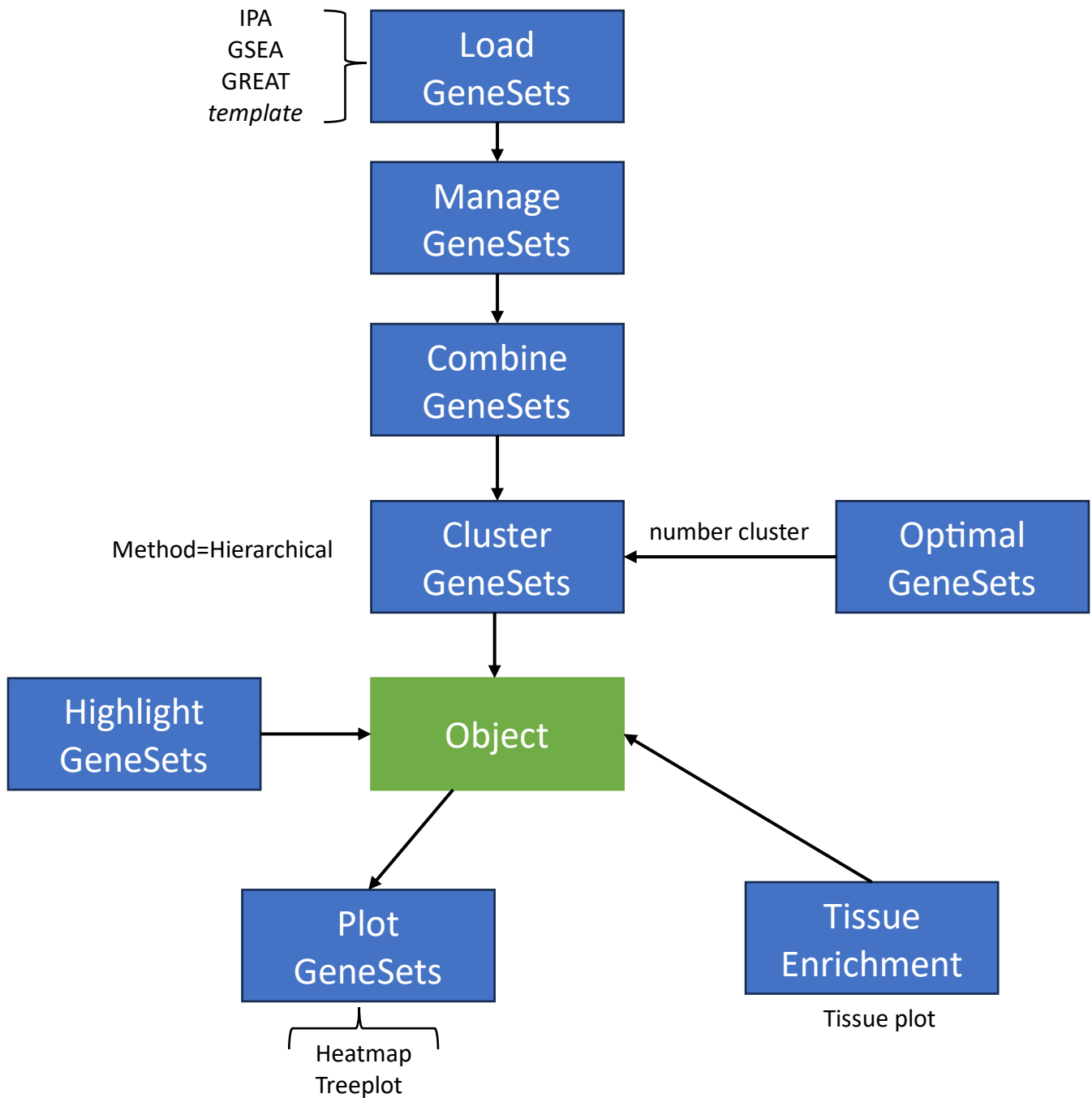


Table of Contents

1. General	2
2. Inputs	4
2.1 Source:	4
2.2 Gene ID:	4
2.3 Organism:	4
2.4 Upload files:	4
2.5 Run example:	4
3. Outputs	5
3.1 Plots:	5
3.2 Data:	5
4. Additional functionalities	8
4.1 HighlightGeneSets:	8
4.2 Data filtering:	8
4.3 Recalculate clustering:	9
4.4 Downloads:	9

1. General

This is a schematic representation of the workflow implemented by default in the shiny:



Here you can see some general screenshots:

[Main](#) [Help](#)

[Download template](#) [Run example](#)

Source

☒ GREAT ☐ IPA ☐ GSEA ☐ Template

Gene ID

☐ Ensembl ID ☒ Symbol ☐ Entrez ID

Organism

☐ Homo sapiens ☒ Mus musculus

Upload files

[Browse...](#) No file selected

name	group
MM10.GREAT.KO.uGvsMac.bed.tsv	KO
MM10.GREAT.WT.uGvsMac.bed.tsv	WT

[Run analysis](#) [Reset](#)

Heatmap [Treeplot](#)

[Download data](#) [Download image](#)

Clustering

Set number of clusters:

[Recalculate](#)

☒ Cluster_1 ☒ Cluster_2

Data [ORA](#) [Genes](#)

Show entries

Search:

ID	Term	Type	Group	Cluster	Pval	Ratio	Definition
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	GO Biological Process	KO	Cluster_1	1.632557e-06	0.6	The process in which an antigen-presenting cell expresses a peptide antigen of exogenous origin on its cell surface in association with an MHC class II protein complex. The peptide antigen is typically, but not always, processed from a whole protein.
GO:0002478	antigen processing and presentation of exogenous peptide antigen	GO Biological Process	KO	Cluster_1	0.0001018555	0.375	The process in which an antigen-presenting cell expresses a peptide antigen of exogenous origin on its cell surface in association with an MHC protein complex. The peptide is typically a fragment of a larger exogenous protein which has been degraded within the cell.
GO:0048002	antigen processing and presentation of peptide antigen	GO Biological Process	KO	Cluster_1	6.307226e-05	0.233	The process in which an antigen-presenting cell expresses peptide antigen in association with an MHC protein complex on its cell surface, including proteolysis and transport steps for the peptide antigen both prior to and following assembly with the MHC protein complex. The peptide antigen is typically, but not always, processed from an endogenous or

Showing 1 to 25 of 522 entries

Previous

1

2

3

4

5

...

21

Next

Mouse databases:

[Calculate](#)

2. Inputs

All the following input fields are mandatory:

The screenshot shows a web application interface for data input. At the top, there are two buttons: 'Download template' (green) and 'Run example' (blue). Below these are four input sections, each with an information icon (i):

- Source:** Radio buttons for GREAT (selected), IPA, GSEA, and Template.
- Gene ID:** Radio buttons for Ensembl ID, Symbol (selected), and Entrez ID.
- Organism:** Radio buttons for Homo sapiens and Mus musculus (selected).
- Upload files:** A 'Browse...' button and a file list showing '2 files'. Below the list is a progress bar labeled 'Upload complete'.

Below the upload section is a table with two columns: 'name' and 'group'. It contains two rows of data:

name	group
MM10.GREAT.KO.uGvsMac.bed.tsv	1
MM10.GREAT.WT.uGvsMac.bed.tsv	2

At the bottom, there are two buttons: 'Run analysis' (with a magnifying glass icon) and 'Reset' (with a circular arrow icon).

2.1 Source: Choose the data source for your analysis from three options: GREAT, IPA or GSEA. In the case of using another gene set enrichment tool, a template Excel can be downloaded (clicking *Download template* button), and uploaded into the app by selecting *Source = Template*. All the following template's fields (columns) have to be filled:

ID: gene set id

Count: number of your genes mapped in this gene set

GeneRatio: "Count" column value divided by the total of genes of the gene set.

p.adjust: adjusted p-value

geneID: genes id (Ensembl ID, Symbol or Entrez ID) separated by comma

2.2 Gene ID: Specify the format of gene representation in your dataset.

2.3 Organism: Either Homo sapiens or Mus musculus

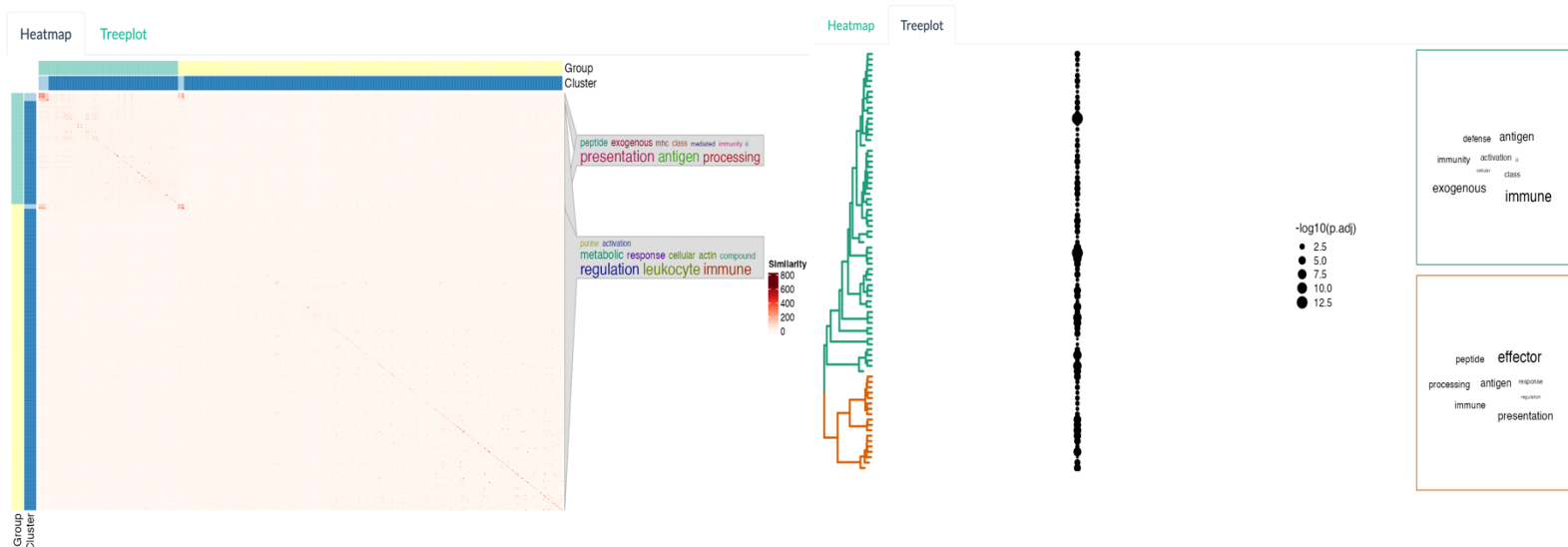
2.4 Upload files: Upload your data files, which must be in .txt, .csv, or .xls format. Once uploaded, a table will appear below. By default, each file represents a separate group (1, 2, etc.). You can modify the group names by double-clicking on the respective cell.

2.5 Run example: Show an example of the application using GREAT dataset

3. Outputs

During execution, pop-up notifications will notify you through the progress of the analysis.

3.1 Plots:



- **Heatmap:** Visualize the RR matrix as a heatmap, annotated with cluster and group info.
- **Tree plot:** Display the RR matrix as a tree plot, annotated with cluster and group info.
Note: only tree plot is performed with size < 200.

3.2 Data:

- **Data table:** View your original input data with an added column for cluster information. For inputs from GREAT or GSEA, a "Description" column provides GO term descriptions. By clicking in "ID", a hyperlink will direct you to the QuickGO database for further details

Data

ORA

Genes

Tissue enrichment

Show

25

entries

Search:

ID	Term	Type	Group	Cluster	Pval	Ratio	Definition
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	GO Biological Process	1	Cluster_1	1.63e-06	0.6	The process in which an antigen-presenting cell expresses a peptide antigen of exogenous origin on its cell surface in association with an MHC class II protein complex. The peptide antigen is typically, but not always, processed from a whole protein.

- **ORA:** Explore ORA results per cluster, showing the top 5 enrichments by default (also top 10, 15, and 20 are allowed). Clicking on the “ID” column links to QuickGO.

Data **ORA** Genes Tissue enrichment

Top per cluster:

Show entries

Search:

ID	Description	GeneRatio	BgRatio	pvalue	p.adjust	qvalue	Count	Cluster
GO:0048002	antigen processing and presentation of peptide antigen	15/23	79/28943	0	0	0	15	Cluster_1
GO:0019882	antigen processing and presentation	15/23	128/28943	0	0	0	15	Cluster_1
GO:0002440	antigen processing and presentation	15/23	128/28943	0	0	0	15	Cluster_1

- **Gene:** Gene frequency distribution per cluster. For instance, if Cluster_1 has 10 pathways and a gene appears in only one pathway, its frequency is $1/10=0.1$. Filter genes and frequencies as needed. Clicking on a gene name leads to GeneCards (<https://www.genecards.org/>) for additional information, in case of human data, otherwise to Mouse Genome Informatics (<https://www.informatics.jax.org/>)

Data ORA **Genes** Tissue enrichment

Table of gene frequency distribution per cluster:

Show entries

gene	Cluster_1 (Pathways=10)	Cluster_2 (Pathways=98)
<input type="text" value="Cd74"/>	<input type="text" value="0.4 ... 1.0"/>	<input type="text" value="All"/>
Cd74	<div><div></div></div>	0.694

Showing 1 to 1 of 1 entries (filtered from 2,590 total entries)

Previous Next

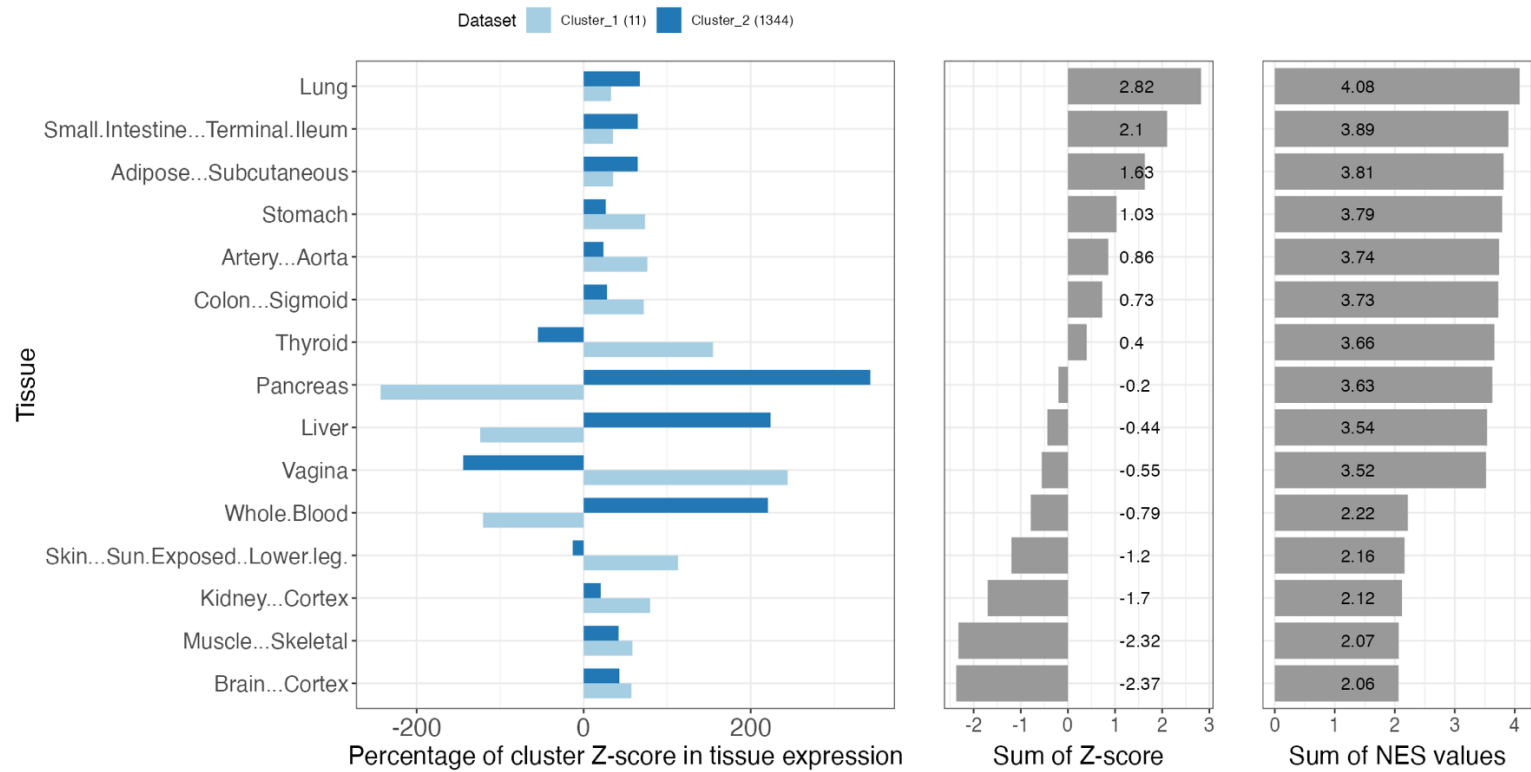
- **Tissue enrichment** (optional, for Organism=Homo sapiens): Calculate tissue enrichment analysis per cluster using 15 representative tissues from <https://gtexportal.org/home/>. Click "Calculate" to execute. View results as an NES value table and a corresponding plot. The value in the column next to the cluster name indicates the number of genes in this cluster.

Tissue enrichment results:

Show25entries

Search:

Tissue	Cluster_1..11.	Cluster_2..2335.
Adipose...Subcutaneous	1.517	2.314
Artery...Aorta	1.554	2.263
Brain...Cortex	0	2.15
Colon...Sigmoid	1.464	2.253
Kidney...Cortex	0	2.24
Liver	1.461	2.181



4. Additional functionalities

4.1 HighlightGeneSets: Select a gene of interest from the Mammalian Phenotype (Mus musculus) or Human Phenotype Ontology (Homo sapiens). View the corresponding scores per cluster in a table below.

Calculate highlight gene sets per cluster:

Mouse databases:

Mammalian Phenotype (MP) ▼

Choose MP (info):

MP:0000600_liver hypo... ▼

liver hypoplasia

Rb1 Hgf Tsc2 Nf1 Lhx2 Sptbn1 Sp3 Spei

Q Calculate

Cluster	Mean
Cluster_1	0
Cluster_2	0.006

4.2 Data filtering: Focus on specific clusters by filtering out the rest. Unselecting clusters will hide their associated data.

☐ Cluster_1 ☒ Cluster_2

Data ORA Genes Tissue enrichment

Show 25 ▼ entries

ID	Term	Type	Group	Cluster	Pval	Rat
GO:0048583	regulation of response to stimulus	GO Biological Process	1	Cluster_2	3.77e-14	0

4.3 Recalculate clustering: Modify the default number of cluster set by the OptimalGeneSets function. All plots and data will update corresponding to the new value. Click “Recalculate” to execute it.

Clustering

Set number of clusters:

🔍 Recalculate

4.4 Downloads: downloads plots in png, jpg, or pdf format (in a zip), as well as raw data with cluster and RR info in .csv format.

