

Table S1. Comparison of features of omics data visualization tools. The table in excel format is available at <https://bit.ly/3qax8VZ> for better view.

| Tools | Quickomics | Ideal | BEAVR | WiSON | IRIS-EDA | DEBrowser | PIVOT | PaintOmics 3 | iSEE | IDEP | START |
|---|---|---|---|---|--|---|---|---|--|---|---|
| Year | 2020 | 2020 | 2020 | 2019 | 2019 | 2019 | 2018 | 2018 | 2018 | 2018 | 2017 |
| Main function | | | | | | | | | | | |
| Data analysis | Y | Y | Y | | Y | Y | Y | | Y | Y | Y |
| Data visualization | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Data type supported | | | | | | | | | | | |
| RNAseq | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Other datatypes | Proteomics | | | Multi-omics | | Other sequencing | | Multi-omics | Multi-omics | | |
| Key features | Comprehensive analytical tasks, generates both interactive and publication-ready figures. | Interactive differential expression analysis for RNAseq. Generate publication-ready outputs. | Simplifies RNAseq analysis for RNAseq users/experts | Provides customized and dynamic visualization for multi-omics data | Performs RNAseq data analysis and provides framework to expedite data submission to NCBI's Gene Expression Omnibus | Interactive differential expression analysis and visualization for count data | Interactive analysis and visualization of transcriptomics data, automatic report generation, publication-quality plots. | Integrated visualization of multiple omic data types onto KEGG pathway diagrams. | Provides a general visual interface for exploring data in a SummarizedExperiment object. | Integrated web application for differential expression and pathway analysis of RNA-Seq data | Web-based RNAseq analysis and visualization |
| Online/Standalone | Both | Both | Standalone | Standalone | Online | Standalone | Standalone | Online | Standalone | Both | Both |
| Source Code Link | https://github.com/integrative-report/Quickomics | https://github.com/idealeideneomics/rnideal | https://github.com/developper/BEAVR | https://github.com/molgen-mpg/deiloslab/wilson-n-apps | https://github.com/UMMS-BMBL/IRIS | https://github.com/qinzhuhu/PIVOT | https://github.com/UMMS-BioCore/debrouter | https://github.com/qinzhuhu/PIVOT | https://github.com/SDSU/idep | https://github.com/miminiere/STARTapp | |
| Demo Link | http://quickomics.bx.psu.edu/mainz.de:3838/ | http://shiny.lmb.ei.uni-mainz.de:3838/ | http://loosolab.mpi-bn.mpg.de/wilson/ | http://bmbi.lsdsite.edu/IRIS/ | | https://kim.bio.upenn.edu/software/pivot.html | http://www.pajotware/pivot.html | http://www.paintomics.org/ | http://bioinformatics.state.edu/dep/ | https://kcvishwanathapps.io/START | |
| Demo using example data | Y | Y | | Y | Y | | | | Y | Y | Y |
| Analysis tool | | | | | | | | | | | |
| PCA-2D | Y | | Y | Y | Y | Y | Y | | Y | Y | Y |
| PCA-3D | Y | | | | | | Y | | | | |
| PCA 3D interactive plot | Y | | | | | | | | | | |
| Sample-sample distance heatmap | Y | Y | Y | Y | Y | | Y | | | | Y |
| Sample to sample scatter plots | | Y | | | | Y | Y | Y | | | Y |
| Dendograms | Y | | Y | Y | Y | | Y | | | | Y |
| Rank-frequency plot/Mean- Variability Plot/Rank-Sd plot | | | | | | | | | | | |
| Gene expression per sample in box plot | Y | | | | | Y | Y | | | | Y |
| CV Distribution plot | Y | | | | | Y | Y | Y | | | |
| Multidimensional scaling plot | | | | | | Y | | | | | Y |
| t-distributed Stochastic Neighbor Embedding plot | | | | | | Y | | Y | | | Y |
| Histogram for read counts by sample/Bar plot of total reads by sample | | | | | | Y | Y | Y | | | Y |
| Differential expression analysis | | Y | Y | | | Y | Y | Y | | | Y |
| Bar plot for DEG number overview | | | | | | Y | | | | | Y |
| Static volcano plot | Y | Y | Y | Y | Y | Y | | | | | Y |
| Interactive volcano plot | Y | | | | | Y | | | | | Y |
| MA plots | | Y | | Y | Y | Y | Y | | | | Y |
| Scatter plot | | | | | | Y | | | Y | | Y |
| Rectangle plot | | | | | | | | | Y | | Y |
| Histograms for unadjusted p-values/histogram for estimated log fold changes/Schwarzer-Sjøtvedt plot | | | Y | | | | | | | | |
| Fold change vs fold change from two comparisons plot | Y | | | Y | | | | | | | |
| Static gene expression heatmap | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Interactive gene expression heatmap | Y | Y | | Y | | Y | | | | | Y |
| Gene expression box plot | Y | Y | Y | Y | | | | | | | Y |
| Gene expression jitter plot | | | Y | | | | | | | | |
| Gene expression bar plot | Y | | | Y | | | Y | | | | Y |
| Gene expression violin plot | Y | | | Y | | | Y | | | | Y |
| Gene expression line plot | Y | | | Y | | | | | | | |
| Rank Abundance Curve | Y | | | | | | | | | | |
| Minimum Spanning Tree and Community Detection | | | | | | | | | | | |
| Functional enrichment analysis | Y | | Y | Y | | Y | Y | Y | | | Y |
| Interactive pathway network | | | | | | | | | | | Y |
| Multi-omic visualization of single pathways | | | | | | | | | | | Y |
| Gene expression pattern clustering | Y | | | | | Y | | | | | Y |
| Co-expression network | Y | | | | | | Y | | | | Y |
| Visualizing expression profiles on chromosomes | | | | | | | | | | | Y |
| Venn Diagram for single dataset | Y | | Y | | | | | | | | Y |
| Venn Diagram across datasets | Y | | Y | | | | | | | | Y |
| UpSet plots | | Y | | | | | | | | | |
| GEO Submission | | | | | Y | | | | | | |

Table S2. Available modules and analysis in Quickomics.

| Modules/Analysis | Description |
|----------------------------|---|
| QC Plots | |
| PCA Plot | To perform principal component analysis (PCA) and generate a plot that can be colored, shaped and sized by attributes of samples. It has options to emphasize mean points and/or add marginal rugs to the plot. |
| Eigenvalues | To generate a plot of variances explained by the first 10 PCs in the dataset to allow users to make educated decisions of which PCs to plot in 2D or 3D scatter plot. |
| PCA 3D Plot | To display 3D PCA representation on PC1, PC2 and PC3. Users can zoom in and out to focus on different areas of the plot or rotate the plot in different axes to find the best separation. Users can change color, include an ellipsoid and add labels to the samples. |
| PCA 3D Interactive | To create interactive 3D PCA plot to enable users to see detailed sample annotation information by hovering over samples on the plot. It allows users to change color and shape of the dots representing samples in the plot. |
| Sample-sample Distance | To generate a distance matrix for every sample pair and plot it as a heatmap to show pair-wise similarity between all samples. The rows and columns are clustered based on similarity scores. |
| Dendrogram | To generate a Dendrogram plot to help visualize hierarchical clustering relationships between samples. The default plot is circular and divided into four parts. Users have the option to visualize the dendrogram plot in tree horizontally or vertically and divide the plot in multiple regions. |
| Box Plot | To generate a plot to visualize the distribution of the normalized expression values in all samples. This identifies the minimum, first quartile, median, third quartile, and maximum values in the dataset. |
| CV Distribution | To generate a plot to show the histogram of coefficient of variation (CV) and a dotted line for each group indicating the median CV. |
| Order Groups | To allow users to select and/or re-order the groups. The order will be stored in current session and will be displayed in a preferred order in all plots. Users can either drag-n-drop group names to reorder or use the select panel to delete and add. |
| Volcano Plot | |
| Volcano Plot (Statics) | To generate a volcano plot to visualize the differentially expressed genes/proteins provided in the input data file. Users have the option to change cutoffs of fold change and P value and more advanced display options. |
| Volcano Plot (Interactive) | To create an interactive volcano plot allowing users to see detailed annotation information of genes/proteins (dots in plot) by hovering over the dots on the plot. |

| | |
|----------------------------|--|
| DEGs in Two Comparisons | To generate a scatter plot to visualize the relationship between differentially expressed genes/proteins in two comparisons. |
| Data Table | To view the differentially expressed genes/proteins in a tabular format with a searchable feature. Users could download the table as a CSV file. |
| Heatmap | |
| Static Heatmap Layout 1 | To generate a heatmap of gene/protein-by-sample gene expression matrix by using ComplexHeatmap package. Users can customize the gene list and pick annotation categories to be shown on the heatmap. |
| Static Heatmap Layout 2 | To generate a heatmap using the heatmap.2 function from gplots package. The same customization of the above function is applied here. |
| Interactive Heatmap | To generate an interactive heatmap of gene/protein-by-sample gene expression matrix by using package 'heatmaply'. |
| Expression Plot | |
| Browsing | To generate Violin Plots to visualize the expression of the differentially expressed genes/proteins identified in various comparisons. |
| Searched Expression Data | To generate Box plots to visualize the expression of user entered genes/proteins list or user selected genes/proteins list from databases like KEGG, MSigDB etc. provided in Quickomics. |
| Data Table | A searchable table for the genes/proteins selected in "Searched Expression Data" tab with their normalized expression values. |
| Result Table | A searchable table for the genes/proteins selected in "Searched Expression Data" tab with their statistics results including P value, adjusted P value and fold change. |
| Rank Abundance Curve | To generate a plot to visualize the relative abundance of a list of genes within a dataset. This plot helps interpret the distribution of abundance and expression levels of a set of genes. |
| Gene Set Enrichment | |
| Gene Set Enrichment | To perform gene set enrichment analysis based on user selected gene set and functional databases. |
| Gene Expression | To provide logarithm transformed fold change values for genes in the enriched functional pathways identified from the above analysis. |
| Gene Set Heatmap | To generate expression heatmap for genes in the enriched functional pathways identified in the "Gene Set Enrichment" analysis. |
| KEGG Pathway View | To allow users to view the fold change levels of genes in the KEGG pathways for up to 5 comparisons together. |
| Pattern Clustering | |

| | |
|---------------------------------|--|
| Clustering of Centroid Profiles | To perform genes/proteins clustering based on their expression values across different groups and generate plots to visualize the identified co-expression clusters. Three clustering algorithms including soft (fuzzy) clustering, k-means and partitioning around medoids are available. |
| Data Table | A table to show the expression values for genes/proteins with their co-expression cluster IDs assigned in the "Clustering of Centroid Profiles" section. |
| Correlation Network | |
| visNetwork | To build co-expression networks based on gene-gene or protein-protein correlation using R package visNetwork. Users can choose the correlation coefficient and P value cutoffs to select expression correlated genes/proteins. |
| networkD3 | To build co-expression networks based on gene-gene or protein-protein correlation using R package networkD3. Users could drag, zoom, and highlight gene/protein nodes in the network. |
| Data Table | A table to show genes/proteins in the co-expression networks identified in the "visNetwork" section. Users can perform searches and sort the table based on correlation statistics. |
| Venn Diagram | |
| Venn Diagram | To generate a Venn Diagram to show the intersections of differentially expressed genes/proteins from up to 5 different comparisons. Users can use all differentially expressed genes/proteins or only up- or down-regulated ones. |
| Venn Diagram (black & white) | To generate a black and white Venn Diagram to show the intersections of differentially expressed genes/proteins from up to 5 different comparisons. |
| Intersection Output | Gene lists from different intersecting regions of the Venn Diagram generated in the "Venn Diagram" section. |
| DEG Table | A table to show fold changes and P values of genes from the Venn Diagram in the comparisons being queried. |
| Venn Across Projects | |
| Venn Diagram | To generate a Venn Diagram to show the intersections of differentially expressed genes/proteins across projects. This function could be used to identify common differentially expressed genes/proteins present in RNAseq and Proteomics datasets generated from the same samples. |
| Venn Diagram (black & white) | To generate a black and white Venn Diagram to show the intersections of differentially expressed genes/proteins across projects. |
| Intersection Output | Gene lists from different intersecting regions of the Venn Diagram generated in the "Venn Diagram" section. |

Quickomics Supplementary Tutorial

Table of Contents

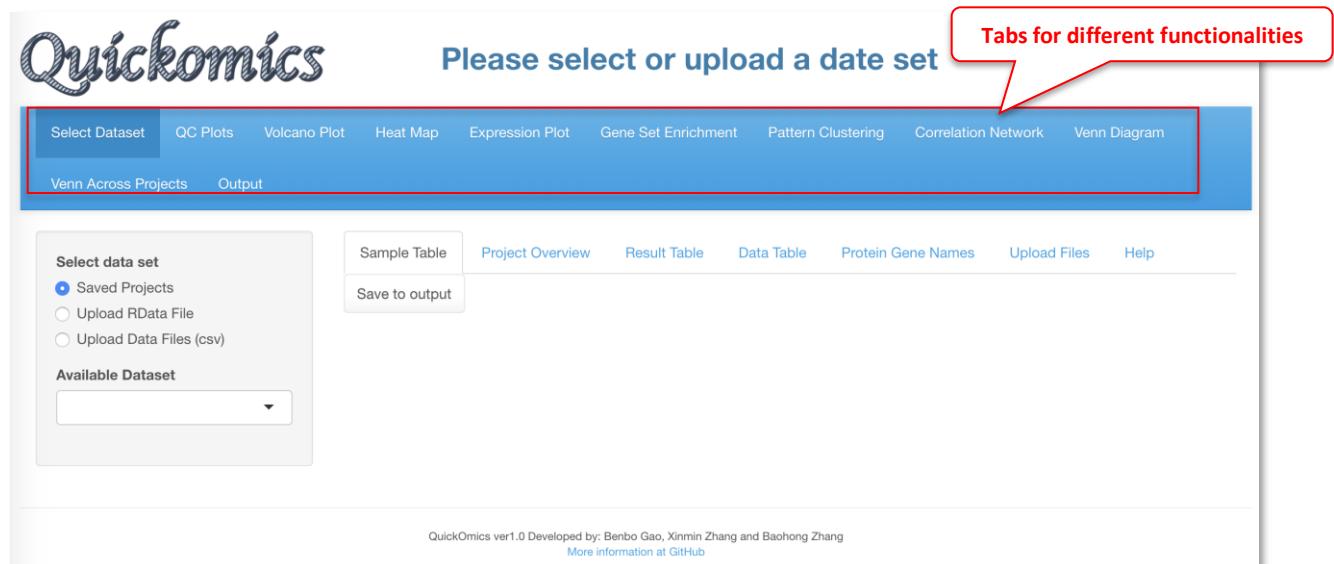
| | | |
|----------|--|-----------|
| 1 | <i>Introduction</i> | 8 |
| 1.1 | Integrated Gene Set Query | 8 |
| 2 | <i>Select Dataset Module</i> | 10 |
| 2.1 | Upload Files | 10 |
| 2.2 | Prepare R Data Files by Computational Biologists..... | 11 |
| 2.2.1 | Example R script to prepare R data files from RNA-Seq results | 13 |
| 2.2.2 | Example R script to prepare R data files from proteomics results | 13 |
| 2.3 | Sample Table | 13 |
| 2.4 | Project Overview..... | 14 |
| 2.5 | Result Table | 15 |
| 2.6 | Data Table..... | 15 |
| 2.7 | Protein Gene Names | 15 |
| 2.8 | Help..... | 16 |
| 3 | <i>QC Plots Module</i> | 17 |
| 3.1 | PCA Plot..... | 17 |
| 3.2 | Eigenvalues..... | 20 |
| 3.3 | PCA 3D Plot..... | 21 |
| 3.4 | PCA 3D Interactive | 22 |
| 3.5 | Sample-sample Distance..... | 23 |
| 3.6 | Dendrograms | 23 |
| 3.7 | Box Plot | 24 |
| 3.8 | CV Distribution..... | 24 |
| 3.9 | Order Groups | 25 |
| 3.10 | Help..... | 26 |
| 4 | <i>Volcano Plot Module</i> | 27 |
| 4.1 | Volcano Plot (Static) | 27 |
| 4.2 | Volcano Plot (Interactive) | 29 |

| | | |
|-----------|--|-----------|
| 4.3 | DEGs and DEPs in Two Comparisons | 29 |
| 4.4 | Data Table..... | 31 |
| 5 | <i>Heatmap Module</i>..... | 32 |
| 5.1 | Static Heatmap Layout 1..... | 32 |
| 5.2 | Static Heatmap Layout 2..... | 33 |
| 6 | <i>Expression Plot Module</i>..... | 35 |
| 6.1 | Browsing..... | 35 |
| 6.2 | Searched Expression Data..... | 35 |
| 6.3 | Data Table..... | 37 |
| 6.4 | Result Table | 38 |
| 6.5 | Rank Abundance Curve..... | 38 |
| 7 | <i>Gene Set Enrichment Module</i> | 40 |
| 7.1 | Gene Set Enrichment..... | 40 |
| 7.2 | Gene Expression..... | 41 |
| 7.3 | Gene Set Heat Map | 41 |
| 7.4 | KEGG Pathway View..... | 42 |
| 8 | <i>Pattern Clustering Module</i> | 43 |
| 8.1 | Clustering of Centroid Profiles | 43 |
| 8.2 | Data Table..... | 43 |
| 9 | <i>Correlation Network Module</i>..... | 44 |
| 9.1 | visNetwork | 44 |
| 9.2 | Data Table..... | 44 |
| 10 | <i>Venn Diagram Module</i>..... | 46 |
| 10.1 | Venn Diagram | 46 |
| 10.2 | Venn Diagram(black & white) | 46 |
| 10.3 | Intersection Output..... | 47 |
| 10.4 | DEG Table | 47 |
| 11 | <i>Venn Across Projects Module</i>..... | 49 |
| 11.1 | Venn Diagram | 49 |
| 11.2 | Intersection Output..... | 49 |

| | | |
|-----------|-----------------------------------|-----------|
| 12 | <i>Output Module</i> | 50 |
| 13 | <i>References</i> | 51 |

1 Introduction

The Quickomics tool can be accessed via the link <http://quickomics.bxgenomics.com>. Implemented through R Shiny, it helps with visualizing statistical analysis results for RNAseq and Proteomics datasets. This supplemental tutorial provides a detailed guide on using the different functionalities and customizing the tools to best fit individual analysis and plotting needs, using published RNAseq dataset and proteomics dataset as examples (Gyoneva *et al.*, 2019; Connor-Robson *et al.*, 2019).



The interface contains multiple tabs, corresponding to different functional modules, that can be accessed on top panel of the webpage. Users have the option to upload their own dataset or choose from existing demo datasets for visualization.

1.1 Integrated Gene Set Query

To supply genes in pre-defined gene sets to plotting functions in Quickomics, we developed xGenesets API for Quickomics to query and retrieve a list of genes/proteins when it is needed. The lists can be defined based on pathways, and gene sets from various sources, including KEGG Pathways (Kanehisa and Goto, 2000), WikiPathways (Martens *et al.*, 2020), Small Molecule Pathway Database (Frolkis *et al.*, 2010), Reactome (Jassal *et al.*, 2020), Gene Ontology (The Gene Ontology Consortium, 2019), Molecular Signatures (Subramanian *et al.*, 2005; Liberzon *et al.*, 2015), and LIPID MAPS Proteome Database (Cotter *et al.*, 2006). xGenesets database currently has 175,537 gene sets stored in MySQL tables, covering human (115,497 records), mouse (30,554 records) and rat (29,486 records) gene sets. xGenesets features personalized settings, dynamic drop-down list upon typing for quick gene set selection, full gene set database browsing, and a list of genes from any gene sets. Throughout Quickomics, users interact with this convenient tool in appropriate interface when a gene set is needed by checking “Geneset” option and then clicking on “Select Geneset” to pick a set from a table of available gene sets in a popup window as shown in the following screenshot.

The screenshot illustrates the workflow for selecting a gene set:

- 1. Check "Geneset"**: A red callout points to the "Geneset" radio button in the "Label Genes" section of the left panel.
- 2. Click "Select Geneset" button**: A red callout points to the "Select Geneset" button in the "List of genes (Protein.ID)" section.
- 3. Pick the list**: A red callout points to the "Select - List" button in the "Actions" column of the "Select Gene Set" table.
- 4. Genes in the set is listed**: A red callout points to the table listing KEGG Pathways entries.

| ID | DB Name | DB ID | Species | Category | Name | Count | Actions |
|-------|---------------|-------|---------|------------|--|-------|---------------|
| 20437 | KEGG Pathways | 01200 | human | Metabolism | Carbon metabolism | | Select - List |
| 20438 | KEGG Pathways | 01210 | human | Metabolism | 2-Oxocarboxylic acid metabolism | 49 | Select - List |
| 20439 | KEGG Pathways | 01212 | human | Metabolism | Fatty acid metabolism | | Select - List |
| 20440 | KEGG Pathways | 01230 | human | Metabolism | Biosynthesis of amino acids | 74 | Select - List |
| 20441 | KEGG Pathways | 00010 | human | Metabolism | Glycolysis / Gluconeogenesis | 85 | Select - List |
| 20442 | KEGG Pathways | 00020 | human | Metabolism | Citrate cycle (TCA cycle) | 34 | Select - List |
| 20443 | KEGG Pathways | 00030 | human | Metabolism | Pentose phosphate pathway | 37 | Select - List |
| 20444 | KEGG Pathways | 00040 | human | Metabolism | Pentose and glucuronate interconversions | 35 | Select - List |
| 20445 | KEGG Pathways | 00051 | human | Metabolism | Fructose and mannose metabolism | 35 | Select - List |
| 20446 | KEGG Pathways | 00052 | human | Metabolism | Galactose metabolism | 34 | Select - List |

For developers of web-based applications (general user can skip this paragraph), xGenesets can be easily embedded in any modern websites and online applications, e.g., R Shiny web apps, by leveraging jQuery, DataTables, and bootstrap (either v3 or v4) JavaScript libraries. xGenesets API is freely available for all usages. Examples of using xGenesets API and guidance on setting up the API in your own applications can be found at <https://bxf.net/genesets>.

2 Select Dataset Module

This first module is for selecting dataset. Users can either select from pre-loaded example datasets or upload their own dataset in a pre-defined format as detailed in section 2.1 and in our GitHub page (<https://github.com/interactivereport/Quickomics>). We have pre-loaded four datasets from published studies that users can use for demonstrative purposes. These includes two RNAseq datasets from Gyoneva *et al.*, 2019 and Connor-Robson *et al.*, 2019; and two proteomics datasets from Connor-Robson *et al.*, 2019. Users can walk through multiple sub-tabs to visualize, select and download figures or data.

Please select or upload a date set

Select Dataset QC Plots Volcano Plot Heatmap Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn

Select data set
● Saved Projects
○ Upload RData File
○ Upload Data Files (csv)

Available Dataset

Mouse Microglia RNA
AD/PD Brain Proteome
LRRK2 Neuron RNA
LRRK2 Neuron Proteome

Sample Table Project Overview Result Table Data Table Protein Gene Names Upload Files
Save to output

QuickOomics ver1.0 Developed by: Benbo Gao, Xinmin Zhang and Baohong Zhang
[More information at GitHub](#)

2.1 Upload Files

For a data set, the "Upload Files" tool allows users to upload three required files, namely sample metadata, normalized expression data and statistical comparison results in csv format (Comma Separated Values) to Quickomics directly. Example data sets are provided in GitHub for both RNAseq (<https://bit.ly/2MRkFcb>) and proteomics (<https://bit.ly/3rn4i6a>). Detailed formatting guidance is outlined below,

1. **Sample Metadata File:** It should have “sampleid” and “group” columns, with additional columns optional. Sample identifiers must match those used in the expression data file.
2. **Expression Data File:** It should be a matrix of expression values with genes/proteins as rows, and samples as columns. The unique IDs for genes/proteins are in the first column. We recommend using log of normalized expression values, e.g. $\log_2(\text{TPM}+1)$ for RNAseq data or normalized intensity or ratio for proteomics data.
3. **Comparison Data File:** It should have five columns, “UniqueId”, “test”, “Adj.P.Value”, “P.Value” and “logFC”. The comparison names are listed in “test” column. Please note that wrongly named column headers will cause issues.
4. **Optional Gene/Protein Name File:** The system has built-in function to convert unique IDs in the data files to gene symbols and create the Gene/Protein Name file, so most users don't need to prepare the file. Nevertheless, if provided by users, it must have four columns: “id” (sequential numbers like 1,2,3), “UniqueId” (matching IDs used in the expression and comparison data file), “Gene.Name” (official gene symbols), “Protein.ID” (UniProt protein IDs, or keep it empty for RNA-Seq data). Additional columns (e.g. gene biotype) are optional.

Quickomics

AD PD Proteomics Demo Test

Select Dataset OC Plots Volcano Plot Heatmap Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects Output

Select data set
 Saved Projects
 Upload RData File
 Upload Data Files (csv)

Available Dataset

Sample Table Project Overview Result Table Data Table Protein Gene Names Upload Files Help

Prepare your own data files in Excel, save them as csv files and upload here. The system will automatically process the files and create the R data files. You need sample metadata file, expression data file, and comparison data file. The system can create gene/protein annotation based on the IDs from data files, or you can upload your own Gene/Protein Name file.

Download RNA-Seq example csv files (200 genes from mouse microglia dataset)
Download Proteomics example csv files (200 proteins from AD PD dataset)

Project Name
AD PD Proteomics Demo Test

Select Species
 human mouse rat

Sample MetaData must have sampleID and group columns, with additional columns optional. The sample names in sampleID column must match the expression data file.

Sample MetaData File
Browse... AP_PD_Proteomics_Sample.csv
Upload complete

Expression data should be matrix of expression values with genes/proteins as rows, and samples as columns. The unique IDs for genes/proteins are in the first column. We recommend using log of normalized expression values (e.g. log2(TPM+1)). Upload csv file, can be compressed as .gz or .zip file.

Expression Data File
Browse... AP_PD_Proteomics_Exp_data.csv
Upload complete

Comparison data should have five columns, UniqueID, test, Adj.PValue, PValue and logFC. The comparison names are listed in test column. Upload csv file, can be compressed as .gz or .zip file.

Comparison Data File
Browse... AP_PD_Proteomics_Comparis.csv
Upload complete

Create Gene/Protein Name File automatically (or uncheck to upload your own file)

Unique ID Type in the Data Files
 Ensembl Gene ID Gene Symbol NCBI GeneID UniProtKB Protein ID UniProt Protein Name
 Show ID type examples
 Fill in uniqueID when Gene.Name not found
 Add gene/protein description

Submit Data

The direct URL for the uploaded dataset is: http://quickomics.bxgenomics.com/?unlisted=PRJ_A.D.PD.Proteomics.Demo.Test_dVR7nE

Annotations on the Quickomics upload page:

- A red box labeled "Tool to upload text files directly" points to the "Upload Files" button.
- A red box labeled "Uploaded files" points to the "Upload complete" status message for each of the three uploaded files (Sample MetaData, Expression Data, and Comparison Data).
- A red box labeled "ID handling options" points to the "Unique ID Type in the Data Files" section.
- A red box labeled "Link to access uploaded project" points to the direct URL provided at the bottom of the form.

After the data files are uploaded and processed, the system will automatically generate all required R data files and provide a link for the user to explore the dataset in Quickomics immediately.

Behind the scene, Bioconductor biomaRt package

(<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>) has been used to convert gene IDs (Ensembl gene, NCBI gene ID, etc.) into gene symbols by querying Ensembl databases. For protein IDs, we generated a custom lookup table using information downloaded from UniProt Knowledgebase to convert UniProt IDs to gene symbols and protein names. We didn't use biomaRt for proteins as Ensembl databases only cover about 60-80% protein IDs in a typical proteomics data set.

2.2 Prepare R Data Files by Computational Biologists

We recommend uploading csv files, which is convenient for general users who can skip section 2.2 entirely. Nevertheless, experienced R programmers can create R data files to be uploaded through “Upload RData File” option. Section 2.2.1 and 2.2.2 provide example R scripts to prepare such R data files for RNA-Seq and proteomics data, respectively.

Two R data files are required for each data set, one contains the main data and the other contains gene co-expression network information. For the pre-loaded datasets, main data files are located in the “data” folder, <https://github.com/interactivereport/Quickomics/tree/master/data>, and gene co-expression network files are located in the “networkdata” folder, <https://github.com/interactivereport/Quickomics/tree/master/networkdata>. One can review the content of a R data file (e.g. Mouse_microglia_RNA-Seq.RData) in the “data” folder by loading it into R. The main R data file contains the following R data frame objects.

1. **MetaData:** It must have “sampleid”, “group”, “Order” and “ComparePairs” columns. Additional metadata columns about samples are optional. “sampleid” should match those used in expression data. “group” holds group names of samples. “Order” is ordered group names used on plotting. “ComparePairs” are names of comparisons performed.
2. **ProteinGeneName:** It must have “UniqueId”, “Gene.Name” and “Protein.ID” columns. “UniqueId” matches gene ID in below data_wide and data_long objects. “Gene.Name” should be official gene symbols. “Protein.ID” is UniProt protein IDs, or empty for RNA-Seq data. Additional columns about proteins or genes are optional.
3. **data_wide:** This is the expression matrix in which rows are genes and columns are samples. Samples must match “sampleid” values in MetaData and gene IDs must match “UniqueId” values in ProteinGeneName.
4. **data_long:** Gene expression matrix in long format with four columns, “UniqueId”, “sampleid”, “expr” and “group”. “group” values must match those listed in MetaData.
5. **results_long:** The comparison results in long format with five columns, “UniqueId”, “test”, “Adj.P.Value”, “P.Value” and “logFC”. “UniqueId” matches “UniqueId” in ProteinGeneName. “test” column has the comparison names that must match “ComparePairs” values in MetaData. The other values are typically computed from statistical analysis, but the data headers must be changed to “Adj.P.Value”, “P.Value” and “logFC”.
6. **data_results:** This is a summary table starting with “UniqueId” and “Gene.Name” columns, then the intensity (max or mean expression value from data_wide for each gene), mean and SD expression values for each group, and finally comparison data (comparison name added as prefix of columns).

The network data object is computed from “data_wide” expression matrix by using Hmisc R package exemplified by the code snippet below.

```
cor_res <- Hmisc::rcorr(as.matrix(t(data_wide)))
cormat <- cor_res$r
pmat <- cor_res$P
ut <- upper.tri(cormat)
network <- tibble::tibble (
  from = rownames(cormat)[row(cormat)[ut]],
  to = rownames(cormat)[col(cormat)[ut]],
  cor = signif(cormat[ut], 2),
  p = signif(pmat[ut], 2),
  direction = as.integer(sign(cormat[ut]))
)
```

2.2.1 Example R script to prepare R data files from RNA-Seq results

We have provided example input files (TPM and count matrix files, sample grouping file, comparison list file) and the R scripts to generate the main data and network R data files at

https://github.com/interactivereport/Quickomics/tree/master/demo_files/Example_RNA_Seq_data.

Please note that you may need to modify RNA_Seq_raw2quickomics.R to fit your input files.

- rsem TPM.txt: The TPM matrix. One can also use RPKM matrix if needed.
- rsem_expected_count.txt: The gene count matrix. We used RSEM counts in this case, but gene count results from other methods can be used as well.
- grpID.txt: This file lists the group information for each sample.
- comparison.txt: This list lists the comparisons to perform (group 1 vs group 2 in each row).

The following command will read the above data files, run differential gene expression analysis using DESeq2, and create main and network R data files.

```
$ Rscript RNA_Seq_raw2quickomics.R
```

2.2.2 Example R script to prepare R data files from proteomics results

We have provided the example input files (normalized protein expression, comparison data, sample information, protein and gene names) and the R script to generate the main data and network R data files at

https://github.com/interactivereport/Quickomics/tree/master/demo_files/Example_Proteomics_data.

Please note that you may need to modify Proteomics2Quickomics.R to fit your input files.

- NormalizedExpression.csv: Normalized protein expression (log2 transformed).
- ComparisonData.csv: Comparison results. The statistic values are: logFC, P.Value and Adj.P.Value. This can be created using R packages like limma.
- Sample.csv: Sample information file.
- ProteinID_Symbol.csv: This file lists the proteinIDs and associate gene symbols.

The following command will read the above data files and create main and network R data files.

```
$ Rscript Proteomics2Quickomics.R
```

2.3 Sample Table

A sample table with metadata/details appears upon completion of dataset loading. Here we have selected the “Mouse Microglia RNA” dataset from the previously published paper (Gyoneva *et al.*, 2019) as an example to illustrate all the functionalities. This publication used RNAseq to conclude that the knockout of the Cx3cr1 gene altered the microglial transcriptome in a manner similar to what ageing does, albeit in a milder extent. Using Quickomics, we can visualize the original results from the paper, as well as provide novel visualizations to support new results.

The Sample Table contains detailed info of all study samples. A total of 93 samples that were sequenced in this project are listed in the table with attributes like sampleid, group, Age, Genotype and Gender, shown in different columns.

Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data

Sample Table Project Overview Result Table Data Table Protein Gene Names Upload Files Help

Show 15 entries CSV Excel Print

Select data set
 Saved Projects
 Upload RData File
 Upload Data Files (csv)

Available Dataset
 Mouse Microglia RNA

| sampleid | group | Age | Genotype | Gender |
|----------|-------------|-----|----------|--------|
| 1 | 2mo_WT-10F | 2mo | WT | F |
| 2 | 2mo_WT-11F | 2mo | WT | F |
| 3 | 2mo_WT-1M | 2mo | WT | M |
| 4 | 2mo_WT-2M | 2mo | WT | M |
| 5 | 2mo_WT-3M | 2mo | WT | M |
| 6 | 2mo_WT-4M | 2mo | WT | M |
| 7 | 2mo_WT-5M | 2mo | WT | M |
| 8 | 2mo_WT-6F | 2mo | WT | F |
| 9 | 2mo_WT-7F | 2mo | WT | F |
| 10 | 2mo_WT-8F | 2mo | WT | F |
| 11 | 2mo_WT-9F | 2mo | WT | F |
| 12 | 2mo_Het-10M | 2mo | Het | M |
| 13 | 2mo_Het-1M | 2mo | Het | M |
| 14 | 2mo_Het-2M | 2mo | Het | M |
| 15 | 2mo_Het-3F | 2mo | Het | F |

Showing 1 to 15 of 93 entries

Previous 1 2 3 4 5 6 7 Next

Metadata attributes as columns

Showing the total number of samples

2.4 Project Overview

This tab gives an overview of all samples in the project by providing a summary from the metadata. This includes information like which species was used, the different comparisons run and total number of samples in each group. This tab also helps identify how many groups were present in the dataset, like in this project there are 9 different groups that correspond to three genotypes in each of three time points (WT, Het, KO; 2mo, 1yr, 2yr).

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects Output

Sample Table Project Overview Result Table Data Table

Select data set
 Saved Projects
 Upload RData File
 Upload Data Files (csv)

Available Dataset
 Mouse Microglia RNA

Project Mouse Microglia RNA

- Species: mouse
- Description: Cx3cr1-Deficient Mouse Microglia RNA-Seq
- Number of Samples: 93
- Number of Groups: 9 (please see group table below)
- Number of Genes/Proteins: 15402
- Number of Comparison Tests: 9
 - 2mo_Hetvs2mo_WT
 - 2mo_KOvs2mo_WT
 - 1yr_Hetvs1yr_WT
 - 1yr_KOvs1yr_WT
 - 2yr_Hetvs2yr_WT
 - 2yr_KOvs2yr_WT
 - 2yr_WTvs1yr_WT
 - 2yr_WTvs2mo_WT
 - 1yr_WTvs2mo_WT

Quick summary of data structure in groups

Number of Samples in Each Group

| Tyr_Het | 1yr_KO | 1yr_WT | 2mo_Het | 2mo_KO | 2mo_WT | 2yr_Het | 2yr_KO | 2yr_WT |
|---------|--------|--------|---------|--------|--------|---------|--------|--------|
| 11 | 12 | 10 | 9 | 11 | 11 | 8 | 11 | 10 |

Details about the project and listing the comparisons as defined during data upload

2.5 Result Table

This tab contains statistical analysis results like Ensemble ID (for RNAseq data) and Accession ID (for proteomics data), gene name, mRNA or protein abundance (count or intensity), mean and SD values for each group, log2 fold change, P value and adjusted P value.

Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects Output

Sample Table Project Overview Result Table Data Table Protein Gene Names Upload Files

Show 15 entries CSV Excel Print

Search:

| Gene.Name | Intensity | Protein.ID | 2mo_WT_Mean | 2mo_WT_sd | 2mo_Het_Mean | 2mo_Het_sd |
|---------------------|-----------|------------|-------------|-----------|--------------|------------|
| ENSMUSG000000000001 | Gna13 | | 6.09 | 6.08 | 0.152 | 6.1 |
| ENSMUSG000000000028 | Cdc45 | | 1.54 | 1.51 | 0.547 | 1.42 |
| ENSMUSG000000000056 | Narf | | 5.78 | 5.81 | 0.175 | 5.96 |
| ENSMUSG000000000058 | Cav2 | | 3.84 | 3.68 | 0.231 | 3.73 |
| ENSMUSG000000000078 | Klf6 | | 6.99 | 6.75 | 0.219 | 6.74 |
| ENSMUSG000000000085 | Scmh1 | | 5.28 | 5.25 | 0.196 | 5.35 |
| ENSMUSG000000000088 | Cox5a | | 7.28 | 7.28 | 0.0998 | 7.2 |
| ENSMUSG000000000127 | Fer | | 4.84 | 4.81 | 0.159 | 4.93 |

2.6 Data Table

The Data table contains normalized expression values (TPM for RNAseq, intensity for label free proteomic quantification, or ratio for isobaric label proteomics quantification) per sample for each gene.

Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects Output

Sample Table Project Overview Result Table Data Table Protein Gene Names Upload Files

Show 15 entries CSV Excel Print

Search:

| | 2mo-WT-10F | 2mo-WT-11F | 2mo-WT-1M | 2mo-WT-2M | 2mo-WT-3M | 2mo-WT-4M | 2mo-WT-5M | 2mo-WT-6F | 2mo-WT-7F | 2mo-WT-8F | 2mo-WT-9F |
|---------------------|------------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| ENSMUSG000000000001 | 6.129 | 6.295 | 6.269 | 6.293 | 6.064 | 5.996 | 5.865 | 6.044 | 6.063 | 6.035 | 5.872 |
| ENSMUSG000000000028 | 1.692 | 2.036 | 1.575 | 1.614 | 1.007 | 0.422 | 2.205 | 1.345 | 2.018 | 1.807 | 0.895 |
| ENSMUSG000000000056 | 5.752 | 5.783 | 5.884 | 5.83 | 5.869 | 5.64 | 6.1 | 5.945 | 5.616 | 5.492 | 5.964 |
| ENSMUSG000000000058 | 3.794 | 3.377 | 3.374 | 3.587 | 3.781 | 3.979 | 3.411 | 3.876 | 3.534 | 3.974 | 3.782 |
| ENSMUSG000000000078 | 6.607 | 6.75 | 6.959 | 7.052 | 6.739 | 6.559 | 6.512 | 6.93 | 6.87 | 6.36 | 6.934 |
| ENSMUSG000000000085 | 5.527 | 5.409 | 5.155 | 5.17 | 5.443 | 5.087 | 5.288 | 4.946 | 4.968 | 5.409 | 5.309 |

2.7 Protein Gene Names

This tab contains all quantified gene or protein IDs and gene names.

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network

Output

For protein data, Protein.ID can be matched to Gene.Name

| Select data set | | Sample Table Project Overview Result Table Data Table Protein Gene Names Upload Files Help | | | | |
|---|---|---|-----------|------------|----------------|---------|
| <input checked="" type="radio"/> Saved Projects | <input type="radio"/> Upload RData File | <input type="radio"/> Upload Data Files (csv) | | | | |
| Save to output | | | | | | Search: |
| Show 15 entries CSV Excel Print | | | | | | |
| | id | UniqueID | Gene.Name | Protein.ID | GeneType | |
| 0 | ENSMUSG000000000001 | | Gnai3 | | protein_coding | |
| 1 | ENSMUSG000000000028 | | Cdc45 | | protein_coding | |
| 2 | ENSMUSG000000000056 | | Narf | | protein_coding | |
| 3 | ENSMUSG000000000058 | | Cav2 | | protein_coding | |
| 4 | ENSMUSG000000000078 | | Klf6 | | protein_coding | |
| 5 | ENSMUSG000000000085 | | Scmh1 | | protein_coding | |
| 6 | ENSMUSG000000000088 | | Cox5a | | protein_coding | |
| 7 | ENSMUSG000000000127 | | Fer | | protein_coding | |
| 8 | ENSMUSG000000000131 | | Xpo6 | | protein_coding | |

2.8 Help

Finally, the “Help” feature summarizes what data is being displayed in each tab under Select Dataset. The Help tab is available for all sections of the subsequent analysis.

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects

Output

Help

| Select data set | | Sample Table Project Overview Result Table Data Table Protein Gene Names Upload Files Help | | | | |
|---|---|---|-------|--|----------------|--|
| <input checked="" type="radio"/> Saved Projects | <input type="radio"/> Upload RData File | <input type="radio"/> Upload Data Files (csv) | | | | |
| Result Table: Statistics result by using LIMMA package, include log2 Fold Change, p value, p-values adjusted (Benjamini-Hochberg) | | | | | | |
| Data Table: Normalized data | | | | | | |
| Sample Table: sample group and comparison information | | | | | | |
| Protein Gene Names: protein accession number and gene symbol | | | | | | |
| 0 | ENSMUSG000000000001 | | Gnai3 | | protein_coding | |
| 1 | ENSMUSG000000000028 | | Cdc45 | | protein_coding | |
| 2 | ENSMUSG000000000056 | | Narf | | protein_coding | |
| 3 | ENSMUSG000000000058 | | Cav2 | | protein_coding | |
| 4 | ENSMUSG000000000078 | | Klf6 | | protein_coding | |
| 5 | ENSMUSG000000000085 | | Scmh1 | | protein_coding | |
| 6 | ENSMUSG000000000088 | | Cox5a | | protein_coding | |
| 7 | ENSMUSG000000000127 | | Fer | | protein_coding | |
| 8 | ENSMUSG000000000131 | | Xpo6 | | protein_coding | |

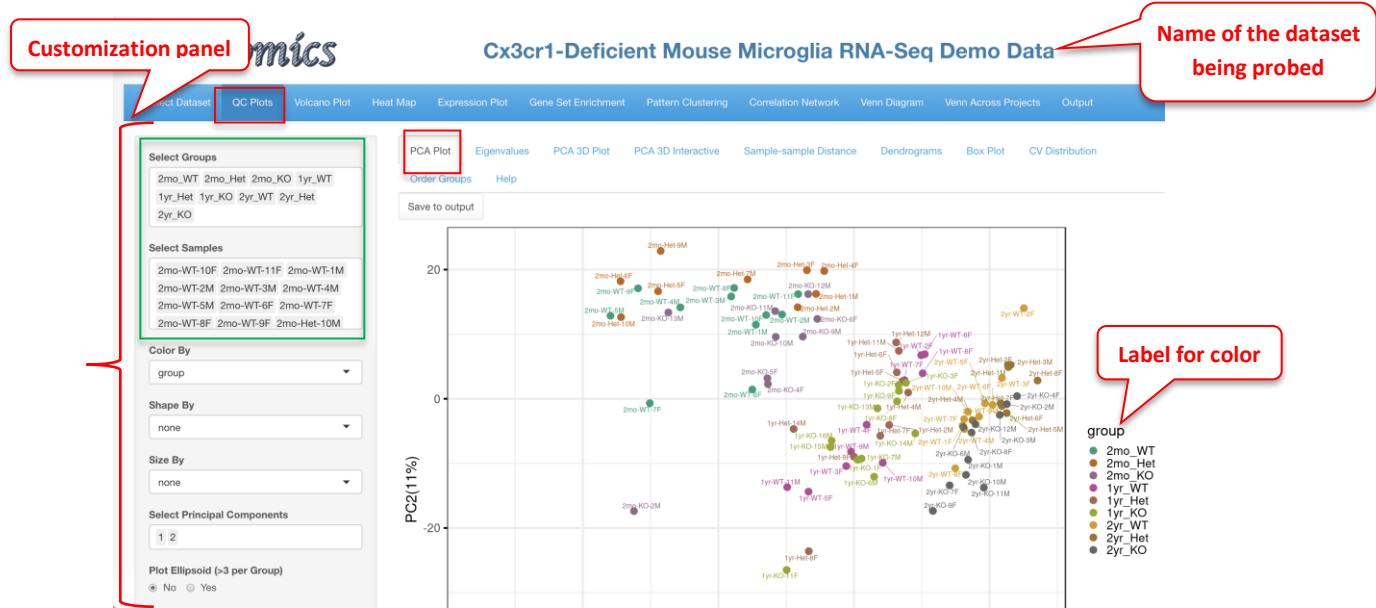
QuickOomics ver1.0 Developed by: Benbo Gao, Xinmin Zhang and Baohong Zhang
[More information at GitHub](#)

3 QC Plots Module

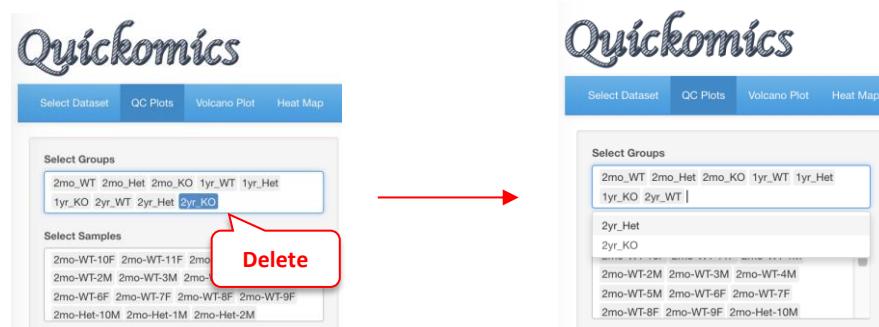
The “QC Plots” module performs basic QC analyses, as described in following paragraphs, on an uploaded dataset to evaluate integrity and quality of the data. The plots were generated using ggplot, plotly, ComplexHeatmap and pheatmap R packages.

3.1 PCA Plot

The principle component analysis (PCA) plot displays, by default, the first 2 PC's in the dataset and colors study groups with “none” selected in shape and size settings. This plot is highly customizable. Users can change the display by annotating the plot with different colors, shapes, size scheme, as well as label selected subsets data points.



Additionally, the “Select Groups” and “Select Samples” feature (Green square) is present in all the QC Plots tabs. The User is able to remove any particular group or sample from the plots.



To illustrate the strengths of customization, we changed a few display elements as shown below, to visualize the influence of Age and Genotype on sample clustering:

Select Groups

2mo_WT 2mo_Het 2mo_KO 1yr_WT 1yr_Het 1yr_KO
 2yr_WT 2yr_Het 2yr_KO

Select Samples

2mo-WT-10F 2mo-WT-11F 2mo-WT-1M 2mo-WT-2M
 2mo-WT-3M 2mo-WT-4M 2mo-WT-5M 2mo-WT-6F
 2mo-WT-7F 2mo-WT-8F 2mo-WT-9F 2mo-Het-10M
 2mo-Het-1M 2mo-Het-2M 2mo-Het-3F 2mo-Het-4F

Color By

Age

Shape By

Genotype

Size By

none

Select Principal Components

1 2

Plot Ellipsoid (>3 per Group)

No Yes

Show Mean Point

No Yes

Show Marginal Rugs

No Yes

Select palette

Set3

Dot Size (when size by not used):

A horizontal slider with a blue track and a grey circular handle. The numbers 1 through 20 are labeled below the track. The handle is positioned between 3 and 5, with a small blue box above it containing the number 4.

Label Samples:

All None Subset

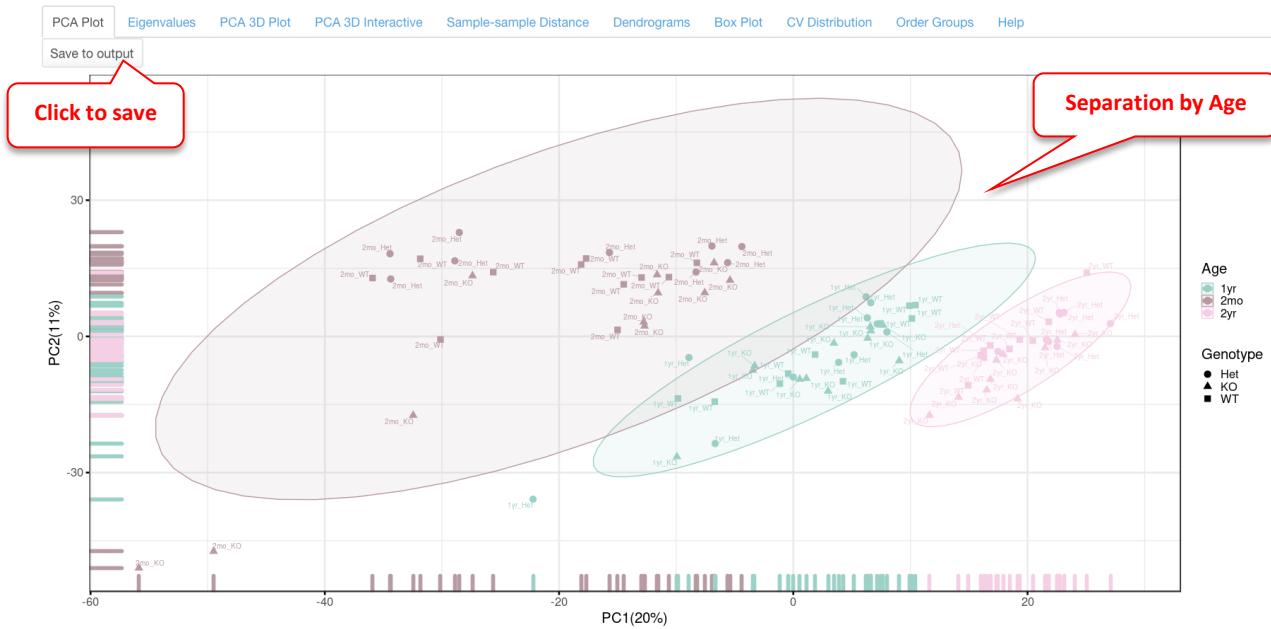
Label Font Size:

A horizontal slider with a blue track and a grey circular handle. The numbers 1 through 20 are labeled below the track. The handle is positioned between 9 and 11, with a small blue box above it containing the number 10.

Select Sample Label

sampleid group Age Genotype Gender

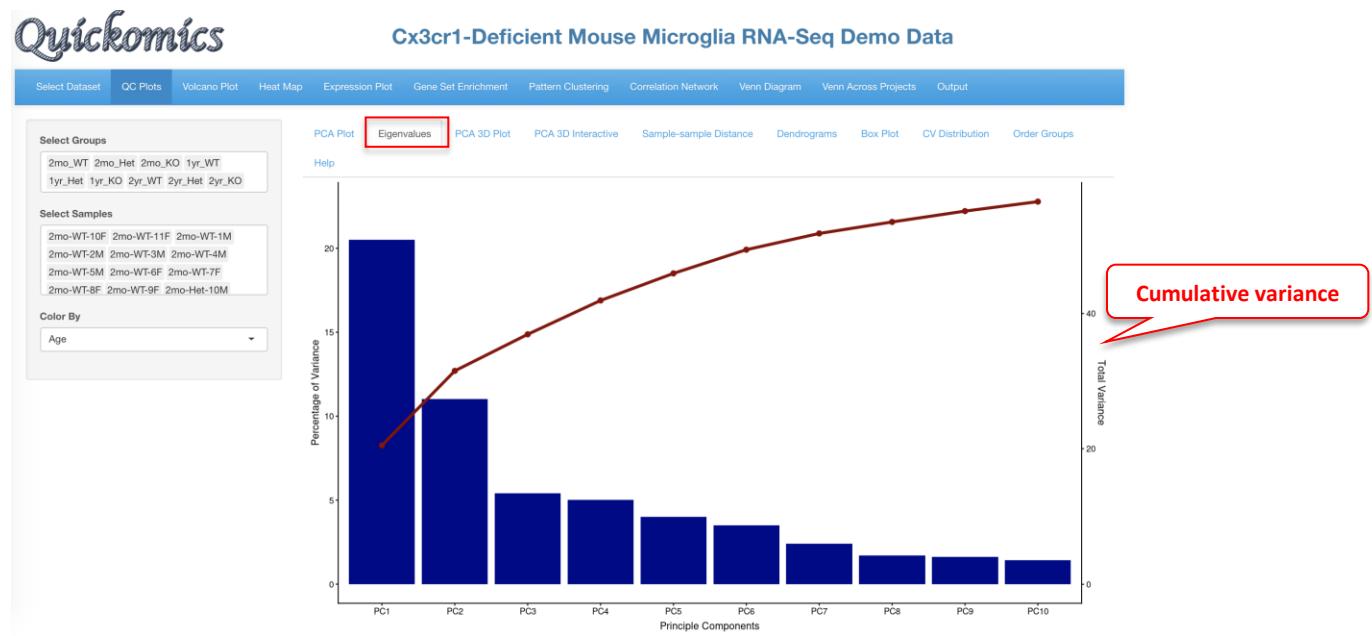
These changes results in a PCA plot that clearly showed that Age is the factor that drives the biggest separation, especially along PC1. There is some separation driven by the Genotype as well. The Ellipsoid feature helps group samples by the color attribute, while the Marginal Rug feature helps understand the sample distribution along the axis.



Furthermore, the plot can be saved in high resolution by clicking on “Save to Output”. Section 12 of this document describes the next steps of downloading and obtaining the saved plots.

3.2 Eigenvalues

By default, the PCA plot displays PC1 and PC2 as described in Section 3.1. However, in many cases other PC's may be important in explaining additional variance in the dataset. This sub-tab generates a plot of the variances versus the first 10 PCs in the dataset, enabling Users to make educated decisions of which PCs to plot in 2D scatter plot. In this dataset PC1 and PC2 together explain around 30% of the variance.

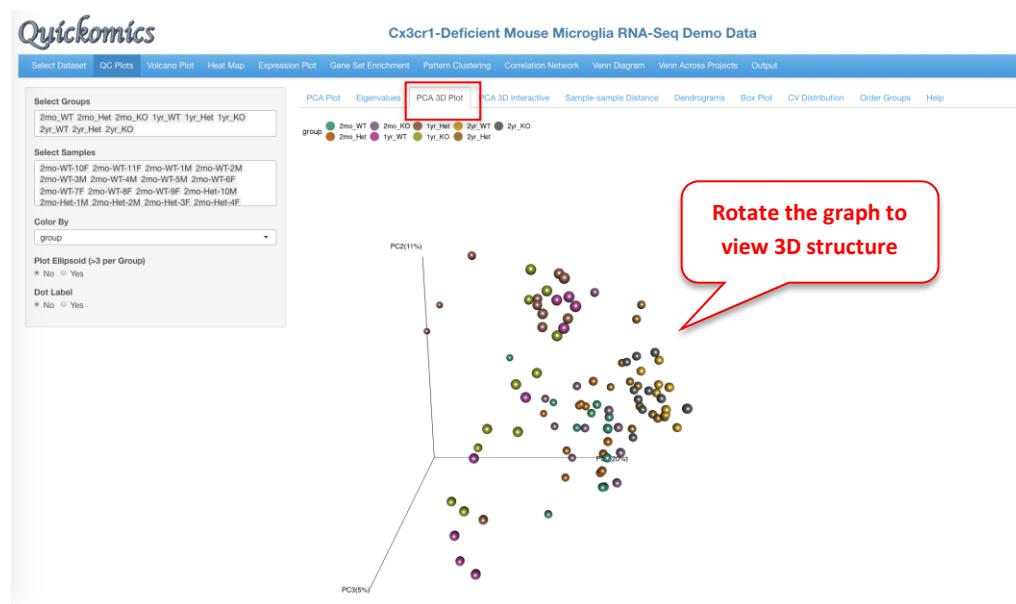


PC2 and PC3 were used to redo the PCA plot, but here the factors driving the separation of the samples is not as clear.

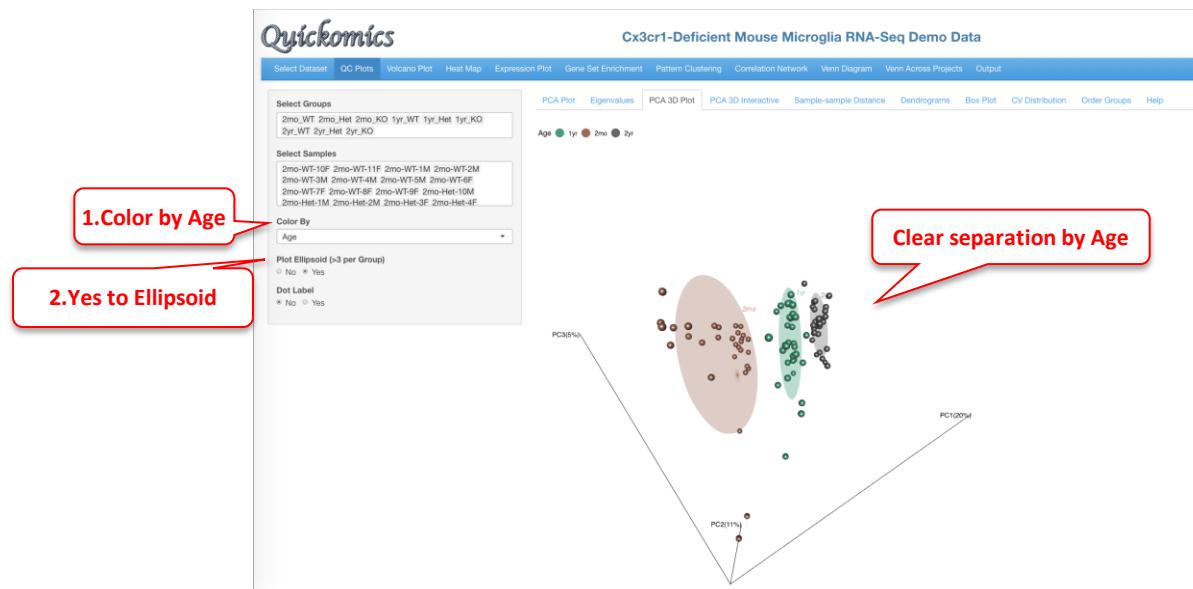


3.3 PCA 3D Plot

This sub-tab contains a way to perform 3D PCA representation. The functionality is limited to doing so only on PC1, PC2 and PC3. Users can zoom in and out to focus on different areas of the plot or rotate the plot in different axes to find the best separation by age as shown below. There are 3 attributes that can be changed: color, ellipsoid inclusion, and labels of the samples.

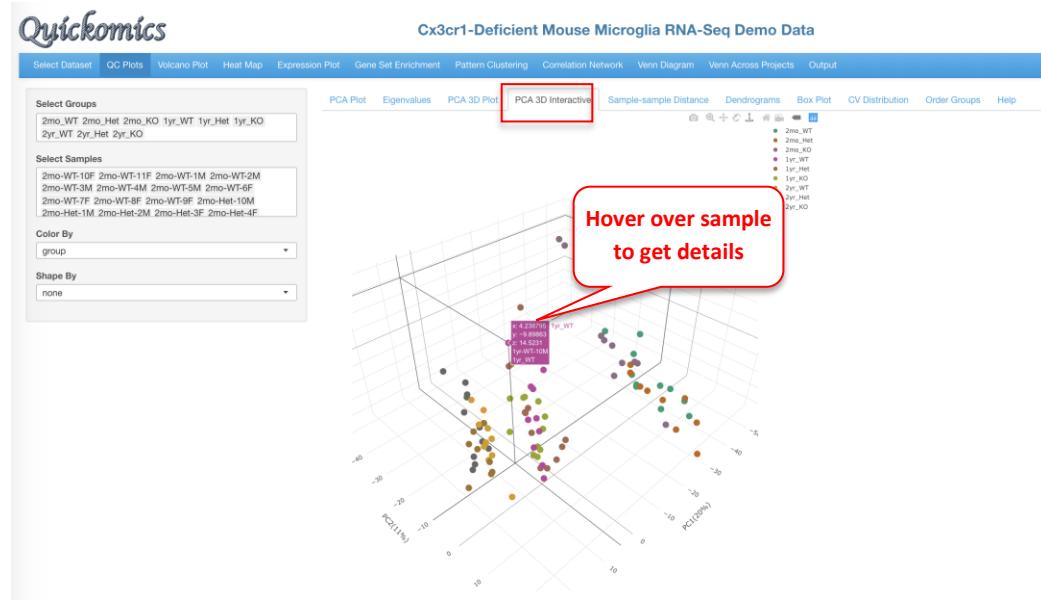


The “Color By” and “Plot Ellipsoid” attributes were changed in the following plot. This highlights the clear separation of samples by Age and not Genotype, contrary to the original biological hypothesis.

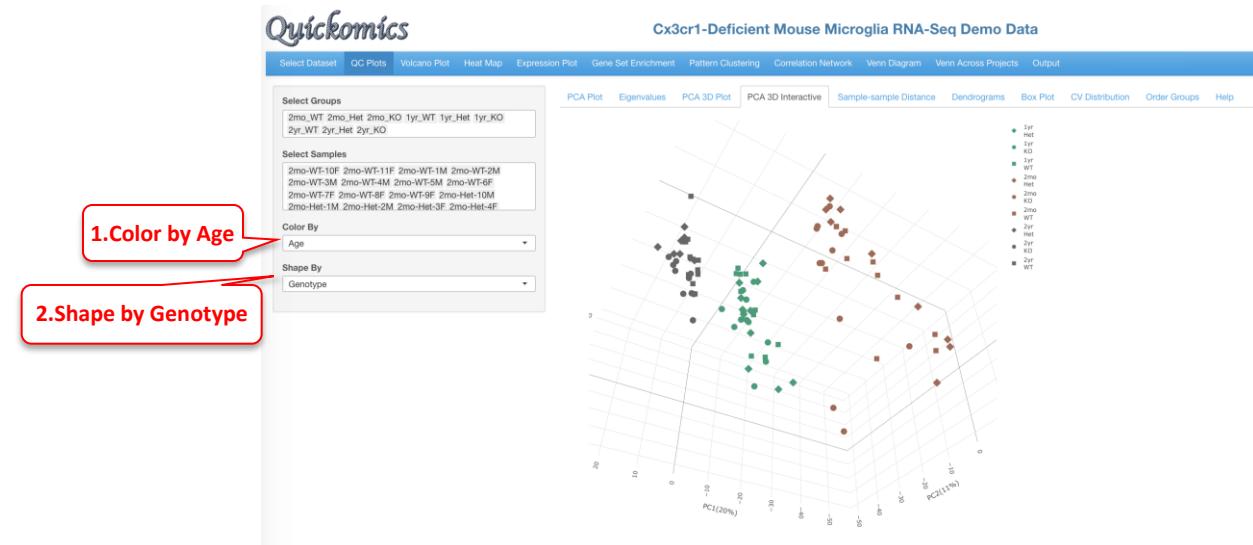


3.4 PCA 3D Interactive

This tab is similar to the previous one (3.2 PCA 3D Plot), but here users have the additional capability to hover over the samples on the plot to identify more details, see the pointer below. This was implemented through the plotly package in R. Users can change 2 attributes here, the color and shape.



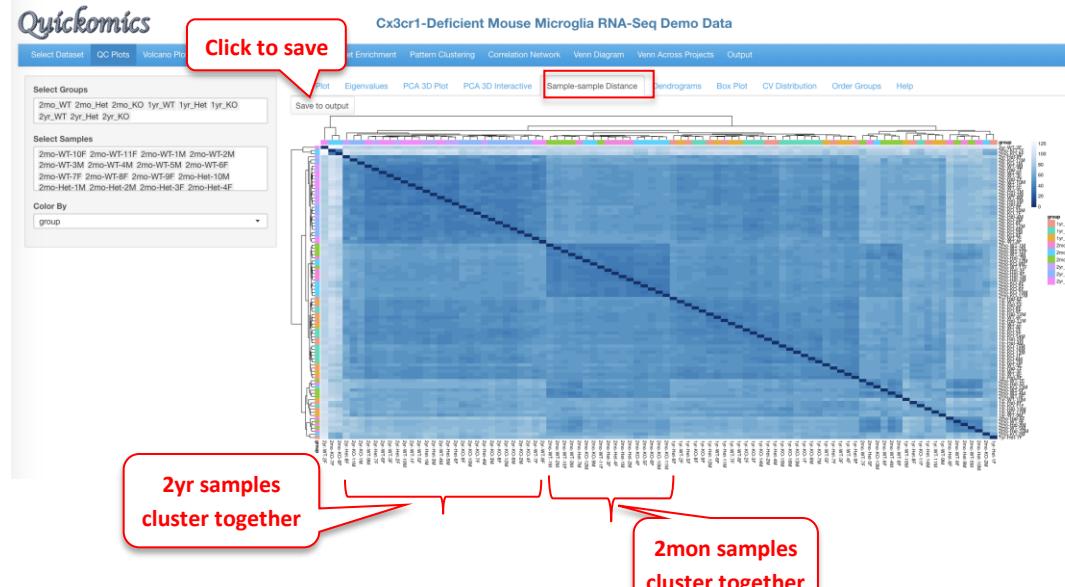
In the plot below, the “Color By” and “Shape By” attributes have been changed to highlight the drivers of separation.



3.5 Sample-sample Distance

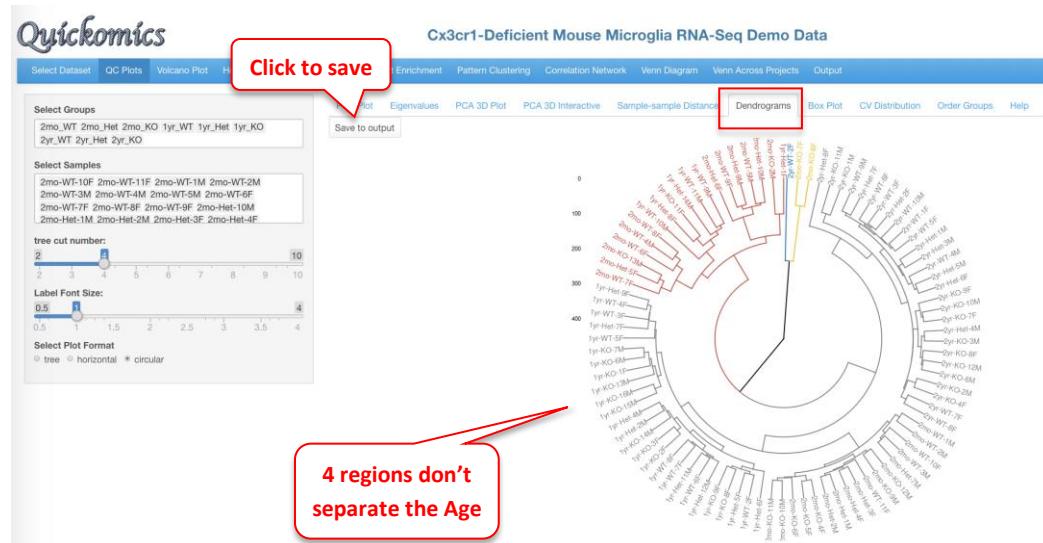
This tab helps identify pair-wise similarity between all samples. A distance matrix is generated for every sample pair and is plotted as a heatmap. The rows and columns are clustered based on similarity.

It is clear from this plot that samples of the same Age have the smaller distances, that is, they have closer gene expression patterns. Here again, Users have the option to save to output.



3.6 Dendrograms

Like the previous plot (3.5 Sample-sample Distance), the Dendrogram plot helps visualize hierarchical clustering relationships between samples. The Default plot is Circular and cut into four parts. Users have the option to visualize in two other ways (tree or horizontal) and cut the plot in multiple regions. Here again, Users can save the plot to output.

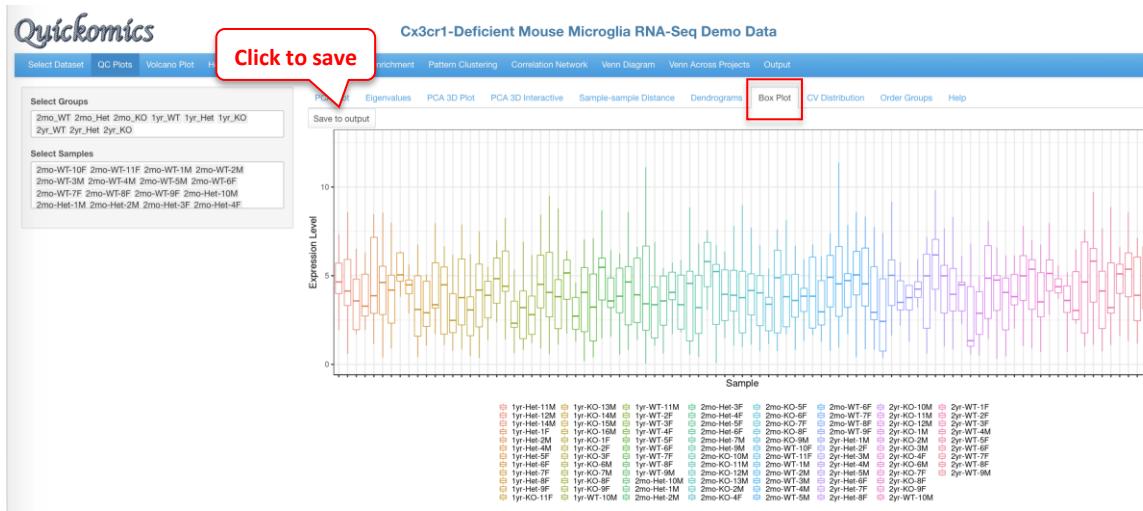


In the plot below, we have selected to visualize as a tree and cut into eight parts. This shows the relationship between the samples to help understand the hierachal clustering. Overall, the samples are arranged by age, but there are some outliers that appear. It is also interesting how the 2mo samples cluster with the 1yr samples.



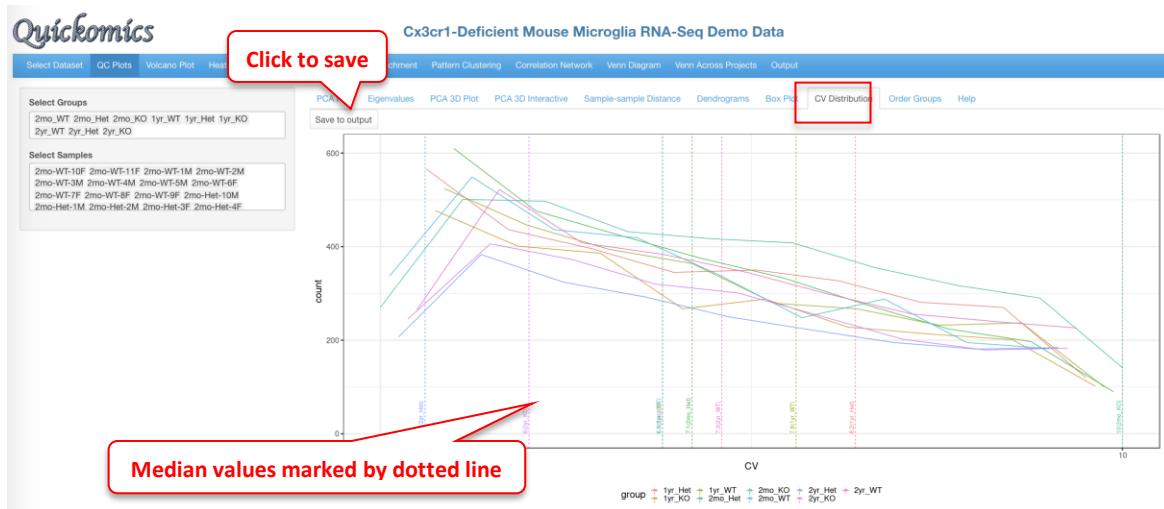
3.7 Box Plot

The Boxplot is a visualization to understand the distribution of the normalized expression levels in all samples. This identifies the minimum, first quartile, median, third quartile, and maximum values in the dataset. In this demo dataset, most of the samples have the same range of expression, indicating that there are no outliers.



3.8 CV Distribution

This plot shows the histogram of CVs (coefficient of variation) and a dotted line for each group indicates the median CV. In this dataset, the 2mon_KO samples have the highest CV while the 2yr_Het samples have the lowest variability, which corresponds to what is visually seen on the PCA plot for these samples in Section 3.1.



3.9 Order Groups

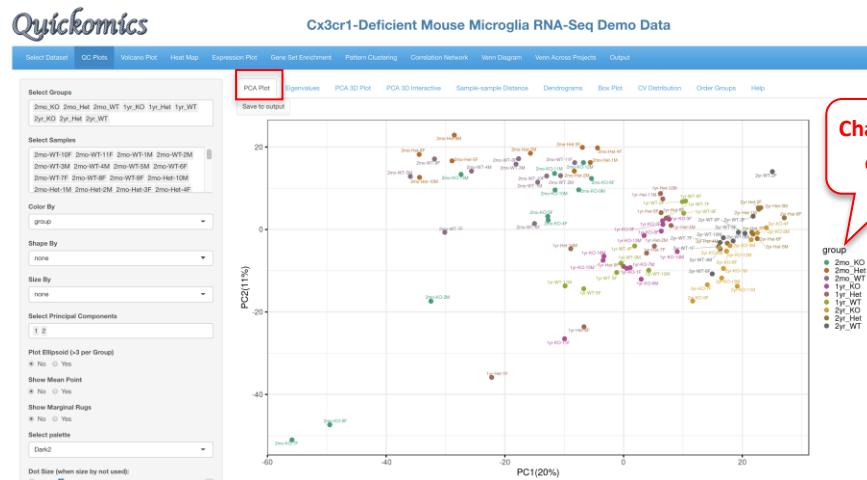
In this sub-tab, Users can select and/or re-order the groups in the dataset. The order will be stored in the current session, and labels will be displayed in this preferred order for all plots. Users can either drag and drop a group to reorder or use the select panel to delete and add. In this example the KO genotype was of highest interest and so we have changed the order of the group so that the KO samples are before Het and WT.

| Drag and drop | Delete and Add |
|---------------|----------------|
| | |

The changed groups are now visible. To revert back to the default, Users have the option to click “Reset Groups”.

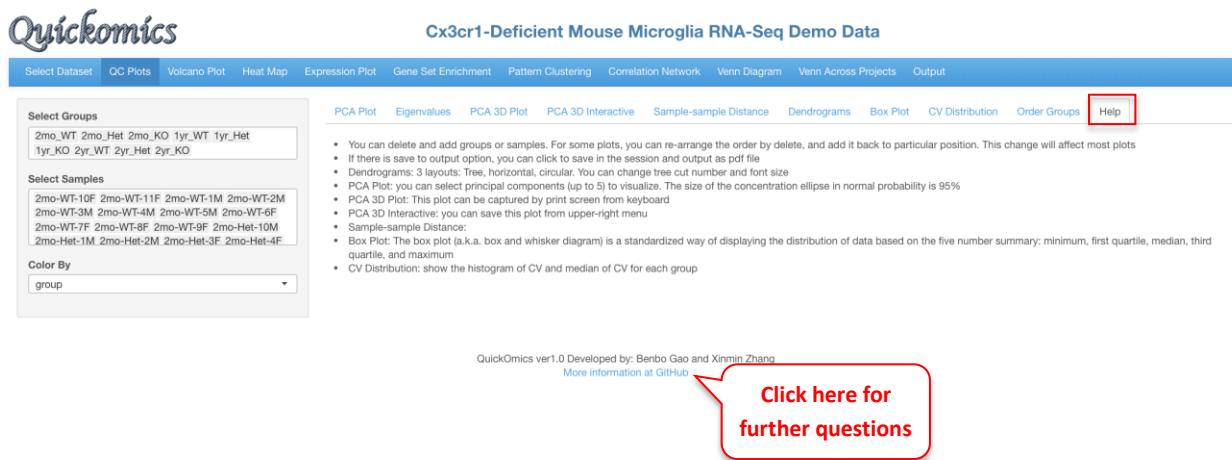
The figure shows the 'Reset Groups' feature in the Quickomics interface. It displays the 'Select Groups' and 'Select Samples' panels. A red box highlights the message 'Changed order' at the bottom of the page, indicating that the group order has been modified.

As noted above, that the new order of groups will dynamically change all other related visualizations. As an example, a 2D PCA plot reflecting a newly changed group order is shown below.



3.10 Help

The Help section describes each tab of the QC Plot and what is being visualized. For further help please visit the GitHub containing the source code for this tool (<https://github.com/interactivereport/Quickomics>) or click on the link at the bottom of the page.



4 Volcano Plot Module

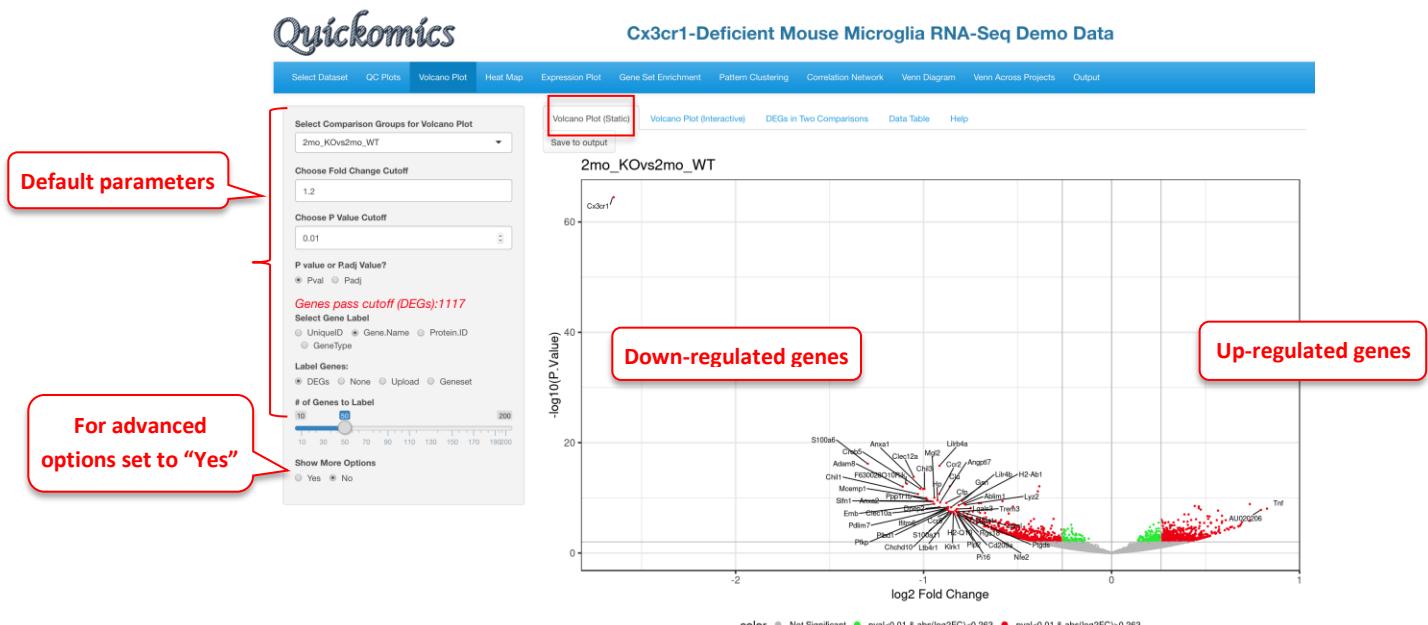
This section is designed to explore the differentially expressed genes (DEGs) and differentially expressed proteins (DEPs) in the dataset. The comparisons are defined by the User when they input the dataset. Please refer to our GitHub for more details on uploading datasets.

The “Volcano Plot (Static)” described in Section 4.1 and “Volcano Plot (Interactive)” described in Section 4.2 are ways to visualize each pair-wise comparison one by one. “DEGs/DEPs in Two Comparisons” described in Section 4.3 is a novel way to visualize the similarities and differences in two comparisons. Finally, “Data Table” described in Section 4.4 displays a searchable table containing all genes and their fold changes and P values in each comparison.

4.1 Volcano Plot (Static)

This plot reveals the differentially expressed genes provided in the input data file. By default, the Fold change Cutoff is set to 1.2 and the P value Cutoff is set to 0.01. Users have the option to change any of these parameters. For more advanced display options, they can set the “Show More Options” to “Yes” to customize other settings.

The red dots are indicative of all genes that pass the Fold Change and P value cutoff, while the green dots are indicative of genes that only pass the P value cutoff. Fifty random genes are labelled for both the up-regulated and down-regulated genes by default.



For demonstration purposes, we have changed the following attributes to highlight a few genes in this dataset. We have selected the 2mon_KOvs2mon_WT comparison, where Gyoneva et al., 2019 report immune response genes as the top differentially expressed category. To highlight these genes, in the example below, we select an “Immune Response” dataset from the Molecular Signatures Database (MSigDB) to highlight the genes by adding labels to the volcano plot. As MSigDB contains only human gene lists and the demo dataset is from mouse, Quickomics does a quick conversion of the names of

Human genes to Mouse genes by changing the letters after the first one to lowercase. Additionally, in this example we removed some of the genes from the list to reduce size. This visualization was not available in the original publication, but clearly supports the finding that the KO genotypes altered the expression of immune-related genes.

Quickomics

Select comparison
Change Fold Change
Select Padj
Select Geneset
Add or remove any gene
For customizing the axis limits in the plot

Volcano Plot (Interactive) DEGs in Two Comparisons Data Table Help

Select Comparison Groups for Volcano Plot 2mo_KOvs2mo_WT

Choose Fold Change Cutoff 1.5

Choose P Value Cutoff 0.01

P value or Padj Value? Padj

Genes pass cutoff (DEGs): 146

Select Gene Label UniqueID Gene.Name Protein.ID GeneType

Label Genes: DEGs None Upload Geneset

of Genes to Label 10

Q Select Genes List of genes to label (UniqueId, Gene.Name or Protein.ID)

CD55
CD14
EMP2
MDK
FCGR2
DAB2

Show More Options Yes No

Max abs(logFC) in plot (use 0 for full range) 1.5

Max -log10(Stat Value) in plot (use 0 for full range) 15

Label Font Size: 1 2 3 4 5 6 7 8 9 10

Legend Font Size: 8 10 12 14 16 18 20 22 24

Select Database

Search for a particular keyword

Immune response

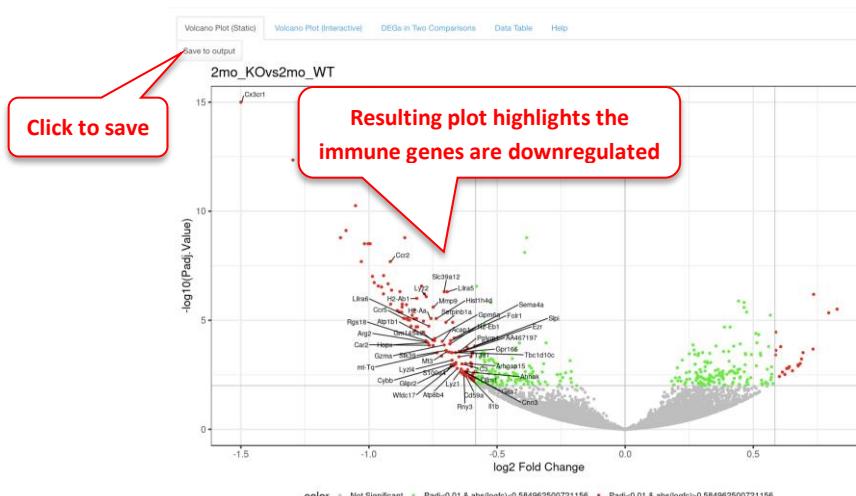
GeneSet Details

Display Method:

- Gene IDs, one per row
- Gene IDs, comma separated
- Gene Names, one per row
- Gene Names, comma separated

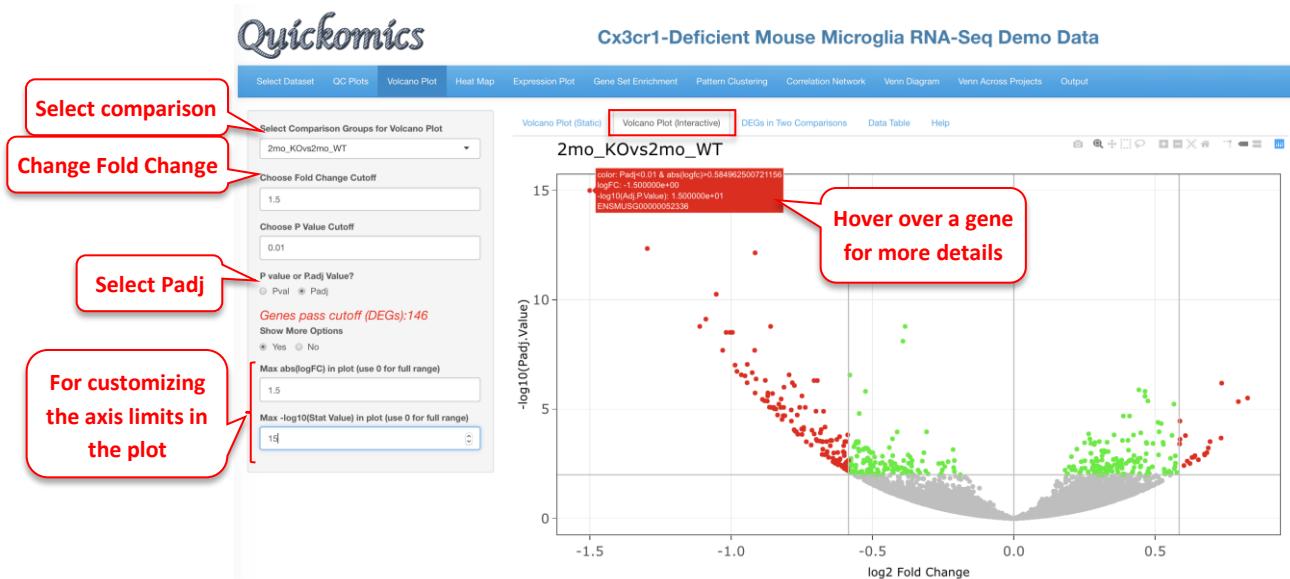
CD55
CD14
EMP2
MDK
FCGR2
DAB2
MYLK
SELENOP

To view genes in a group



4.2 Volcano Plot (Interactive)

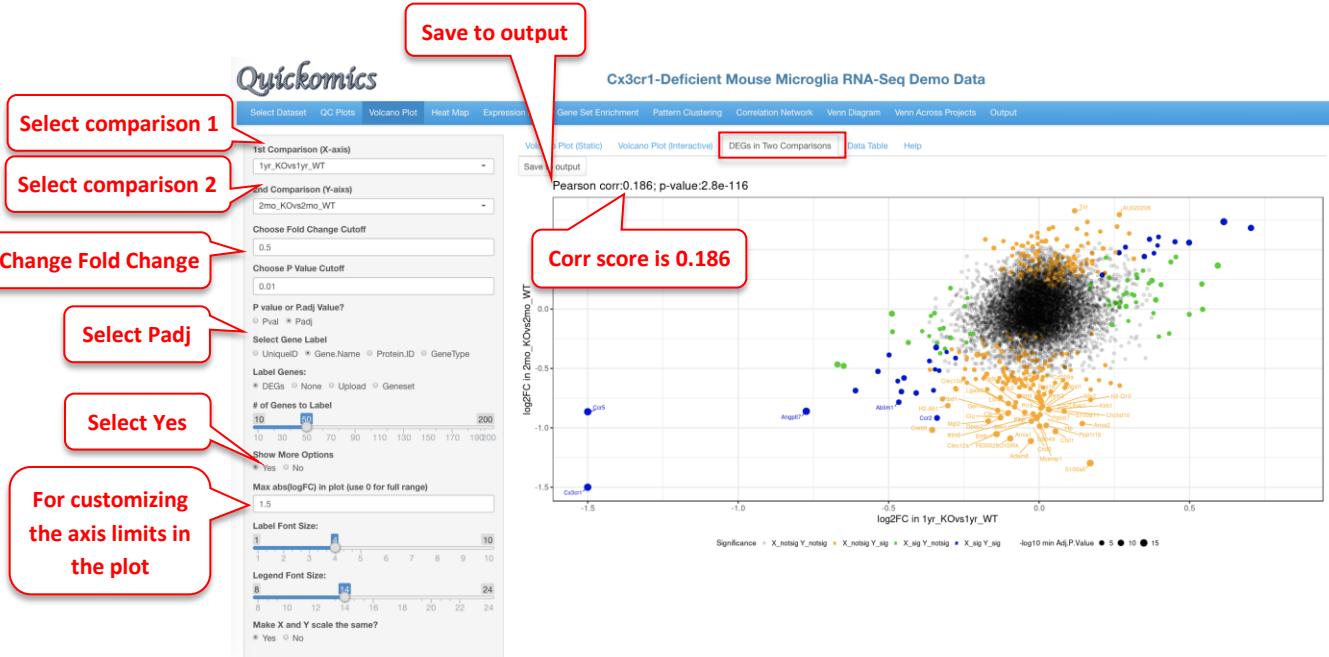
This next sub-tab is a very similar visualization as the previous one (4.1 Volcano Plot (Static)) with an additional capability where users can hover over a particular dot/gene on the plot to see more details. This plot has a few attributes that can be changed, and in the example below we have used the same cutoffs and values as in section 4.1.



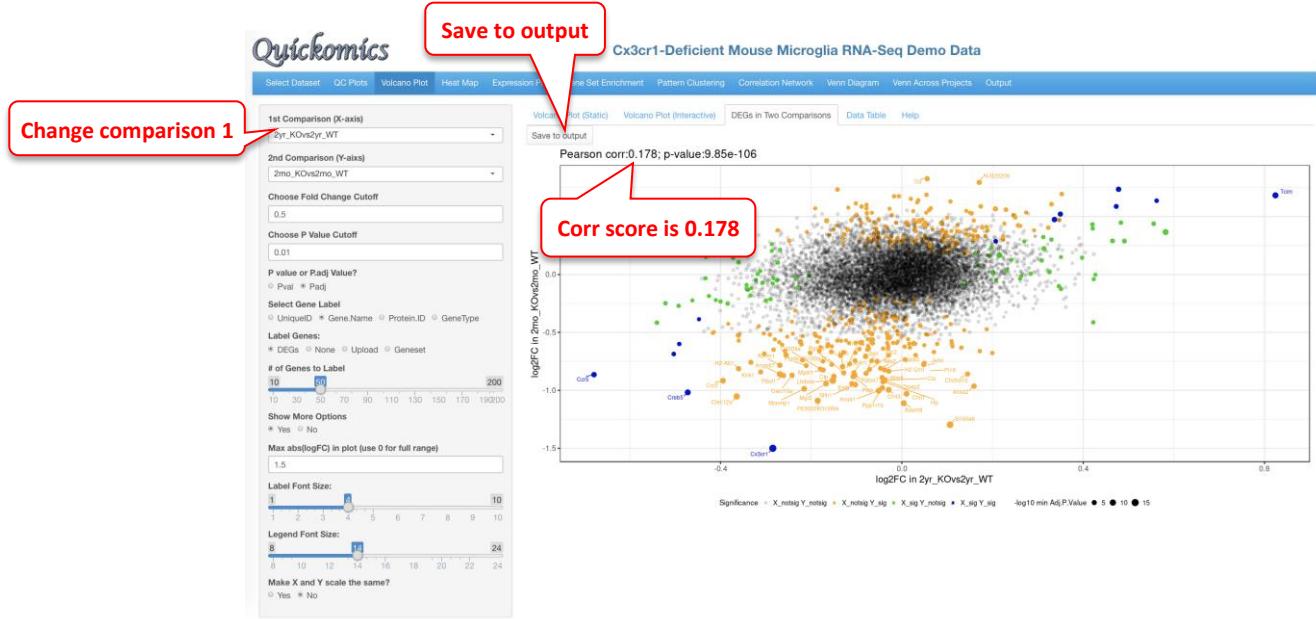
4.3 DEGs and DEPs in Two Comparisons

This next tab helps with identifying DEGs/DEPs that follow a similar trend in two comparisons. Gyoneva et al., 2019, the source of this demo dataset, report that the genes differentially expressed in Cx3cr1 KO microglia in the 2mon KO are quite different from WT, while the differences from WT are smaller for the 1yr KO and 2yr KO, and the two older ages are similar to each other. We have plotted out the following comparisons that illustrate this conclusion.

2mo_KOvs2mo_WT vs 1yr_KOvs1yr_WT has a low correlation score. There are many genes that are colored yellow, meaning those genes were significant in 2mo_KOvs2mo_WT (Y axis) but not in 1yr_KOvs1yr_WT (X axis).



Similarly, we changed the X axis to 2yr_KOvs2yr_WT to compare to 2mo_KOvs2mo_WT, and it shows a very similar plot with a low correlation score of 0.178



Next, we changed the Y axis comparison to 1yr_KOvs1yr_WT comparing it to 2yr_KOvs2yr_WT on the X axis. This plot has a higher correlation score of 0.371 than the previous one, suggesting that these two comparisons are more similar. These comparisons represent additional insights supported by statistical analyses that were gained from the data and were not available in the original manuscript by Gyoneva et al.



4.4 Data Table

This tab lets the users view the DEGs/DEPs in a tabular format with a searchable feature. Users also have the ability to download the gene list as a CSV.

Select comparison

Choose cutoffs

Save whole list to output

Search for any gene

Click on any column to sort the list

Download the current page as CSV, Excel or Pdf

| ID | UniqueID | Gene.Name | Protein.ID | Genotype | test | AdJ.PValue | PValue | logFC |
|-------|--------------------|-------------|------------|----------------|----------------|------------|----------|--------|
| 9310 | ENSMUSG00000052336 | Cx3cr1 | | protein_coding | 2mo_KOvs2mo_WT | 4.43e-61 | 3.18e-65 | -2.65 |
| 106 | ENSMUSG0000001025 | S100b | | protein_coding | 2mo_KOvs2mo_WT | 4.51e-13 | 6.46e-17 | -1.3 |
| 3450 | ENSMUSG00000054743 | Aden8 | | | | 1.64e-9 | 9.42e-13 | -1.11 |
| 11600 | ENSMUSG00000078122 | Filoblastin | | | | 7.65e-10 | 2.74e-13 | -1.09 |
| 9390 | ENSMUSG00000053063 | Clec12a | | | | 5.55e-11 | 1.59e-14 | -1.05 |
| 10400 | ENSMUSG00000064246 | Ch11 | | | | 2.04e-8 | 2.04e-11 | -1.03 |
| 9380 | ENSMUSG00000053007 | Greb5 | | | | 3.09e-9 | 2.2e-12 | -1.02 |
| 3130 | ENSMUSG00000024658 | Anxa1 | | protein_coding | 2mo_KOvs2mo_WT | 3.09e-9 | 2.44e-12 | -1 |
| 7940 | ENSMUSG00000040809 | Chn3 | | protein_coding | 2mo_KOvs2mo_WT | 3.09e-9 | 2.27e-12 | -0.995 |
| 1010 | ENSMUSG00000013974 | Mcmpl1 | | protein_coding | 2mo_KOvs2mo_WT | 9.71e-8 | 1.11e-10 | -0.986 |
| 1020 | ENSMUSG00000061718 | Ppp1rlb | | protein_coding | 2mo_KOvs2mo_WT | 1.87e-7 | 2.29e-10 | -0.98 |
| 5900 | ENSMUSG00000032231 | Anxa2 | | protein_coding | 2mo_KOvs2mo_WT | 2.69e-7 | 3.85e-10 | -0.964 |
| 11700 | ENSMUSG00000078763 | Sifm1 | | protein_coding | 2mo_KOvs2mo_WT | 2.99e-7 | 4.71e-10 | -0.951 |
| 2250 | ENSMUSG00000021728 | Emb | | protein_coding | 2mo_KOvs2mo_WT | 6.18e-7 | 1.24e-9 | -0.943 |
| 7970 | ENSMUSG00000044950 | Mg2 | | protein_coding | 2mo_KOvs2mo_WT | 8.95e-8 | 9.62e-11 | -0.943 |
| 5700 | ENSMUSG00000031722 | Hp | | protein_coding | 2mo_KOvs2mo_WT | 2.14e-7 | 2.76e-10 | -0.925 |
| 898 | ENSMUSG00000049103 | Ccr2 | | protein_coding | 2mo_KOvs2mo_WT | 2.04e-8 | 1.97e-11 | -0.917 |
| 9440 | ENSMUSG00000053687 | Dpep2 | | protein_coding | 2mo_KOvs2mo_WT | 0.00000182 | 4.7e-9 | -0.915 |
| 15100 | ENSMUSG00000112148 | Lrb4a | | protein_coding | 2mo_KOvs2mo_WT | 7.11e-13 | 1.53e-16 | -0.915 |
| 2370 | ENSMUSG00000020307 | Clu | | protein_coding | 2mo_KOvs2mo_WT | 4.16e-7 | 6.86e-10 | -0.912 |

Showing 1 to 20 of 420 entries

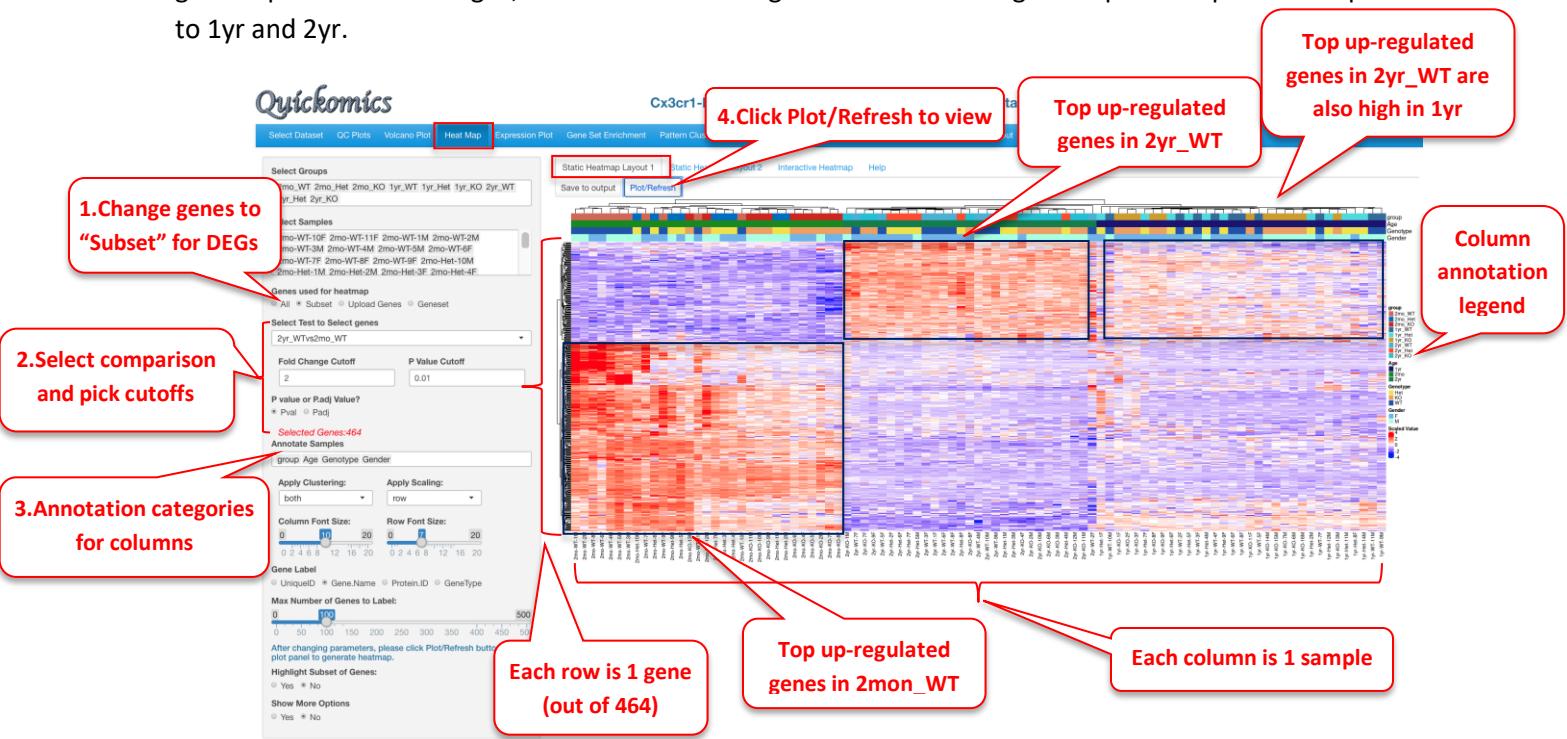
5 Heatmap Module

This module generates highly customized publication-quality heatmaps with the ComplexHeatmap package in R. Users can enter automated list of differentially expressed genes or enter a custom list of genes to highlight on the heatmap. This module differs from other modules in that each time Users change the attributes; they need to click on the “Plot/Refresh” button to generate the plot. This is to avoid changing the output before all the attribute are added.

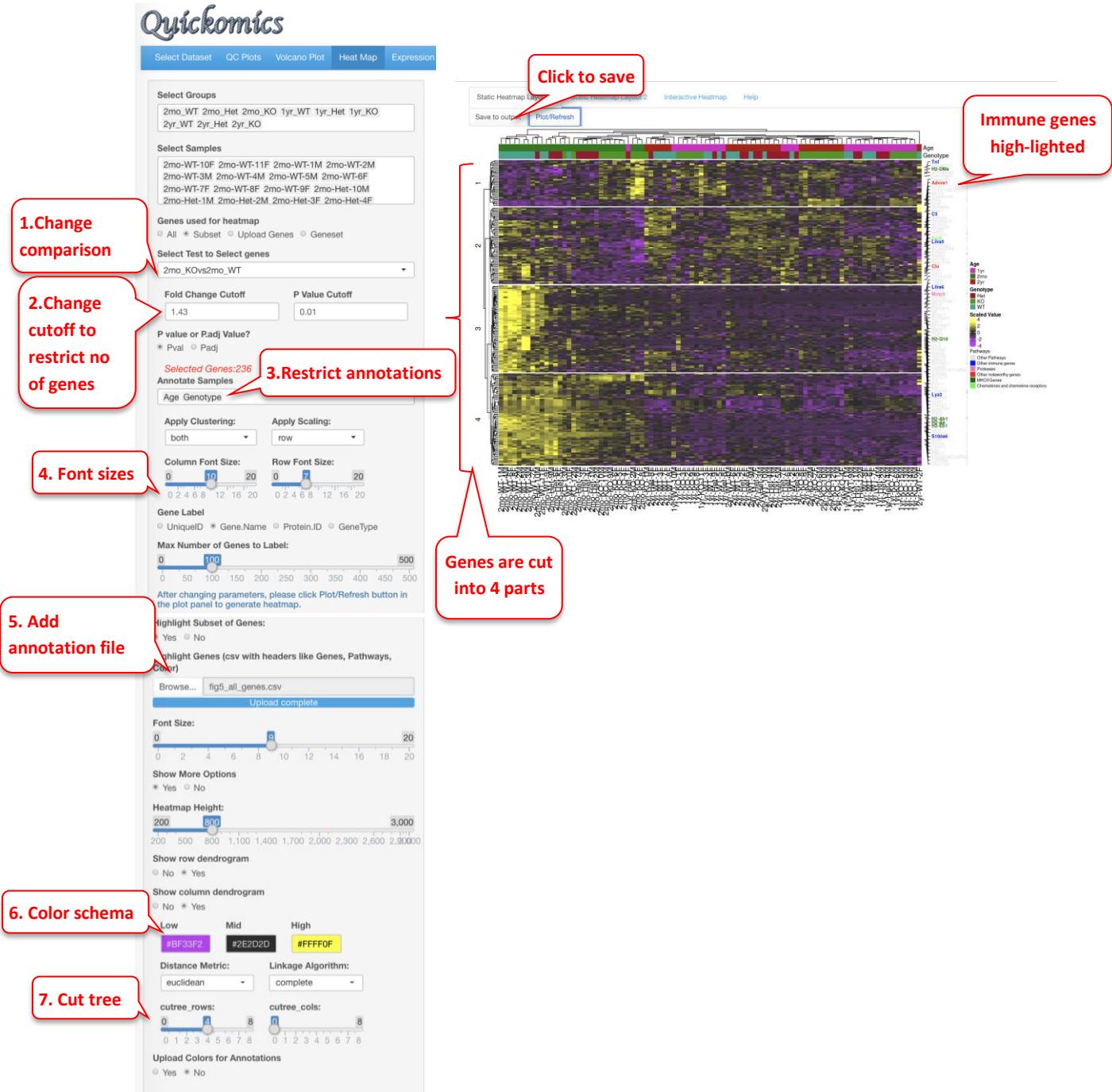
5.1 Static Heatmap Layout 1

The “Static Heatmap Layout 1” uses the ComplexHeatmap package (Gu *et al.*, 2016) for generating the plot. For demo purposes, we have selected options that highlight some capabilities to generate informative heatmap. For gene selection the default option is the selection of 100 random genes which can also be changed to variable genes. Here we selected DEGs/DEPs from the 2yr_WTvs2mon_WT comparison. Users also have the option to change the cutoff criteria for filtering the genes. They can also add/remove annotation categories for the samples on the columns.

In this heatmap, the top up-regulated genes in 2mon_WT samples are seen as high in 2mon samples but low in 2yr and 1yr samples, while the top up-regulated genes in the 2yr_WT samples are high in the 2yr and 1yr samples but low in the 2mon samples. This is another example that age has a major effect on gene expression in microglia, and the 2mon microglia have a different gene expression profile compared to 1yr and 2yr.



In this next example, we customize another Heatmap to highlight immune related genes that were called out in Gyoneva *et al.*, 2019. This is done by uploading a csv file consisting of the gene names with pathway information and the color for display. An example of this file is on our GitHub (Quickomics/demo_files/fig5_all_genes.csv)



5.2 Static Heatmap Layout 2

The “Static Heatmap Layout 2” uses the `heatmap.2` function from the `gplots` package for generating the plot. Here is a demo of the same set of genes with dendrograms for both rows and columns.

Quickomics

Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data

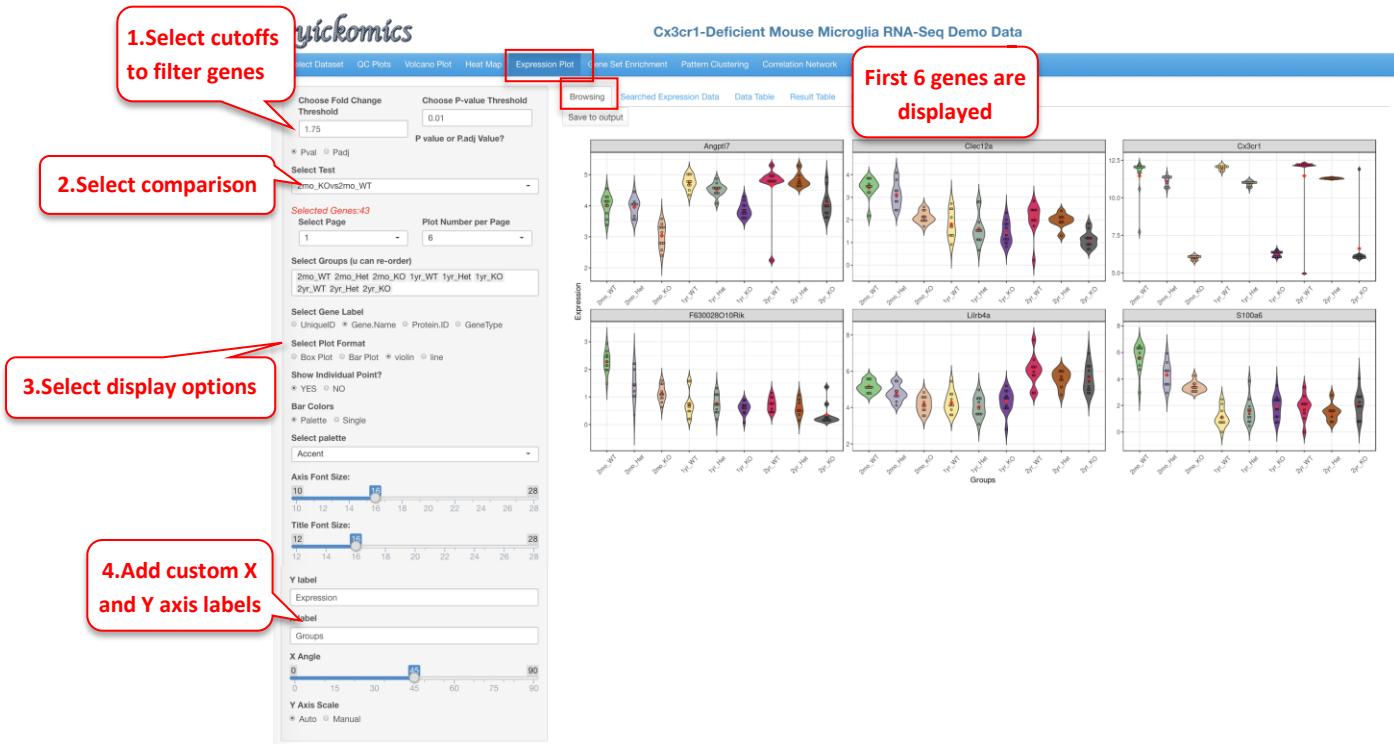


6 Expression Plot Module

This functional module allows Users to browse through a subset of genes/proteins. It also offers an option to plot expression of individually selected genes. Four different displays: Box Plot, Bar Plot, Violin Plot and Line Plot are made available to fit the Users' preference.

6.1 Browsing

This tab lets the Users plot out the expression of the DEGs/DEPs identified in the different comparisons. In this example below, we demonstrate how to display a few top genes as Violin Plots.



6.2 Searched Expression Data

Another way to plot genes/proteins is using this functionality. Users can either enter a custom list of genes/proteins or select a list from online databases like KEGG, MSigDB etc. The selected genes/proteins can be plotted on a single plot or in multiple plots. Here is an example of plotting genes involved in MHCII antigen processing.

Quickomics

Cx3cr1-Deficient Mouse Micr

1. Select Geneset

2. Select Mus musculus

3. Search "MHC"

4. Select Separate

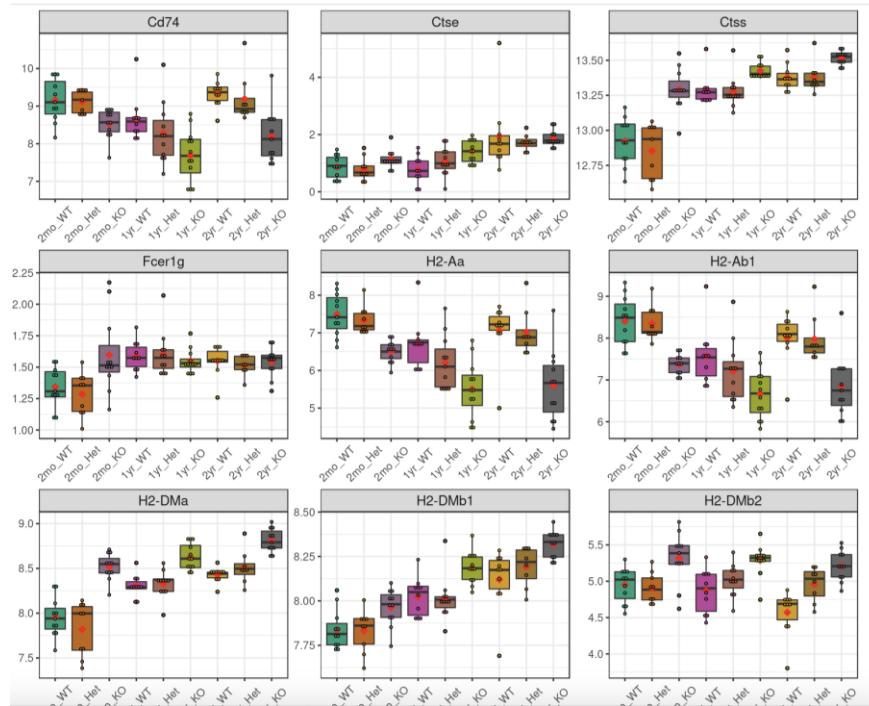
5. Select display options

6. Click Plot/Refresh

4. Select list

The screenshot shows the Quickomics web application. In the top right, a search bar contains 'MHC' with a dropdown menu open, showing several entries related to MHC class II antigen processing. A red box highlights this dropdown menu with the label '4. Select list'. In the center, there's a search results table for 'Mus musculus' containing 25 entries, each with details like ID, DB Name, DB ID, Species, Category, and Name. A red box highlights the 'Name' column with the label '4. Select list'. On the left, a detailed plot configuration panel is shown with various settings like 'Select Geneset', 'Separate or One Plot', and 'Display Options'. A red box highlights the 'Separate' option with the label '4. Select Separate'. At the bottom right of the main interface, a button labeled 'Plot/Refresh' is highlighted with a red box and the label '6. Click Plot/Refresh'.

The resulting plot shows the expression of a subset of genes from the MHCII antigen presentation family of genes. Many of them appear altered by gene KO in 2mon microglia, but also in all genotypes in 2yr microglia, as indicated in the publication.



6.3 Data Table

This sub-tab lists the normalized data of the genes/proteins selected in “Searched Expression Data” (Section 6.2) as a searchable table. For a demo of this feature, we select a proteome dataset from published paper (Connor-Robson *et al.*, 2019).

The screenshot shows the Quickomics web application. At the top, there's a navigation bar with tabs like 'Select Dataset', 'QC Plots', 'Volcano Plot', etc. Below the navigation bar, there's a search bar labeled 'Select data set' with options for 'Saved Projects' and 'Upload RData File'. A dropdown menu labeled 'Available Dataset' shows 'LRRK2 Neuron Proteome' as the selected item. To the right, there's a table with columns 'sampleid', 'group', 'Sex', and 'TimePoint'. Two rows are shown: row 1 has sampleid 'HFSet10', group 'CtlD35', Sex 'F', and TimePoint 'D35'; row 2 has sampleid 'HFSet11', group 'DiseaseD35', Sex 'M', and TimePoint 'D35'.

Once the dataset is selected, we go into the “Searched Expression Data” sub-tab to select a gene set.

The screenshot shows the 'Searched Expression Data' tab selected in the Quickomics interface. On the left, there's a search interface for 'Select Gene Set' with a search bar and a dropdown menu. A red box highlights the 'Select Geneset' button. In the center, there's a list of genes: ECSIT, IRAK1, IRAK2, IRAK4, MAP3K1, MYD88, and others. A red box highlights the 'Selected Expression Data' section. On the right, there are three box plots for genes ECSIT_Q9BQ95, IRAK1_D3YT85, and IRAK4_Q9NWZ3, comparing expression levels between CtlD35 and DiseaseD35 groups. A red box highlights the 'Select list' button.

Finally, we click the “Data Table” tab to look at the expression of the protein selected for a searchable table of values.

Quickomics LRRK2 Neuron Proteome

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects Output

Browsing Searched Expression Data Data Table Result Table

Show 15 entries CSV Excel Print

Protein expression values

| | id | UniqueID | Gene.Name | Protein.ID | sampleId | expr | labelgeneid |
|----|-----|------------------|-----------|------------|----------|-------|------------------|
| 1 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet10 | 0.883 | MYD88_A0A0A0MS70 |
| 2 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet11 | 1.11 | MYD88_A0A0A0MS70 |
| 3 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet12 | 0.863 | MYD88_A0A0A0MS70 |
| 4 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet13 | 1.18 | MYD88_A0A0A0MS70 |
| 5 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet14 | 0.988 | MYD88_A0A0A0MS70 |
| 6 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet15 | 0.926 | MYD88_A0A0A0MS70 |
| 7 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet16 | 0.892 | MYD88_A0A0A0MS70 |
| 8 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet17 | 0.99 | MYD88_A0A0A0MS70 |
| 9 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet18 | 1.06 | MYD88_A0A0A0MS70 |
| 10 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet19 | 1.1 | MYD88_A0A0A0MS70 |
| 11 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet20 | 0.935 | MYD88_A0A0A0MS70 |
| 12 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet21 | 1.23 | MYD88_A0A0A0MS70 |
| 13 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet22 | 0.812 | MYD88_A0A0A0MS70 |
| 14 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet23 | 1.3 | MYD88_A0A0A0MS70 |
| 15 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | HFSet24 | 0.966 | MYD88_A0A0A0MS70 |

Showing 1 to 15 of 456 entries

Previous 1 2 3 4 5 ... 31 Next

6.4 Result Table

This tab lists the statistics results of the genes/proteins selected in “Searched Expression Data” as a searchable table. Each comparison defined is probed for the genes in the list and the P value and Fold changes are listed. This is specifically designed for the gene values, unlike “Data Table” in the previous section (6.3) that is designed for protein values.

Quickomics LRRK2 Neuron Proteome

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects Output

Browsing Searched Expression Data Data Table Result Table

Show 15 entries CSV Excel Print

Differential gene exp values

| | id | UniqueID | Gene.Name | Protein.ID | test | Adj.PValue | PValue | logFC |
|----|-------|------------------|-----------|------------|--------------------|------------|----------|---------|
| 1 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | DiseaseD35vsCtlD35 | 0.0195 | 0.0111 | 0.141 |
| 2 | 411 | MYD88_A0A0A0MS70 | MYD88 | A0A0A0MS70 | DiseaseD56vsCtlD56 | 0.00805 | 0.0031 | 0.175 |
| 3 | 1240 | IRAK1_D3YTBS | IRAK1 | D3YTBS | DiseaseD35vsCtlD35 | 0.0327 | 0.0196 | -0.169 |
| 4 | 1240 | IRAK1_D3YTBS | IRAK1 | D3YTBS | DiseaseD56vsCtlD56 | 0.748 | 0.673 | -0.0385 |
| 5 | 2790 | IRAK2_O43187 | IRAK2 | O43187 | DiseaseD35vsCtlD35 | | | |
| 6 | 2790 | IRAK2_O43187 | IRAK2 | O43187 | DiseaseD56vsCtlD56 | | | |
| 7 | 5674 | MAP3K1_Q13233 | MAP3K1 | Q13233 | DiseaseD35vsCtlD35 | 0.182 | 0.135 | 0.102 |
| 8 | 5674 | MAP3K1_Q13233 | MAP3K1 | Q13233 | DiseaseD56vsCtlD56 | 0.282 | 0.2 | 0.131 |
| 9 | 9285 | ECSIT_Q9BQ95 | ECSIT | Q9BQ95 | DiseaseD35vsCtlD35 | 0.102 | 0.0696 | 0.111 |
| 10 | 9285 | ECSIT_Q9BQ95 | ECSIT | Q9BQ95 | DiseaseD56vsCtlD56 | 0.00431 | 0.00151 | 0.211 |
| 11 | 10428 | IRAK4_Q9NWZ3 | IRAK4 | Q9NWZ3 | DiseaseD35vsCtlD35 | 3.8e-8 | 4.67e-9 | 0.568 |
| 12 | 10428 | IRAK4_Q9NWZ3 | IRAK4 | Q9NWZ3 | DiseaseD56vsCtlD56 | 0.00076 | 0.000202 | 0.406 |
| 13 | 11350 | TRAF6_Q9Y4K3 | TRAF6 | Q9Y4K3 | DiseaseD35vsCtlD35 | 0.00434 | 0.00205 | -0.0898 |
| 14 | 11350 | TRAF6_Q9Y4K3 | TRAF6 | Q9Y4K3 | DiseaseD56vsCtlD56 | 0.00514 | 0.00185 | -0.0963 |

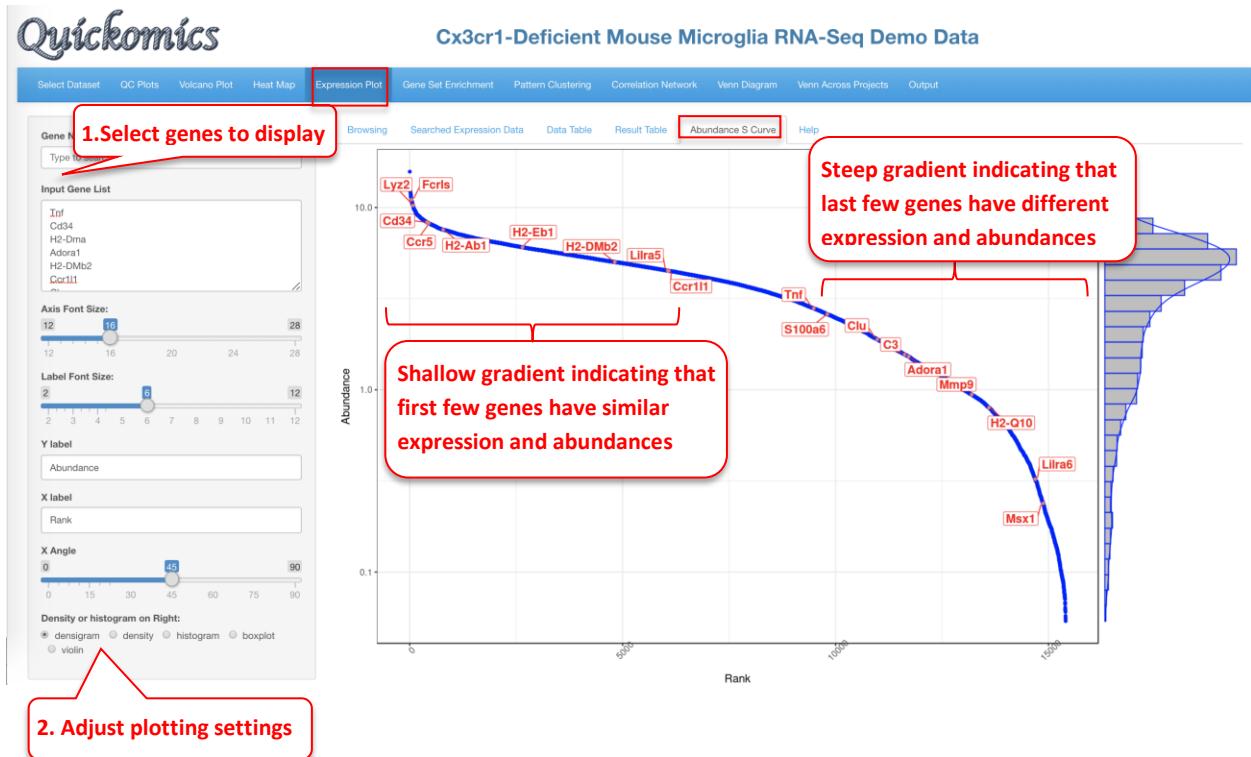
Showing 1 to 14 of 14 entries

Previous 1 Next

6.5 Rank Abundance Curve

This functionality helps identify the relative abundance of a list of genes within a dataset. This plot helps interpret the distribution of abundance and expression levels of a set of genes. A steep gradient

indicates the genes are very different in their expression levels as the high-ranking species have much higher abundances than the low-ranking species. A shallow gradient indicates that the genes have very similar expression levels as the abundances of different species are similar.



7 Gene Set Enrichment Module

This functionality in Quickomics is an extension to the differential gene identification. Users are able to use different cutoffs to filter out DEGs/DEPs in different comparisons and probe enriched pathways. This is a useful way to make biological sense of the data.

7.1 Gene Set Enrichment

This first tab lets the users pick a particular comparison, apply different cutoff criteria and select from a list of different databases for enrichment analysis. To demonstrate usage of this functionality, we will use the Mouse Microglia RNA dataset to help answer the question which pathways were altered by Genotype and/or Age. Very similar to the results from Gyoneva et al., 2019, this analysis points to immune related terms being enriched in the down-regulated genes for 2mon_KOvs2mon_WT.

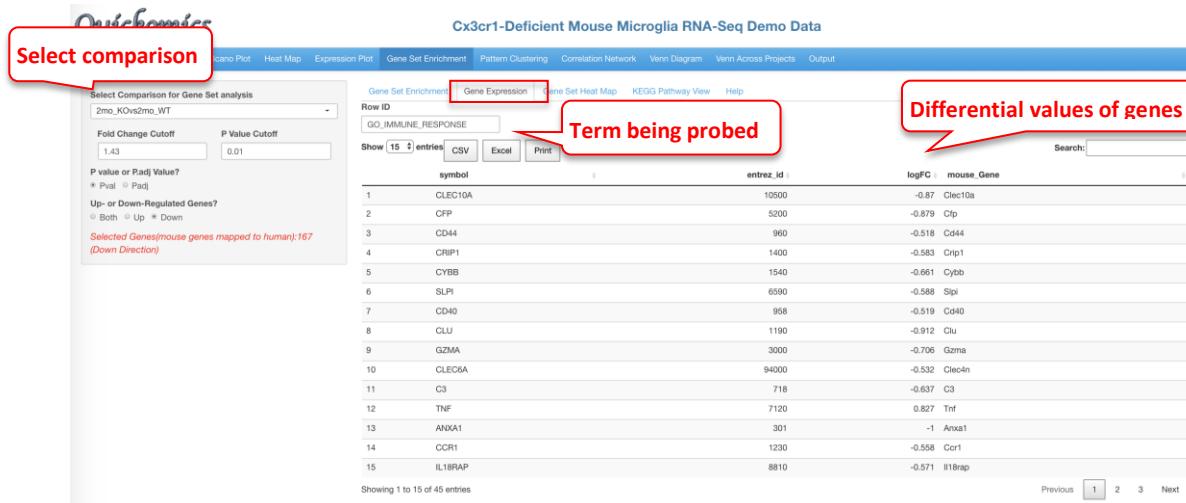
| ID | Rank | p.value | p.adj | DeGeneNum | UpGene | DownGene | SetNum | |
|----|---|---------|-----------|-----------|--------|----------|--------|------|
| 1 | GO_DEFENSE_RESPONSE | 1 | 2.71e-11 | 1.2e-7 | 46 | 7 | 39 | 1230 |
| 2 | GO_IMMUNE_RESPONSE | 2 | 9.89e-11 | 2.18e-7 | 41 | 5 | 36 | 1100 |
| 3 | GO_NEUTROPHIL_RESPONSE | | | 2.84e-7 | 20 | 1 | 19 | 259 |
| 4 | GO_IMMUNE_SYSTEM_PROCESS | | | 0.0000207 | 59 | 7 | 52 | 1980 |
| 5 | GO_INFLAMMATORY_RESPONSE | | | 0.0000158 | 25 | 5 | 20 | 454 |
| 6 | GO_NEUTROPHIL_CHEMOTAXIS | | | 0.000034 | 11 | 0 | 11 | 117 |
| 7 | GO_REGULATION_OF_IMMUNE_SYSTEM_PROCESS | | | 0.0000359 | 44 | 4 | 40 | 1400 |
| 8 | GO_RESPONSE_TO_WOUNDING | | | 0.0000571 | 23 | 1 | 22 | 563 |
| 9 | GO_REGULATION_OF_CELL_CELL_ADHESION | | | 0.0000571 | 19 | 1 | 18 | 380 |
| 10 | GO_REGULATION_OF_HOMOTYPIC_CELL_CELL_ADHESION | 10 | 2.45e-7 | 0.000109 | 16 | 0 | 16 | 307 |
| 11 | GO_POSITIVE_REGULATION_OF_CELL_CELL_ADHESION | 11 | 3.93e-7 | 0.000158 | 15 | 1 | 14 | 243 |
| 12 | GO_RESPONSE_TO_BACTERIUM | 12 | 6.92e-7 | 0.000253 | 22 | 3 | 19 | 528 |
| 13 | GO_DENDRITIC_CELL_CHEMOTAXIS | 13 | 7.41e-7 | 0.000253 | 5 | 0 | 5 | 16 |
| 14 | GO_CELL_CHEMOTAXIS | 14 | 8.65e-7 | 0.000274 | 11 | 0 | 11 | 162 |
| 15 | GO_REGULATION_OF_CELL_ACTIVATION | 15 | 0.0000151 | 0.000447 | 20 | 1 | 19 | 484 |

Similarly, we pull out the GO enrichment for up-regulated genes.

| ID | Rank | p.value | p.adj | DeGeneNum | UpGene | DownGene | SetNum | |
|----|--|---------|--------|-----------|--------|----------|--------|----|
| 1 | GO_ACTIVATION_OF_MAPKKK_ACTIVITY | 1 | 0.0158 | 1 | 2 | 2 | 0 | 11 |
| 2 | GO_NEGATIVE_REGULATION_OF_LIPID_CATABOLIC_PROCESS | 2 | 0.0219 | 1 | 3 | 2 | 1 | 21 |
| 3 | GO_NEGATIVE_REGULATION_OF_BLOOD_PRESSURE | 3 | 0.0252 | 1 | 2 | 2 | 0 | 43 |
| 4 | GO_REGULATION_OF_SYNAPTIC_TRANSMISSION_GLUTAMATERGIC | 4 | 0.0287 | 1 | 2 | 2 | 0 | 50 |
| 5 | GO_POSITIVE_REGULATION_OF_OSTEOCLAST_DIFFERENTIATION | 5 | 0.0287 | 1 | 4 | 2 | 2 | 24 |
| 6 | GO_NEGATIVE_REGULATION_OF_BLOOD_CIRCULATION | 6 | 0.0364 | 1 | 2 | 2 | 0 | 36 |

7.2 Gene Expression

For the enriched terms identified in the previous tab (Gene Set Enrichment, section 7.1), Users have the ability to look at the differential values of the genes present in the set. Here we are probing the GO_IMMUNE_RESPONSE identified in the down-regulated category.



7.3 Gene Set Heat Map

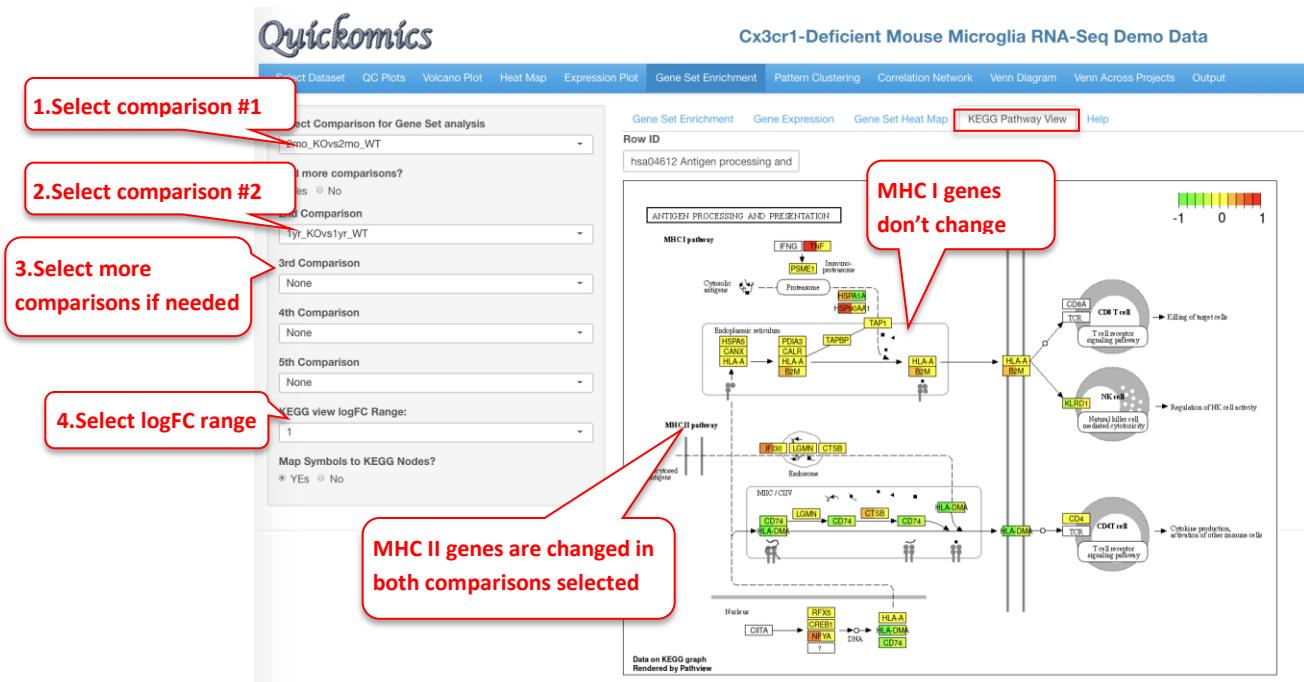
Like the “Gene Expression” tab, Users have the ability to plot expression of the genes in the pathway term selected as a heatmap. This visualization shows that while the Immune Response genes are enriched in the differentially expressed genes list, not all of them change in the same direction.



7.4 KEGG Pathway View

This functionality is available only to KEGG pathway terms. Users have the option to view the Fold Change levels of different genes in the KEGG pathway selected for up to 5 comparisons. In this example we demo the KEGG pathway “hsa04612 Antigen processing and presentation” as identified as a down-regulated term in the 2mon_KOvs2mon_WT and 1yr_KOvs1yr_WT comparisons.

This figure identifies all the genes that are differentially expressed in this pathway. It is also clear that genes in the MHC I pathway do not change, but many genes in the MHC II pathway are down regulated in both the comparisons. The visualization is also useful to highlight whether or not a large proportion of genes in a pathway are altered by the perturbation.

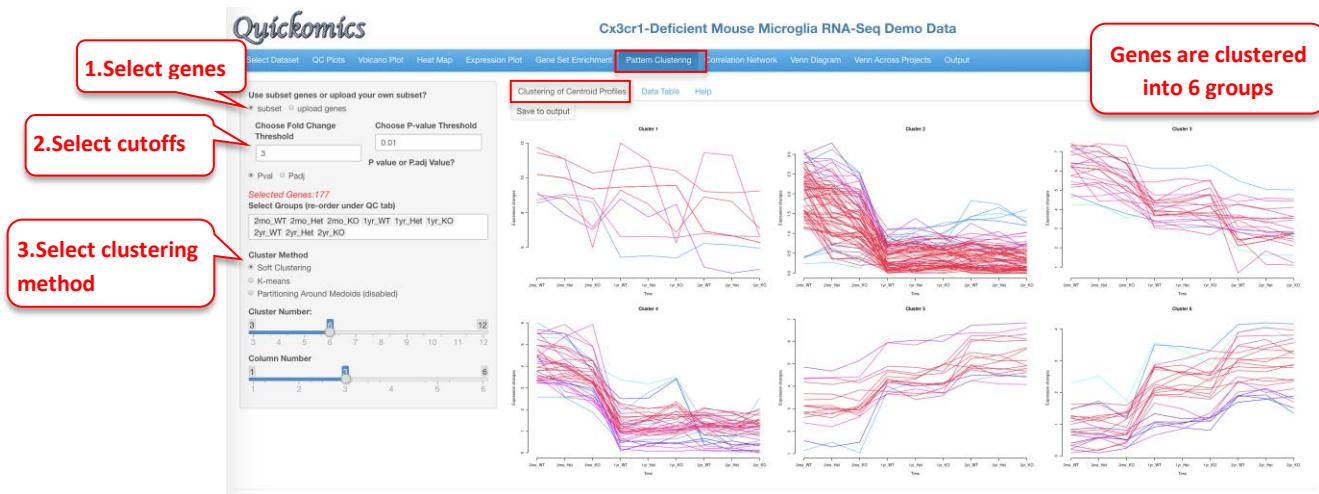


8 Pattern Clustering Module

This functional module in Quickomics helps cluster genes/proteins based on their expression profiles across different groups. Three different algorithms, soft (fuzzy) clustering, k-means and partitioning around medoids (PAM), are implemented.

8.1 Clustering of Centroid Profiles

Users have the option to either select the DEGs/DEPs as the list of genes/proteins to cluster based off on or upload a custom gene list. DEGs/DEPs are selected based off all the comparisons, and cutoffs can be applied to limit the number of genes. In this example below, we use the top DEGs/DEPs with a Fold Change cutoff of 3 to identify 6 clusters of genes.



Please note that this visualization for this dataset is novel and was not previously reported in Gyoneva et al., 2019. The genes in each cluster can be viewed in the “Data Table” tab.

8.2 Data Table

The data table contains genes from the previous clusters, along with their expression values

The figure shows the Quickomics interface for the 'Data Table' tab. A red box highlights the 'Data Table' tab. A callout box points to the 'Cluster numbers as seen in previous tab' (6) in the table header. The table lists 10 genes (1-10) with their expression values across 12 samples (2mo_WT, 2mo_Het, 2mo_KO, 1yr_WT, 1yr_Het, 1yr_KO, 2yr_WT, 2yr_Het, 2yr_KO) and their corresponding cluster numbers (4, 4, 4, 6, 4, 2, 2, 2, 4, 5).

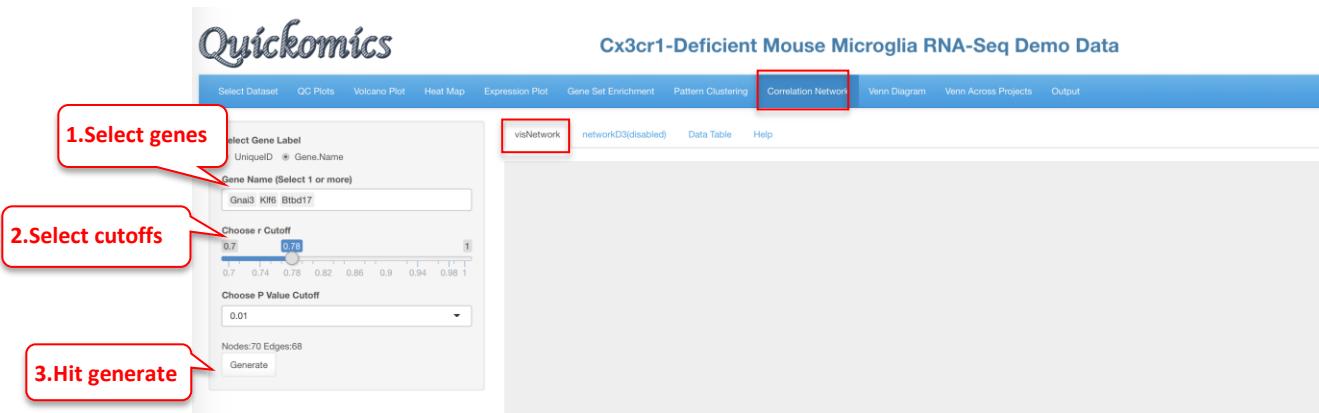
| Gene | ID | 2mo_WT | 2mo_Het | 2mo_KO | 1yr_WT | 1yr_Het | 1yr_KO | 2yr_WT | 2yr_Het | 2yr_KO | cluster |
|------|---------------------|--------|---------|--------|--------|---------|--------|--------|---------|--------|---------|
| 1 | ENSMUSG00000000318 | 3.33 | 3.16 | 2.23 | 1.02 | 1.53 | 1.04 | 1.38 | 1.07 | 0.943 | 4 |
| 2 | ENSMUSG000000001020 | 4.55 | 3.92 | 3.26 | 1.11 | 1.58 | 2.37 | 1.34 | 1.24 | 1.85 | 4 |
| 3 | ENSMUSG000000001025 | 5.6 | 4.31 | 3.39 | 1.12 | 1.56 | 1.91 | 1.85 | 1.48 | 2.05 | 4 |
| 4 | ENSMUSG000000001120 | 1.27 | 1.13 | 0.895 | 2.78 | 2.87 | 2.6 | 3.48 | 3.36 | 3.3 | 6 |
| 5 | ENSMUSG000000001128 | 3.53 | 2.59 | 2.4 | 1.46 | 1.12 | 1.19 | 1.64 | 1.33 | 1.2 | 4 |
| 6 | ENSMUSG000000001281 | 2.64 | 1.88 | 1.72 | 0.687 | 0.893 | 0.857 | 1.01 | 0.925 | 1.02 | 2 |
| 7 | ENSMUSG000000001588 | 1.59 | 0.35 | 0.401 | 0.147 | 0 | 0.0158 | 0.144 | 0.022 | 0.0989 | 2 |
| 8 | ENSMUSG000000002058 | 2.55 | 1.71 | 1.34 | 0.437 | 0.653 | 0.6 | 0.582 | 0.648 | 0.579 | 2 |
| 9 | ENSMUSG000000002204 | 3.75 | 3.24 | 2.86 | 0.108 | 0.441 | 0.414 | 0.56 | 0.109 | 0.096 | 4 |
| 10 | ENSMUSG000000002602 | 4.34 | 4.37 | 4.16 | 4.73 | 4.57 | 4.95 | 6.25 | 6.28 | 6.11 | 5 |

9 Correlation Network Module

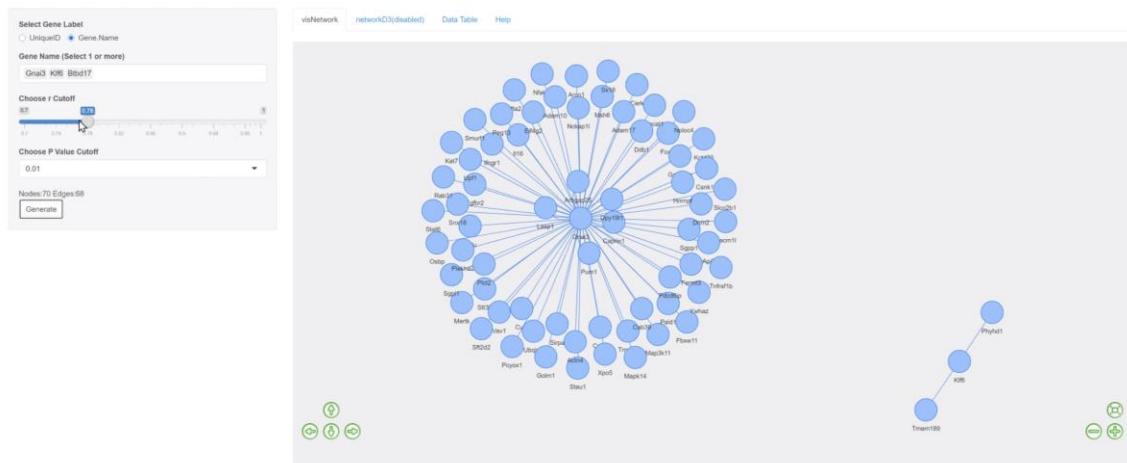
This module helps build co-expression networks based on gene-gene or protein-protein correlations. This is especially useful during co-immunoprecipitation or pulldown proteomics experiments to identify protein partners. Users have the ability to enter 1 or more genes/proteins to identify if their expression is correlated.

9.1 visNetwork

This uses the R package visNetwork to visualize the expression of correlated genes/proteins. Users have the ability to choose the R and P value cutoff for selecting the correlated genes/proteins. In this example below, we probe 3 genes and identify the expression of other genes that are correlated with them.



The following network plot was produced. Gnai3 is part of a large network with multiple genes correlated with it. On the other hand, Klf6 was correlated with 2 other genes, while Btbd17 was not correlated with any genes.



9.2 Data Table

The genes identified in the previous tab as part of the network plot are available as a table in this tab. Users can perform searches and sort based on correlation statistics.

Quickomics

Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data

Select Dataset QC Plots Volcano Plot Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network

Gene queried

Genes identified as correlated

| from | to | cor | p | direction |
|------|--------------------|------|---|-----------|
| 1 | ENSMUSG00000000001 | 0.81 | 0 | 1 |
| 2 | ENSMUSG00000000001 | 0.79 | 0 | 1 |
| 3 | ENSMUSG00000000001 | 0.81 | 0 | 1 |
| 4 | ENSMUSG00000000001 | 0.8 | 0 | 1 |
| 5 | ENSMUSG00000000001 | 0.79 | 0 | 1 |
| 6 | ENSMUSG00000000001 | 0.84 | 0 | 1 |
| 7 | ENSMUSG00000000001 | 0.8 | 0 | 1 |
| 8 | ENSMUSG00000000001 | 0.86 | 0 | 1 |
| 9 | ENSMUSG00000000001 | 0.79 | 0 | 1 |
| 10 | ENSMUSG00000000001 | 0.79 | 0 | 1 |
| 11 | ENSMUSG00000000001 | 0.79 | 0 | 1 |
| 12 | ENSMUSG00000000001 | 0.81 | 0 | 1 |
| 13 | ENSMUSG00000000001 | 0.79 | 0 | 1 |
| 14 | ENSMUSG00000000001 | 0.84 | 0 | 1 |
| 15 | ENSMUSG00000000001 | 0.79 | 0 | 1 |

Showing 1 to 15 of 68 entries

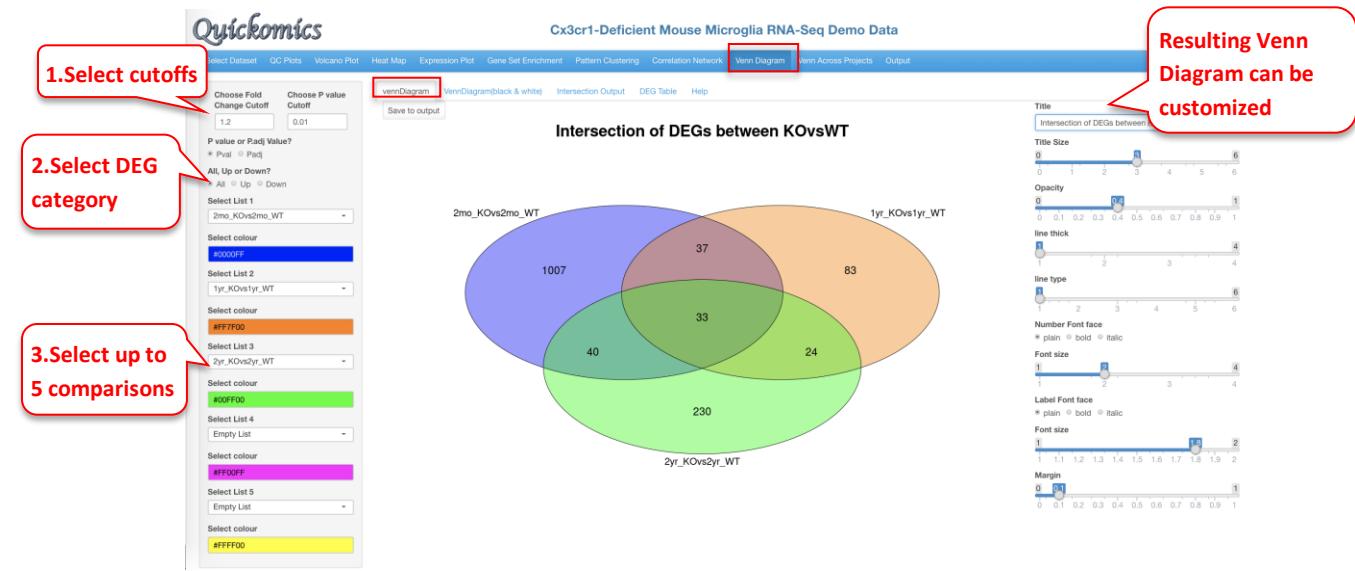
Previous 1 2 3 4 5 Next

10 Venn Diagram Module

This module in Quickomics helps Users identify DEGs/DEPs that are common or unique to different comparisons within one project. Users have the option to either visualize these DEGs/DEPs as a highly customizable Venn Diagram or download a table with the gene names.

10.1 Venn Diagram

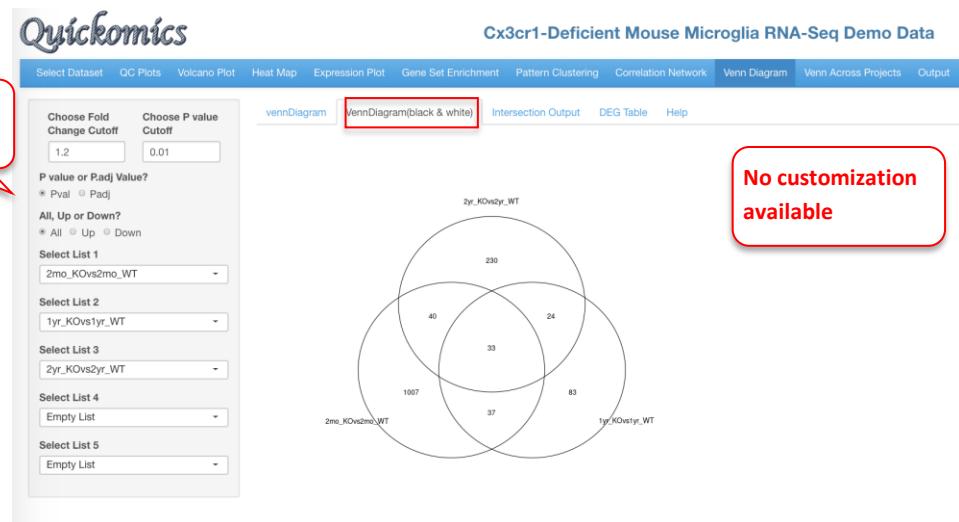
To view the DEGs/DEPs as a Venn Diagram, Users have the option to see the intersection of DEGs/DEPs for up to 5 different comparisons. Users can also use all DEGs/DEPs or restrict it to the ones that are Up or Down regulated.



In this example there were 33 DEGs that were identified in the three comparisons probed here, which represent the set of genes altered by the KO genotype at each of the time points.

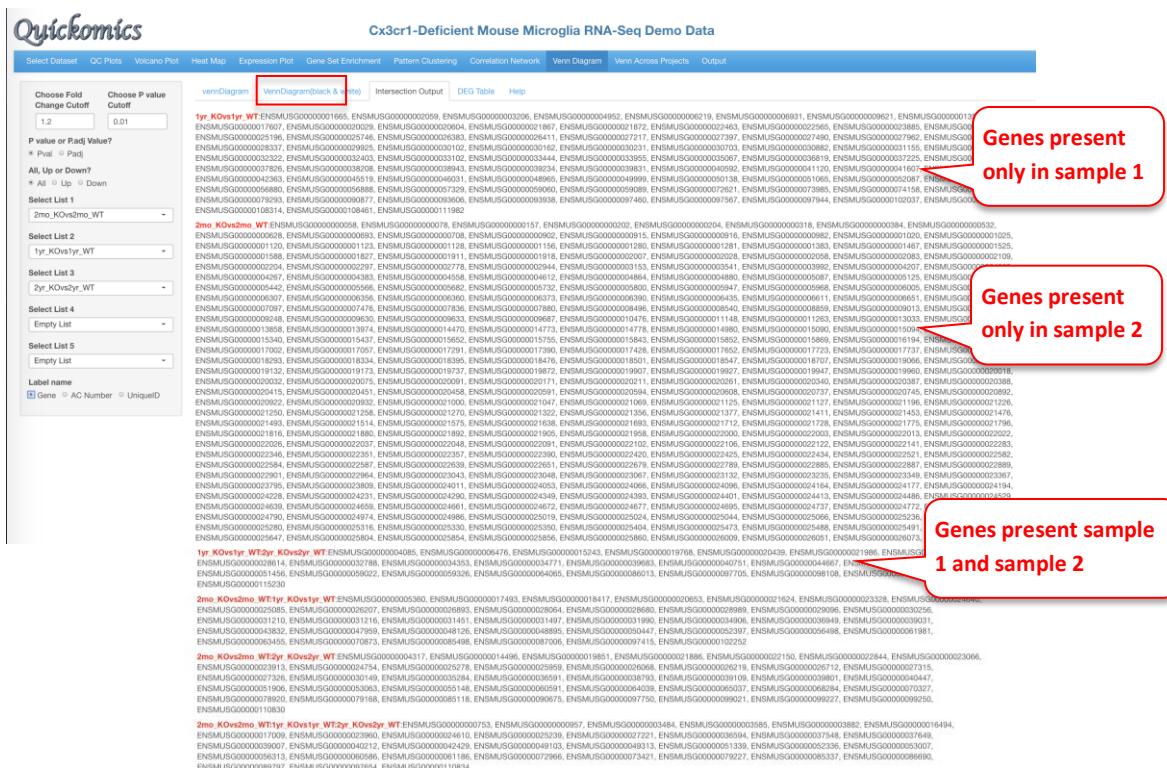
10.2 Venn Diagram(black & white)

This tab has the same plot seen before in Section 10.1 but highly simplified and in black and white.



10.3 Intersection Output

This tab lets the User view the individual genes that were present in the Venn Diagram and the distribution of the genes in intersection regions.



10.4 DEG Table

This tab lets the Users view the Fold Change and P value of the genes identified in the Venn Diagram in the comparisons being queried.

Quickomics

Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data

Heat Map Expression Plot Gene Set Enrichment Pattern Clustering Correlation Network Venn Diagram Venn Across Projects Out

Choose Fold Change Cutoff Choose P value Cutoff
1.2 0.01

P value or Padj Value? * Padj = Padj

All, Up or Down? * All = Up = Down

Select List 1 2mo_KOvs2mo_WT

Select List 2 1yr_KOvs1yr_WT

Select List 3 2yr_KOvs2yr_WT

Select List 4 Empty List

Select List 5 Empty List

vermDiagram VermDiagram(black & white) Intersection Output DEG Table Help

Show 20 entries CSV Excel Print

Save to output

| UniqueID | Gene.Name | 2mo_KOvs2mo_WT_DESeq.JogFC | 2mo_KOvs2mo_WT_DESeq.Pvalue | 2mo_KOvs2mo_WT_DESeq.AdjPvalue | 1yr_KOvs1yr_WT_DESeq.JogFC | 1yr_KOvs1yr_WT_DESeq.Pvalue |
|---------------------|-----------|----------------------------|-----------------------------|--------------------------------|----------------------------|-----------------------------|
| ENSMUSG00000000058 | Cav2 | -0.433 | 0.0000442 | 0.00264 | -0.0228 | 0.772 |
| ENSMUSG00000000078 | K06 | 0.355 | 9.92e-7 | 0.00015 | 0.0885 | 0.143 |
| ENSMUSG000000000157 | Igfb2l | -0.421 | 0.000057 | 0.00311 | 0.0028 | 0.811 |
| ENSMUSG000000000184 | Cord2 | -0.359 | 0.0215 | 0.149 | -0.0922 | 0.188 |
| ENSMUSG000000000202 | Btbd17 | -0.452 | 0.00327 | 0.0494 | -0.0956 | 0.22 |
| ENSMUSG000000000203 | Sfrn4 | -0.551 | 0.000031 | 0.000344 | 0.0163 | 0.268 |
| ENSMUSG000000000318 | Clec10a | -0.87 | 4.56e-9 | 0.0000182 | -0.0476 | 0.332 |
| ENSMUSG000000000384 | Tbrg4 | 0.298 | 0.0046 | 0.0607 | 0.0925 | 0.327 |
| ENSMUSG000000000489 | Sept1 | -0.25 | 0.0852 | 0.325 | 0.212 | 0.0225 |
| ENSMUSG000000000532 | Aovt1b | 0.42 | 0.000505 | 0.0644 | -0.0438 | 0.643 |
| ENSMUSG000000000628 | Hk2 | 0.332 | 0.000129 | 0.00547 | -0.126 | 0.0685 |
| ENSMUSG000000000693 | Lox3 | -0.308 | 0.00619 | 0.0725 | -0.0521 | 0.529 |
| ENSMUSG000000000707 | Kat2b | 0.329 | 0.00307 | 0.0471 | 0.118 | 0.177 |
| ENSMUSG000000000753 | Serpinf1 | 0.588 | 1.78e-7 | 0.000035 | 0.367 | 9.52e-9 |
| ENSMUSG000000000902 | Smartb1 | 0.265 | 8.52e-7 | 0.000135 | 0.045 | 0.371 |
| ENSMUSG000000000915 | Hip1r | -0.561 | 0.000254 | 0.00887 | 0.0258 | 0.768 |
| ENSMUSG000000000916 | Neuro5 | 0.31 | 0.000122 | 0.00525 | -0.0581 | 0.445 |
| ENSMUSG000000000957 | Mmp14 | 0.323 | 0.00429 | 0.0583 | 0.345 | 0.0000342 |
| ENSMUSG000000000962 | Ctb | 0.455 | 0.00112 | 0.0246 | 0.0905 | 0.335 |
| ENSMUSG000000001020 | S100a4 | -0.678 | 0.0000143 | 0.00117 | 0.172 | 0.00749 |

Showing 1 to 20 of 1,454 entries

Previous 1 2 3 4 ... 73 Next

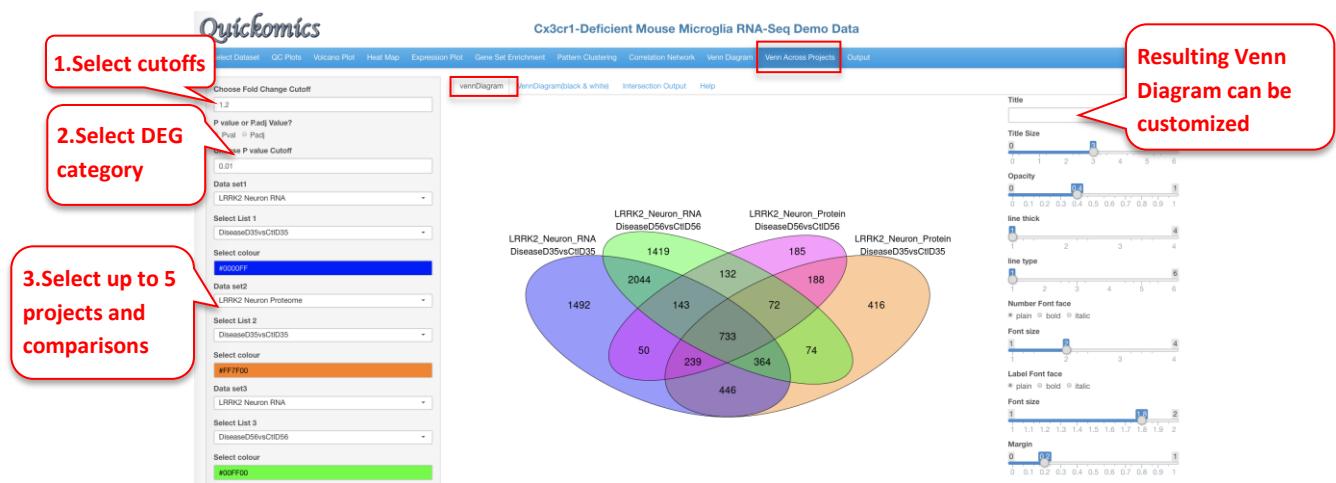
Fold Change and P value statistics in each comparison

11 Venn Across Projects Module

This functionality in Quickomics is the only one that lets users probe more than one project/dataset together. The DEGs/DEPs identified in different comparisons across projects/datasets loaded in the interface can be compared here. By taking advantage of the “Protein Gene name” function, this module can be used to identify common DEGs/DEPs present in both RNAseq and Proteomics datasets of the same project.

11.1 Venn Diagram

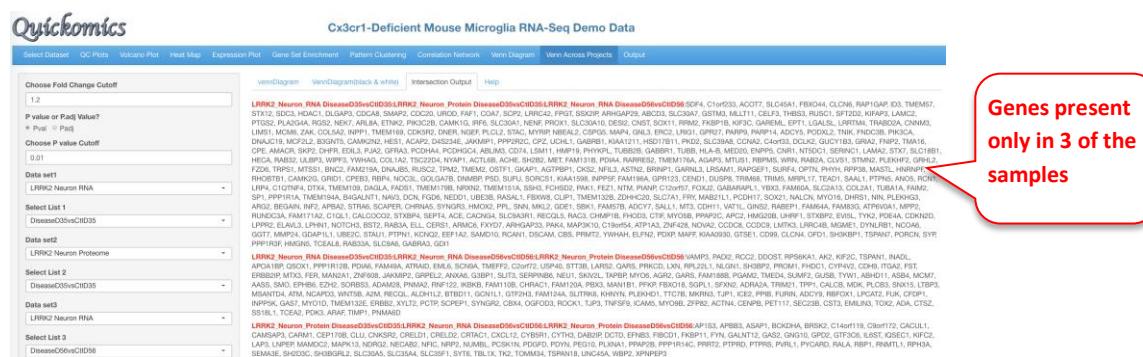
To illustrate this feature, we compare the two RNAseq and two Proteomic comparisons from Connor-Robson et al., 2019. We load 2 comparisons from each dataset/project and view the number of DEGs/DEPs that are common and different between the 4 comparisons.



Users have the option to view this Venn Diagram as a Black & White plot.

11.2 Intersection Output

Like in 10.3 Intersection Output, Users can view the individual genes present in different regions of the Venn Diagram, like the ones identified in all the samples, or the ones only present in 3 of them.



12 Output Module

This final tab in Quickomics is available for Users to download plots in high resolution, including for use in publications. Users can save different plots by clicking on “Save to Output” present on the top left corner of the plot during analysis and exploration of the datasets in all previous tabs. Upon clicking on the “Output” tab as seen in the image below, Users can see a list of plots that are available to download to their local machine.

The screenshot shows the Quickomics interface with the title "Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data". The "Output" tab is selected. On the left, there are two sliders for "Plot File Page Width" (ranging from 3 to 30) and "Plot File Page Height" (ranging from 3 to 50). A red callout box labeled "Select width and height of plots to save" points to these sliders. Below the sliders is a "Clear all saved plots" button. A red callout box labeled "Select all the plots available to save" points to a list of checked checkboxes under "Plots to Save". The list includes: PCA Plot, Sample Dendrogram, Volcano Plot (2mo_KOvs2mo_WT), Volcano Plot (2yr_KOvs2yr_WT vs 1yr_KOvs1yr_WT), Heatmap, Browsing Plot (1), vennDiagram (1), vennDiagram (2), and Pattern Clustering (Soft Clustering). At the bottom are three download buttons: "Download PDF", "Download SVG (for the first selected plot)", and "Download tables in .xlsx". A red callout box labeled "Plots can be downloaded as PDF or SVG" points to the "Download PDF" and "Download SVG" buttons.

13 References

- Cotter,D. *et al.* (2006) LMPD: LIPID MAPS proteome database. *Nucleic Acids Res.*, **34**, D507–D510.
- Frolikis,A. *et al.* (2010) SMPDB: The Small Molecule Pathway Database. *Nucleic Acids Res.*, **38**, D480–D487.
- Jassal,B. *et al.* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **48**, D498–D503.
- Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Liberzon,A. *et al.* (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
- Martens,M. *et al.* (2020) WikiPathways: connecting communities. *Nucleic Acids Res.*
- Subramanian,A. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.*, **102**, 15545 LP – 15550.
- The Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.