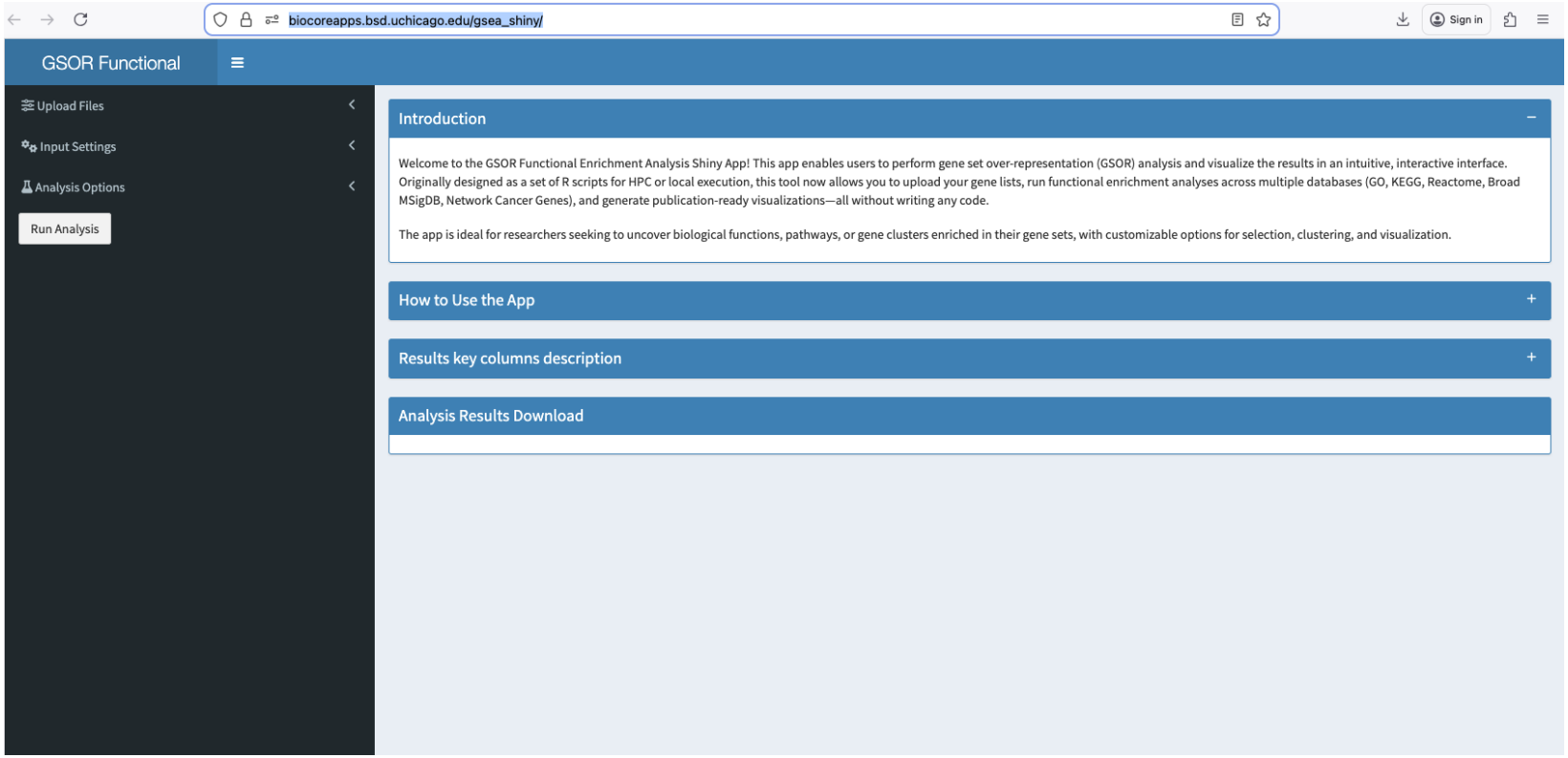
Functional Enrichment Analysis Reference Genome

- 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/

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

GSOR analysis application

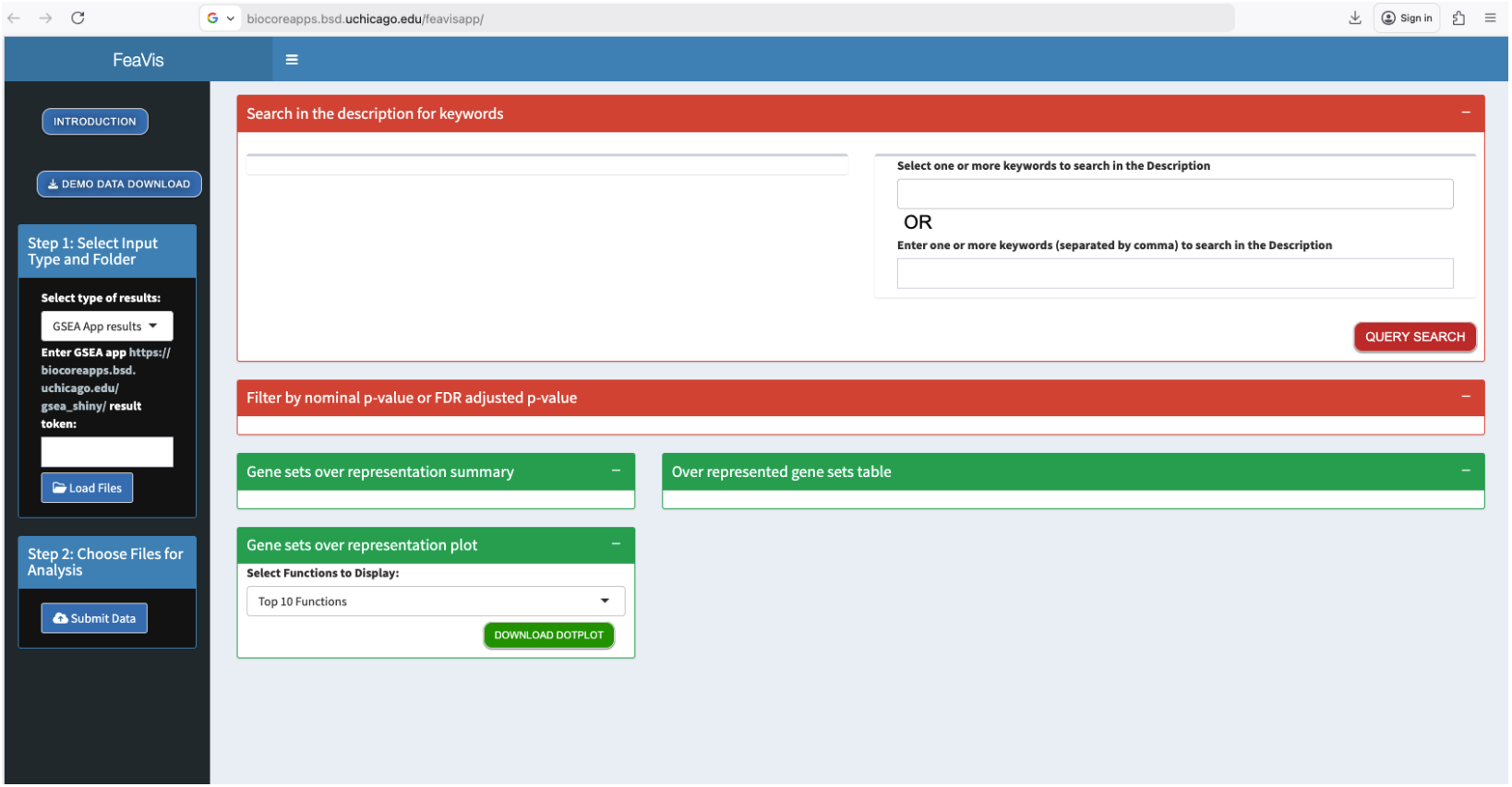
- https://biocoreapps.bsd.uchicago.edu/gsea_shiny/



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

GSOR visualization application

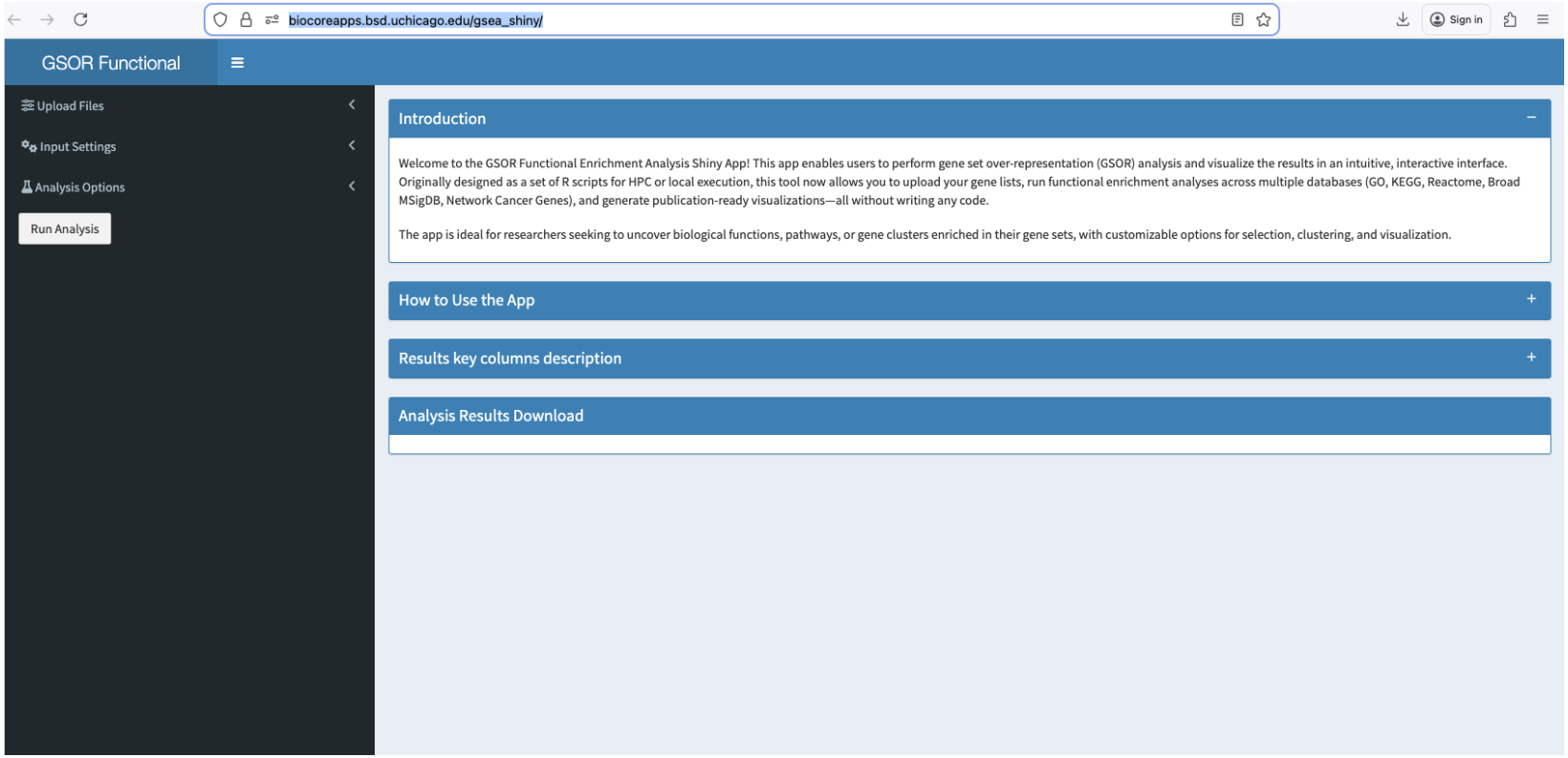
- https://biocoreapps.bsd.uchicago.edu/feavisapp/



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis application

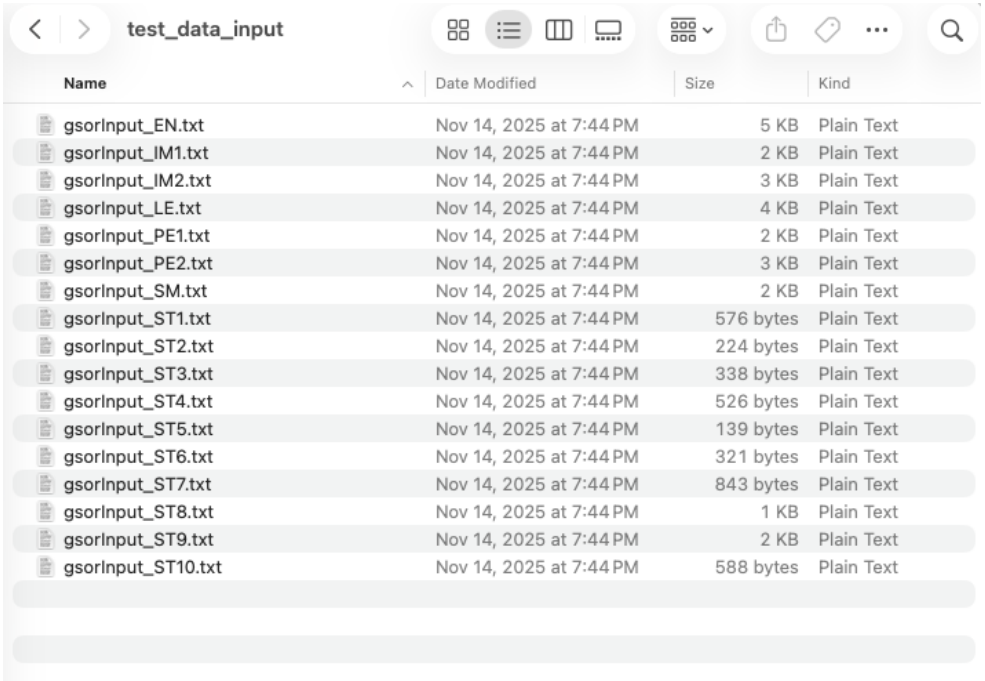
- https://biocoreapps.bsd.uchicago.edu/gsea_shiny/



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis application

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

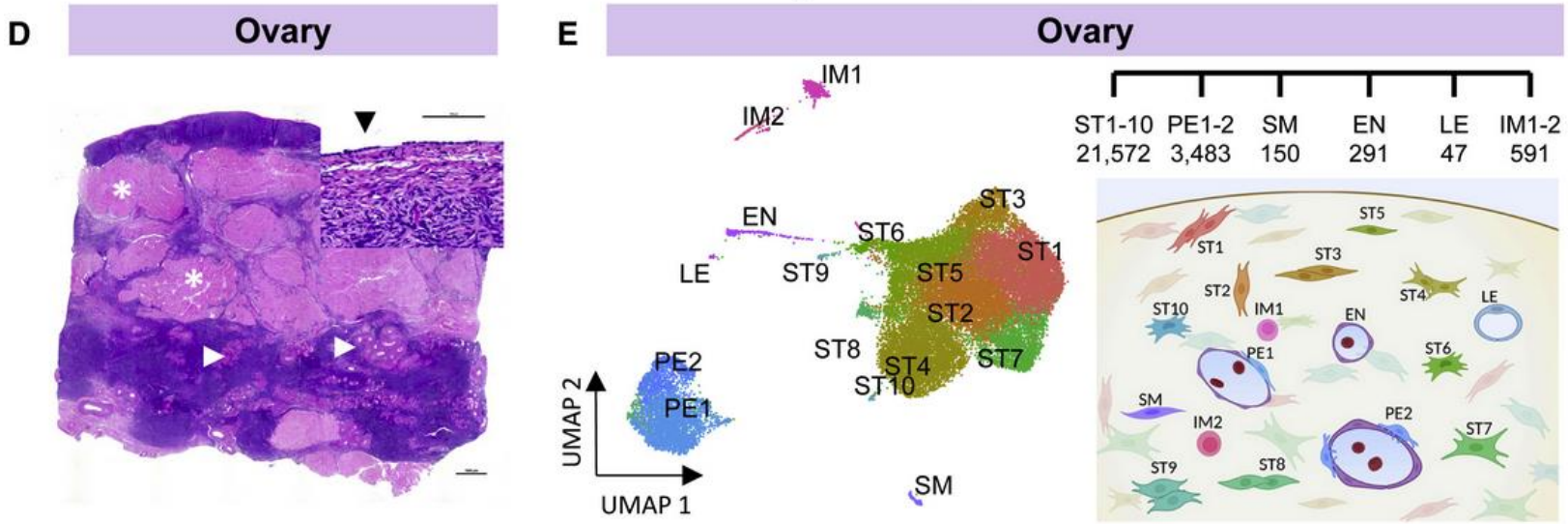


- A molecular atlas of the human postmenopausal fallopian tube and ovary from single-cell RNA and ATAC sequencing, Lengyel E, Li Y et al., Cell Reports 2022

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis application

- ovary scRNA-seq data



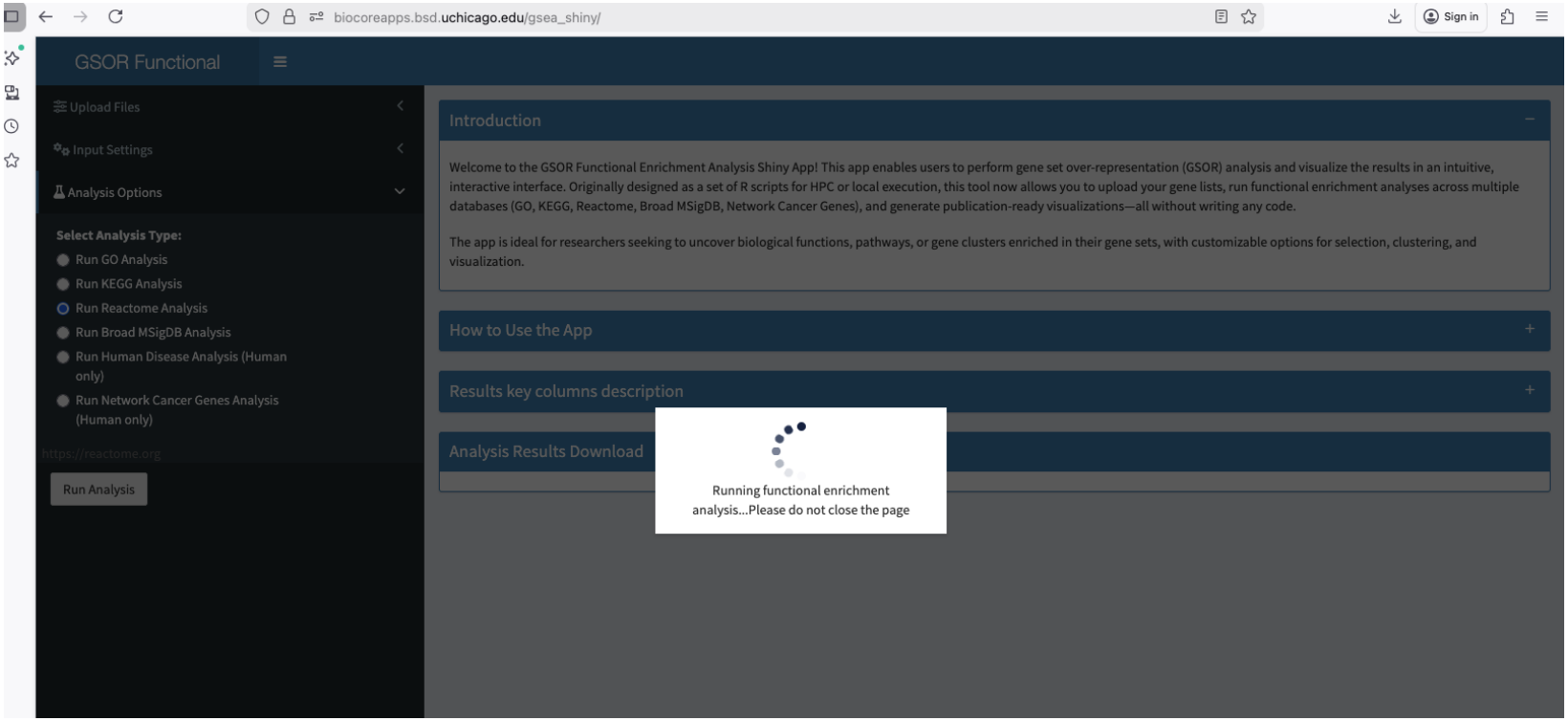
test_data_input

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/

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

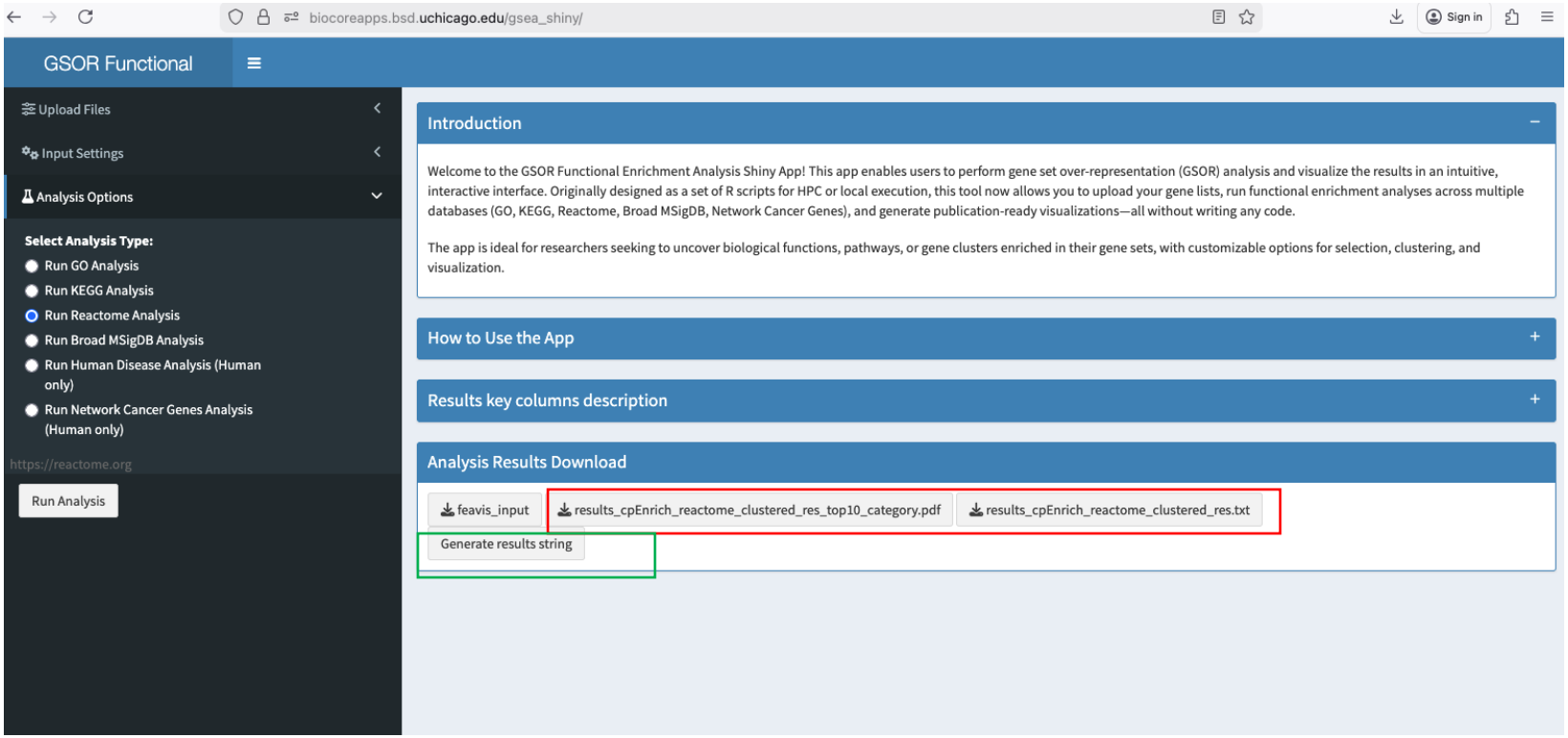
Hands-on: GSOR analysis application



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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

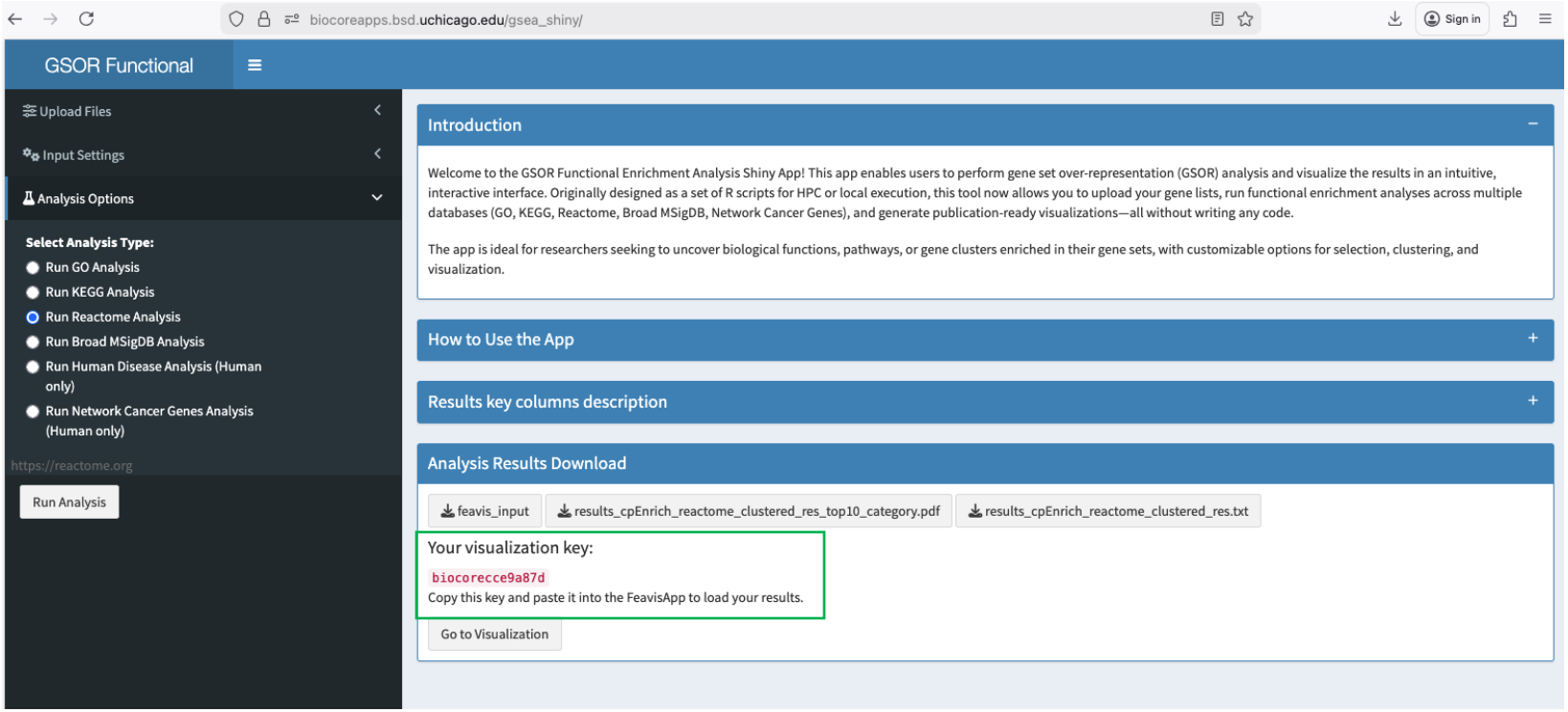
Hands-on: GSOR analysis application



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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis application

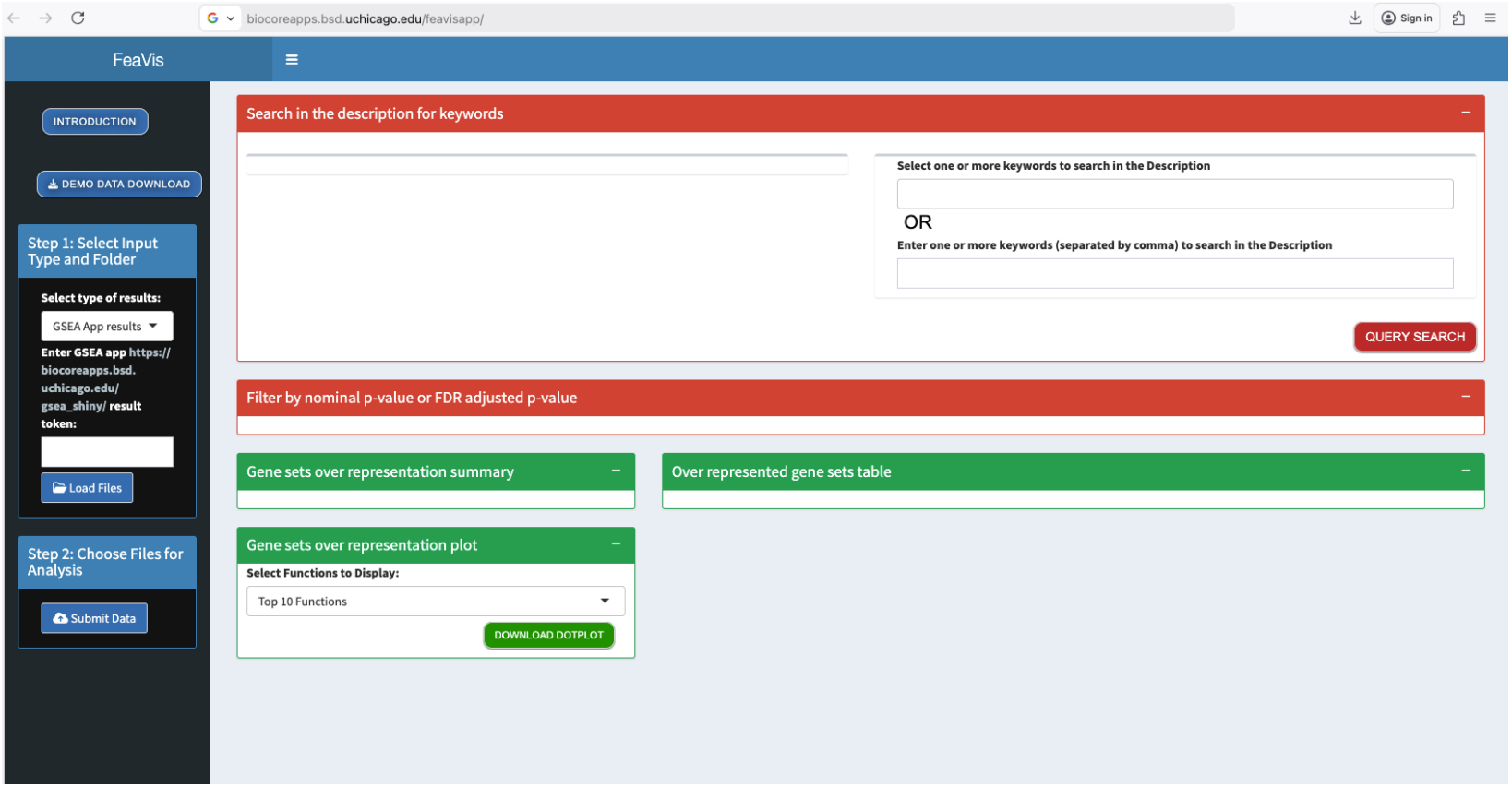


- Red: analysis results to download
- Green: 'generate results string' button click for visualization application
  - Red color string is your token used for visualization application at https://biocoreapps.bsd.uchicago.edu/feavisapp/

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

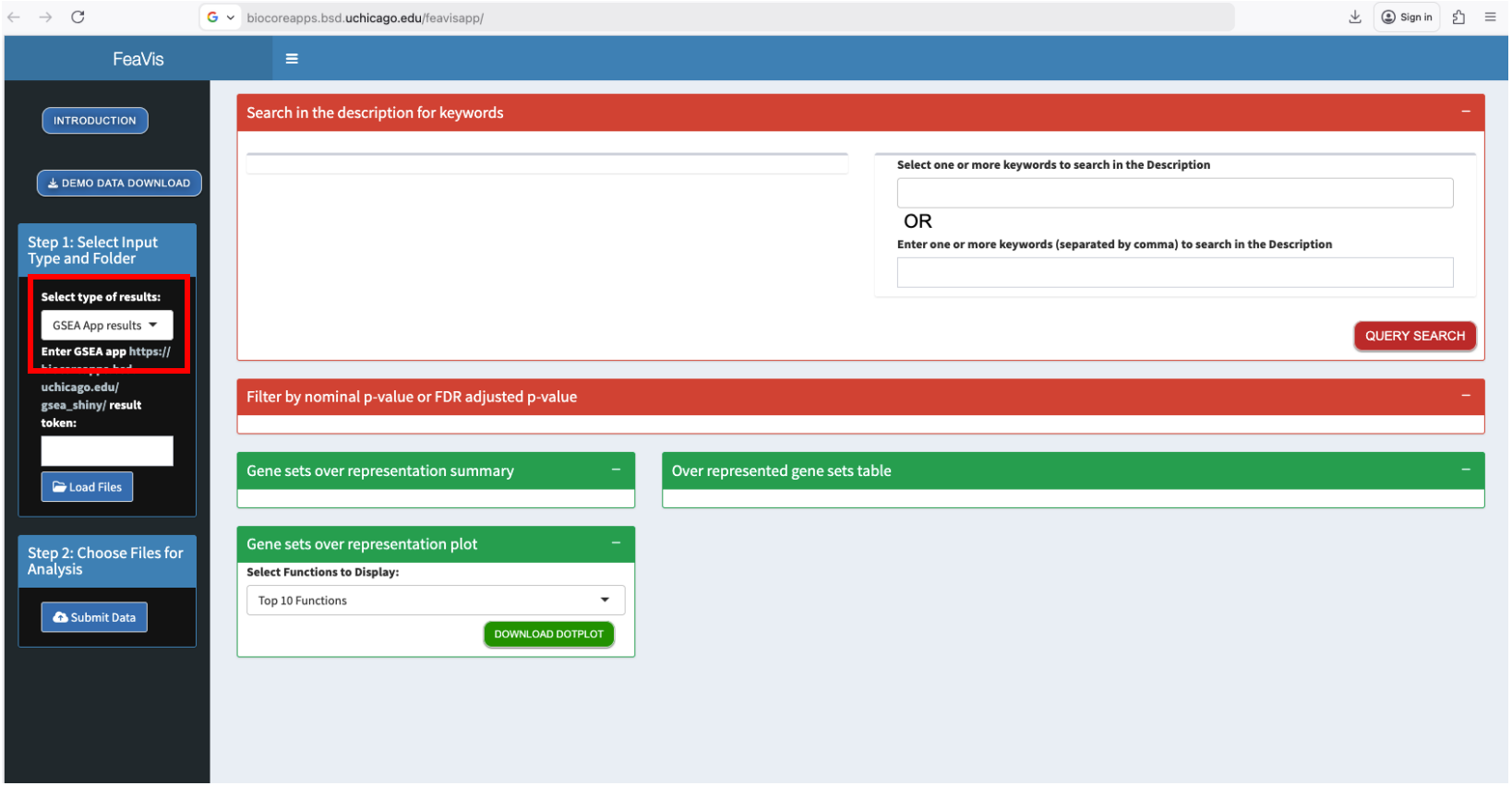
Hands-on: GSOR visualization application

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



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

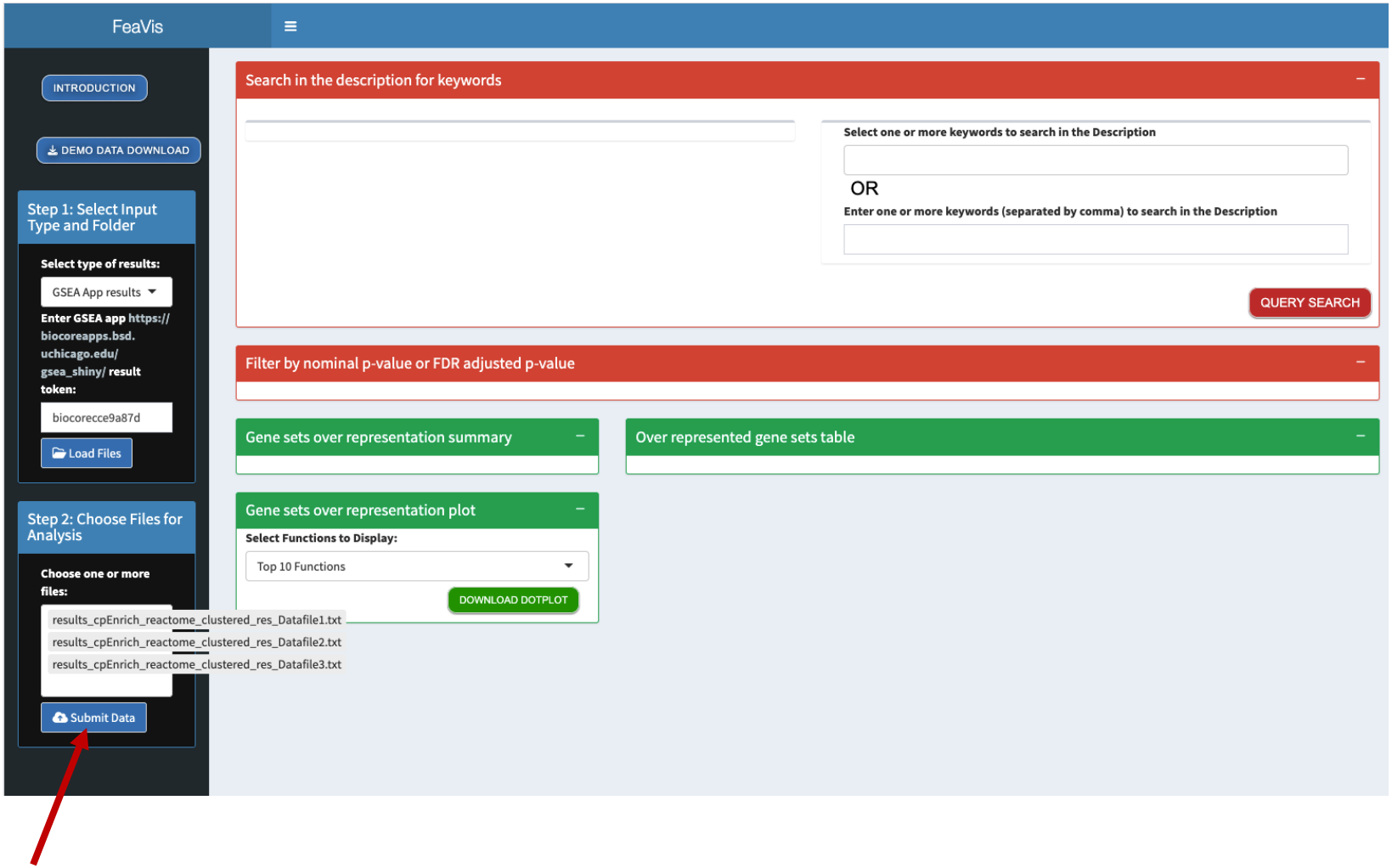
Hands-on: GSOR visualization application



- 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

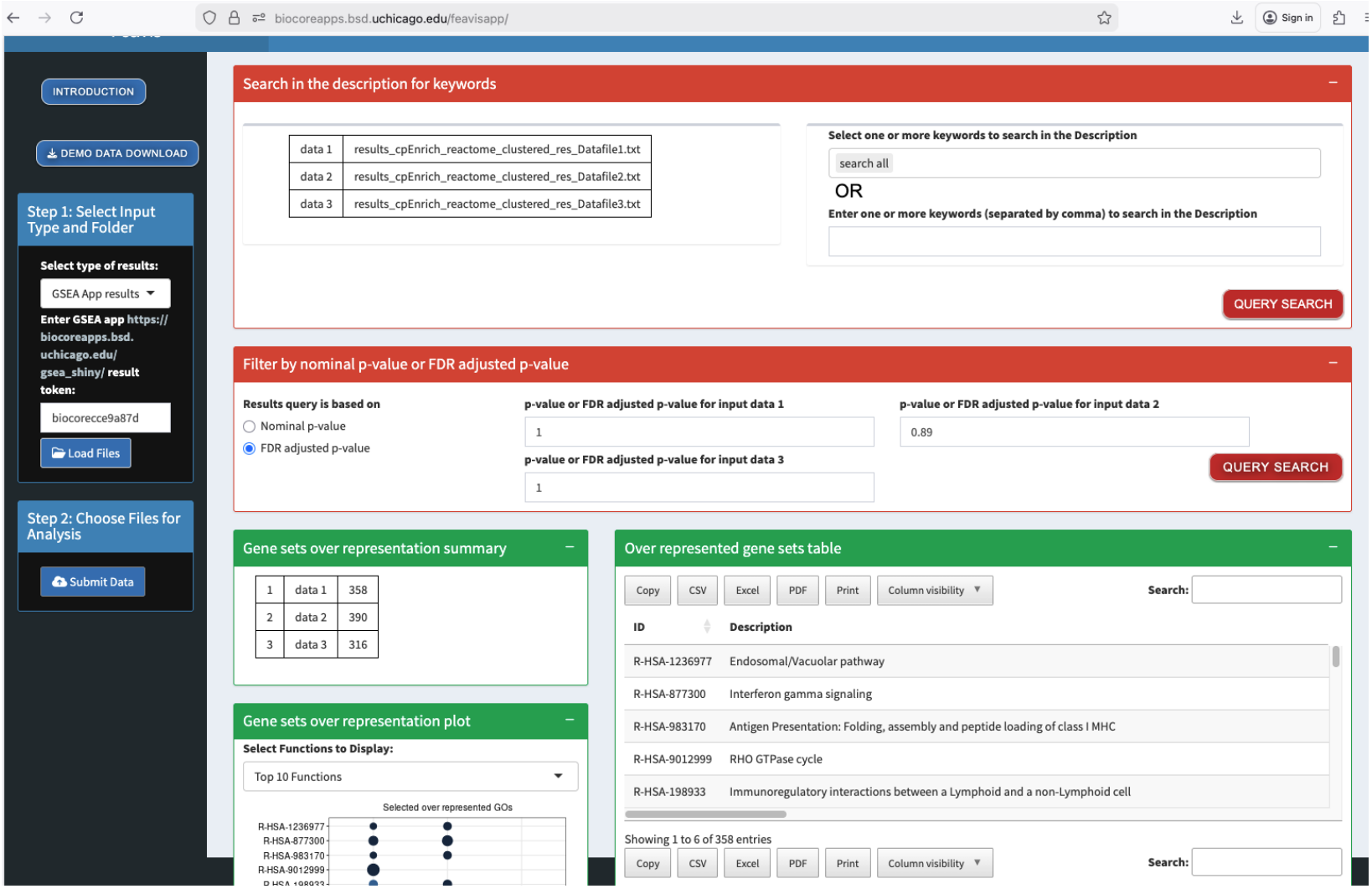
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



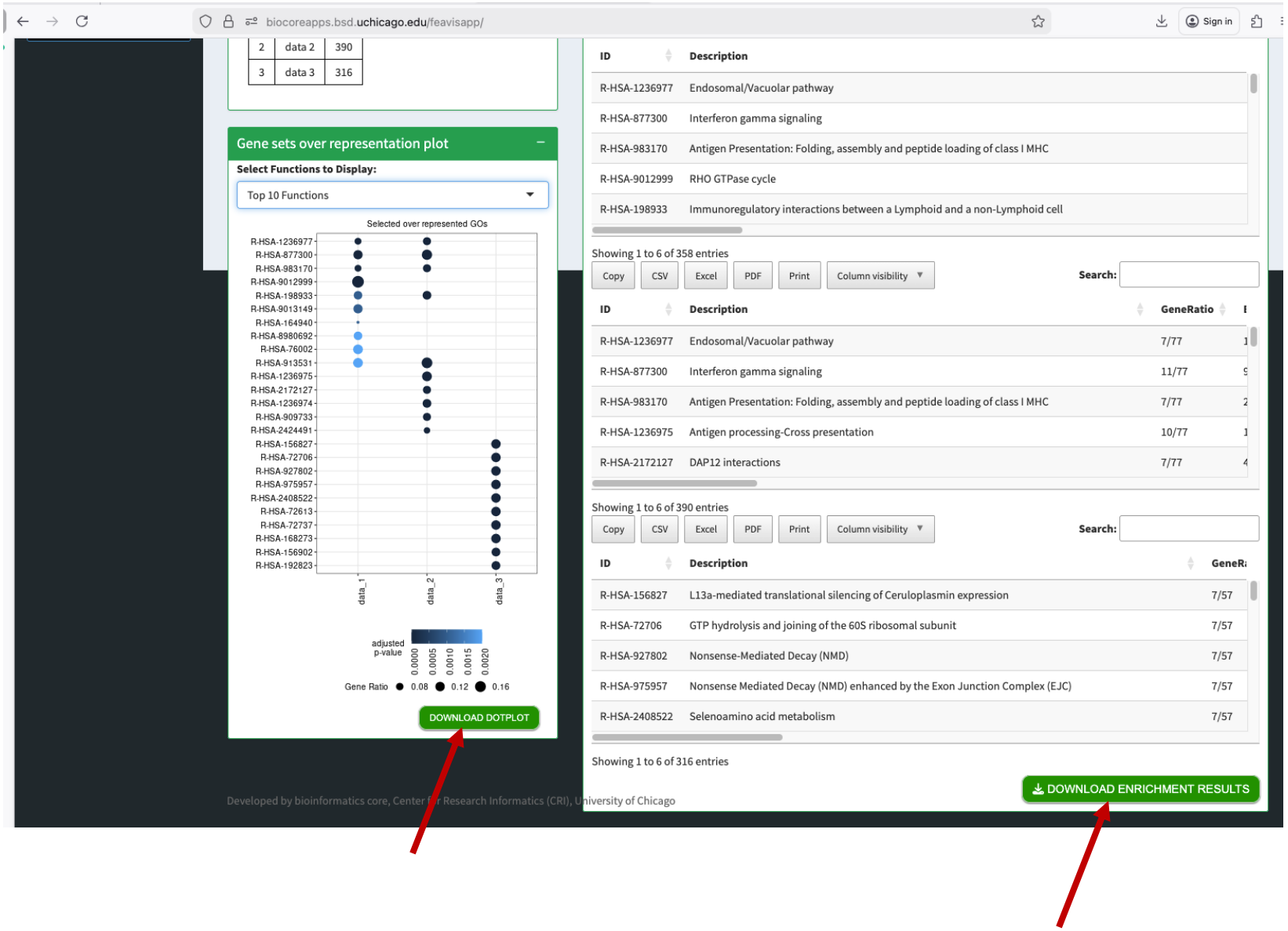
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



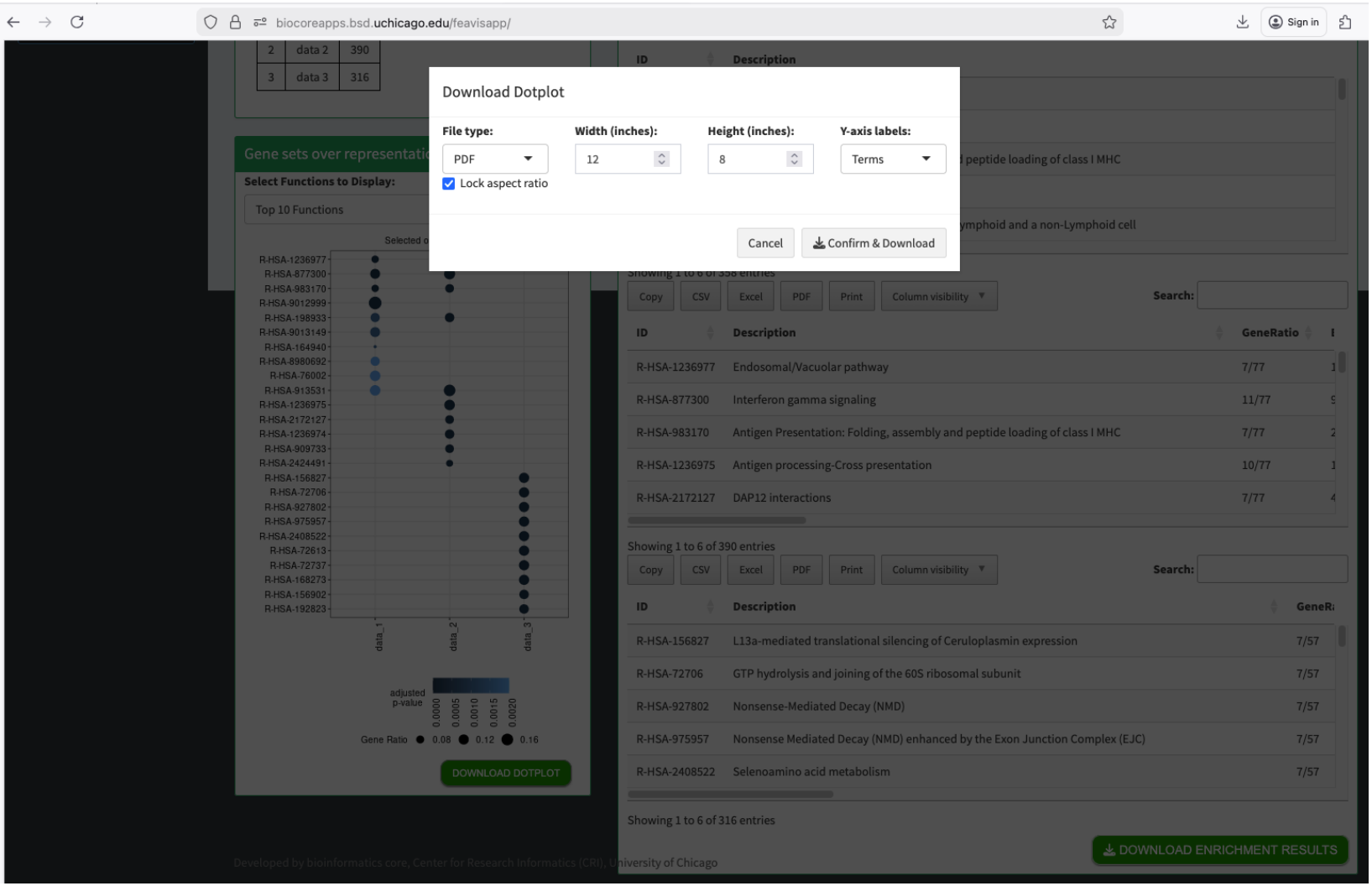
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



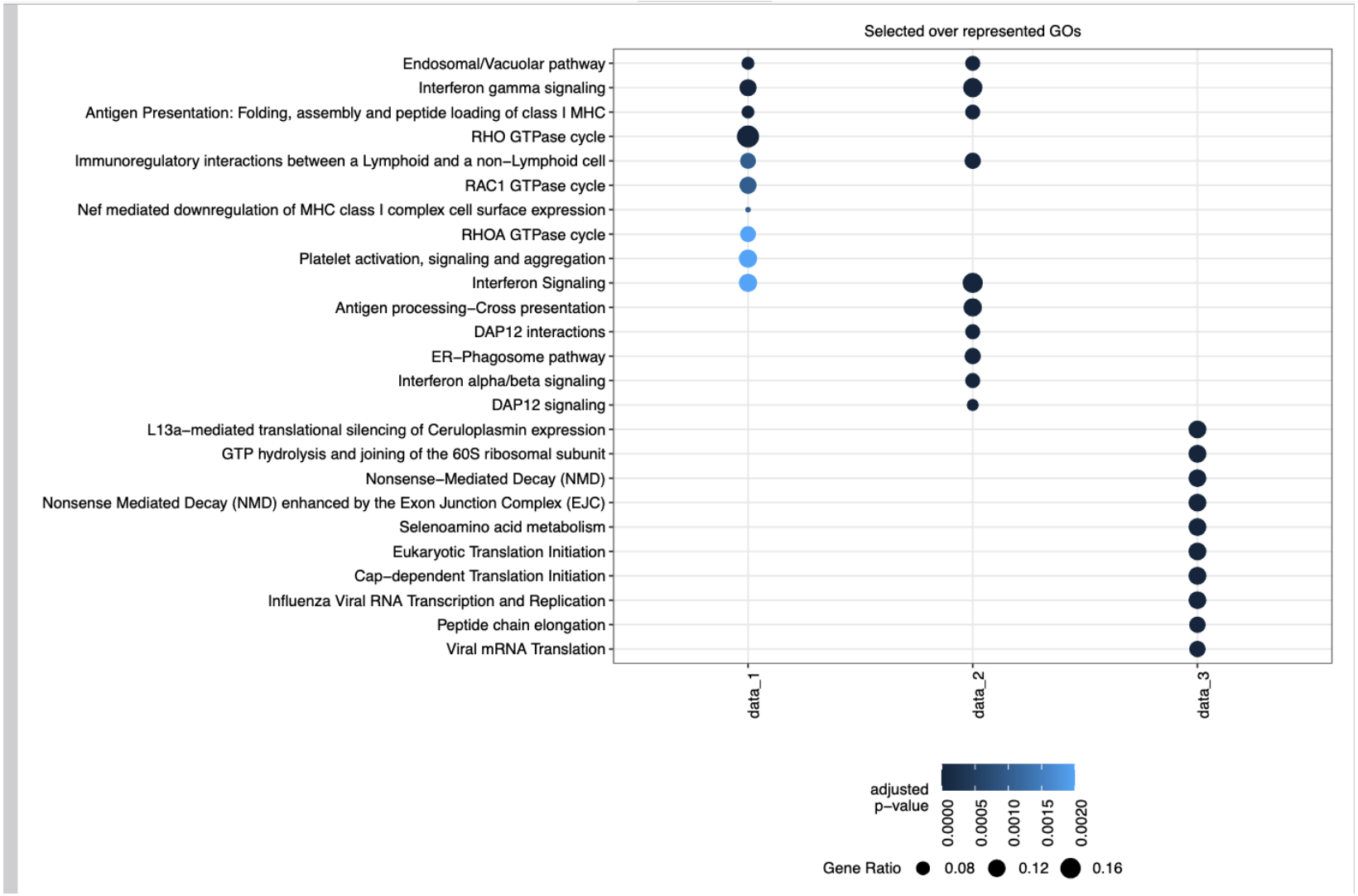
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



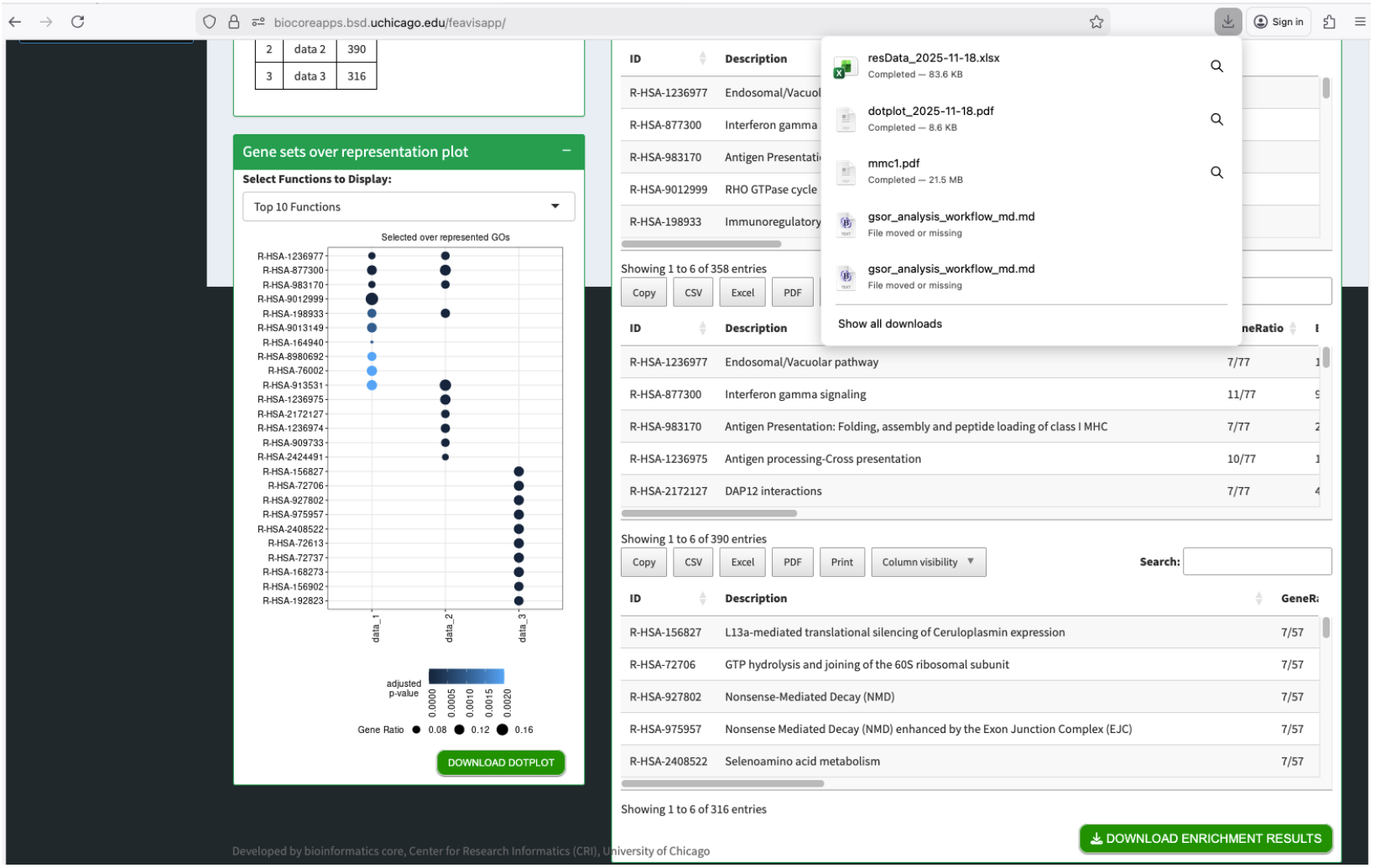
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



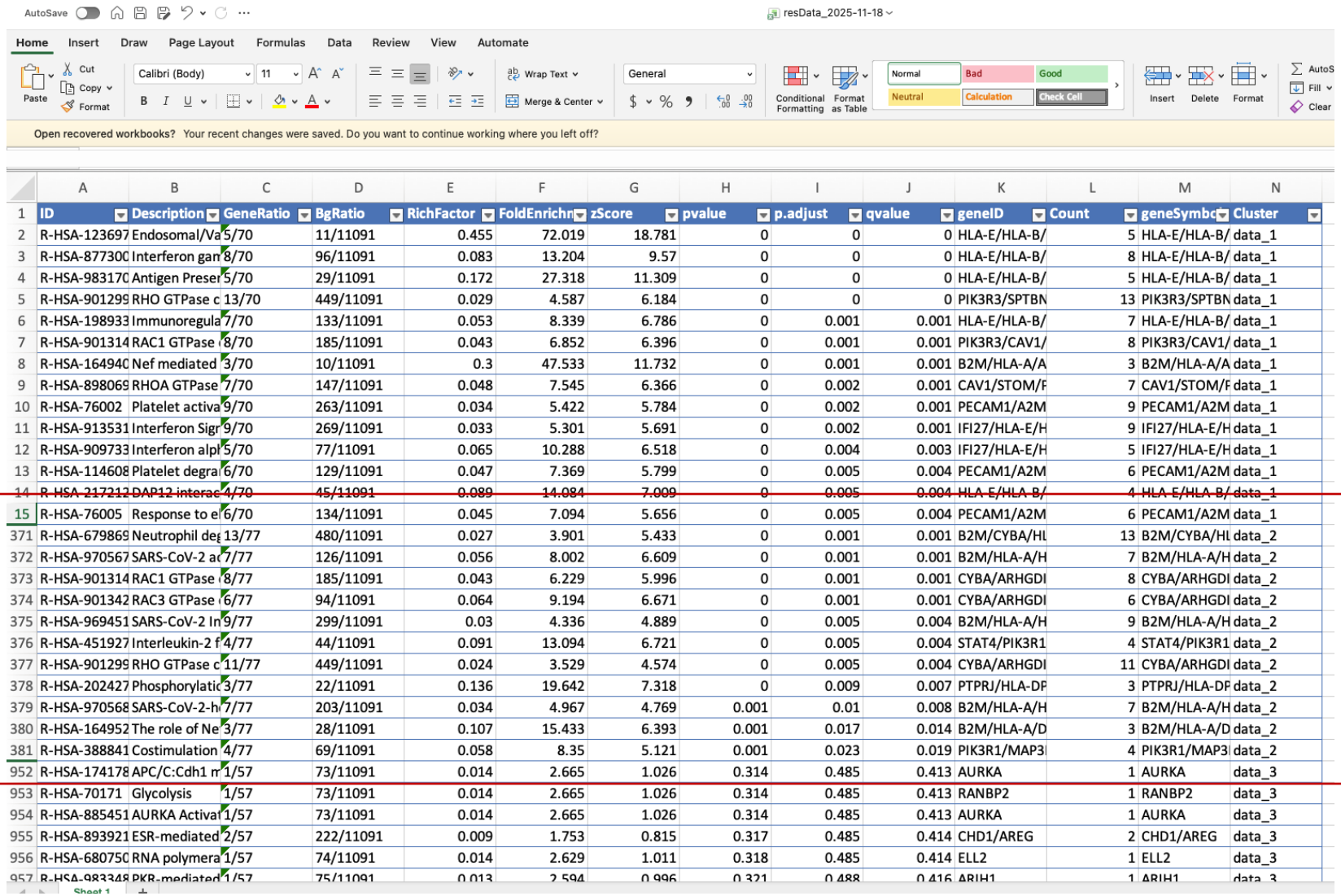
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



AutoSave resData_2025-11-18

Home Insert Draw Page Layout Formulas Data Review View Automate

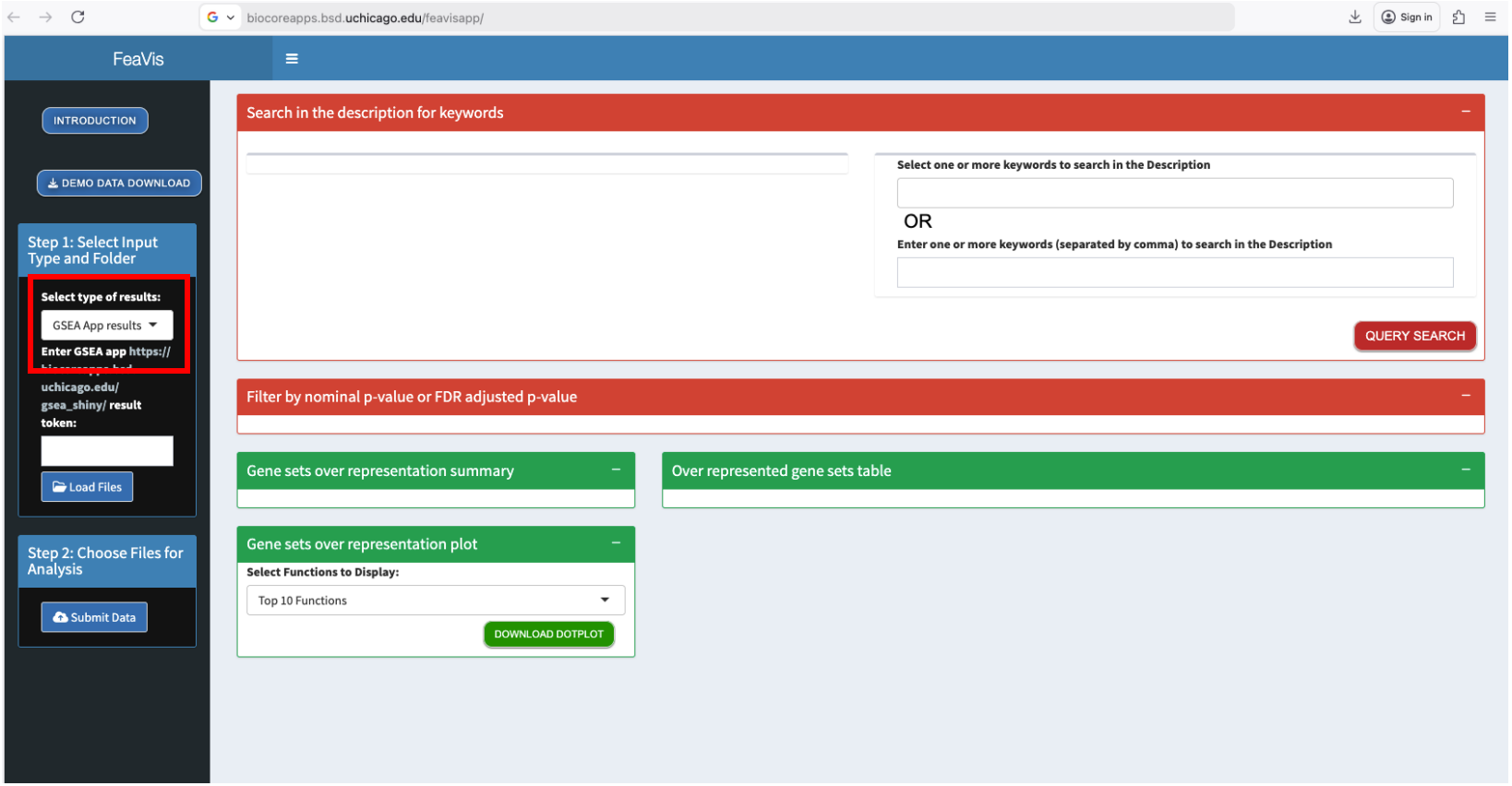
Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

	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	qvalue	geneID	Count	geneSymbo	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-877300	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-983170	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-164940	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 Sigr	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 alpl	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/HL	13	B2M/CYBA/HL	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	Phosphorylatic	3/77	22/11091	0.136	19.642	7.318	0	0.009	0.007	PTPRJ/HLA-DP	3	PTPRJ/HLA-DP	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 m	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-680750	RNA polymera	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

Sheet 1

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR visualization application



- 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**

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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)
------------------------------
```

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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 ~]$
```

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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 ~]$
```

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis randi

3. Check Network Interfaces

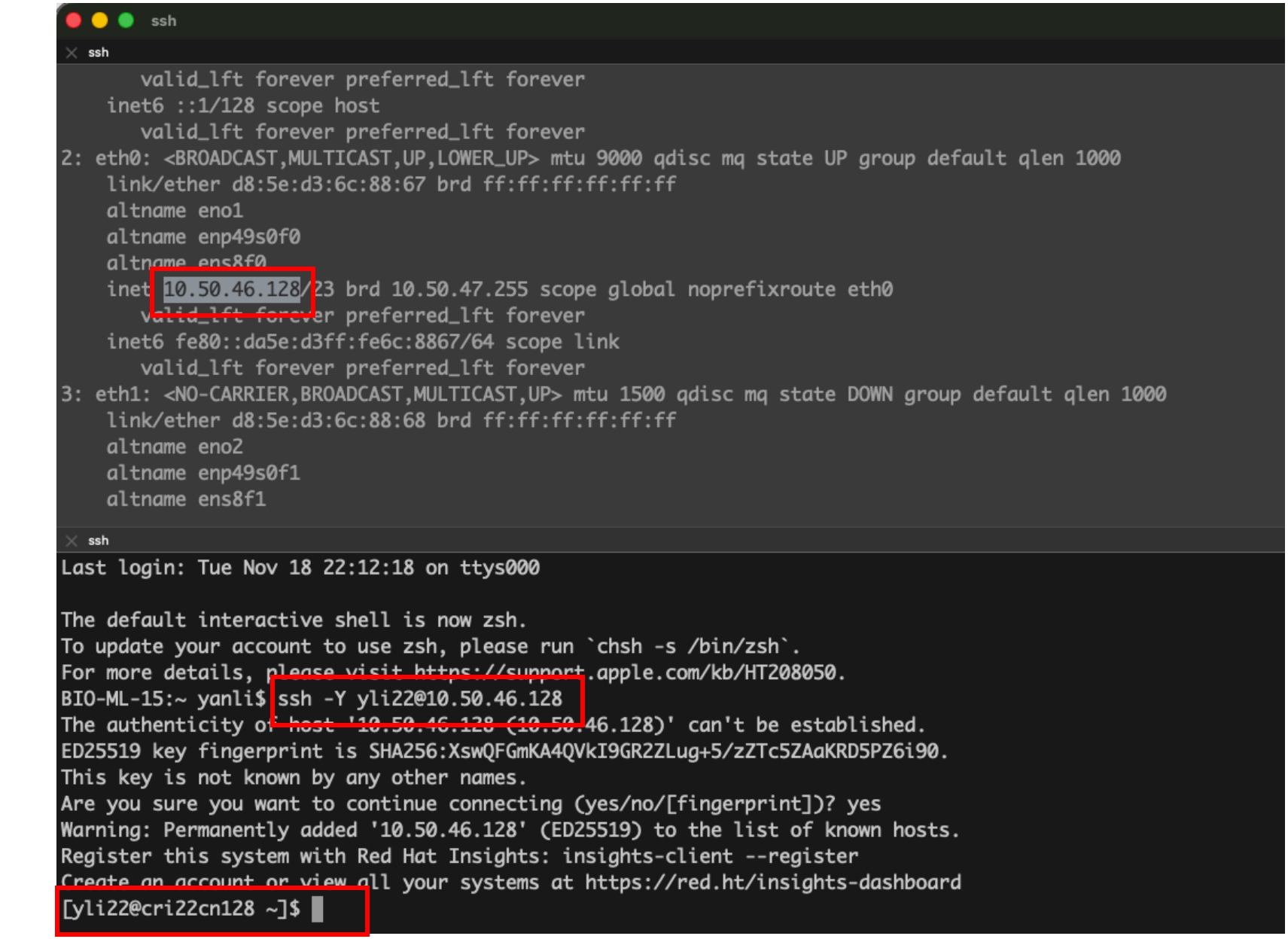
```
ip a
```

- 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
    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: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
    inet6 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:80:00:00:00:00:00:00:00:08:c0:eb:03:00:75:81:6a brd 00:ff:ff:ff:ff:12:40:1b:ff:ff:00:00:00:00:00:00: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
    inet6 fe80::ac0:eb03:75:816a/64 scope link
       valid_lft forever preferred_lft forever
[yli22@cri22cn008 ~]$
```

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis randi



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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 ~]$
```

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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.

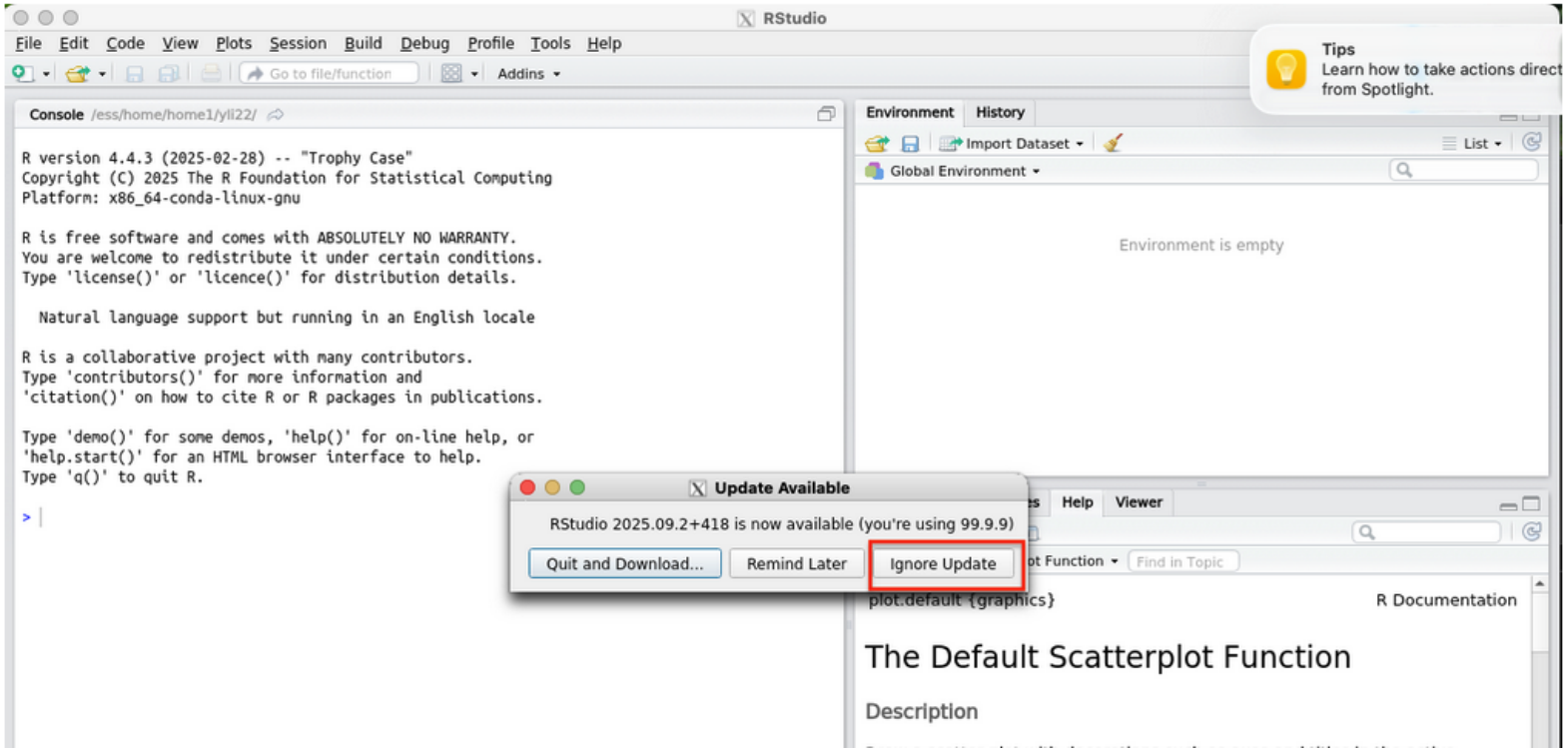
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis randi

7. Launch RStudio

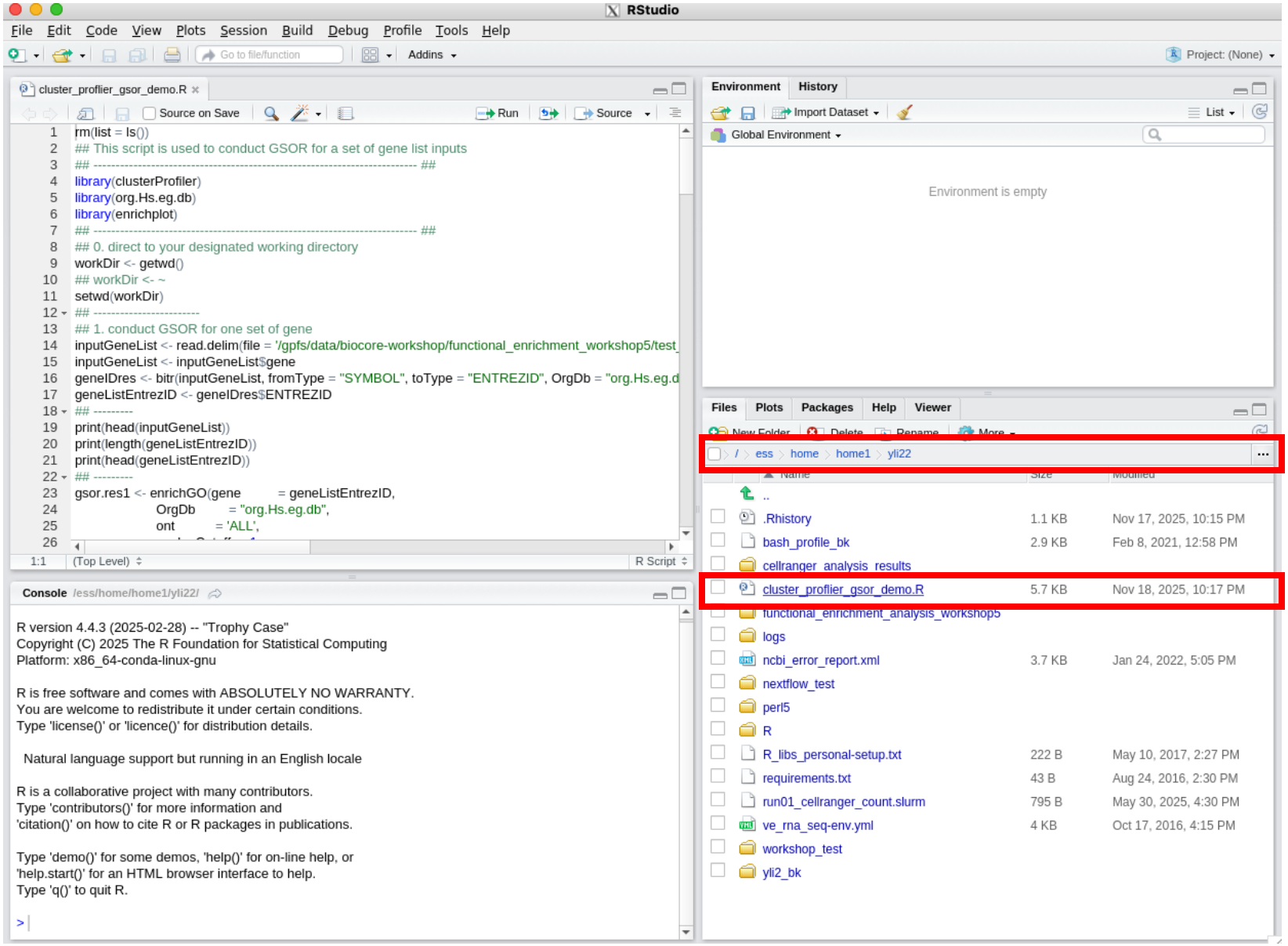
```
rstudio
```

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



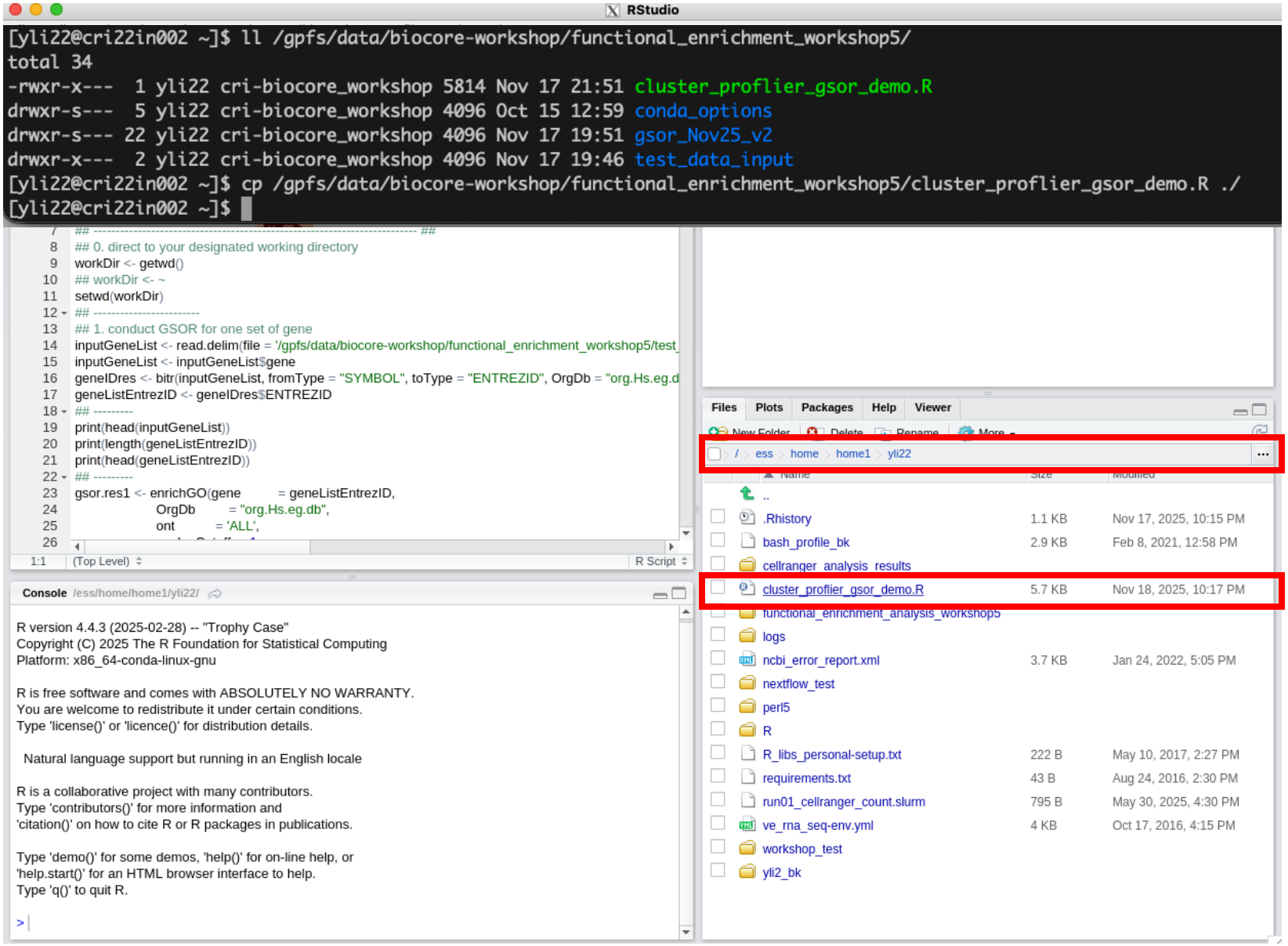
https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis randi



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Hands-on: GSOR analysis randi



https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

Take Home Message

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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

OUR AWESOME TEAM

Mengjie Chen, PhD

Faculty Director

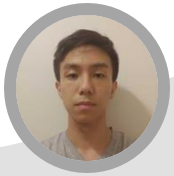
Wenjun Kang, MS

Technical Director

Jason Shapiro, PhD

Bioinformaticians

Diana Vera Cruz, PhD



Evan Wu

Katie Aracena, PhD

David Tieri, PhD

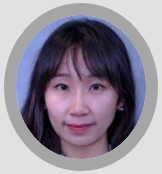
Yan Li, PhD

Associate Director

Houxiang Zhu, PhD

Qiaoshan Lin, PhD

Yildiz Koca, PhD



Zhongyu Li, MS

Geetha Priyanka, MS

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/

https://cri.uchicago.edu/bioinformatics

Biological Sciences Bioinformatics Core

THE UNIVERSITY OF CHICAGO

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/

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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-2kghy4392-1ELPfgn8pL5BcXk4oF9D4g

- 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

https://cri.uchicago.edu/bioinformatics
Biological Sciences Bioinformatics Core
THE UNIVERSITY OF CHICAGO

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

Workship feedback survey:
https://mycri.cri.uchicago.edu/educations/trainings/79/survey/