

Quickomics Supplementary

Table of Contents

1	<i>Introduction</i>	3
2	<i>Select Dataset</i>	4
2.1	Sample Table	4
2.2	Project Overview	5
2.3	Result Table	5
2.4	Data Table	6
2.5	Protein Gene Names	6
2.6	Help	7
3	<i>QC Plots</i>	8
3.1	PCA Plot.....	8
3.2	Eigenvalues.....	11
3.3	PCA 3D Plot.....	12
3.4	PCA 3D Interactive	13
3.5	Sample-sample Distance	14
3.6	Dendograms	14
3.7	Box Plot	15
3.8	CV Distribution	15
3.9	Order Groups	16
3.10	Help	17
4	<i>Volcano Plot</i>	18
4.1	Volcano Plot (Static).....	18
4.2	Volcano Plot (Interactive)	20
4.3	DEGs in Two Comparisons.....	20
4.4	Data Table	22
5	<i>Heat Map</i>	23
5.1	Static Heatmap Layout 1	23

5.2	Static Heatmap Layout 2	24
6	<i>Expression Plot</i>	26
6.1	Browsing.....	26
6.2	Searched Expression Data	26
6.3	Data Table	27
6.4	Result Table	28
7	<i>Gene Set Enrichment</i>	30
7.1	Gene Set Enrichment	30
7.2	Gene Expression	31
7.3	Gene Set Heat Map.....	31
7.4	KEGG Pathway View	32
8	<i>Pattern Clustering</i>	33
8.1	Clustering of Centroid Profiles.....	33
8.2	Data Table	33
9	<i>Correlation Network</i>	34
9.1	visNetwork	34
9.2	Data Table	34
10	<i>Venn Diagram</i>	36
10.1	vennDiagram	36
10.2	VennDiagram(black & white)	36
10.3	Intersection Output	37
10.4	DEG Table	37
11	<i>Venn Across Projects</i>	38
11.1	vennDiagram	38
11.2	Intersection Output	38
12	<i>Output</i>	39

1 Introduction

The Quickomics interface can be accessed via the url <http://quickomics.bxgenomics.com/>. It has been implemented through RShiny and helps with visualizing secondary and tertiary analysis for RNAseq and Proteomics datasets. This supplemental text will serve as a detailed guide on using the different functionalities and customizing the tools to best fit individual datasets.

The screenshot shows the Quickomics web application. At the top, there is a logo and a navigation bar with several tabs: Select Dataset, QC Plots, Volcano Plot, Heat Map, Expression Plot, Gene Set Enrichment, Pattern Clustering, Correlation Network, and Venn Diagram. A red box highlights the 'Select Dataset' tab, and a red callout bubble points to it with the text 'Tabs for different functionailties'. Below the tabs, there are two buttons: 'Venn Across Projects' and 'Output'. The main content area contains three sections: 'Select data set' (with 'Saved Projects' selected), 'Available Dataset' (a dropdown menu), 'Choose data file' (with 'Browse...' and 'No file selected' buttons), and 'Choose network file' (with 'Browse...' and 'No file selected' buttons). To the right of these sections is a horizontal menu bar with links: Sample Table, Project Overview, Result Table, Data Table, Protein Gene Names, and Help.

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

2 Select Dataset

This is the first tab is for selecting datasets. Users can either select from pre-loaded datasets or upload their own dataset in a pre-defined format as detailed in our github page (<https://github.com/interactivereport/Quickomics>). We have added four pre-loaded datasets from published studies that Users can use for demonstrative purposes.

Please select or upload a date set

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

Select data set
Saved Projects
Upload RData File

Available Dataset

Pick from pre-loaded datasets

2.1 Sample Table

Once the dataset is selected the data on the right panel gets auto populated with metadata/details. Here we have selected the “Mouse Microglia RNA” dataset, previously published in Gyoneva et al., 2019 to illustrate all the functionalities.

The Sample Table contains details of all the samples that were sequenced. In this project there were 93 total samples with attributes like sampleid, group, Age, Genotype and Gender.

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

Select data set
Saved Projects
Upload RData File

Available Dataset

Mouse Microglia RNA

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

Show 15 entries CSV Excel Print

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

Metadata attributes as columns

Showing the total number of samples

2.2 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, for example in this project there are 9 different groups.

The screenshot shows the 'Project Overview' tab selected in the navigation bar. A red box highlights the 'Project Overview' tab. A callout bubble points to the 'Number of Samples in Each Group' table, which lists 9 groups with sample counts: 1yr_Het (11), 1yr_KO (12), 1yr_WT (10), 2mo_Het (9), 2mo_KO (11), 2mo_WT (11), 2yr_Het (8), 2yr_KO (11), and 2yr_WT (10). Another callout bubble points to the detailed project description section, which includes the species (mouse), dataset description (Cx3cr1-Deficient Mouse Microglia RNA-Seq Demo Data), number of samples (93), number of groups (9), number of genes/proteins (15402), and comparison tests (9). A red box highlights the 'Quick summary of data structure in groups' section.

2.3 Result Table

This tab contains normalized expression value, which is TPM in this demo RNA-seq data set, per group for each gene. Each group has the mean normalized count value calculated from replicates of the group and a standard deviation value to denote the variation across the replicates.

The screenshot shows the 'Result Table' tab selected in the navigation bar. A red box highlights the 'Result Table' tab. A callout bubble points to the 'Mean and SD per group' section, which shows the mean and standard deviation for each group across the columns: 2mo_WT_Mean, 2mo_WT_sd, 2mo_Het_Mean, 2mo_Het_sd, and 2mo_KO_Me. The main table displays gene expression data for various genes, including Gna3, Cdc45, Narf, Cav2, Klf6, Scrmh1, Cox5a, Fer, Xpo6, Tfie3, Brat1, Gna12, Itgb2l, and Pih1d2, with their corresponding Gene.Name, Intensity, Protein.ID, and group-specific mean and standard deviation values.

2.4 Data Table

The Data table on the other hand contains normalized expression values (TPM) per sample for each gene.

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 Venn Diagram Venn Across Projects Output

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

Save to output 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	2mo-Het-10M	2mo-Het-11
ENSMUSG000000000001	6.129	6.295	6.269	6.293	6.064	5.996	5.865	6.044	6.063	6.035	5.872	5.877	6
ENSMUSG000000000028	1.692	2.036	1.575	1.614	1.007	0.422	2.205	1.345	2.018	1.807	0.895	1.373	1.8
ENSMUSG000000000056	5.752	5.783	5.884	5.83	5.868	5.64	6.1	5.945	5.616	5.492	5.964	5.903	5.7
ENSMUSG000000000058	3.794	3.377	3.374	3.587	3.781	3.979	3.411	3.876	3.534	3.974	3.782	4.114	3.4
ENSMUSG000000000078	6.607	6.75	6.959	7.052	6.739	6.559	6.512	6.93	6.87	6.36	6.934	6.397	7.0
ENSMUSG000000000085	5.527	5.409	5.155	5.17	5.443	5.087	5.288	4.946	4.968	5.409	5.309	5.147	5.5
ENSMUSG000000000088	7.273	7.3	7.484	7.208	7.23	7.282	7.211	7.409	7.338	7.129	7.23	7.044	7.1
ENSMUSG000000000127	4.851	4.825	4.787	4.927	4.755	4.821	4.711	4.418	4.886	5.011	4.962	5.156	4.8
ENSMUSG000000000131	5.612	5.842	5.515	5.72	5.409	5.417	5.32	5.2	5.269	5.55	5.096	5.118	5.5
ENSMUSG000000000134	7.152	7.292	7.131	7.264	7.005	7.066	6.652	6.994	6.923	7.11	6.746	6.971	7.3
ENSMUSG000000000148	5.417	5.549	5.586	5.661	5.536	5.586	5.201	5.504	5.395	5.578	5.144	5.401	5.5
ENSMUSG000000000149	8.368	8.598	8.473	8.468	8.236	8.183	7.782	8.13	7.824	8.256	7.954	7.82	8.6

2.5 Protein Gene Names

This tab contains details about each gene, and if a corresponding protein ID was associated with it in the experiment. Quickomics gives the ability to take in proteomics dataset and this is a useful feature to associate protein name with gene id.

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 Venn Diagram Venn Across Projects Output

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

Save to output Show 15 entries CSV Excel Print

Search:

	id	UniqueID	Gene.Name	Protein.ID	GeneType
0	ENSMUSG000000000001		Gna3		protein_coding
1	ENSMUSG000000000028		Cdc45		protein_coding
2	ENSMUSG000000000056		Narf		protein_coding
3	ENSMUSG000000000058		Cav2		protein_coding
4	ENSMUSG000000000078		Kif6		protein_coding
5	ENSMUSG000000000085		Scmh1		protein_coding
6	ENSMUSG000000000088		Cox5a		protein_coding
7	ENSMUSG000000000127		Fer		protein_coding
8	ENSMUSG000000000131		Xpo6		protein_coding
9	ENSMUSG000000000134		Tfe3		protein_coding
10	ENSMUSG000000000148		Brat1		protein_coding
11	ENSMUSG000000000149		Gna12		protein_coding
12	ENSMUSG000000000157		Itgb2l		protein_coding
13	ENSMUSG000000000167		Pihd2		protein_coding
14	ENSMUSG000000000168		Dlat		protein_coding

Showing 1 to 15 of 15,402 entries

Previous 1 2 3 4 5 ... 1027 Next

2.6 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.

The screenshot shows the Quickomics web application interface. At the top, there is a navigation bar with tabs: Select Dataset (highlighted in blue), QC Plots, Volcano Plot, Heat Map, Expression Plot, Gene Set Enrichment, Pattern Clustering, Correlation Network, Venn Diagram, Venn Across Projects, and Output. Below the navigation bar, the main content area has a title "LRRK2 Neuron Transcriptome". On the left, there is a sidebar titled "Select data set" with two options: "Saved Projects" (selected) and "Upload RData File". Below this is a section titled "Available Dataset" with a dropdown menu showing "LRRK2 Neuron RNA". The main content area contains several tabs: Sample Table, Project Overview, Result Table, Data Table, Protein Gene Names, and Help. The "Help" tab is highlighted with a red box. To the right of the tabs, there is a list of descriptions for each tab:

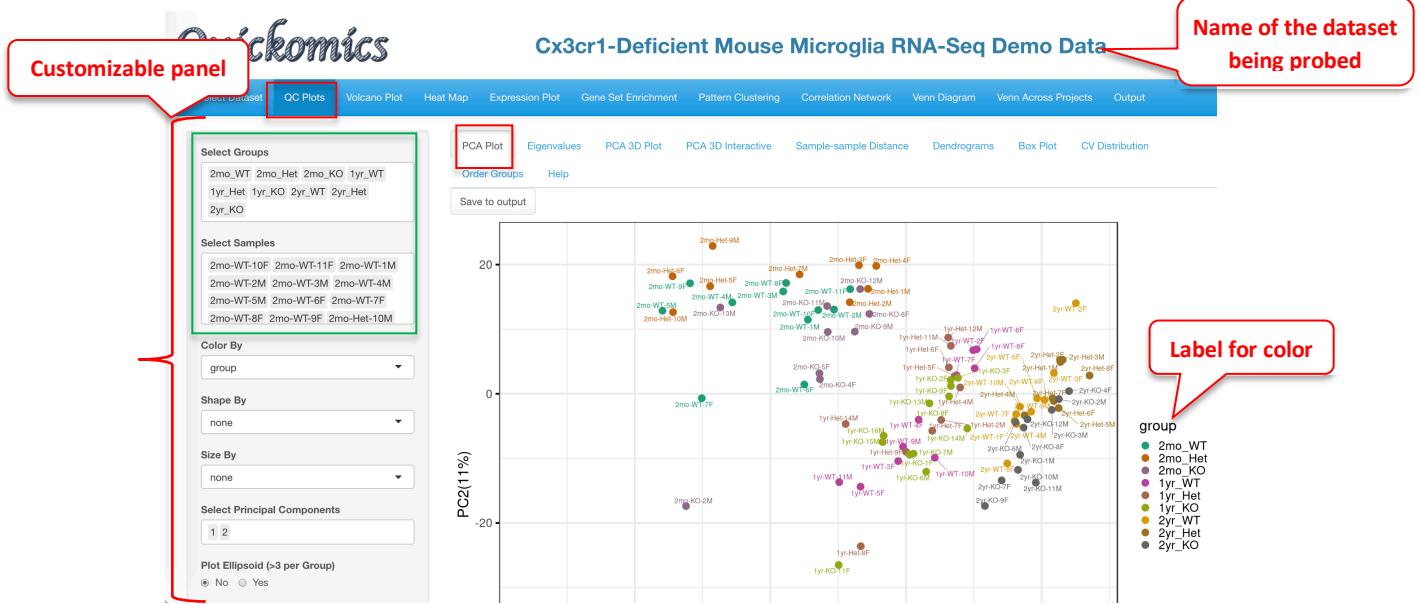
- Result Table: Statistics result by using LIMMA package, include log2 Fold Change, p value, p-values adjusted (Benjamini-Hochberg)
- Data Table: Normalized data
- Sample Table: sample group and comparison information
- Protein Gene Names: protein accession number and gene symbol

3 QC Plots

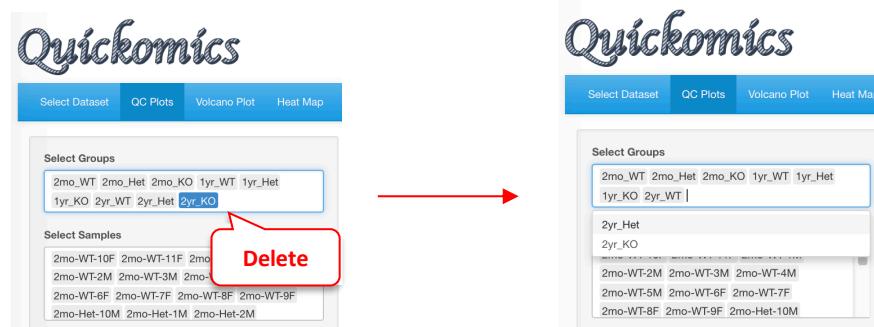
For every dataset, the first step of the analysis is to perform basic QC and test the integrity of the data. "QC Plots" tab performs basic QC analysis on the datasets and helps produce high quality figures. The plots were generated using ggplot2, plotly, ComplexHeatmap and heatmap.2 R packages.

3.1 PCA Plot

The principle component analysis (PCA) plot displays by default the first 2 PC's in the dataset. The default settings for display are to color by the group with no shape or size selected. This plot is highly customizable: depending on the project and samples, different attributes can be highlighted.



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.



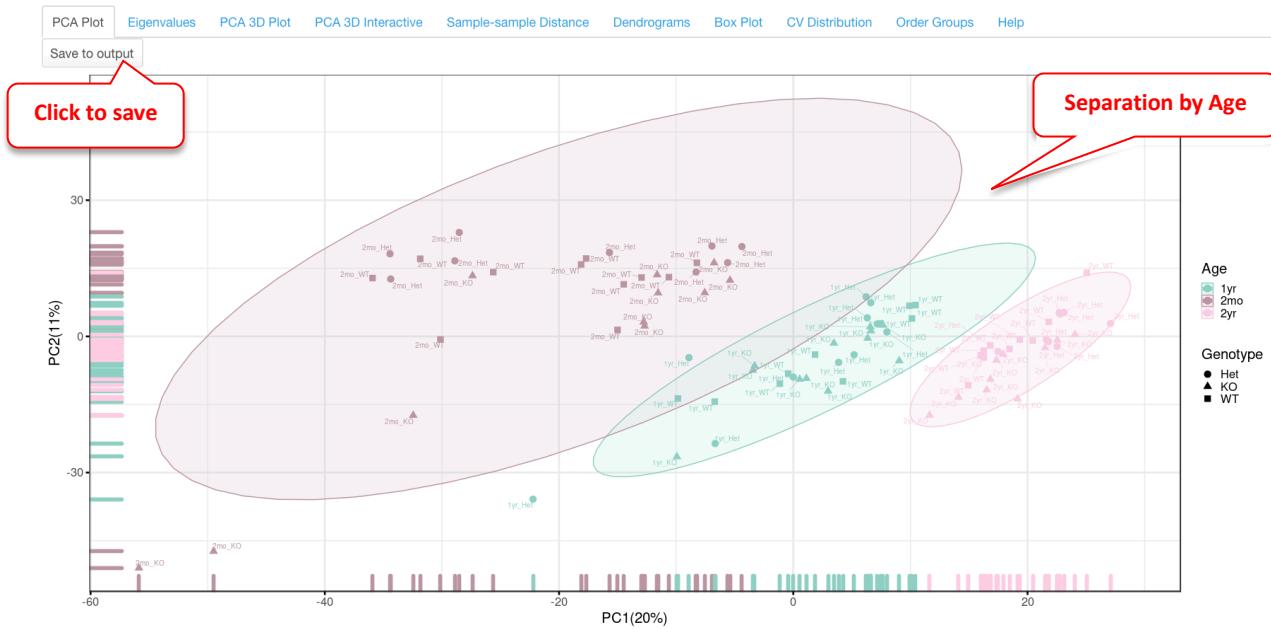
To illustrate the strengths of customization, few different features have been changed below:

The screenshot shows a configuration interface for a PCA plot. Several features are highlighted with red boxes and numbered labels:

- 1. Color by Age**: A red box surrounds the "Color By" dropdown menu set to "Age".
- 2. Shape by Genotype**: A red box surrounds the "Shape By" dropdown menu set to "Genotype".
- 3. Yes to Plot Ellipsoid**: A red box surrounds the "Plot Ellipsoid (>3 per Group)" section, which includes a radio button for "Yes".
- 4. Yes to Margin Rugs**: A red box surrounds the "Show Marginal Rugs" section, which includes a radio button for "Yes".
- 5. Change palette**: A red box surrounds the "Select palette" dropdown menu set to "Set3".
- 6. Change to group**: A red box surrounds the "Select Sample Label" section, which includes a radio button for "group".

The interface also includes other settings such as "Select Groups" (listing sample names), "Select Samples" (listing sample names), "Size By" (set to "none"), "Select Principal Components" (set to 1, 2), "Show Mean Point" (radio button for "Yes"), and "Dot Size" and "Label Font Size" sliders.

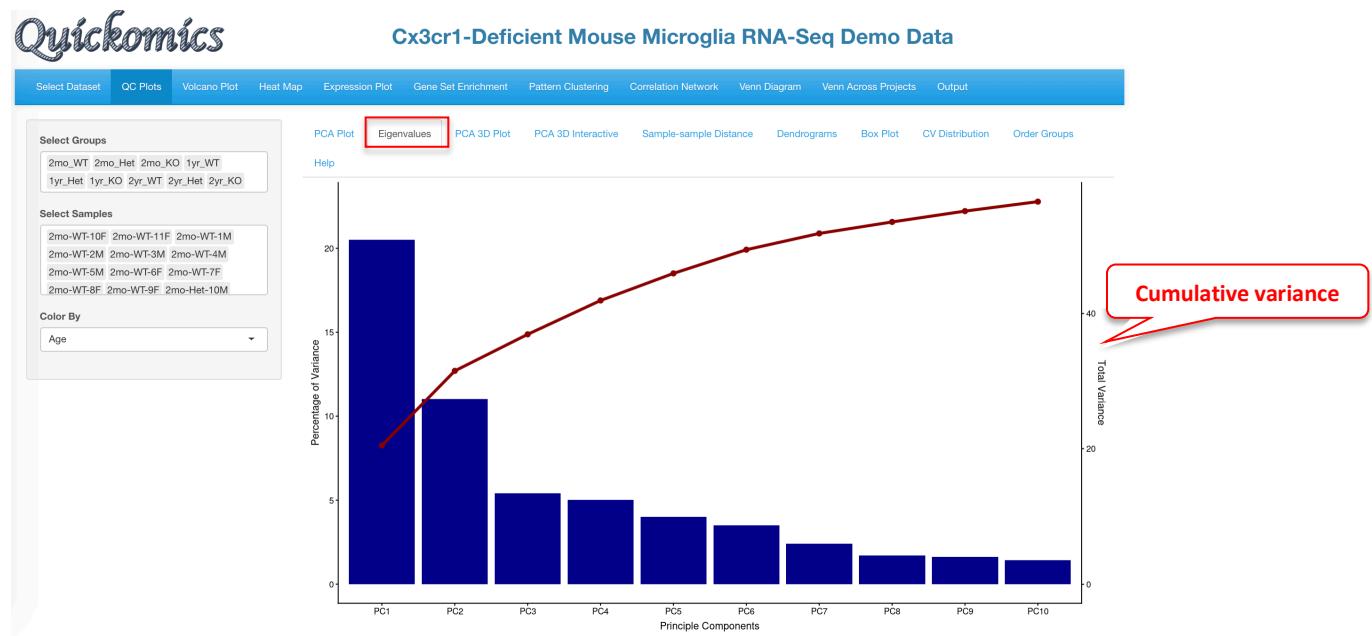
The resulting selections produce this plot. It is very clear that Age is the factor that drives the biggest separation, especially along PC1. There is some separation driven by the Genotype as well. The Ellipsoid feature helps group samples by the color attribute, while the Marginal Rug feature helps understand the distribution of the color attribute along the axis.



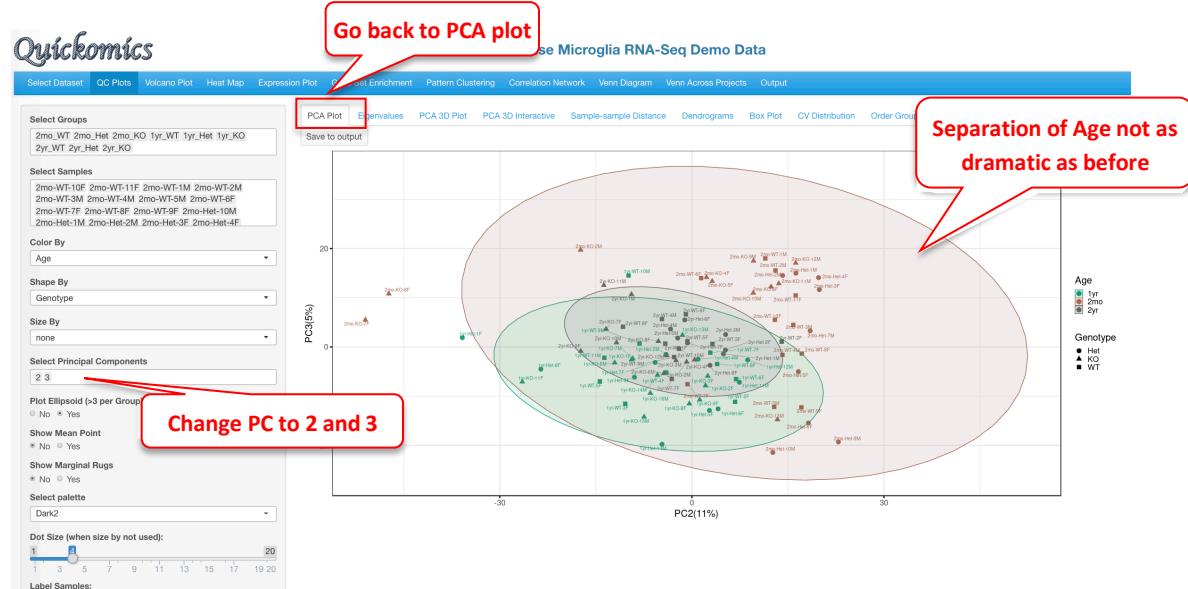
The plot can be saved in high resolution by clicking on “Save to Output”. Section 12 of this document describes the next steps of downloading and obtaining the saved plots.

3.2 Eigenvalues

By default, the PCA plot seen in the previous section (3.1) uses PC1 and PC2. But in many cases, other PC's may be important in explaining the variance in the dataset. This section plots the variance explained by 10 PCs in the dataset. Users can then make educated decisions of which PCs to plot and use in their PCA analysis. In this dataset PC1 and PC2 together explain around 30% of the variance.

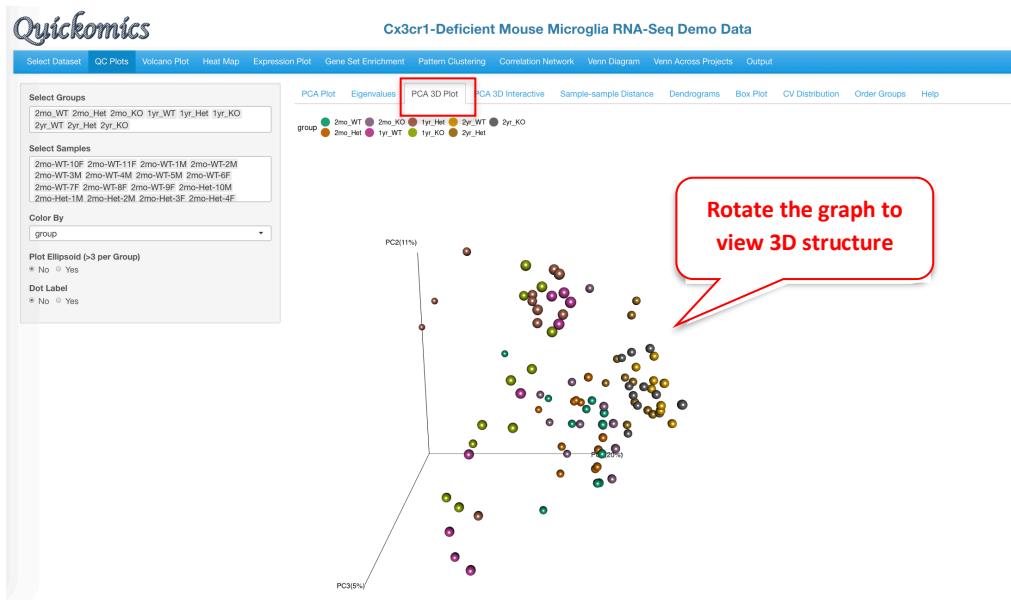


PC2 and PC3 were used to redo the PCA plot, but here the factors driving the separation of the samples is not as clear.

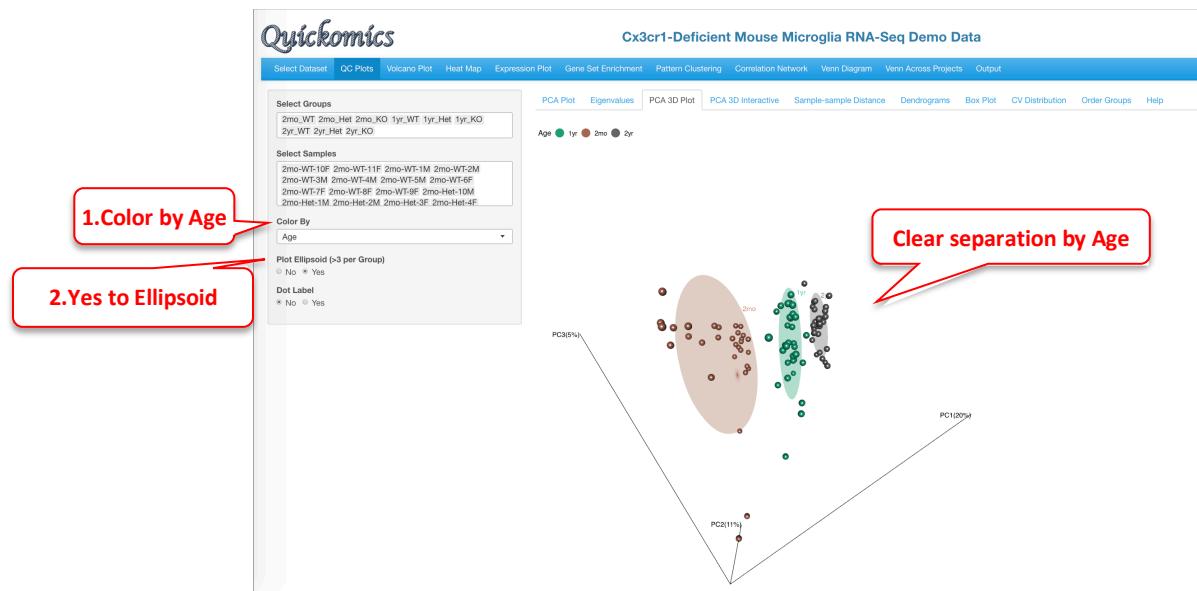


3.3 PCA 3D Plot

This 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 view it from different angles. There are 3 attributes that can be changed: color, including an ellipsoid, and to add labels to the samples.

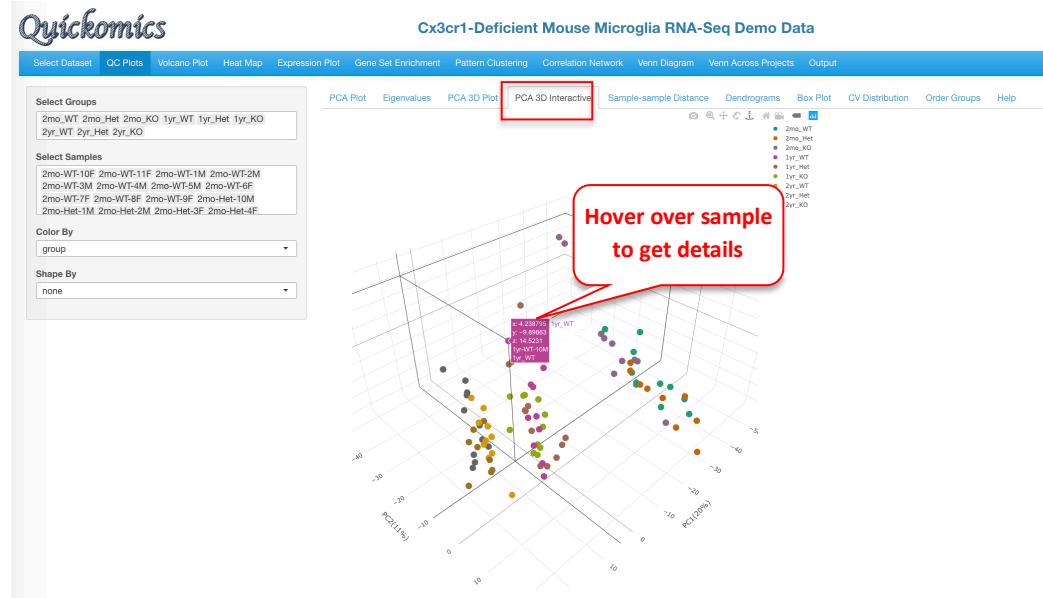


The “Color By” and “Plot Ellipsoid” attributes were changed in the following plot. This highlights the clear separation of samples by Age.

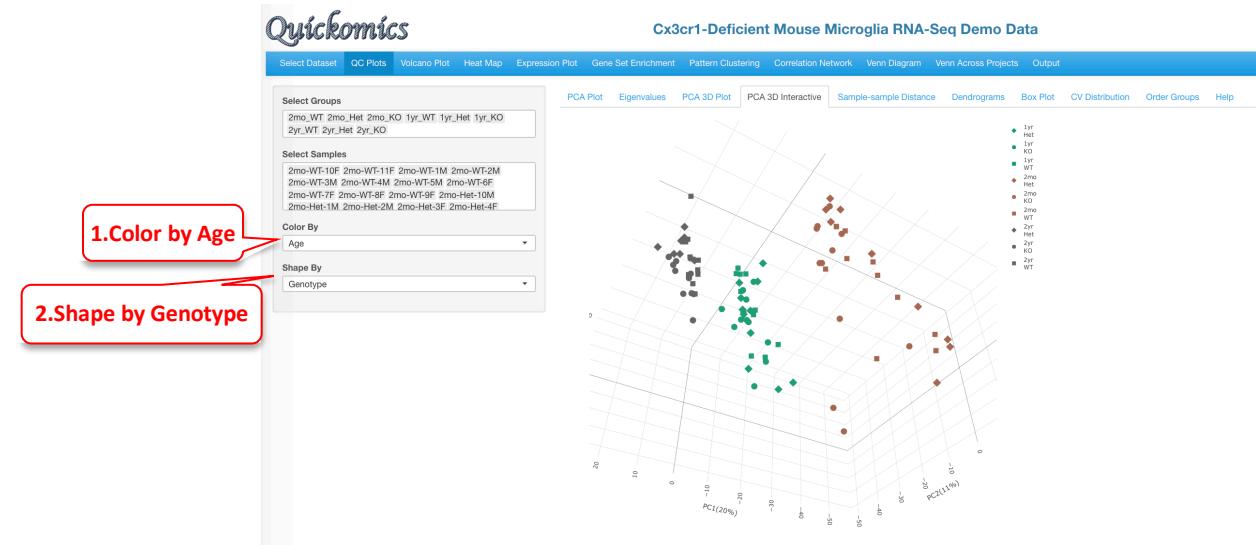


3.4 PCA 3D Interactive

This tab is similar to the previous one (3.2 PCA 3D Plot), but additionally Users have the ability 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.



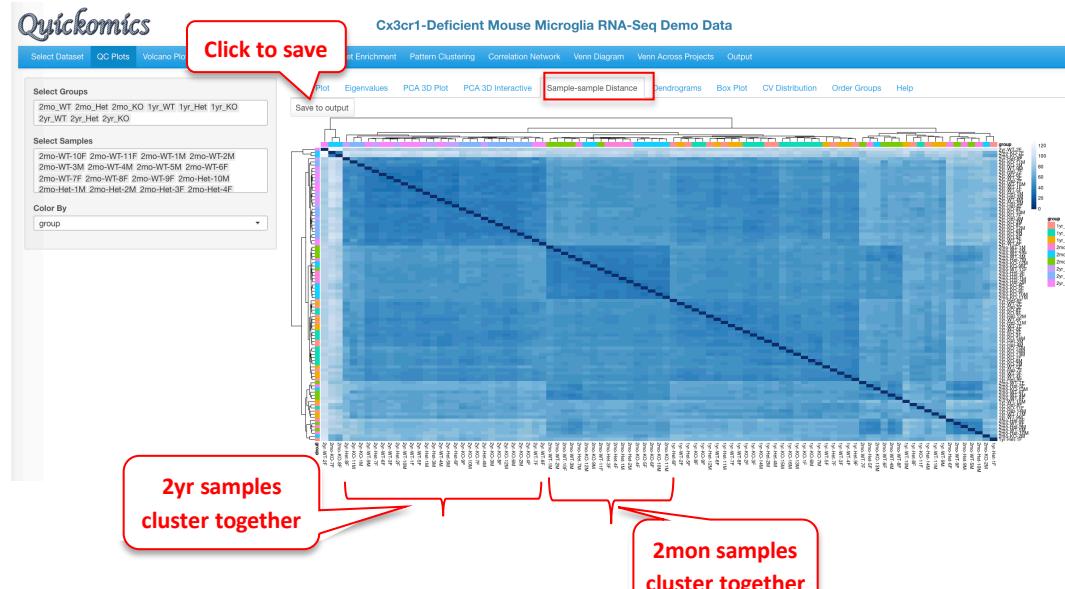
In the plot below, the “Color By” and “Shape By” attributes have been changed to highlight the drivers of separation.



3.5 Sample-sample Distance

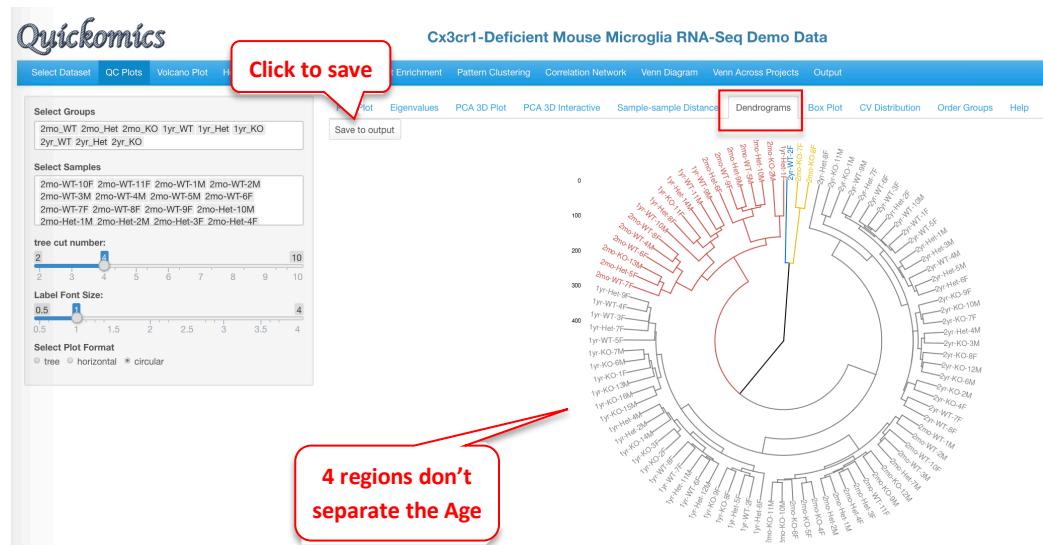
This tab helps identify pair-wise similarity between all samples. A distance matrix is generated for every sample pair and is plotted as a heatmap. The rows and columns are clustered based on similarity.

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.

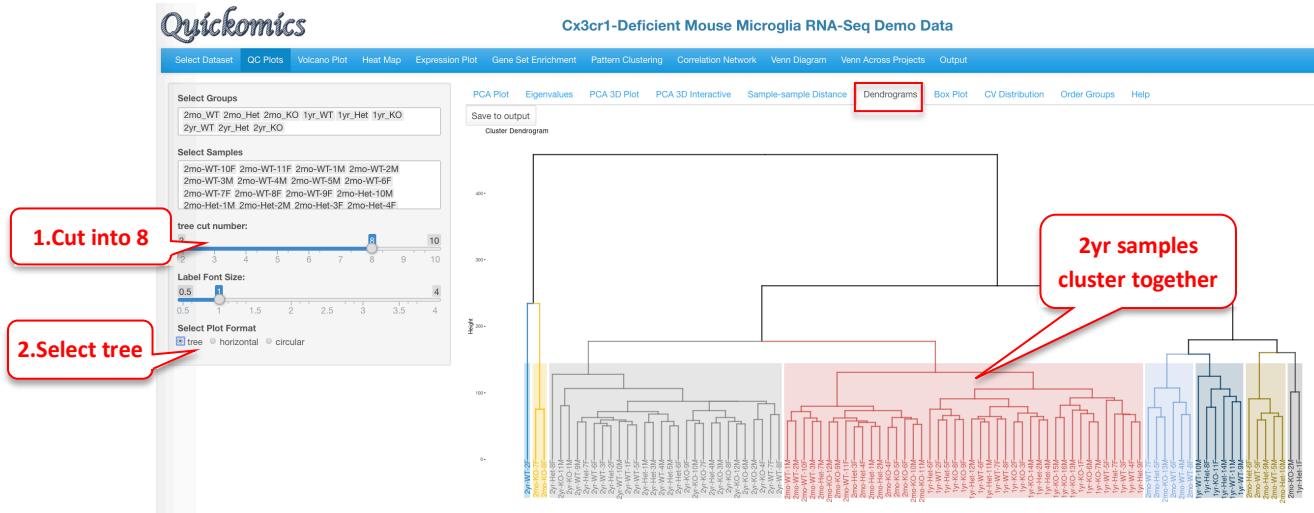


3.6 Dendrograms

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 ability to visualize in two other ways (tree or horizontal) and cut the plot in multiple regions. Here again, users have the option to save to output.

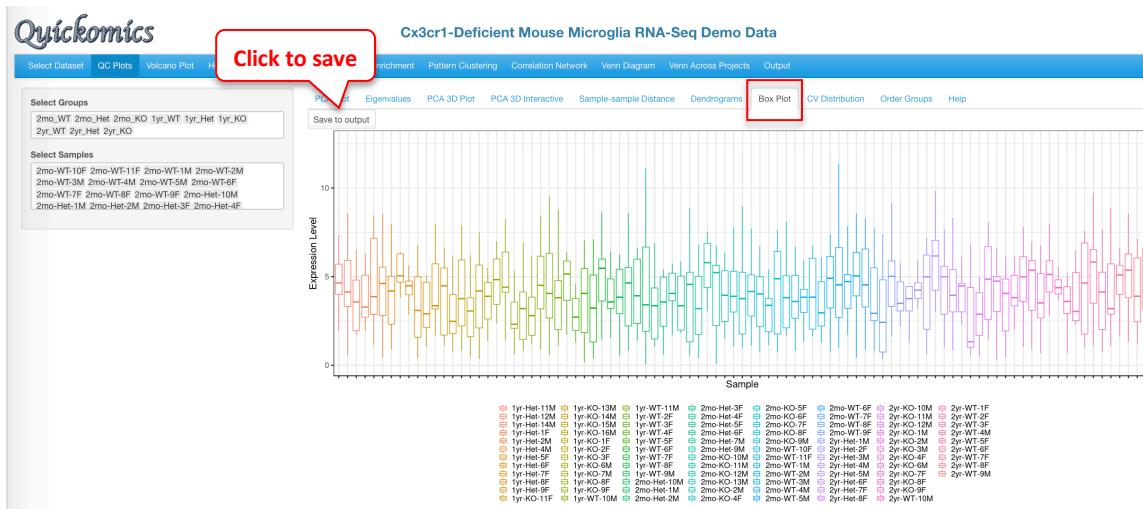


In the plot below, we have selected to visualize as a tree and cut into eight parts. This shows the relationship between the samples to help understand the hierachal clustering. Overall, the samples are arranged by age, but there are some outliers that appear. It is also interesting how the 2mon samples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



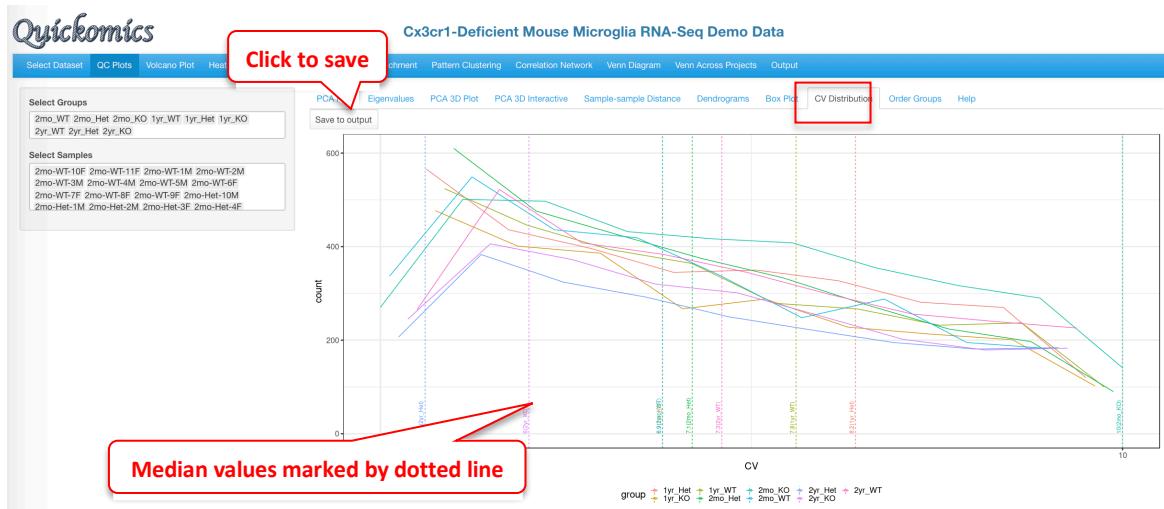
3.7 Box Plot

The BoxPlot is a visualization to understand the distribution of the 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.



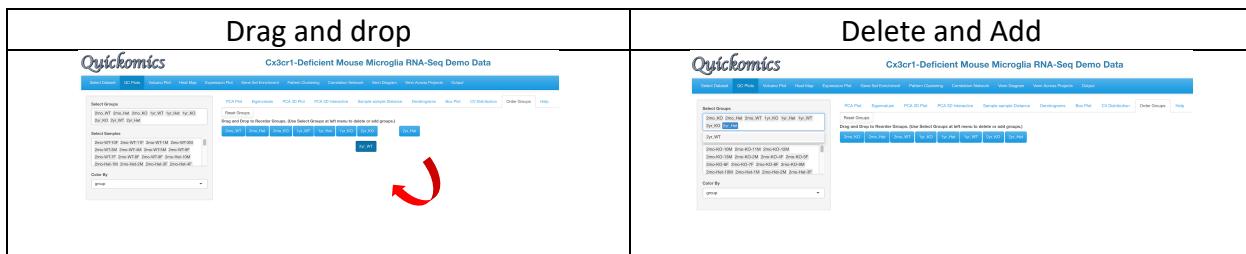
3.8 CV Distribution

This plot shows the histogram of CV (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.

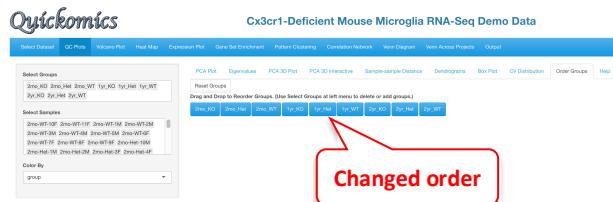


3.9 Order Groups

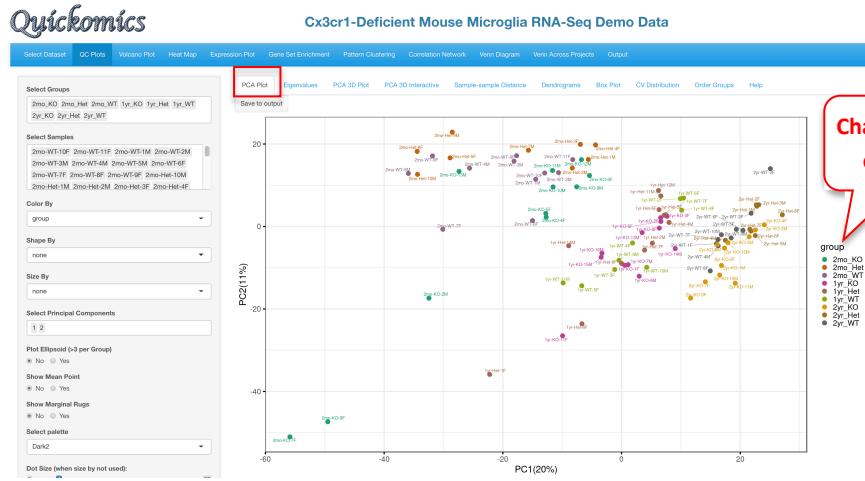
This tab lets the user order the groups, so they are displayed in a preferred order in the other QC plots. Users can either drag and drop the group to reorder, or use the select panel to delete and add. In this example, we have changed the order of the group so that the KO samples are before Het and WT.



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

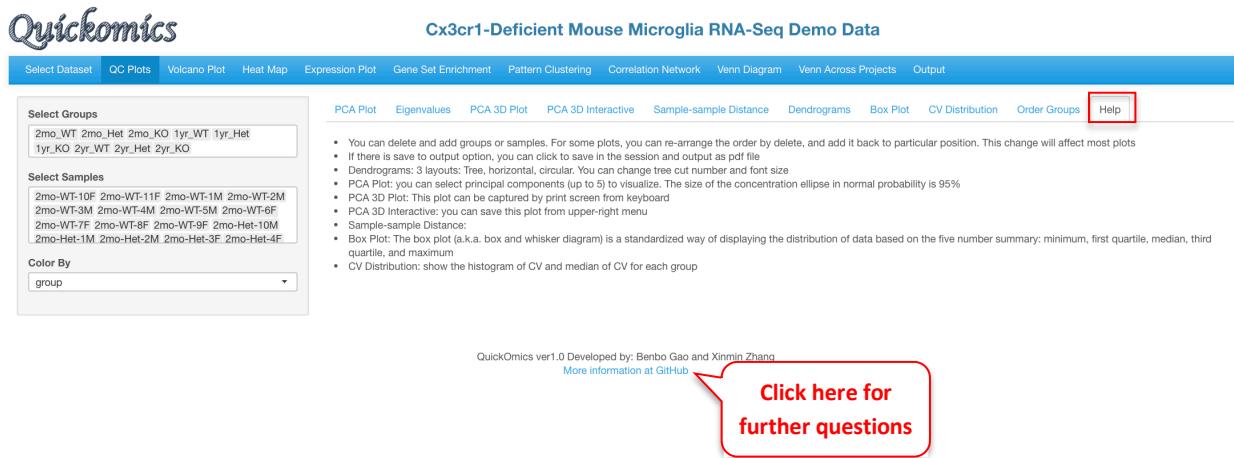


Going back to the PCA plot, the order of the groups is now changed. This order will also be reflected in all other visualizations in the QC plots.



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

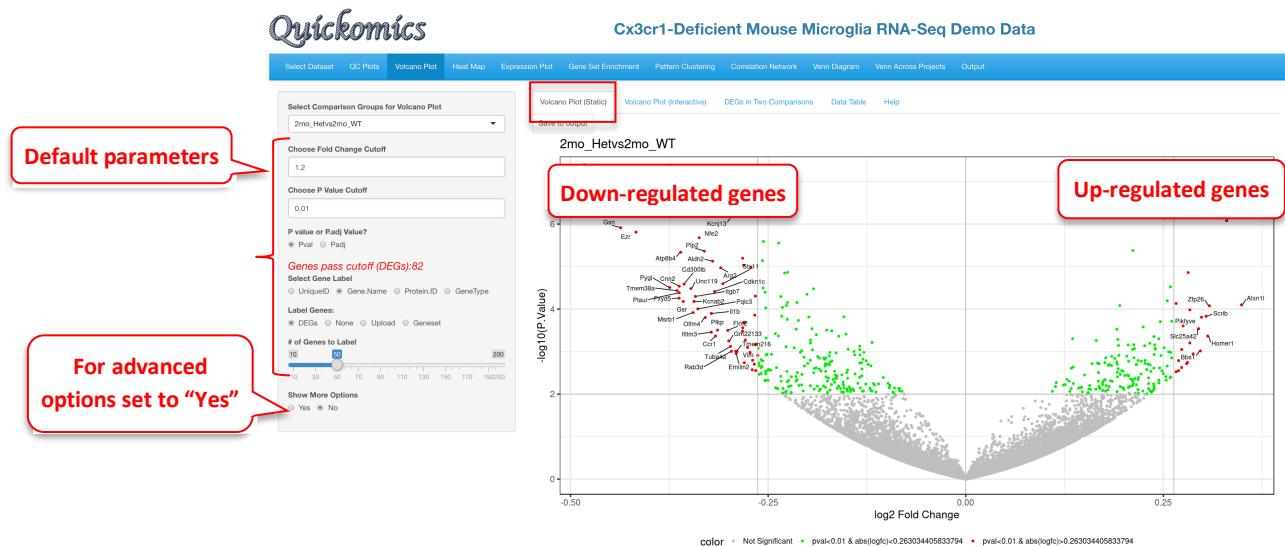
This section is designed to explore the differentially expressed genes (DEGs) in the dataset. The comparisons are defined by the User when they input the dataset. Please refer to our Github for more details on uploading datasets.

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 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.

Quickomics

Select Dataset: QC Plots | Volcano Plot | Heat Map | Expression Plot | Gene Set Enrichment

Select comparison: 2mo_KOvs2mo_WT

Change Fold Change: 1.5

Select Padj: Pval * Padj

Select Geneset: Select Geneset

Add or remove any gene: CD55, CD14, EMP2, MDK, FCGR2, DAB2

For customizing the axis limits in the plot: 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: 10; Legend Font Size: 8

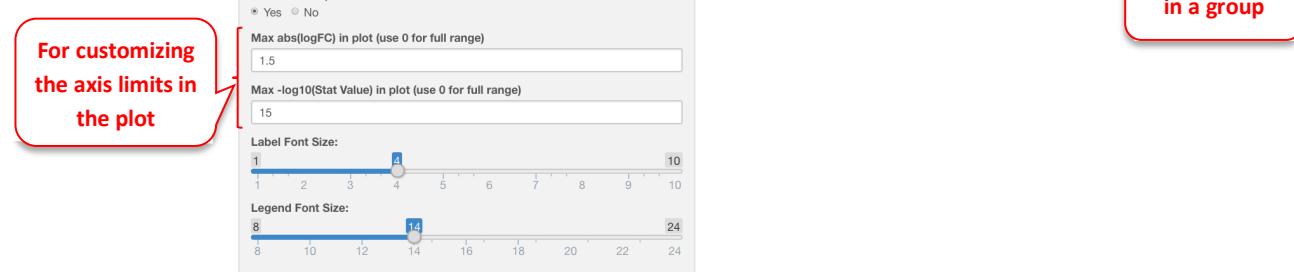
Select Database: GSEA MSigDB

Search for a particular keyword: Immune

Geneset Details:

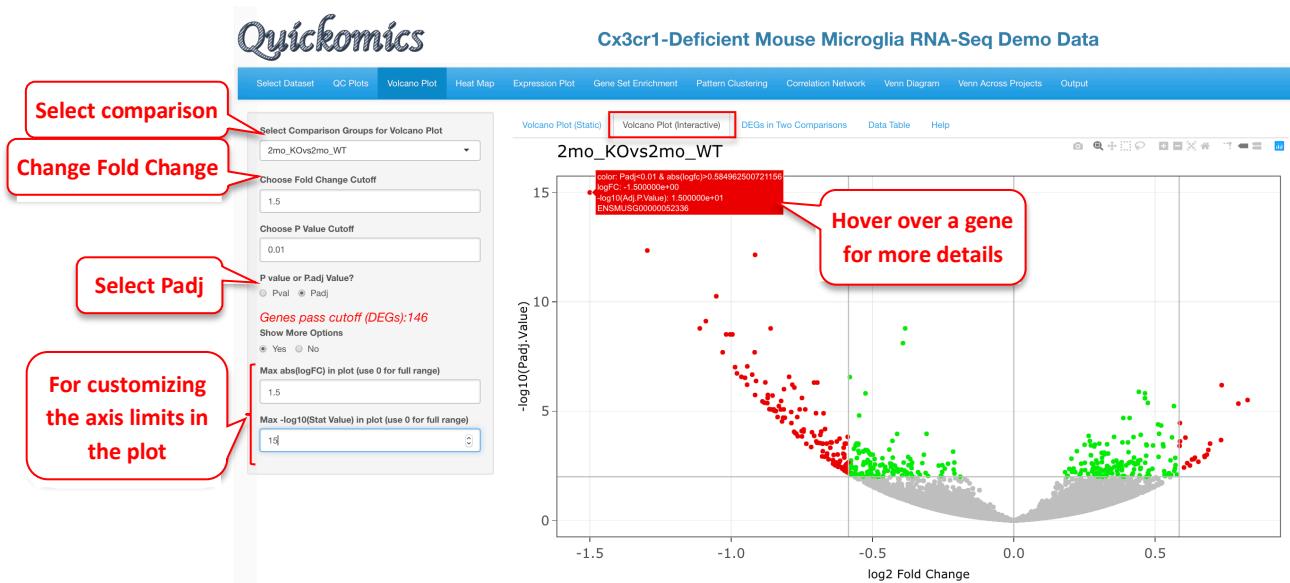
- Immune response, Display Method: Gene IDs, one per row
- Gene IDs, comma separated
- Gene Names, one per row
- Gene Names, comma separated

CD55, CD14, EMP2, MDK, FCGR2, DAB2, MYLK, SLEENOP



4.2 Volcano Plot (Interactive)

This next 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.



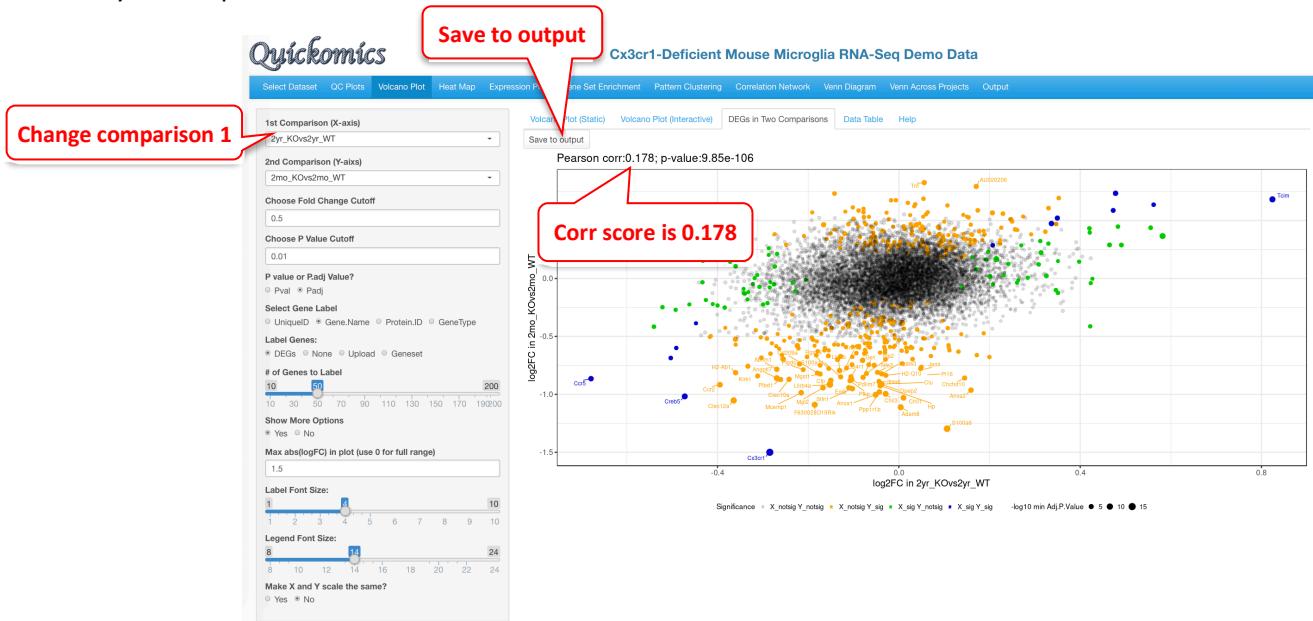
4.3 DEGs in Two Comparisons

This next tab helps with identifying Differentially Expressed Genes (DEGs) 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.

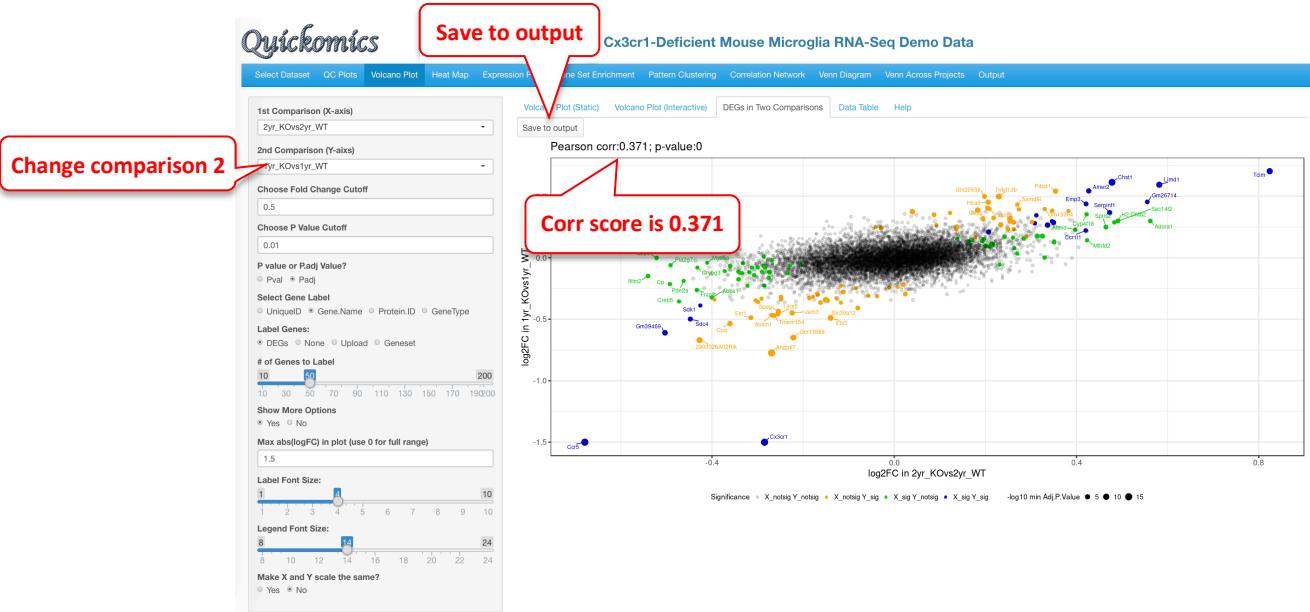
2mo_KOvs2mo_WT vs 1yr_KOvs1yr_WT has a low correlation score. There are many genes that are colored yellow, meaning those genes were significant in 2mo_KOvs2mo_WT (Y axis) but not in 1yr_KOvs1yr_WT (X axis).



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 Differentially Expressed Genes (DEGs) in a tabular format with a searchable feature. Users also have the ability to download the gene list as a CSV.

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	ENSMUSG00000001028	S100a6		protein_coding	2mo_KOvs2mo_WT	4.51e-13	6.46e-17	-1.3
3450	ENSMUSG00000025473	Adam3				1.64e-9	9.42e-13	-1.11
11600	ENSMUSG00000078122	Fil30028D01Rik				7.65e-10	2.74e-13	-1.09
9390	ENSMUSG00000053063	Clec12a				5.55e-11	1.59e-14	-1.05
10400	ENSMUSG00000064246	Chif1				2.04e-8	2.04e-11	-1.03
9380	ENSMUSG00000053007	Creb5				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	Chi3		protein_coding	2mo_KOvs2mo_WT	3.09e-9	2.27e-12	-0.995
1010	ENSMUSG00000013974	Momp1		protein_coding	2mo_KOvs2mo_WT	9.71e-8	1.11e-10	-0.986
10200	ENSMUSG00000061718	Ppp1rb		protein_coding	2mo_KOvs2mo_WT	1.87e-7	2.28e-10	-0.98
5900	ENSMUSG00000032231	Anxa2		protein_coding	2mo_KOvs2mo_WT	2.69e-7	3.86e-10	-0.964
11700	ENSMUSG00000078763	Slim1		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	ENSMUSG00000040950	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	ENSMUSG000000112148	Lrrk4a		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

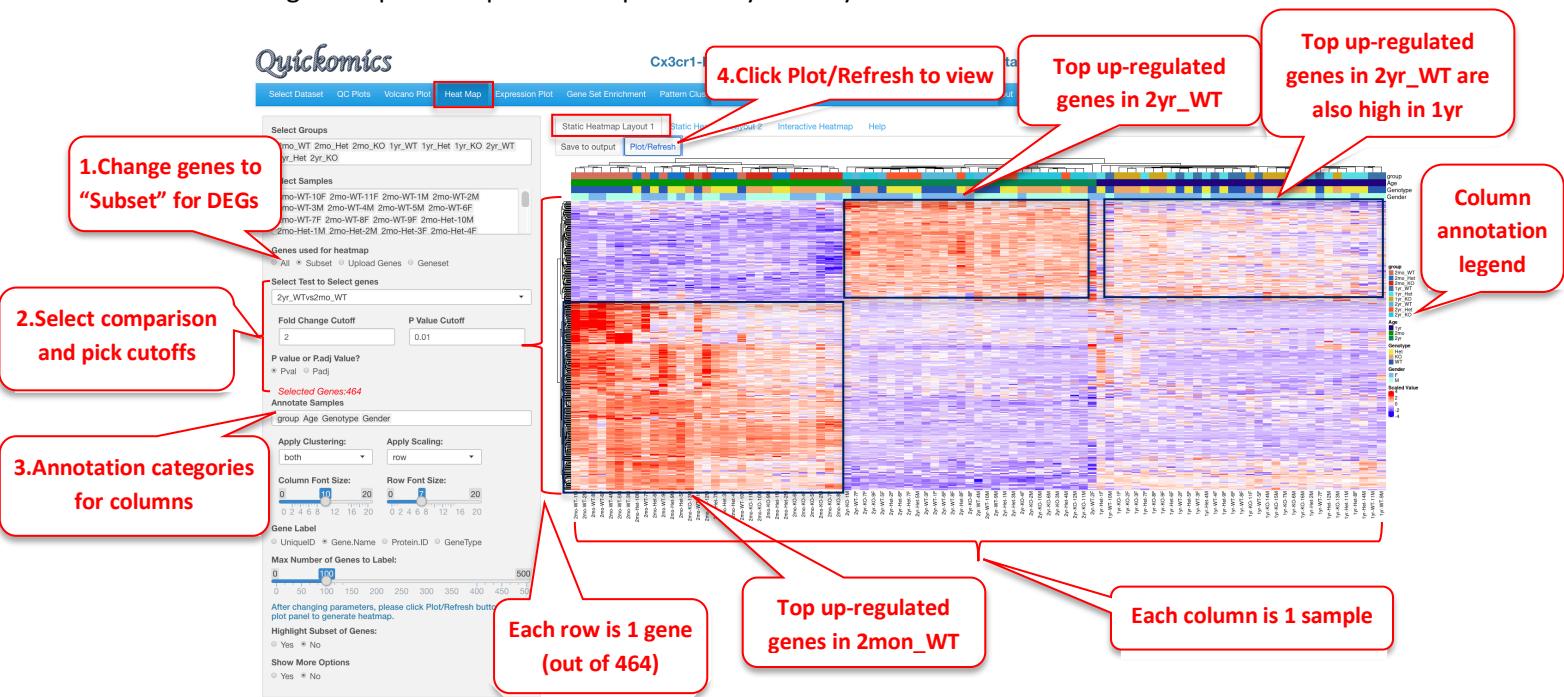
5 Heat Map

This section of Quickomics produces highly customized Heatmaps that are publication quality. We have used pheatmap package from R to generate these plots. Users can enter automated list of differentially expressed genes or enter a list of genes of their own. This section is slightly different from other sections in that each time Users change the attributes; they need to click on the “Plot/Refresh” button to generate the plot.

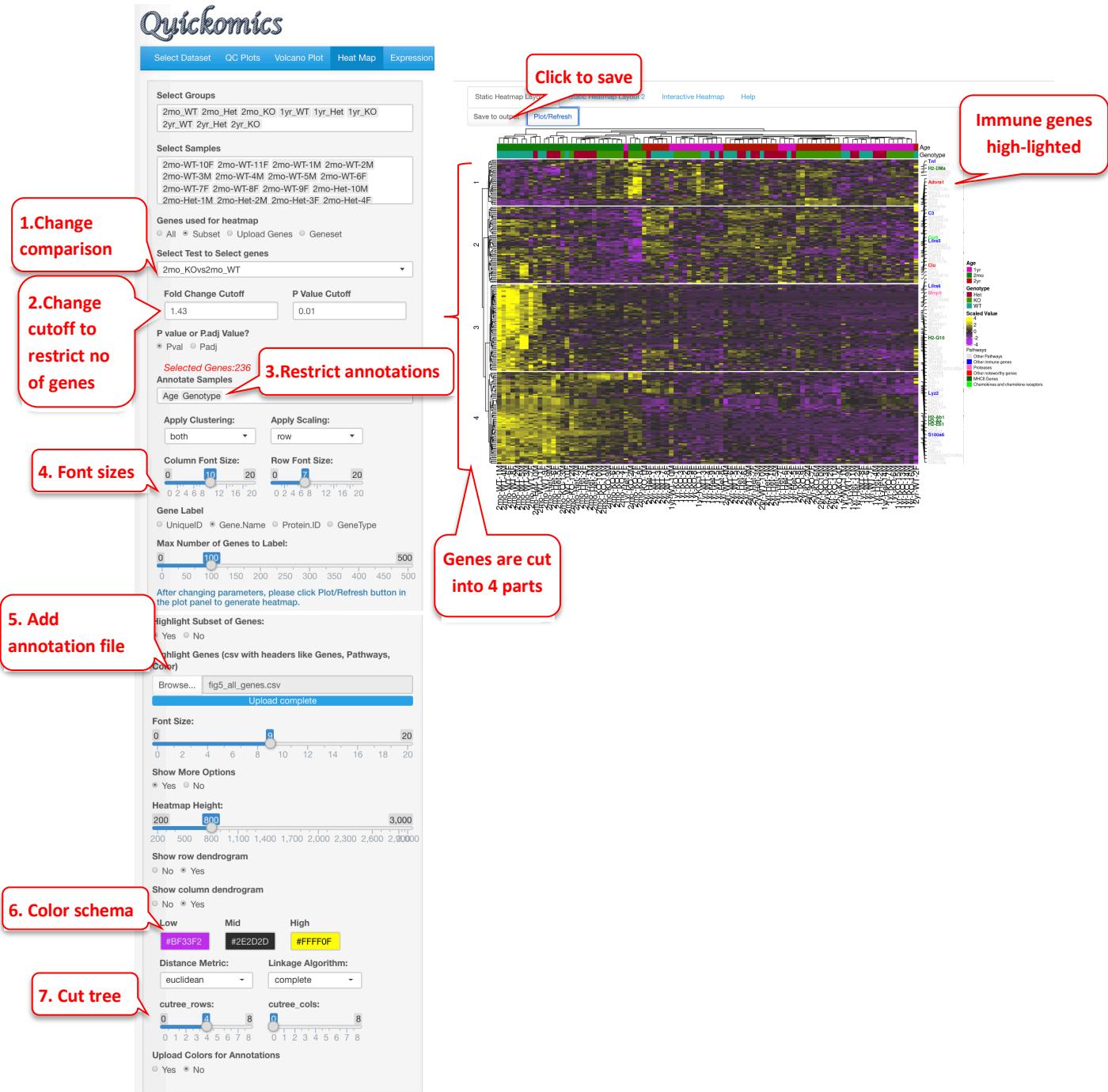
5.1 Static Heatmap Layout 1

The “Static Heatmap Layout 1” uses the ComplexHeatmap package for generating the plot. For demo purposes, we have selected options that highlight some capabilities to generate informative Heatmaps. 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 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 the 2mon microglia have a different gene expression profile compared to 1yr and 2yr.

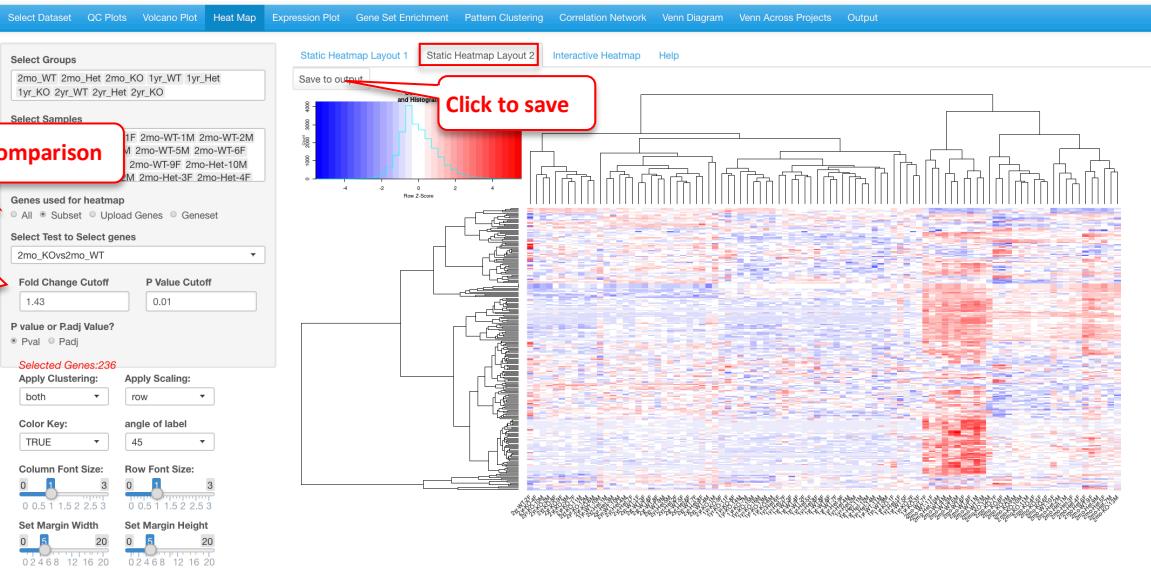


In this next example, we customize another Heatmap to highlight immune related genes that were called out in Gyoneva et al., 2019. This is done by uploading a csv file consisting of the gene names with pathway information and the color for display. An example of this file is on our github (Quickomics/demo_files/fig5_all_genes.csv)



5.2 Static Heatmap Layout 2

The “Static Heatmap Layout 2” uses the `heatmap.2` 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.



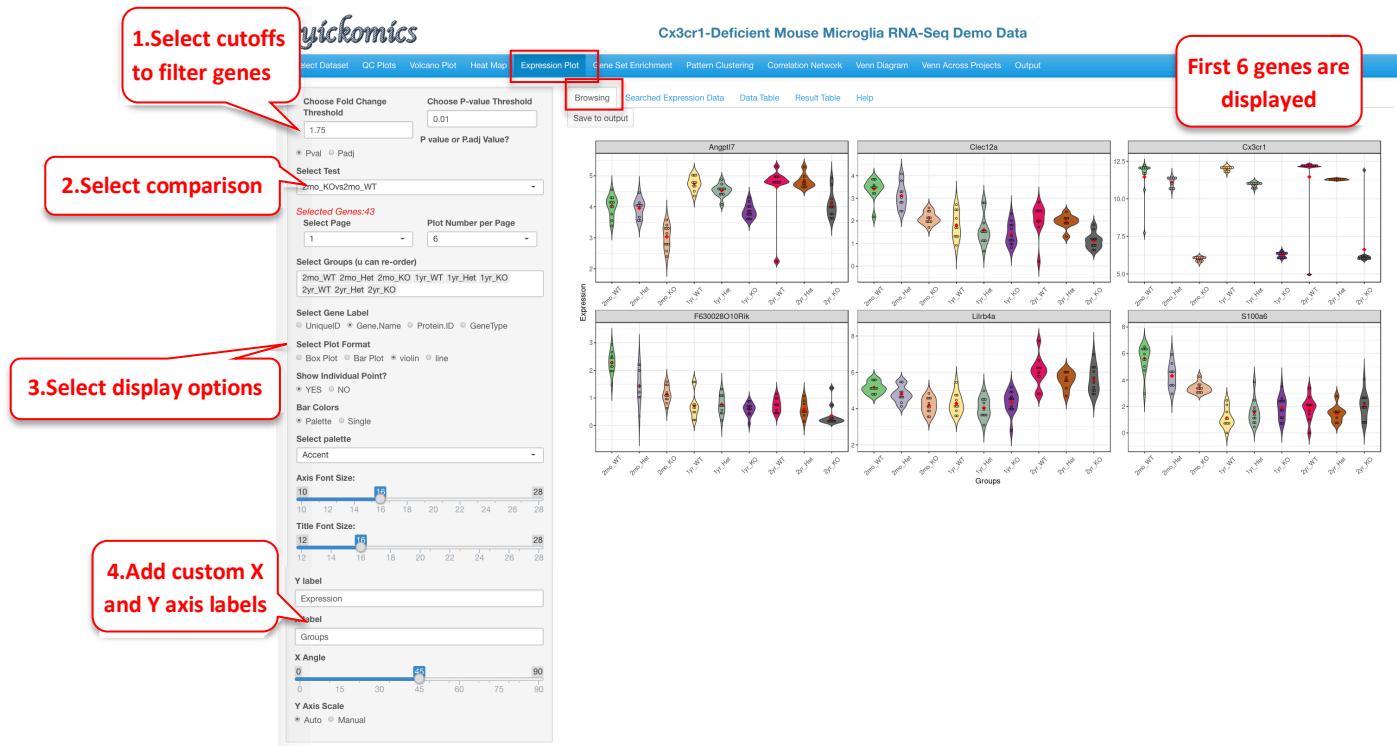
2.Change cutoff to restrict no of genes

6 Expression Plot

This functionality in Quickomics offers Users the ability to plot expression of individual genes in different formats like Box Plot, Bar Plot, Violin Plot or Line Plot.

6.1 Browsing

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



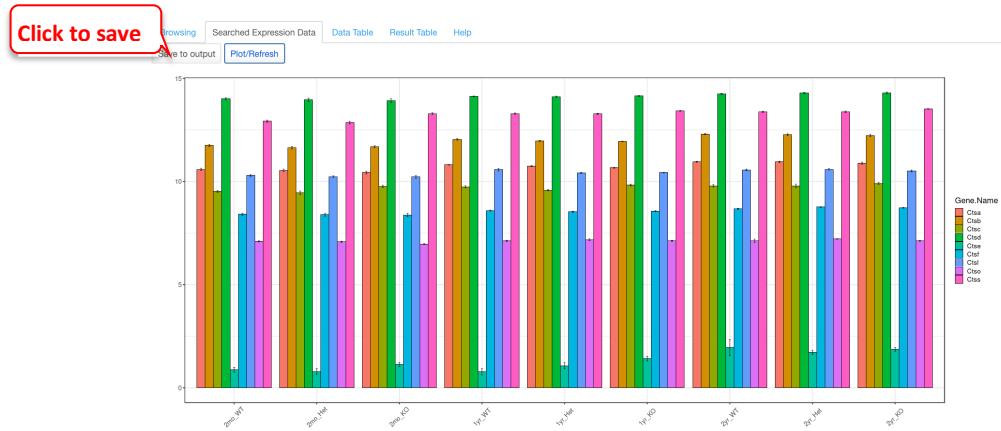
6.2 Searched Expression Data

Another way to plot genes is using this functionality. Users can either enter a list of genes themselves or select a list from online databases like Kegg, mSigDB etc. Here is an example of plotting genes involved in MHCII antigen processing plotted in one bar plot.

The screenshot shows the Quickomics interface with several red boxes and arrows highlighting steps:

- 1. Select Geneset**: A red box highlights the "Select Geneset" button in the top navigation bar.
- 2. Select Mus musculus**: A red box highlights the search bar where "Mus musculus" is typed.
- 3. Search "MHC"**: A red box highlights the search results table where "MHC" is selected from a dropdown menu.
- 4. Select OnePlot**: A red box highlights the "OnePlot" selection in the "Select Plot Format" dropdown.
- 5. Select display options**: A red box highlights the "Display Options" section, which includes font size sliders for axis and title.
- 6. Click Plot/Refresh**: A red box highlights the "Plot/Refresh" button at the bottom of the left sidebar.

The resulting plot shows the expression of a subset of genes from the MHCII antigen presentation family of genes.



6.3 Data Table

This tab is specifically designed to search protein expression values in proteome datasets. For a demo of this feature, we select a proteome dataset from Connor-Robson et al., 2019.

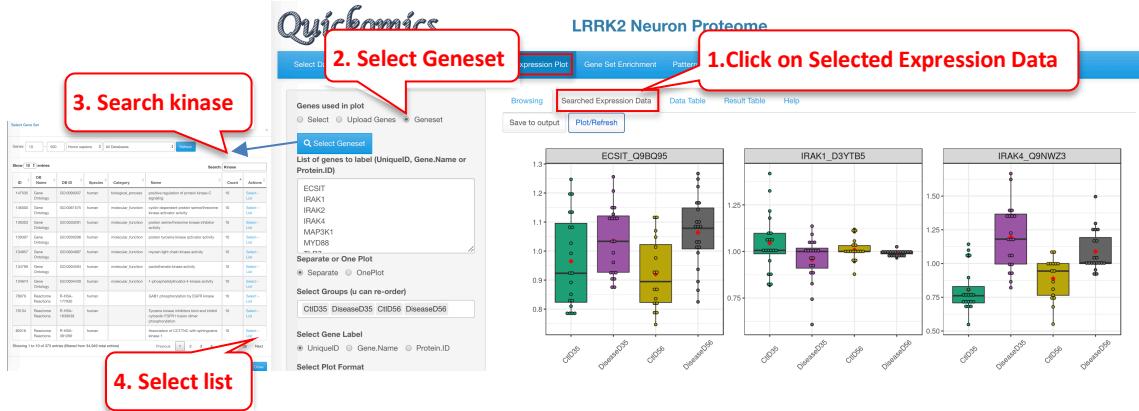
The screenshot shows the Quickomics interface with several red boxes and arrows highlighting steps:

- 1. Select Proteome dataset**: A red box highlights the "Select Proteome dataset" dropdown menu.
- Dataset name**: A red box highlights the "LRRK2 Neuron Proteome" dataset name in the dropdown menu.

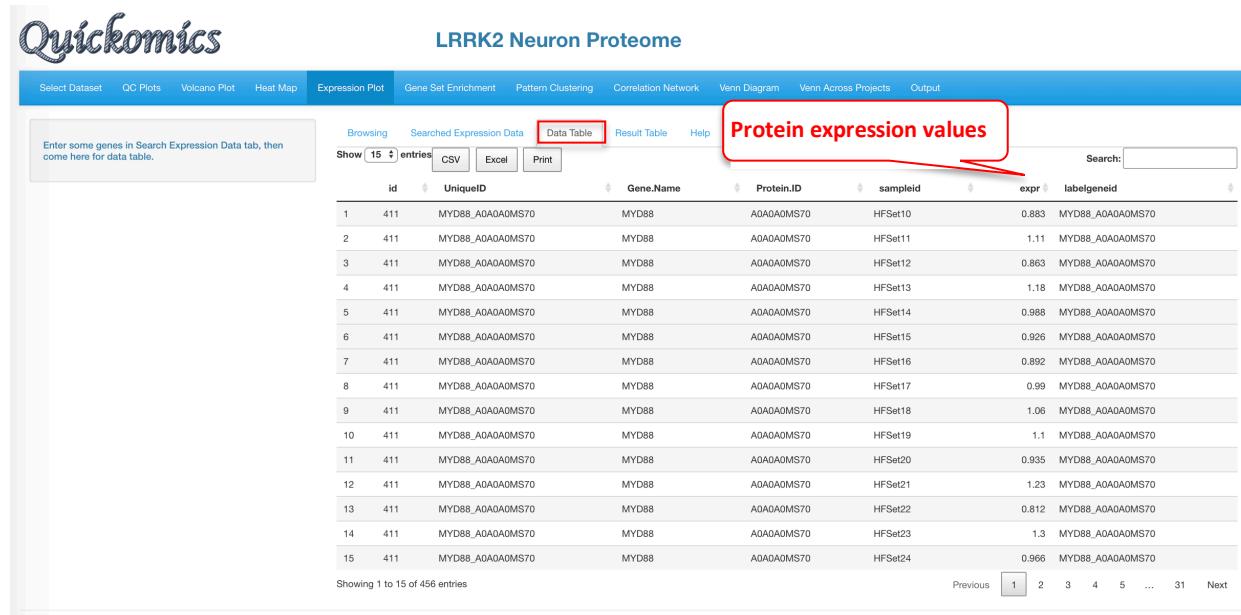
The "Data Table" tab is active, displaying a table of sample data:

sampleid	group	Sex	TimePoint
1	CtlD35	F	D35
2	DiseaseD35	M	D35

Once the dataset is selected, we go into the "Searched Expression Data" tab to select a geneset.



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



6.4 Result Table

This tab lists the log FC of the genes selected in “Selected 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.

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

Browsing Searched Expression Data Data Table Result Table Help

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_D3YTB5	IRAK1	D3YTB5	DiseaseD35vsCtlD35	0.0327	0.0196	-0.169
4	1240	IRAK1_D3YTB5	IRAK1	D3YTB5	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

Search:

Previous **1** Next

7 Gene Set Enrichment

This functionality in Quickomics is an extension to the differential gene identification. Users are able to use different cutoffs to filter out DEGs in different comparisons and probe pathways enriched. This is a useful way to make biological sense from 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. Very similar to the results of the Gyoneva et al., 2019, this analysis points to immune related terms being enriched in the down-regulated genes for 2mon_KOvs2mon_WT.

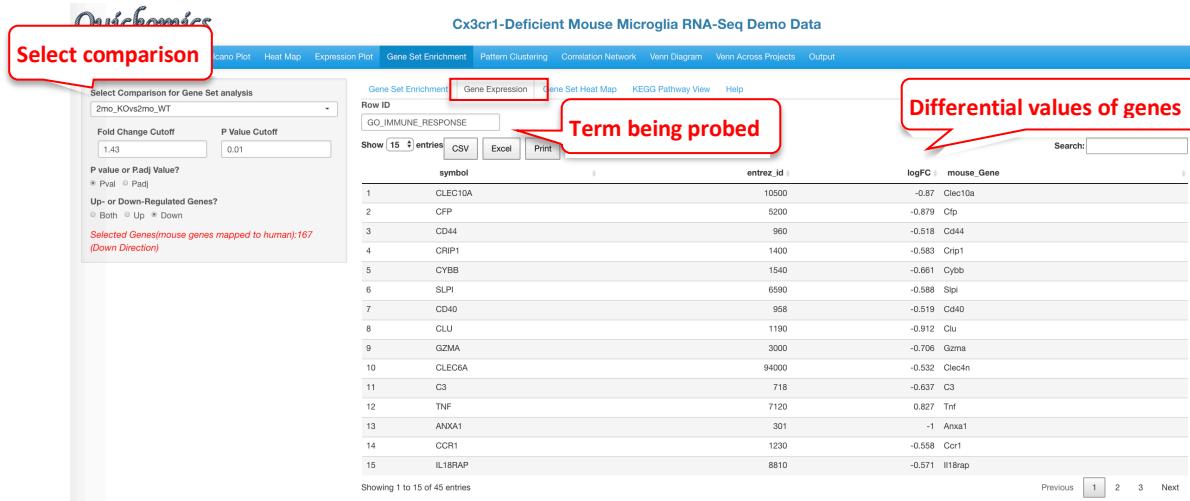
ID	Rank	p.value	p.adj	DeGeneNum	UpGene	DownGene	SetNum
1 GO_DEFENSE_RESPONSE	1	2.71e-11	1.2e-7	46	7	39	1230
2 GO_IMMUNE_RESPONSE	2	9.89e-11	2.19e-7	41	5	36	1100
3 GO_NEUTROPHIL_MIGRATION			2.84e-7	20	1	19	259
4 GO_IMMUNE_SYSTEM_PROCESS			0.0000207	59	7	52	1980
5 GO_INFLAMMATORY_RESPONSE			0.0000158	25	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

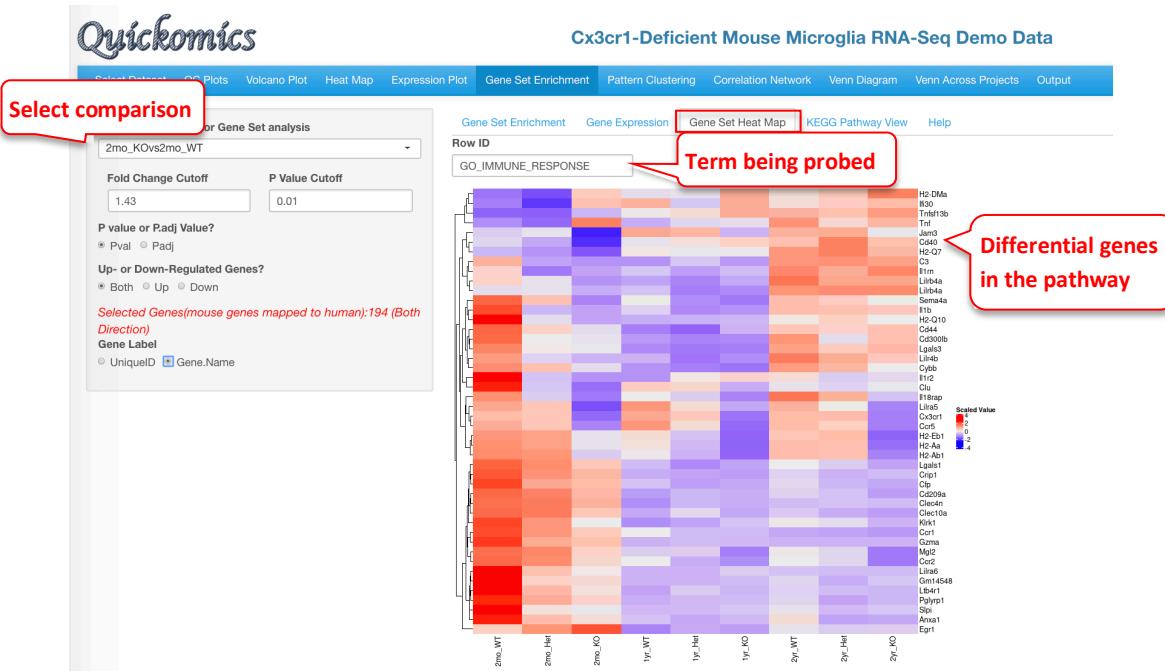
7.2 Gene Expression

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

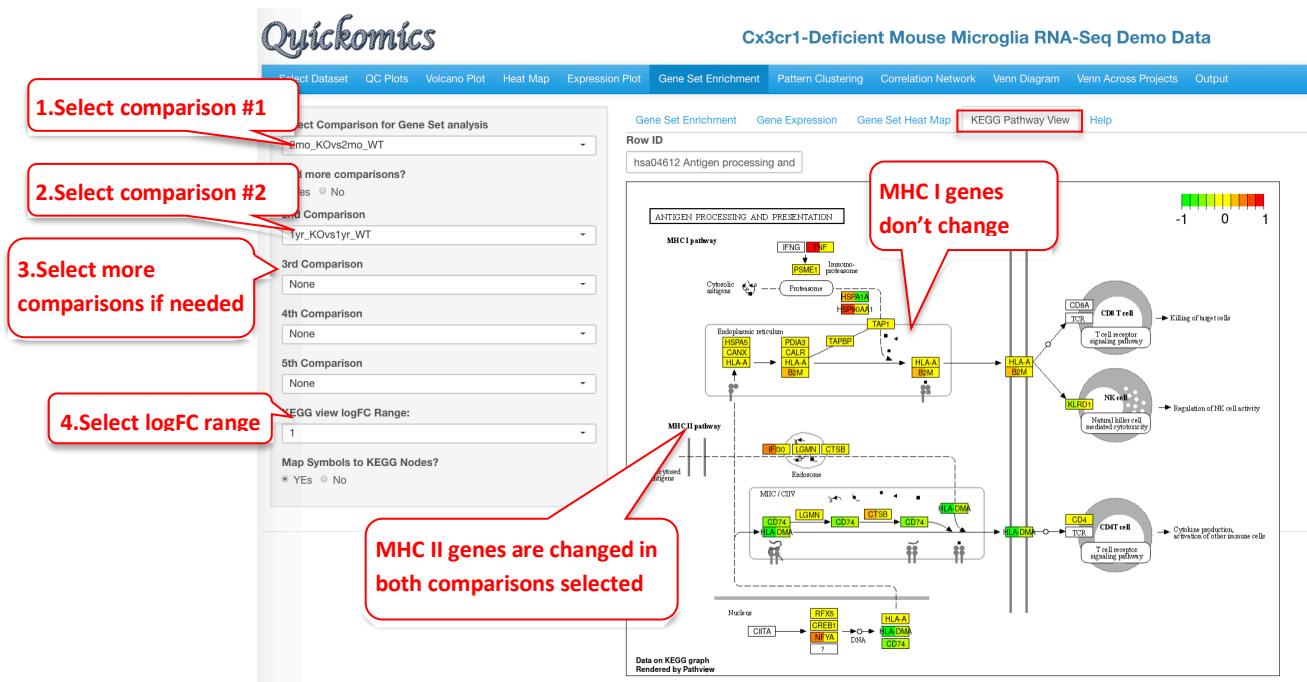
Like the “Gene Expression” tab, Users have the ability to plot expression of the genes in the pathway term selected as a heatmap.



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 don't change, but many genes in the MHC II pathway are down regulated in both the comparisons.

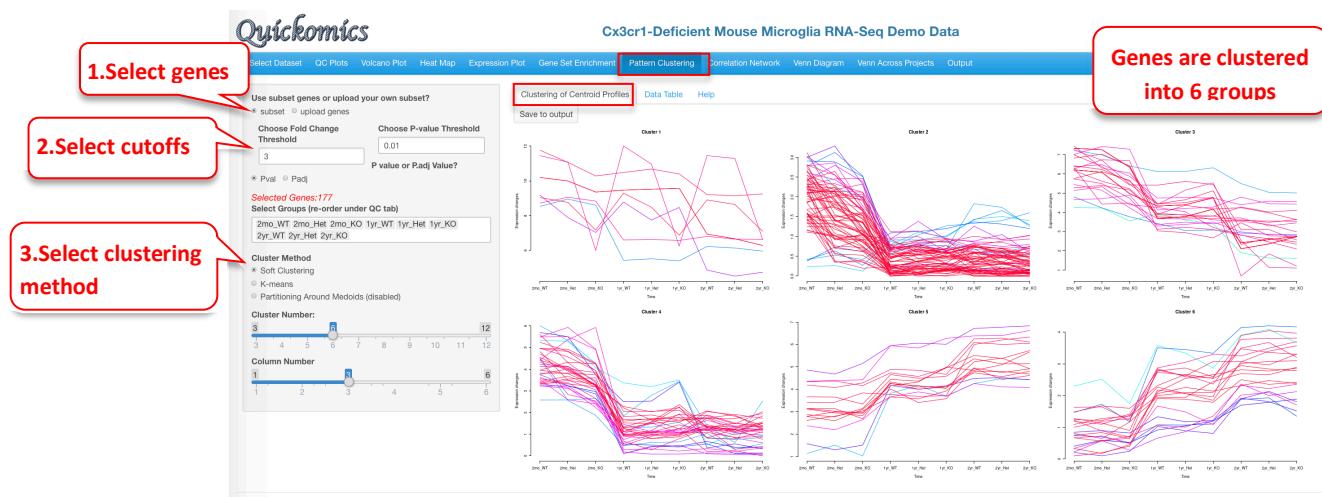


8 Pattern Clustering

This functionality in Quickomics helps cluster genes based on their expression profiles across different groups. There are 3 different methods of clustering the genes: soft (fuzzy) clustering, k-means and partitioning around medoids (PAM).

8.1 Clustering of Centroid Profiles

Users have the ability to either select the DEGs identified before as the list of genes to cluster based off on or upload a custom gene list. DEGs 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 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 clusters and genes identified in the previous tab can be probed in this tab.

The figure shows the 'Data Table' tab of the Quickomics interface. A red box highlights the 'Data Table' tab. A callout box with a red outline points to the 'cluster' column header, with the text 'Cluster numbers as seen in previous tab'. The table lists 10 genes (1-10) with their corresponding ID, expression values for 10 different conditions, and their assigned cluster number. The conditions are grouped by age (2mo, 1yr, 2yr) and genotype (WT, Het, KO). The 'cluster' column shows values 4, 4, 4, 6, 2, 2, 2, 4, 4, 5 respectively. At the bottom, it says 'Showing 1 to 10 of 177 entries' and includes navigation buttons for 'Previous' and 'Next'.

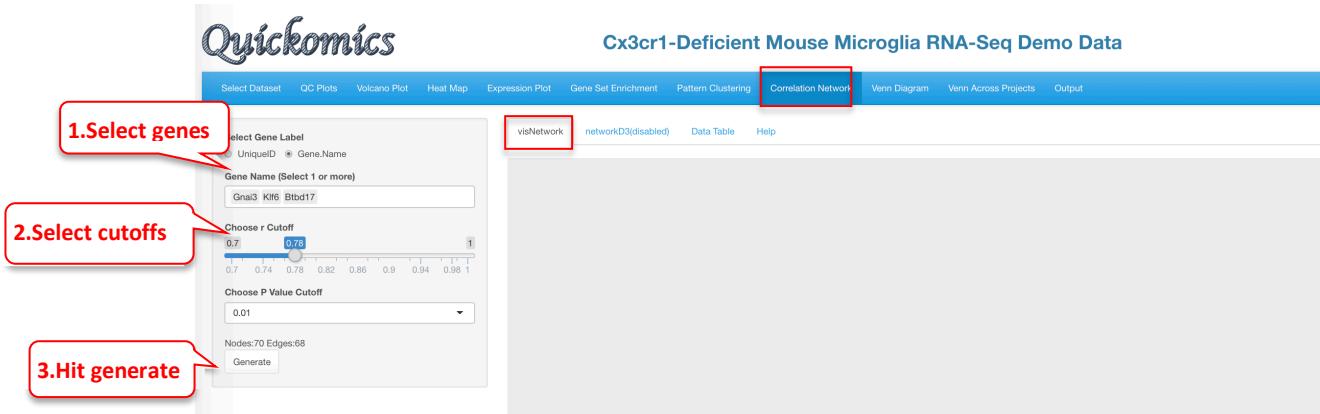
Gene	ID	2mo_WT	2mo_Het	2mo_KO	1yr_WT	1yr_Het	1yr_KO	2yr_WT	2yr_Het	2yr_KO	cluster
1	ENSMUSG000000000318	3.33	3.16	2.23	1.02	1.53	1.04	1.38	1.07	0.943	4
2	ENSMUSG00000001020	4.55	3.92	3.26	1.11	1.58	2.37	1.34	1.24	1.85	4
3	ENSMUSG000000001025	5.6	4.31	3.39	1.12	1.56	1.91	1.85	1.48	2.05	4
4	ENSMUSG000000001120	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

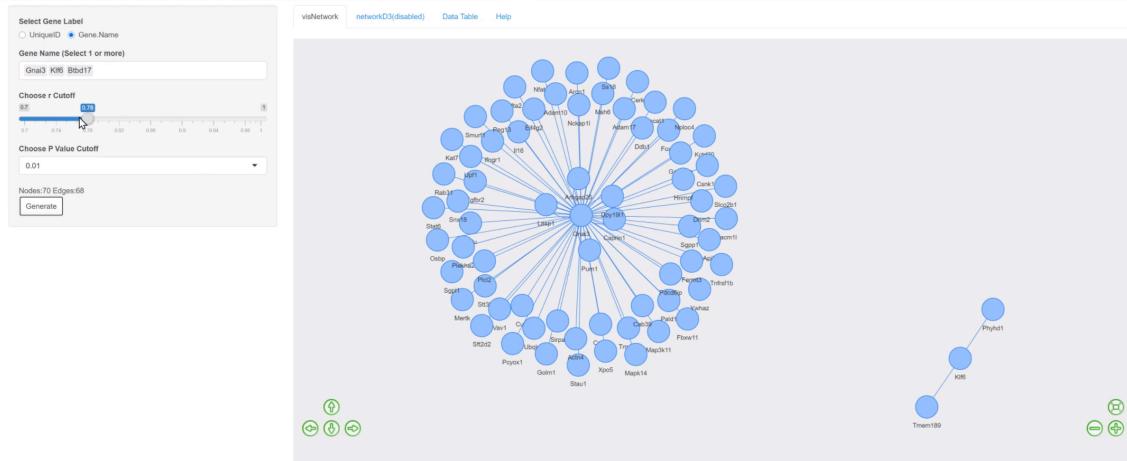
This module helps build co-expression networks based on protein-protein correlation. Users have the ability to enter 1 or more genes to identify if their expression is correlated.

9.1 visNetwork

This uses the R package visNetwork to calculate expression correlation of genes and build the visualization. Users have the ability to choose the R cutoff for the correlation statistics. 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

Data Table Help

Search:

56 entries

	from	to	cor	p	direction
1	ENSMUSG000000000001	ENSMUSG00000001741	0.81	0	1
2	ENSMUSG000000000001	ENSMUSG00000002147	0.79	0	1
3	ENSMUSG000000000001	ENSMUSG00000003154	0.81	0	1
4	ENSMUSG000000000001	ENSMUSG00000004054	0.8	0	1
5	ENSMUSG000000000001	ENSMUSG00000004364	0.79	0	1
6	ENSMUSG000000000001	ENSMUSG00000005312	0.84	0	1
7	ENSMUSG000000000001	ENSMUSG00000005370	0.8	0	1
8	ENSMUSG000000000001	ENSMUSG00000005610	0.86	0	1
9	ENSMUSG000000000001	ENSMUSG00000005936	0.79	0	1
10	ENSMUSG000000000001	ENSMUSG00000014361	0.79	0	1
11	ENSMUSG000000000001	ENSMUSG00000020009	0.79	0	1
12	ENSMUSG000000000001	ENSMUSG00000020092	0.81	0	1
13	ENSMUSG000000000001	ENSMUSG00000020097	0.79	0	1
14	ENSMUSG000000000001	ENSMUSG00000020271	0.84	0	1
15	ENSMUSG000000000001	ENSMUSG000000201054	0.79	0	1

Showing 1 to 15 of 68 entries

Previous 1 2 3 4 5 Next

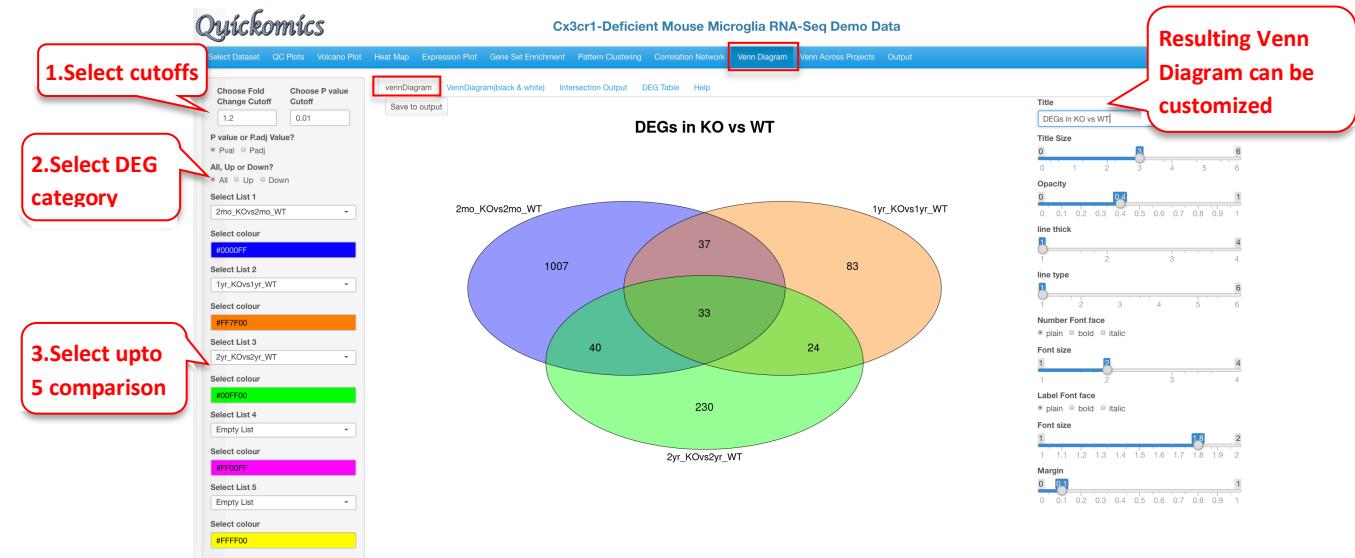
Genes identified as correlated

10 Venn Diagram

This module in Quickomics lets users identify DEGs common or unique to different comparisons within one project. Users have the option to either visualize these DEGs as a highly customisable Venn Diagram or download a table with the gene names.

10.1 vennDiagram

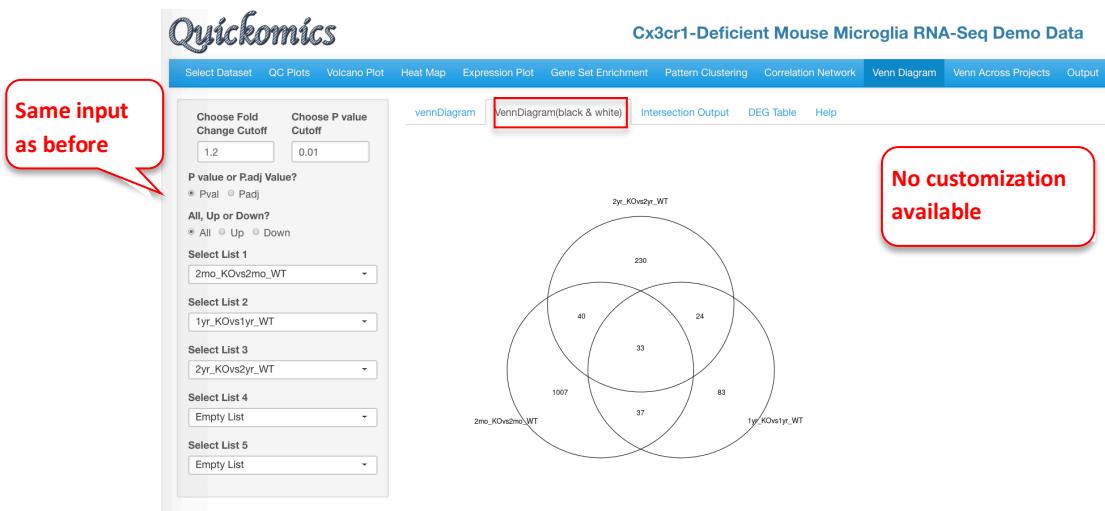
To view the DEGs as a Venn Diagram, Users have the option to see the intersection of DEGs for upto 5 different comparisons. They can also use all DEGs or just the ones that are Up or Down regulated.



In this example there were 33 DEGs that were identified in the three comparisons probed here.

10.2 VennDiagram(black & white)

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

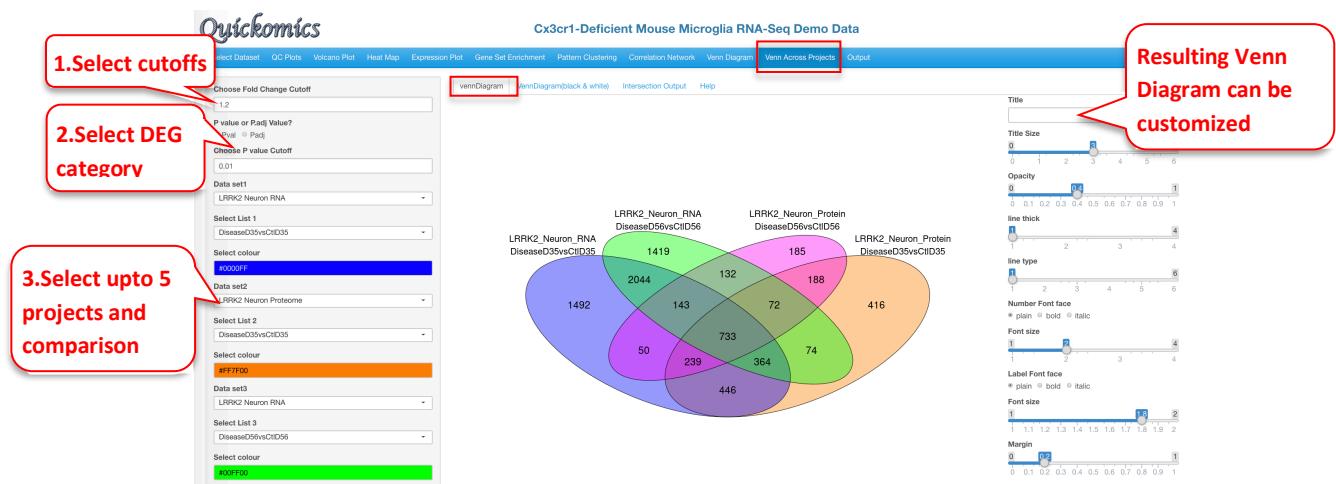


11 Venn Across Projects

This functionality in Quickomics is the only one that lets users probe more than one project/dataset together. The DEGs identified in different comparisons across projects/datasets loaded in the interface can be compared here. Some good use cases are for identifying common DEGs present in RNAseq vs Mass-spec Proteomics of the same samples.

11.1 vennDiagram

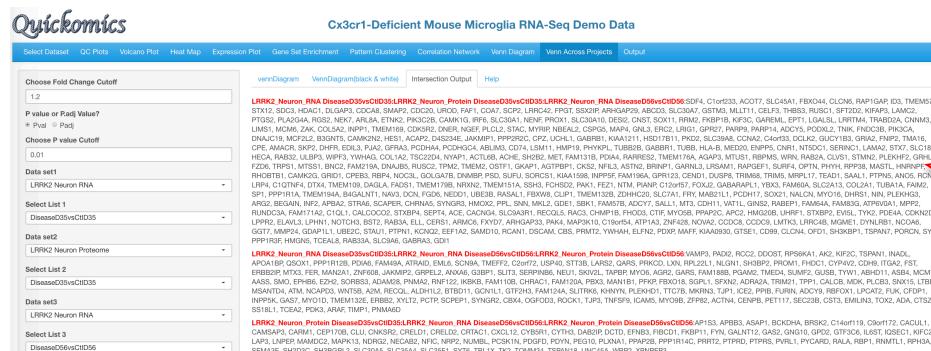
To illustrate this feature, we compare the two RNAseq and two Proteomic comparisons from Connor-Robson et al., 2019. We load 2 comparisons from each dataset/project and view the number of DEGs that are common and different between the 4 comparisons.



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

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.

Select width and height of plots to save

Select all the plots available to save

Plots can be downloaded as PDF or SVG