

Development of a bioinformatics platform for analysis of transcriptomics and quantitative proteomics data: OMnalysis

User manual

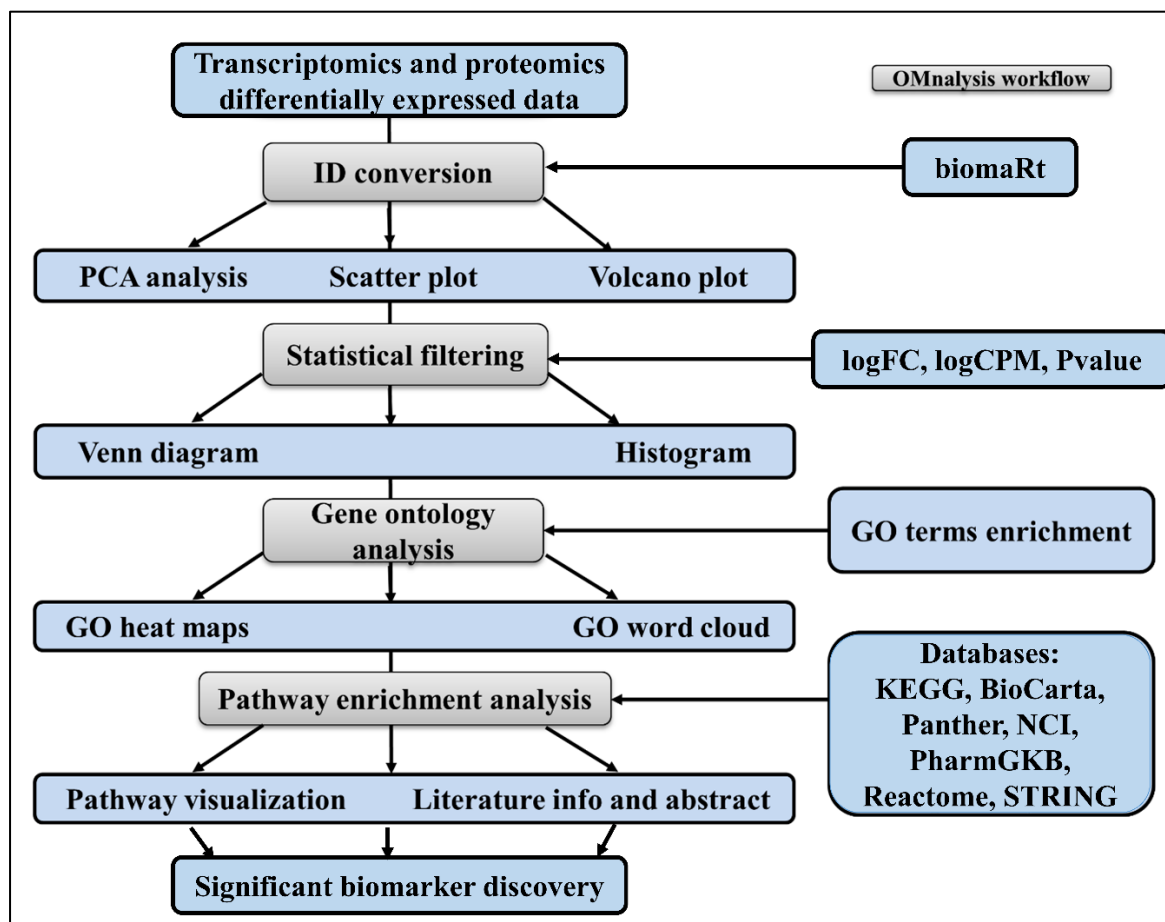
Contents

Introduction.....	3
Instructions.....	4
Data format	4
Upload data	6
PCA.....	8
Plots.....	9
Statistical filtering.....	10
Gene ontology (GO) enrichment analysis.....	11
GO heatmaps.....	12
Pathway enrichment analysis.....	13
Enriched pathway visualization	14
Literature info	17

Introduction

OManalysis is developed using R shiny, flexdashboard and Bioconductor packages. This tool target researchers who are new to RNA-seq technology and proteomics study and often depends on commercial vendors or core facility to sequence and analyze their data. Although exploration of the identified list of genes and proteins is also tedious and challenging (*Fig. 1*).

There are open-source R packages such as DESeq2 and edgeR that are extensively used to identify differentially expressed genes from the count data. OManalysis uses the list of genes produced from one of the differential expression analysis software edgeR's glmTreat function with count per million normalizations, significance value (P value) and log fold change value assigned to each list of a gene in the treatment (*Fig. 2*). On the other hand, label-free relative quantitative proteomics data must contain the columns of UniProt ID, FDR-adjusted P-value and Fold Change in an excel file with each sheet of an experimental condition named Treatment1, Treatment2, Treatment3 and Treatment4 (*Fig. 3*).



The app is hosted on shiny.io link <https://omanalysis.shinyapps.io/OManalysis/>

Instructions

1. You can run this app on your desktop after installing R base version 4.0.3 and R studio version 1.4.1106.
2. Once the R environment is ready, install below packages required for OManalysis.
3. Install R shiny supporting packages.
4. `install.packages(c("flexdashboard", "dplyr", "shiny", "shinydashboard", "DT", "tidyverse", "shinythemes", "tidyr", "gplots", "tibble", "gridExtra", "RColorBrewer", "slickR", "devtools", "ggbiplot", "factoextra", "ggplot2", "data.table", "VennDiagram", "fields", "wordcloud", "SBGNview", "europepmc", "shinyjs", "futile.logger", "rio", "plyr"))`
5. Install Bioconductor packages version 3.12 using `install.packages("BiocManager")`
6. `install.packages(c("AnnotationDbi", "Biobase", "BiocFileCache", "BiocGenerics", "BiocParallel", "BiocVersion", "biomaRt", "Biostrings", "clusterProfiler", "DO.db", "DOSE", "EnhancedVolcano", "enrichplot", "fgsea", "GO.db", "GOSemSim", "graph", "graphite", "IRanges", "KEGGgraph", "KEGGREST", "org.Bt.eg.db", "org.Gg.eg.db", "org.Hs.eg.db", "org.Ss.eg.db", "pathview", "qvalue", "reactome.db", "ReactomePA", "Rgraphviz", "S4Vectors", "XVector", "zlibbioc", "STRINGdb", "SPIA", "SBGNview"))`

Data format

Steps for transcriptomics: -

1. The differential data from the core facility or the commercial vendors will be in edgeR.tsv format.
2. Before going further make copies of edgeR.tsv files in one folder and rename them as per treatments.
3. Open edgeR.tsv file for each treatment in excel. Select and highlight columns belonging to ensemble ID (column A below), logCPM, log2Fold change and pvalue or padj as shown below

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		logCPM	log2FoldChange	pvalue	padj	LR	tgw.Disp	gene_name	gene_biotype	BOR_1_nc	BOR_2_nc	BOR_3_nc	NK_1_nor	NK_2_n
2	ENSG00000000003	7.348127	0.272346958	0.0326007	0.109927	4.566628	0.010863	TSPAN6	protein_cc	153.5335	126.7876	136.1564	160.6326	178.84
3	ENSG000000000419	5.905282	0.097694482	0.5866281	0.745434	0.295641	0.021116	DPM1	protein_cc	45.94695	70.75466	53.42153	80.80457	58.617
4	ENSG000000000457	3.544045	-0.11494675	0.5171282	0.694917	0.419617	0.010899	SCYL3	protein_cc	11.59277	9.313777	11.78013	11.83994	9.8618
5	ENSG000000000460	3.659625	-0.072624137	0.6814747	0.811306	0.168471	0.011396	C1orf112	protein_cc	13.1479	11.41689	9.588479	10.86345	10.083
6	ENSG000000000971	7.036359	-0.240397827	0.0516357	0.153308	3.787534	0.01004	CFH	protein_cc	142.6476	143.312	139.9918	130.1173	124.65
7	ENSG000000001036	7.008691	0.389254138	0.0009193	0.006951	10.98348	0.008857	FUCA2	protein_cc	100.9419	106.0569	108.2128	150.9898	141.05
8	ENSG000000001084	5.254396	-0.134083556	0.3562775	0.557318	0.850972	0.011936	GCLC	protein_cc	44.5332	35.30222	38.49089	39.79198	37.231
9	ENSG000000001167	4.909893	0.11829335	0.553343	0.721251	0.351362	0.024137	NFYA	protein_cc	17.24778	33.95022	27.94357	30.39326	30.472
10	ENSG000000001460	1.37171	0.053013071	0.8851381	0.937965	0.020868	0.040612	STPG1	protein_cc	1.979253	1.051556	3.287478	2.80741	1.4405

4. Delete all other columns.
5. sort column A in A-to-Z (sorting with expanded selection mode).

6. Repeat points 3 to 5 for each treatment edgeR.tsv file.
7. After sorting the information according to column A (yellow), check the first and last ensemble gene ID in all the treatments (it must be the same).



Template.xlsx

8. Open a template excel file () and paste column A (yellow) of the first treatment and name the column header as ENSEMBLGENE.
9. Copy in order log2FoldChnage column C (yellow), logCPM column B and pvalue column D (red) or padj column E (green) from each treatment excel file.
10. Paste the copied columns next to the ENSEMBLGENE in the template excel file.
11. **The master file must look like the below screenshot.** Yellow color columns belong to treatment 1, green color columns belong to treatment 2, red color columns belong to treatment 3 and blue color columns belong to treatment 4.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ENSEMBLGENE	logFC	logCPM	PValue	logFC	logCPM	PValue	logFC	logCPM	PValue	logFC	logCPM	PValue
2	ENSG000000000003	0.561283	4.557846	0.438597	-0.90616	4.557846	0.141092	-0.11978	4.557846	0.880781	-0.35996	4.557846	0.640167
3	ENSG0000000000419	0.602523	4.758842	0.402304	0.096469	4.758842	0.902259	0.001348	4.758842	0.997336	-0.78139	4.758842	0.254466
4	ENSG0000000000460	0.826338	1.782839	0.256336	-0.13903	1.782839	0.8607	0.076887	1.782839	0.923429	0.792963	1.782839	0.277646
5	ENSG0000000000971	0.780918	6.933399	0.223946	-0.32462	6.933399	0.676356	-0.21808	6.933399	0.782282	-0.62059	6.933399	0.380078
6	ENSG000000001036	0.869587	5.462408	0.151536	-0.35845	5.462408	0.642259	-0.29408	5.462408	0.7065	0.171774	5.462408	0.827248
7	ENSG000000001084	0.429974	4.100634	0.56991	-0.62329	4.100634	0.381325	0.636583	4.100634	0.36521	-0.01379	4.100634	0.986338
8	ENSG000000001167	0.413473	3.653166	0.587001	0.34253	3.653166	0.656407	0.252641	3.653166	0.746001	0.178518	3.653166	0.820939
9	ENSG000000001461	-1.51866	5.259942	0.000472	-1.37751	5.259942	0.002377	-0.57455	5.259942	0.426293	-1.1324	5.259942	0.0298
10	ENSG000000001497	0.34803	4.674989	0.652222	0.013774	4.674989	0.985861	-0.6163	4.674989	0.385452	-0.54674	4.674989	0.454072

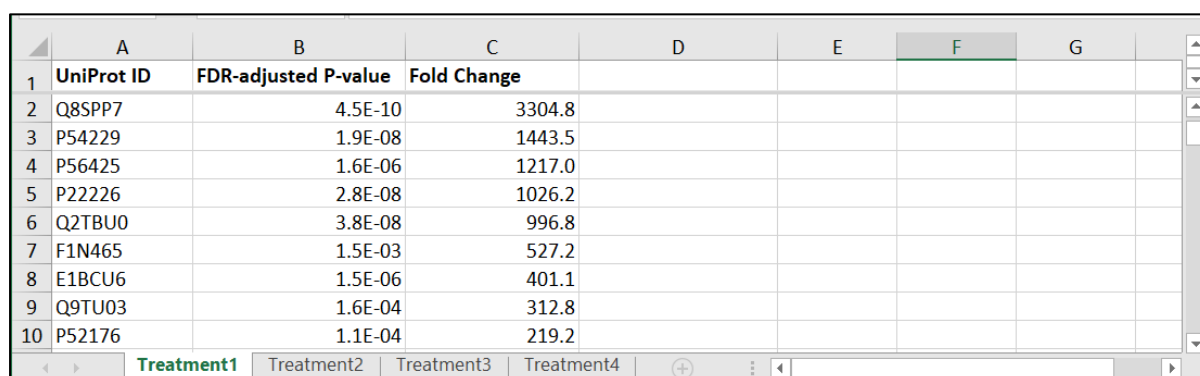
12. Column headers and the order of the columns (logFC, logCPM, Pvalue) must be the same as shown in the above screenshot.
13. The template file of transcriptomics **must be saved as CSV file** to upload to OManalysis.

Steps for Proteomics: -

1. The output from the differential proteomics analysis pipeline for each treatment is in a separate xlsx file.
2. Open each treatment file in excel.
3. Each treatment file must contain the column headers, UniProt ID, FDR-adjusted P-value and Fold Change in the same order presented in the below screenshot.

	A	B	C
1	UniProt ID	FDR-adjusted P-value	Fold Change
2	Q8SPP7	4.5E-10	3304.8
3	P54229	1.9E-08	1443.5
4	P56425	1.6E-06	1217.0
5	P22226	2.8E-08	1026.2
6	Q2TBU0	3.8E-08	996.8
7	F1N465	1.5E-03	527.2
8	E1BCU6	1.5E-06	401.1
9	Q9TU03	1.6E-04	312.8
10	P52176	1.1E-04	219.2

4. Open a new .xlsx file (master file) and name the first excel sheet as Treatment1 (Screenshot below)

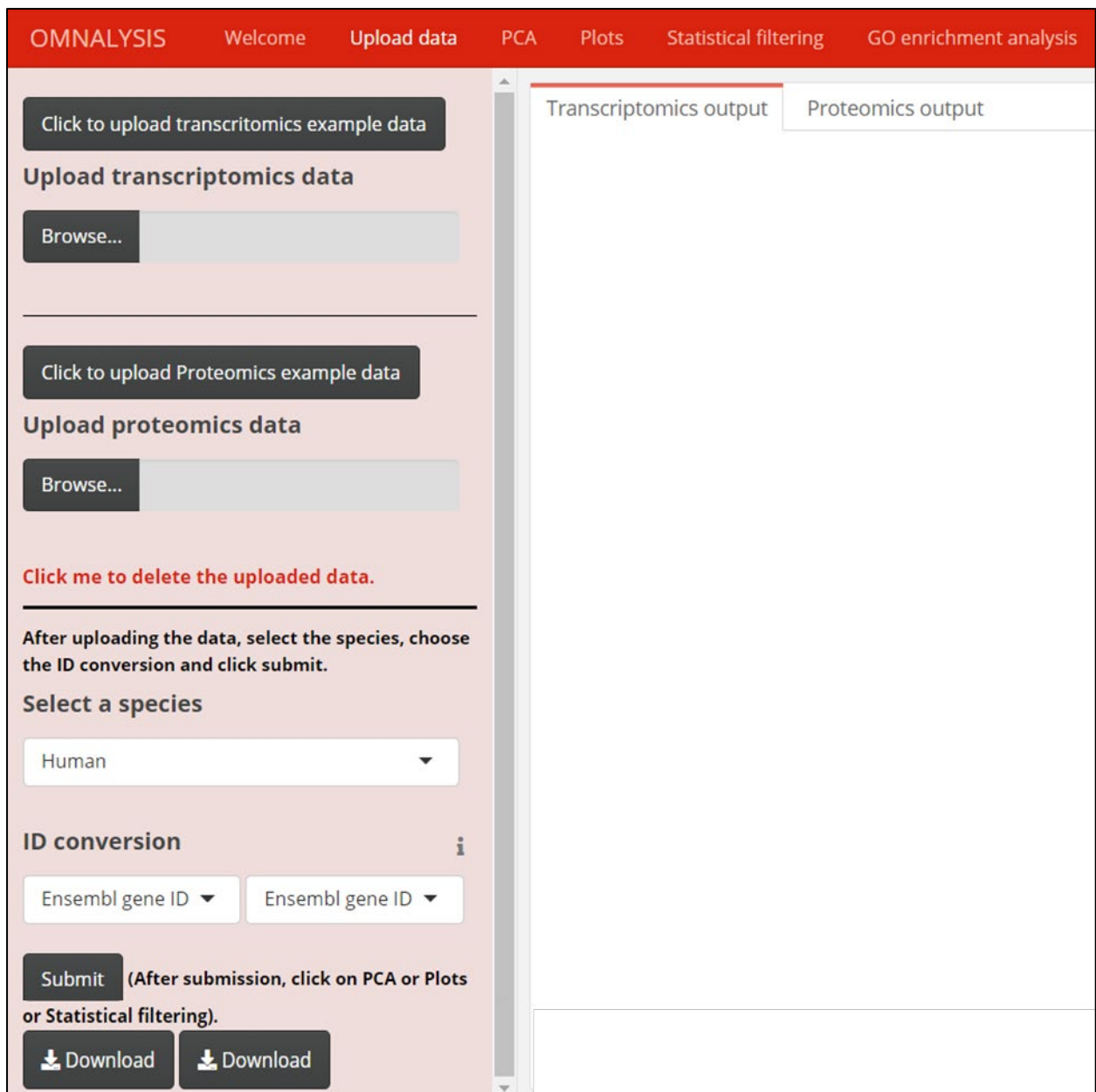


	A	B	C	D	E	F	G
1	UniProt ID	FDR-adjusted P-value	Fold Change				
2	Q8SPP7	4.5E-10	3304.8				
3	P54229	1.9E-08	1443.5				
4	P56425	1.6E-06	1217.0				
5	P22226	2.8E-08	1026.2				
6	Q2TBU0	3.8E-08	996.8				
7	F1N465	1.5E-03	527.2				
8	E1BCU6	1.5E-06	401.1				
9	Q9TU03	1.6E-04	312.8				
10	P52176	1.1E-04	219.2				

5. Paste the columns from each treatment into the master file with a new excel sheet and naming them as Treatment1 for the first treatment, Treatment2 for the second treatment, Treatment3 for the third treatment, Treatment4 for the fourth treatment.
6. Save the master file in .xlsx format.

Upload data

1. The app features can be explored using the [Click to upload transcriptomics example data tab for transcriptomics analysis](#) and [Click to upload Proteomics example data for proteomics data analysis](#).
2. The interactive panel below (left side), use the pre-loaded RNA-Seq data produced by human brain microvascular endothelial cells (hBMEC) induced with various pathogens: *Borrelia burgdorferi* (Treatment1, Table S1), *Neisseria meningitidis* (Treatment2, Table S2), *Streptococcus pneumoniae* (Treatment3, Table S3), and West Nile Virus (Treatment4, Table S4). Tables are available in the supplemental information.
3. In the interactive panel below, use the pre-loaded label-free proteomics example data of milk whey samples collected at a different time point (36, 42, 57 and 81 hours) from a cow with *Streptococcus uberis* infection (Mudaliar, et al., 2016).



OMNANALYSIS Welcome Upload data PCA Plots Statistical filtering GO enrichment analysis

Click to upload transcriptomics example data

Upload transcriptomics data

Browse...

Click to upload Proteomics example data

Upload proteomics data

Browse...

Click me to delete the uploaded data.

After uploading the data, select the species, choose the ID conversion and click submit.

Select a species

Human ▼

ID conversion ⓘ

Ensembl gene ID ▼ Ensembl gene ID ▼

Submit (After submission, click on PCA or Plots or Statistical filtering).

Download Download

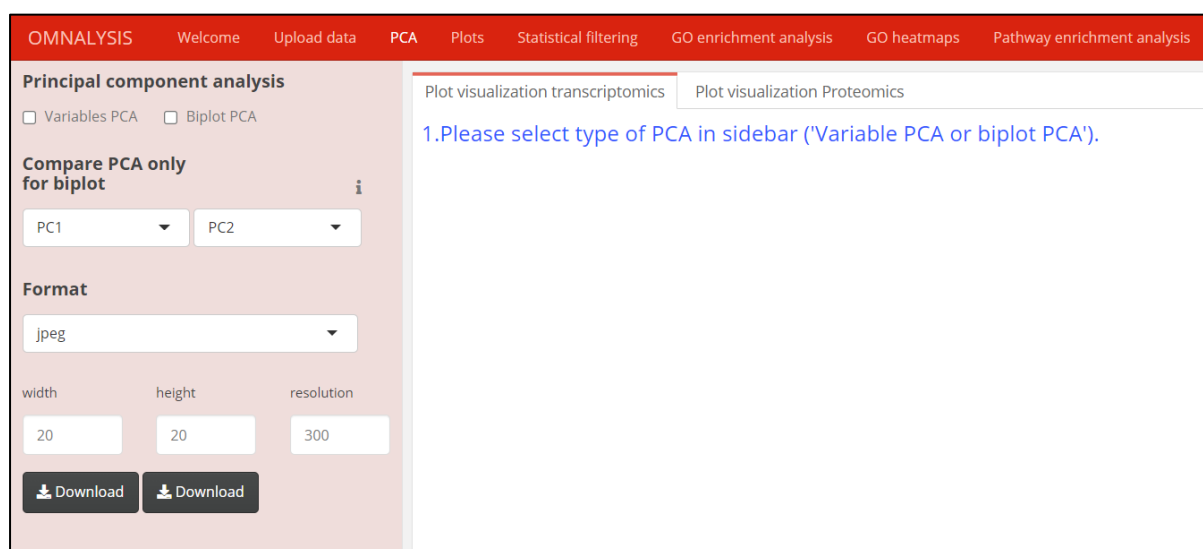
Transcriptomics output Proteomics output

4. Select the **Upload transcriptomics data Browse tab**, to upload your transcriptomics data in CSV file up to four experimental conditions with expression value (log fold change), a measure of significance (P value), count of reads (log counts per million).
5. Select the second **Upload proteomics data Browse tab**, to upload your proteomics data in xlsx format file up to four experimental conditions with UniProt ID, FDR-adjusted P -value and Fold Change.
6. To delete the uploaded file using the red color 'Click me to delete the uploaded data.'.
7. Once the data is loaded (**ONLY ONE OMICS TYPE at a time**) on the window panel, select the species provided in the **select a species tab** (Human, Chicken, Pig, Cow, Mouse, Rat, Dog, *Drosophila melanogaster* and *C. elegance*) with the help of a dropdown menu to perform further analysis.
8. Gene ID conversion is mandatory for further analysis. The converted IDs will be used for the volcano PCA and Volcano plots.

9. After selecting the species user needs to convert the less informative ENSEMBLGENE ID or UniProt ID to more informative IDs using the **ID conversion drop-down menu tab**.
10. Press the **SUBMIT button** to perform the ID conversion.
11. In the **Transcriptomics or Proteomics output** window, the converted IDs will be added to the last column of the uploaded data.
12. Press the first **Download button** to download the converted ID table of transcriptomics and the Second **Download button** for proteomics converted IDs in CSV format.

PCA

1. Principal component analysis (PCA) is an unsupervised dimension reduction method that allows us to understand the relationships among the attributes of expression data. PCA analysis calculates the principal components using the Euclidean distance and linear transformation of the expression data. It calculates the most significant variable in the provided data to calculate the first principal component (PC1), second principal component (PC2) and so on.



OMNANALYSIS Welcome Upload data PCA Plots Statistical filtering GO enrichment analysis GO heatmaps Pathway enrichment analysis

Principal component analysis

☐ Variables PCA ☐ Biplot PCA

Compare PCA only for biplot i

PC1 ▼ PC2 ▼

Format

jpeg ▼

width: 20 height: 20 resolution: 300

Download Download

Plot visualization transcriptomics Plot visualization Proteomics

1. Please select type of PCA in sidebar ('Variable PCA or biplot PCA').

2. Two types of PCA plots are provided, first is a **Variable PCA** plot that provides information about the direction and relationship among the variables (treatments) and the second is **Biplot PCA** that visualize the features of variable PCA plot and observation (genes or proteins) of each treatment in a single plot.
3. Press **Variable PCA** or **Biplot PCA** to generate transcriptomics variable PCA plot, biplot PCA plot in the **Plot visualization transcriptomics window**.
4. Press **Variable PCA** or **Biplot PCA** to generate proteomics variable PCA plot, biplot PCA plot in **Plot visualization proteomic window**.
5. Most of the variables are explained in PC1 and PC2, however, to compare principal components, select PC from **Compare PCA for Biplot only drop-down tab** is to **explore and compare the other available PCs (PC1 vs PC2, PC2 vs PC3 and PC3 vs PC4)**.
6. Use **Format drop-down menu** to download images in **jpeg, png, pdf and tiff** formats.

- Provide **width**, **height** and **resolution** to the **output PCA plot** using **numerical input tabs**. The default values are **20cm**, **20cm** and **300px** for **width**, **height** and **resolution** respectively.
- Press the **first Download button** to download transcriptomics **PCA plots** and the **second Download button** to download proteomics **PCA plots**.

Plots

- Scatter plot uses log fold change versus log counts per million to visualize your expression data in the form of up and down-regulated genes. Each dot in the scatter plot represents up-regulated genes (green colour), non-significant genes (black colour) and downregulated genes (red colour).

- Select from the **Uploaded treatments** checkboxes, Example- Treatment-1, then provide the **numeric input Pvalue/FDRcutoff** for both scatter and volcano plot and **Fold Change-cutoff for volcano plot** to generate plots. The default numeric **Pvalue cutoff** is 0.001 and **Fold Change-cutoff** is 1.2 for transcriptomics and can be changed.
- Provide plot name (Example- “Scatter plot of hbmecc induced with WNV”) in **Plot Title text input option**.
- Press one of the checkboxes in the **Plots heading (one at a time only)** to generate a type of plot and **Table** to visualize each treatment data with converted ids. To display the results click on the windows named **Scatter plot, Volcano plot and Treatment table**.
- Select image format type from the **Format** drop-down menu.
- Provide **width**, **height** and **resolution** to the **output scatter and volcano plots** using **numerical input tabs**. The default values are **20cm**, **20cm** and **300px** for **width**, **height** and **resolution** respectively.
- Press the **Download button** to download the **scatter or volcano plots**.

Statistical filtering

OMNANALYSIS | Welcome | Upload data | PCA | Plots | **Statistical filtering** | GO enrichment analysis | GO heatmaps | Pathway enrichment analysis | Enriched pathway visualization | Literature info | Help

Statistical filtering

LogFC: 0 | LogCPM: 0 | Pvalue/FDR: 0

Omics Type
Transcriptomics

Treatments uploaded
☐ Treatment-1 ☐ Treatment-2
☐ Treatment-3 ☐ Treatment-4

Venn Diagram
☐ Split into Up and Down-regulated
☐ VennDiagram
 width: 20 | height: 20 | resolution: 300
 Download

Histogram
☐ One treatment ☐ Multi treatments
 Title:
 Format: jpeg
 Download

Filtered data | Venn Diagram | Histogram

1. Provide 'logFC cutoff' (usual range is 1 to 1.5), 'logCPM' (usual range is ≥ 3) and 'Pvalue' (usual range is 0.001 to 0.05) or 'FDR cutoff' (usual range is 0.01 to 0.05).
 2. Select 'Omics Type' (Transcriptomics or Proteomics) in the sidebar.
 3. Please select 'Treatment' in the sidebar (one treatment at a time only).

"To generate Venn diagram and Histogram statistical filtering input is required".

1. Provide threshold numerical values in the **Statistical filtering tab input boxes** (LogFC, LogCPM, Pvalue) provided to filter out the genes that are unable to cross the threshold values. The default value is 0 for all components.
2. Select one from the **Omics Type** drop-down tab (Transcriptomics or Proteomics).
3. Use checkboxes provided under **Treatments uploaded** header (Treatment-1, Treatment-2, Treatment-3 and Treatment-4) to visualize the data, Venn diagram and histogram of the selected treatments on **Filter data, Venn Diagram and Histogram** display panel.
4. **Statistical filtering is mandatory for the Venn diagram, histogram and further analysis.**
5. Once the filtered data is visible on the **Filter data panel**, Use the checkboxes provided under the **Venn Diagram header** to visualize the common significantly DEGs or abundance protein in the uploaded treatments. Use **Split into Up and Down-regulated checkbox** to obtain common or different up and down-regulated genes in two groups of uploaded treatments. Use the **VennDiagram checkbox** to obtain common genes in the selected treatments.
6. Provide **width**, **height** and **resolution** to the **output Venn diagram and Histogram** using **numerical input tabs**. The default values are 20cm, 20cm and 300px for width, height and resolution, respectively.
7. Use first **Download tabs**, to download the Venn diagram and second to download the histogram.
8. Use checkboxes provided below the **Histogram header**. Click on the **One treatment checkbox** to generate each treatment histogram showing the number of differentially expressed genes or proteins that falls on the log fold change range or click on the **Multi treatments checkbox** to generate the total number of up and down-regulated genes or proteins available in each treatment in a single histogram.
9. Use the **Title** input tab to provide the title to the generated Venn diagram or histogram (example: - "Venn diagram or Histogram of Treatment-1")

10. Select **image format types** using the drop-down **Format** option.

Gene ontology (GO) enrichment analysis

1. Select the **Omics Type** according to the uploaded data.
2. Use the checkboxes under the **Gene ontology classes header** to perform the categorical analysis of your uploaded data.
3. Clicking on the **GO Biological Process (BP)**-multiple molecular activities integrate to perform a process or **GO Molecular Function (MF)**-activities at the molecular level by gene product or **GO Cellular Component (CC)**-site of function concerning cellular structure.

The screenshot shows the OMNANALYSIS web interface. The top navigation bar includes links for Welcome, Upload data, PCA, Plots, Statistical filtering, GO enrichment analysis (active), GO heatmaps, Pathway enrichment analysis, Enriched pathway visualization, Literature info, and Help. The main content area is divided into a left sidebar and a right panel. The sidebar contains the 'Omics Type' dropdown set to 'Transcriptomics', 'Gene ontology classes' with checkboxes for BP, MF, and CC, 'Pvalue cutoff' and 'q-value cutoff ORA' input fields both set to 0, 'pAdjust Method' dropdown set to 'None', 'Enrichment analysis method' with checkboxes for GO ORA and GO GSEA, and a 'Go!' button. At the bottom of the sidebar are 'Download GO result' checkboxes for Treatment-1, Treatment-2, Treatment-3, and Treatment-4, and a 'Download' button. The right panel has a header 'To construct heatmap from enriched gene ontology terms' and a note: 'Once the result is ready, select only one row (one enriched GO term at a time) in all the enriched result for heatmap visualization in next tab ('GO heatmaps')'. Below this is a tabbed interface with 'Treatment-1' selected, showing a list of instructions: 1. Please select 'Omics Type' in the sidebar. 2. Select only one 'Gene ontology class' at a time (BP or MF or CC). 3. Provide 'Pvalue cutoff' (usual range is 0.001 to 0.05). 4. Provide 'q-value cutoff for ORA' (usual range is 0.01 to 0.05). 5. Select only one 'pAdjustment Method' at a time (BH method is recommended). 6. Select enrichment 'GO ORA or GO GSEA' gene ontology analysis (only one at a time).

4. After selecting one of the **Gene ontology classes**, provide **Pvalue cutoff** and the **q-value cutoff ORA**. The default Pvalue and q-value cutoff is 0 and can be changed.
5. Select one of the p-value adjustment methods from **pAdjust Method drop-down menu**. List are **Holm, Hochberg, Hommel, Bonferroni, Benjamini and Hochberg (BH), Benjamini and Yekutieli (BY)** and **FDR** to control false positive results in the analysis. The first four methods control the family-wise error rate (probability of making one or more false discoveries) and the remaining methods control the expected proportion of discoveries that are rejected falsely (FDR). **We suggest FDR correction methods for a more reliable result.**
6. Select one of the provided **Enrichment analysis methods**, **GO ORA** (based on hypergeometric test and mapping of genes to the annotated biological vocabulary) or **GO GSEA** (based on the Kolmogorov Smirnov test and consider gene set with their sorted log fold change value).
7. After providing the necessary inputs mentioned above, click on the **Go! button** to launch the enrichment analysis. **Please keep in mind that the same input values will be used for all treatments to maintain the standard of enrichment analysis.**
8. The result of the enrichment analysis is visualized on the **Treatment-1, Treatment -2, Treatment -3, and Treatment -4** display panels.
9. From the ontology result table, select one enriched GO term by selecting only one row and the same GO term in all the treatments. The selected information will be used for the generation of heatmaps in the next **GO heatmaps** section.

10. Use checkboxes provided under the **Download GO result header** to download the result in CSV format.
11. Select one **Treatment checkbox** at a time and click on the **Download button**.

GO heatmaps

1. GO heatmaps section of OMnalysis provide visualization of the enrichment analysis using heatmaps (expression values of genes that mapped to the biological term in gene ontology enrichment analysis) and word cloud.

2. It is mandatory to **select at least two enriched GO terms (rows) from the previous tabular output of GO enrichment analysis.**
3. Select one of the **GO ORA** or **GO GSEA** methods from **Heatmap visualization**.
4. Click on at least two checkboxes (**Treatments 1, 2, 3 and 4**) provided under the **Treatments header** to generate a heatmap.
5. Adjust the column text, row text and color key size using **col text size**, **row text size** and **col key size** numerical input tabs, respectively. The default value for **col text size**, **row text size** and **col key size** are 1.2, 1.2 and 0.04, respectively.
6. Input text in the **Title for heatmap input tab** to provide main text on the heatmap.
7. Download **Heatmap and word cloud** using the image formats provided under the **Format** drop-down menu.
8. Adjust **dimension** and **resolution** of **heatmap and word cloud** using **numerical input in tabs Plot width, Plot height and Plot resolution**. The default value is 20cm, 20cm and 300px for width, height and resolution, respectively.
9. After providing the above-mentioned information **download the generated heatmap using the first Download button.**

- Click on the **Word cloud** checkbox of the selected treatment provided under the **Visualization of Ontology header**.
- Provide the numeric input (**maximum number of words you want to visualize in the word cloud**) using **max.words input tab option**. The default value of the maximum words in the **word cloud** is 100.
- Press the second **Download button** to download the generated word cloud diagram.

Pathway enrichment analysis

- This section provides the databases and methods available to perform pathway analysis.

The screenshot shows the OMNANALYSIS web interface. The top navigation bar includes links: OMNANALYSIS, Welcome, Upload data, PCA, Plots, Statistical filtering, GO enrichment analysis, GO heatmaps, Pathway enrichment analysis (active), Enriched pathway visualization, Literature info, and Help.

Omics Type
 Transcriptomics (dropdown menu)

Pathway analysis type
☐ Over-Representation analysis (ORA)
☐ Gene Set Enrichment Analysis (GSEA)
☐ ReactomePA (Human)

Pvalue cutoff (input: 0) **q-value cutoff ORA** (input: 0)

pAdjust Method
 None (dropdown menu)

Network Topology Analysis (NTA)
☐ NTA

Databases for NTA
 biocarta (dropdown menu)

STRING
☐ String

Go! (button)

To download table of enrichment result
☐ Treatment-1 ☐ Treatment-2
☐ Treatment-3 ☐ Treatment-4

Download (button)

Instructions on the right:
 To construct pathway visualization of enriched pathway
 Once the result is ready, select only one row and same enriched pathway in all the enriched result for pathway visualization in next tab ('Enriched pathway visualization').

Treatment-1 Treatment-2 Treatment-3 Treatment-4

1. Please select checkbox 'Omics Type' in the sidebar (only one at a time).
 2. Select 'Pathway analysis type ' (only one at a time).
 3. Provide 'Pvalue cutoff' for 'ORA', 'GSEA' and 'ReactomePA' (usual range is 0.001 to 0.05).
 4. Provide 'q-value cutoff' for 'ORA' and 'ReactomePA' (usual range is 0.01 to 0.05).
 5. Select 'pAdjustment method' one at a time only (BH method is recommended).
 6. Click 'Go' button to perform 'pathway enrichment analysis'.
 OR
 1. Select checkbox 'Network Topology Analysis (NTA)' (only one at a time).
 2. Then select database to perform 'NTA' (one at a time only), available only for human data.
 3. Click 'Go' button to perform 'Network Topology Analysis (NTA)'.
 OR
 1. Select checkbox 'STRING ' (only one at a time).
 2. Click 'Go' button to perform 'STRING'.

To download the result from each enrichment analysis.
 1. Select checkbox 'treatments' to download the result of each treatment (one at a time only).
 2. Click on 'download' button to download result in csv format.

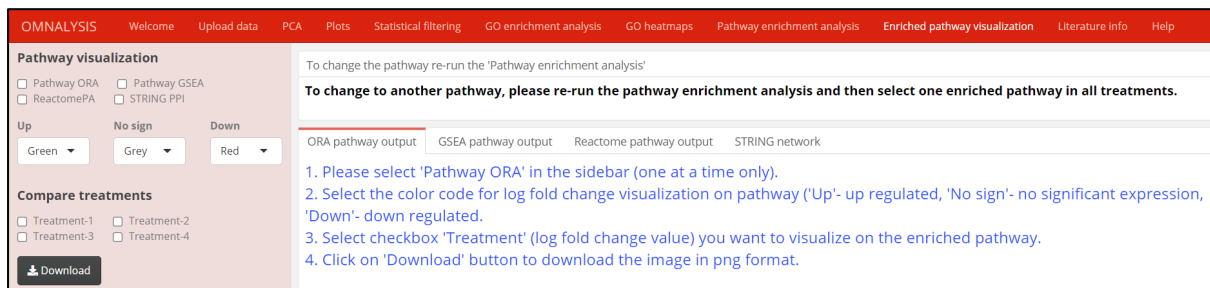
- Select the **Omics Type** according to the uploaded data.
- Select one of the options provided in **Pathway analysis type (Over-represented analysis (ORA), Gene set enrichment analysis (GSEA), ReactomePA (Human))** to perform pathway enrichment analysis.
- Provide numeric input in the **Pvalue cutoff tab** and select one of the **pAdjusted Method tabs** to overcome the chances of getting a false-positive result. Note that stringency cutoff value may result in a fewer number of gene set enrichment pathways.
- If using **Over-represented analysis (ORA)** provide q-value cutoff value in **q-value cutoff ORA numeric input tab**.
- P-value and the q-value cutoff is applicable only for methods under **Pathway analysis type**.
- To perform **Network Topology Analysis (NTA)** use the NTA checkbox provided under the **Network Topology Analysis (NTA)**. This method is supported by **four databases** first, **biocarta** (protein sets participating in the pathway), second, **panther** (a curated and comprehensive database to classify protein and their genes through evolutionary relationship), third, **NCI-Nature Pathway Interaction Database** (Signaling pathways composed of human biomolecular interactions and cellular processes) and fourth,

pharmgkb (a comprehensive resource that provides information about how human genetic variation affects the response to medications).

8. To perform STRING enrichment analysis, click on the **String checkbox** provided under the **STRING header**.
9. Once the above steps are checked, and any one pathway enrichment analysis was selected, click on the **Go! button** to launch the pathway enrichment analysis.
10. Click on the **"Treatment-2, Treatment-3, and Treatment -4"** display panels to perform **pathway enrichment analysis for each treatment**.
11. From the pathway enrichment result table, **select only one row with the same enriched pathway in all treatments**.
12. The selected row or enriched pathway information will be used for the visualization of the expressed gene or proteins in the next section **Enriched pathway visualization**.
13. Click on the **Treatments checkboxes one at a time only** provided under the **To download table of enrichment result** header.
14. Click on the **Download button** to download the result in **CSV format**.

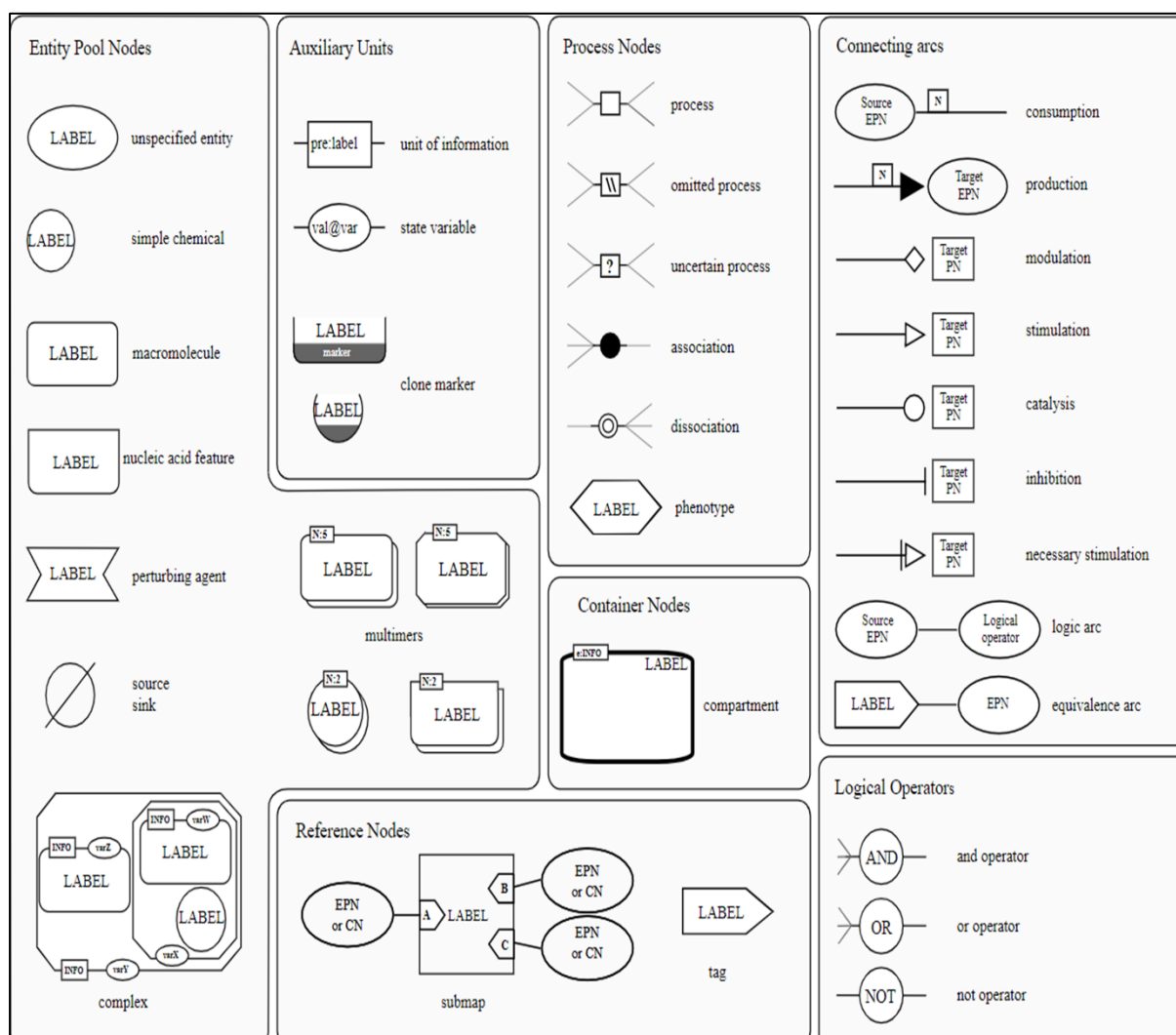
Enriched pathway visualization

1. After selecting the single and same pathway in all the treatments from the **pathway enrichment result table**, you can visualize the pathway using the checkboxes provided under the **Pathway visualization** header and according to the pathway enrichment method performed in the **Pathway enrichment analysis** section of OMAnalysis.



2. Select one of the **Pathway ORA**, **Pathway GSEA**, **ReactomePA** and **STRING PPI** checkboxes provided under the **Pathway visualization** header.
3. If the previous pathway enrichment analysis was ORA, then only the ORA enriched pathway will be visualized in the **ORA pathway output subtab**, if not, it will show an error.
4. Select the color code from three drop-down tabs for **highly expressed (Up)**, **No induced or absent expressed (No sign)** and **suppressed genes or proteins (Down)**.
5. The default colour code for **Up** is green, **No sign** grey and **Down** in red. You can select more colour combinations from the none, red, green, yellow, blue, and grey to visualize the expression values (logFC) on the enriched pathway.
6. Select the **Treatments checkboxes** provided under the **Compare treatments** header to visualize the enriched pathway of **Treatment-1**, **Treatment-2**, **Treatment-3**, **Treatment-4**.
7. Click on **ORA pathway output** or **GSEA pathway output** or **Reactome pathway output** or **STRING network** display panel, according to the enriched pathway method selected.

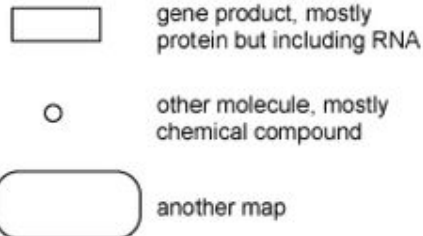
8. The output from the **Reactome pathway analysis** can be interpreted using the below notation image.



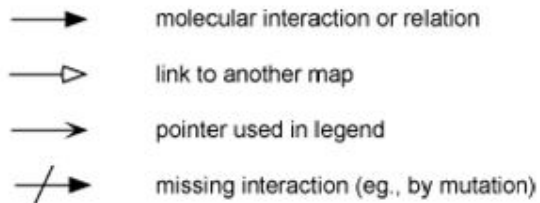
9. Keep in mind that the visualization of pathways depends on the **pathway enrichment analysis method and selection of the same pathway in all treatments.**
10. Click on the **Download button** to download the **pathway image of ORA, GSEA and ReactomePA and STRING output in PNG image format.**

KEGG - notation

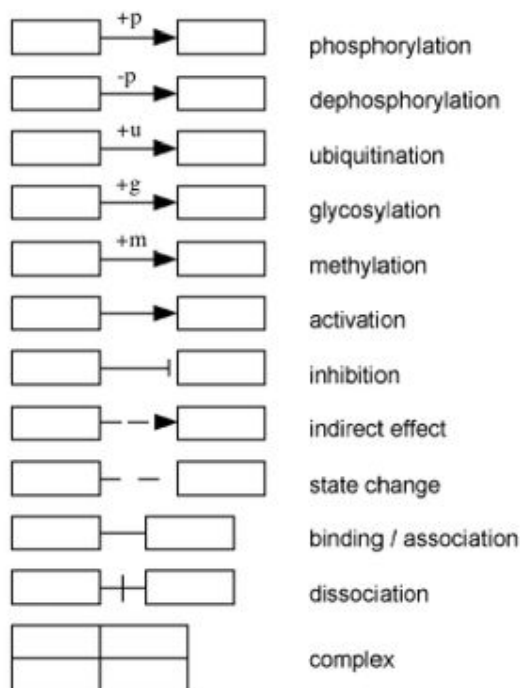
Objects



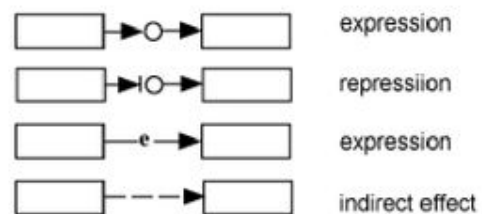
Arrows



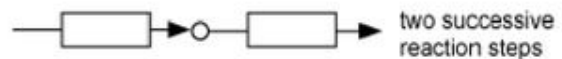
Protein-protein interactions



Gene expression relations



Enzyme-enzyme relations



Literature info

OMNANALYSIS Welcome Upload data PCA Plots Statistical filtering GO enrichment analysis GO heatmaps Pathway enrichment analysis Enriched pathway visualization Literature info Help

Literature search

Literature retrieval limit

100

Submit

ALERT:- Please select only one row from the result table. Selected row will be visualize in the Abstract info window

Literature info Abstract info

1. Please provide input for 'Literature search', example- Gene name (IL6), species (Human), disease type (Cancer), etc.

3. Click on 'Submit' button to perform retrieval!

1. This section provides the option to retrieve the information from the Europe PMC by providing **biomarkers name, species, disease, cell or tissue type** in the **text input tab** below **Literature search**.
2. The **Literature retrieval limit** has an input option to provide a number that will decide the fetching of the literature. The default is 100.
3. Once the literature search and retrieval limit are provided, you can **proceed with the submit button to perform the retrieval of scientific literature**.
4. The result in tabular form will appear in the **literature info subtabs**.
5. **Select one scientific literature row** in the **literature info table** at a time to view the abstract and other information on the **next subtab Abstract info**.
6. If the keyword provided in the Literature search is not correct then the result may produce an error or blank table of literature info.