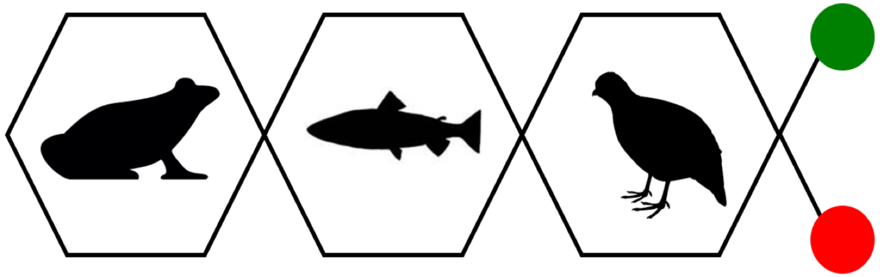
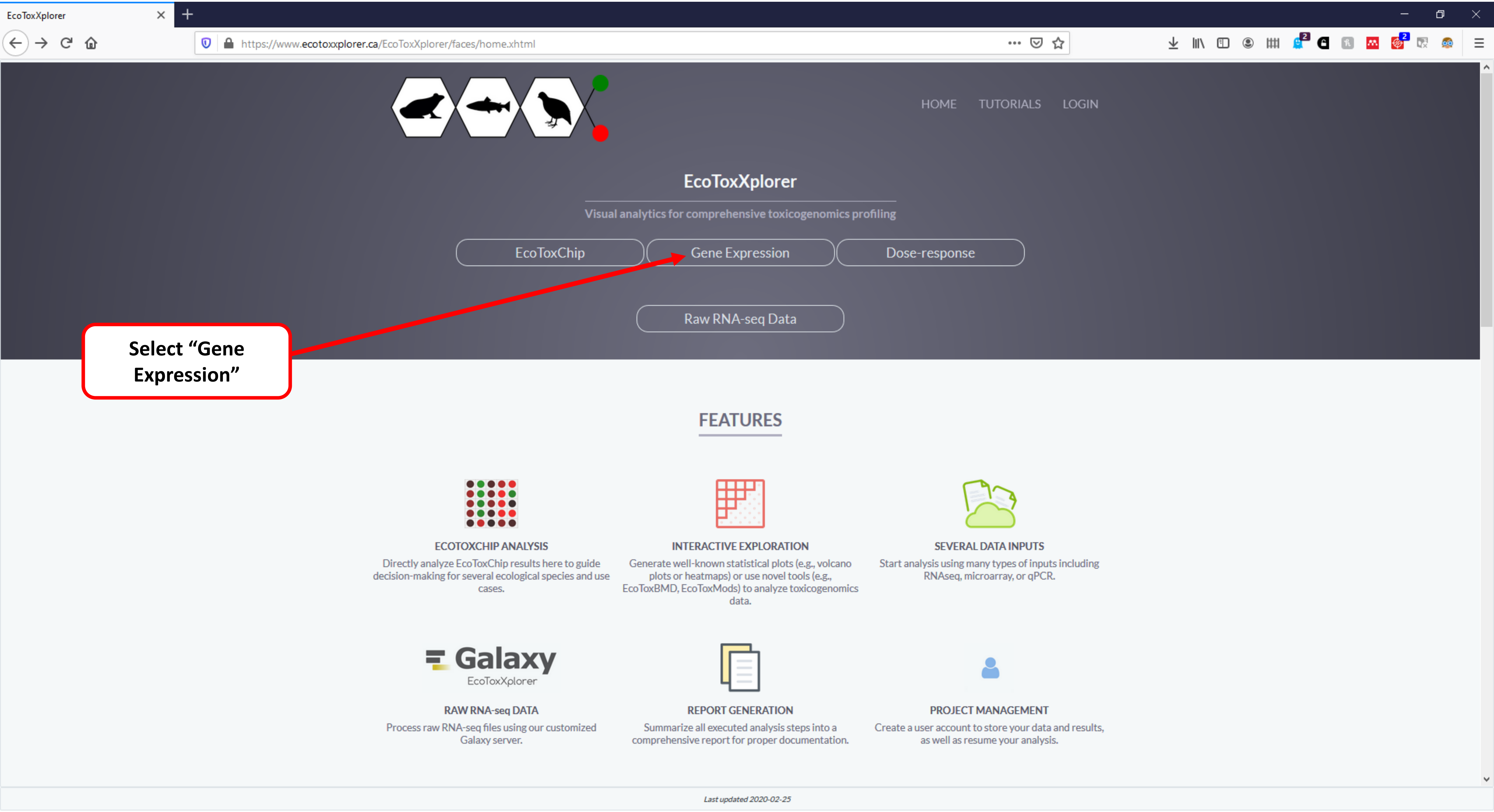


Tutorial #2

RNASeq Gene Expression Module



Select “Gene Expression”

EcoToxXplorer

Visual analytics for comprehensive toxicogenomics profiling

FEATURES

ECOTOXCHIP ANALYSIS

Directly analyze EcoToxChip results here to guide decision-making for several ecological species and use cases.

INTERACTIVE EXPLORATION

Generate well-known statistical plots (e.g., volcano plots or heatmaps) or use novel tools (e.g., EcoToxBMD, EcoToxMods) to analyze toxicogenomics data.

SEVERAL DATA INPUTS

Start analysis using many types of inputs including RNAseq, microarray, or qPCR.

RAW RNA-seq DATA

Process raw RNA-seq files using our customized Galaxy server.

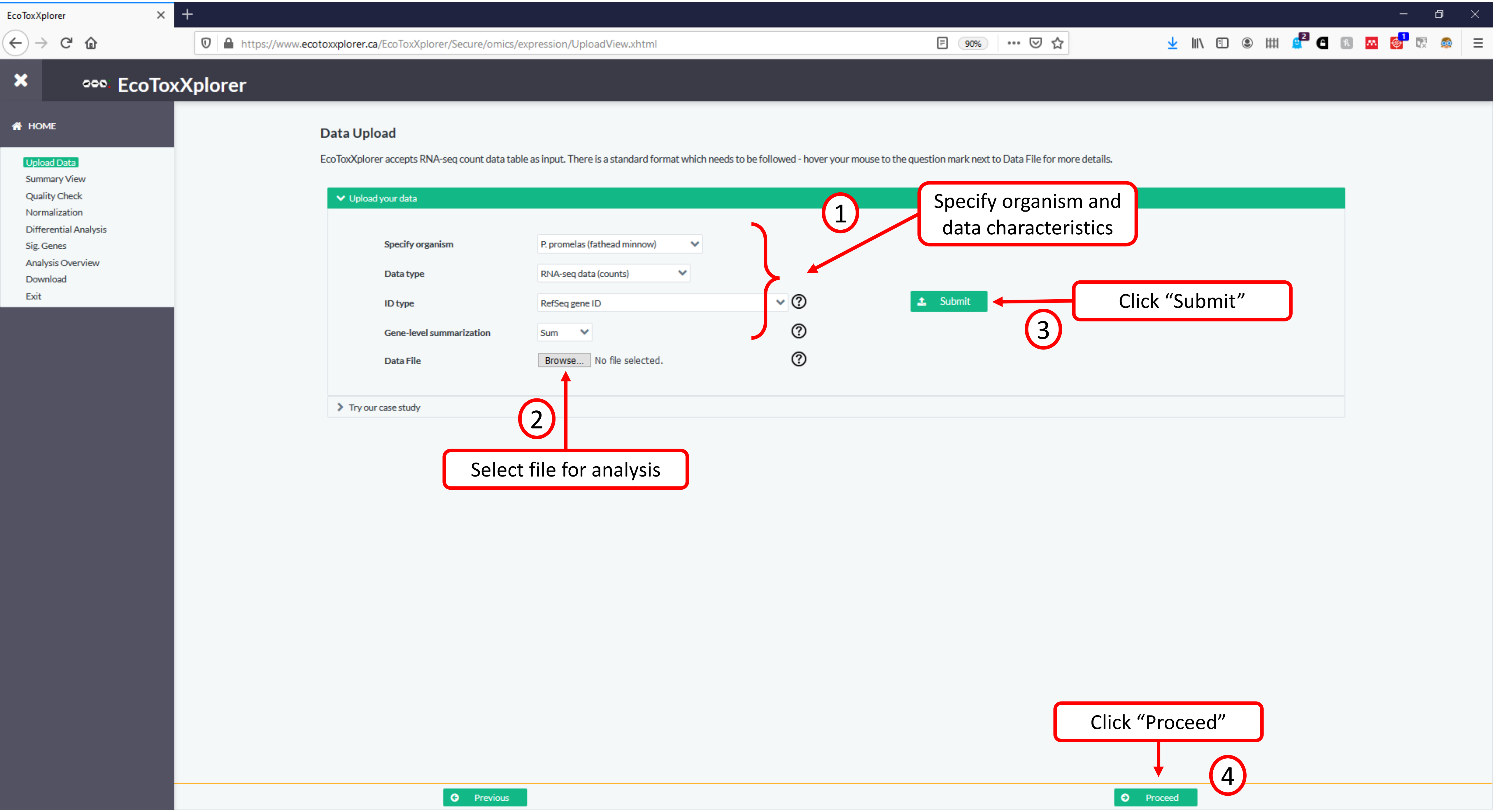
REPORT GENERATION

Summarize all executed analysis steps into a comprehensive report for proper documentation.

PROJECT MANAGEMENT

Create a user account to store your data and results, as well as resume your analysis.

Last updated 2020-02-25



Data Upload

EcoToxXplorer accepts RNA-seq count data table as input. There is a standard format which needs to be followed - hover your mouse to the question mark next to Data File for more details.

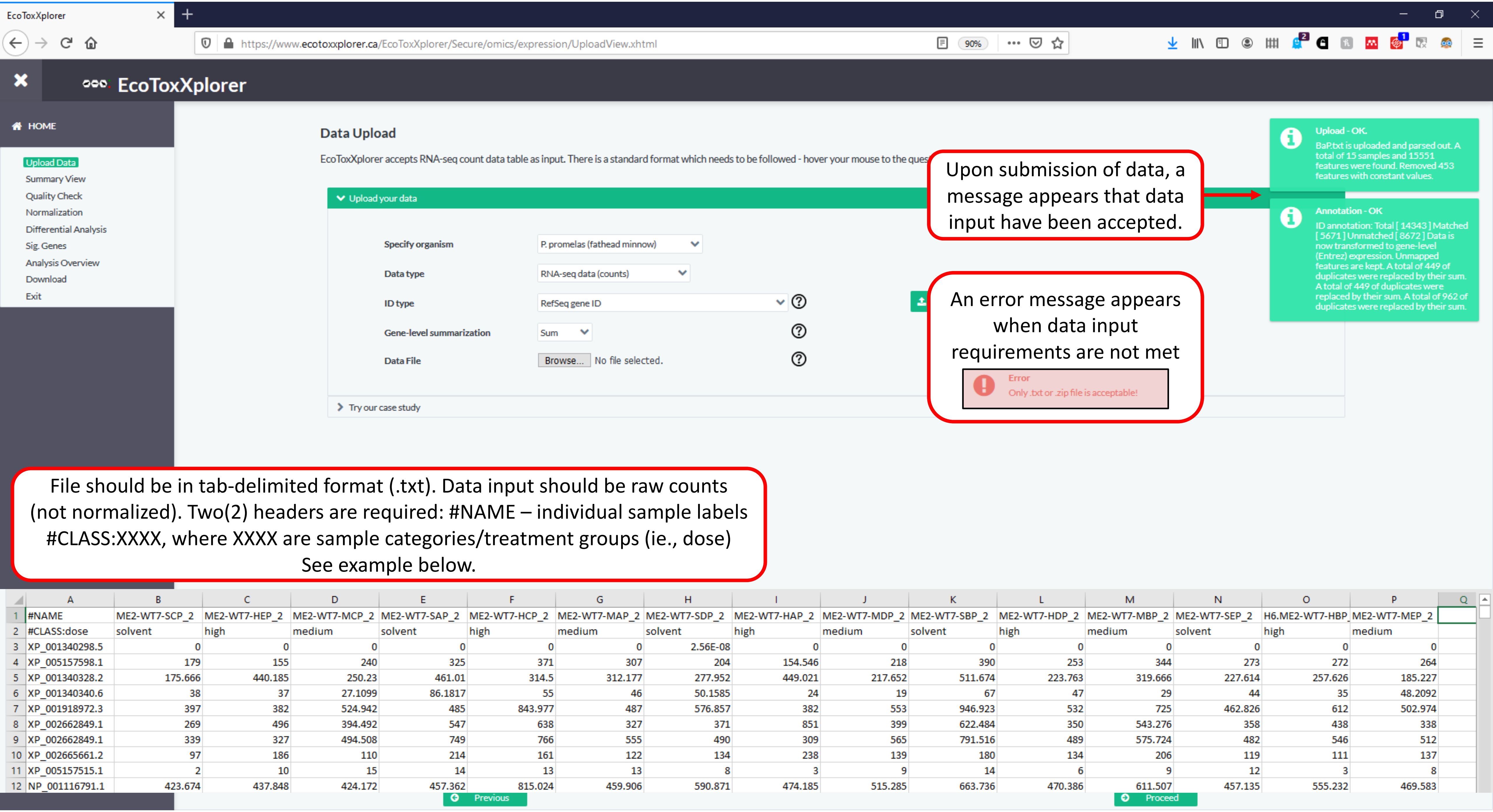
Specify organism	P. promelas (fathead minnow)
Data type	RNA-seq data (counts)
ID type	RefSeq gene ID
Gene-level summarization	Sum
Data File	No file selected.

1 Specify organism and data characteristics

2 Select file for analysis

3 Click “Submit”

4 Click “Proceed”



Data Upload

EcoToxXplorer accepts RNA-seq count data table as input. There is a standard format which needs to be followed - hover your mouse to the ques

Upload your data

Specify organism	P. promelas (fathead minnow)
Data type	RNA-seq data (counts)
ID type	RefSeq gene ID
Gene-level summarization	Sum
Data File	Browse... No file selected.

Try our case study

Upload - OK.

BaP.txt is uploaded and parsed out. A total of 15 samples and 15551 features were found. Removed 453 features with constant values.

Annotation - OK

ID annotation: Total [14343] Matched [5671] Unmatched [8672] Data is now transformed to gene-level (Entrez) expression. Unmapped features are kept. A total of 449 of duplicates were replaced by their sum. A total of 449 of duplicates were replaced by their sum. A total of 962 of duplicates were replaced by their sum.

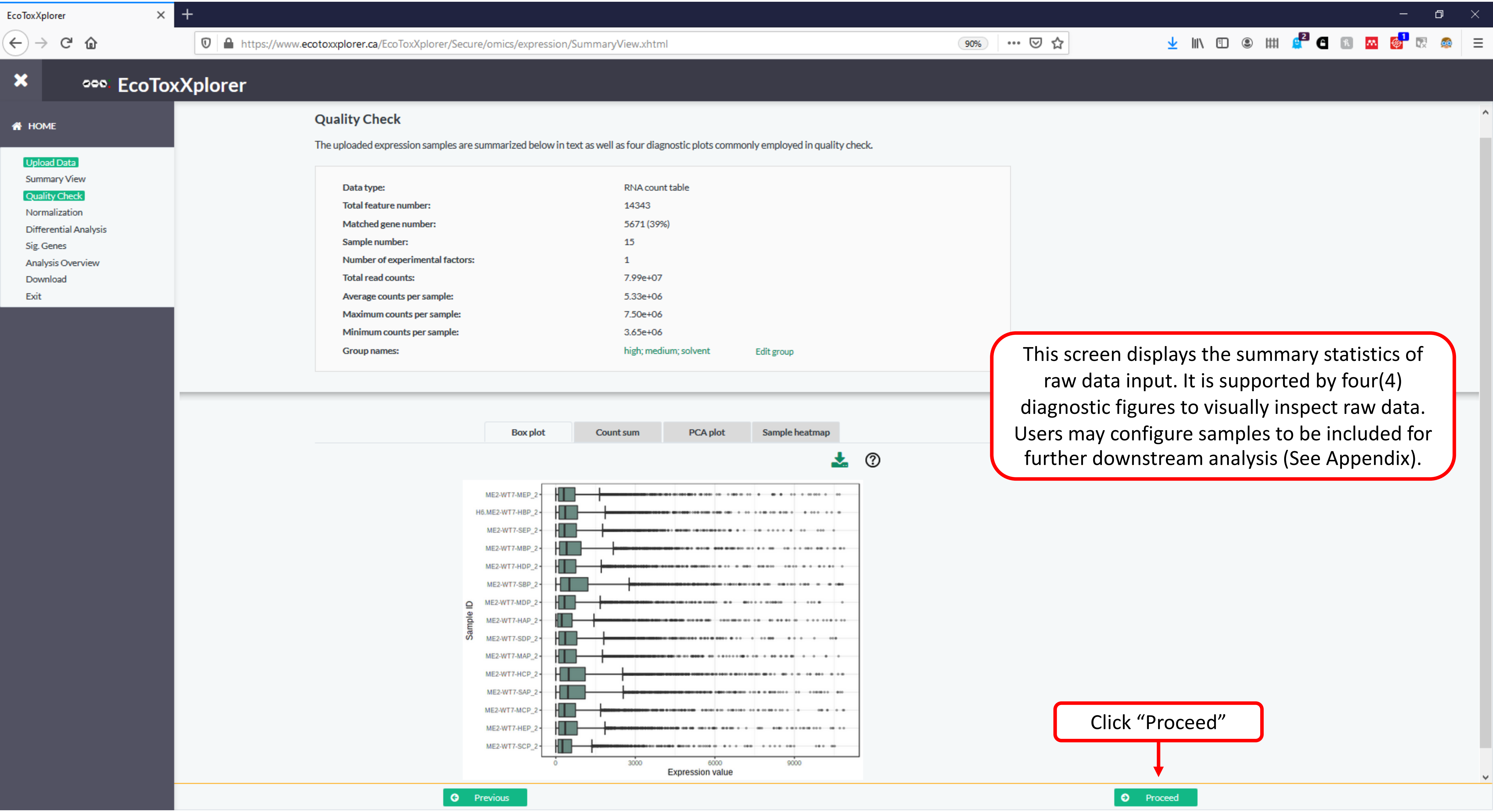
Upon submission of data, a message appears that data input have been accepted.

An error message appears when data input requirements are not met

Error
Only .txt or .zip file is acceptable!

File should be in tab-delimited format (.txt). Data input should be raw counts (not normalized). Two(2) headers are required: #NAME – individual sample labels #CLASS:XXXX, where XXXX are sample categories/treatment groups (ie., dose) See example below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	#NAME	ME2-WT7-SCP_2	ME2-WT7-HEP_2	ME2-WT7-MCP_2	ME2-WT7-SAP_2	ME2-WT7-HCP_2	ME2-WT7-MAP_2	ME2-WT7-SDP_2	ME2-WT7-HAP_2	ME2-WT7-MDP_2	ME2-WT7-SBP_2	ME2-WT7-HDP_2	ME2-WT7-MBP_2	ME2-WT7-SEP_2	H6.ME2-WT7-HBP	ME2-WT7-MEP_2
2	#CLASS:dose	solvent	high	medium	solvent	high	medium	solvent	high	medium	solvent	high	medium	solvent	high	medium
3	XP_001340298.5	0	0	0	0	0	0	2.56E-08	0	0	0	0	0	0	0	0
4	XP_005157598.1	179	155	240	325	371	307	204	154.546	218	390	253	344	273	272	264
5	XP_001340328.2	175.666	440.185	250.23	461.01	314.5	312.177	277.952	449.021	217.652	511.674	223.763	319.666	227.614	257.626	185.227
6	XP_001340340.6	38	37	27.1099	86.1817	55	46	50.1585	24	19	67	47	29	44	35	48.2092
7	XP_001918972.3	397	382	524.942	485	843.977	487	576.857	382	553	946.923	532	725	462.826	612	502.974
8	XP_002662849.1	269	496	394.492	547	638	327	371	851	399	622.484	350	543.276	358	438	338
9	XP_002662849.1	339	327	494.508	749	766	555	490	309	565	791.516	489	575.724	482	546	512
10	XP_002665661.2	97	186	110	214	161	122	134	238	139	180	134	206	119	111	137
11	XP_005157515.1	2	10	15	14	13	13	8	3	9	14	6	9	12	3	8
12	NP_001116791.1	423.674	437.848	424.172	457.362	815.024	459.906	590.871	474.185	515.285	663.736	470.386	611.507	457.135	555.232	469.583



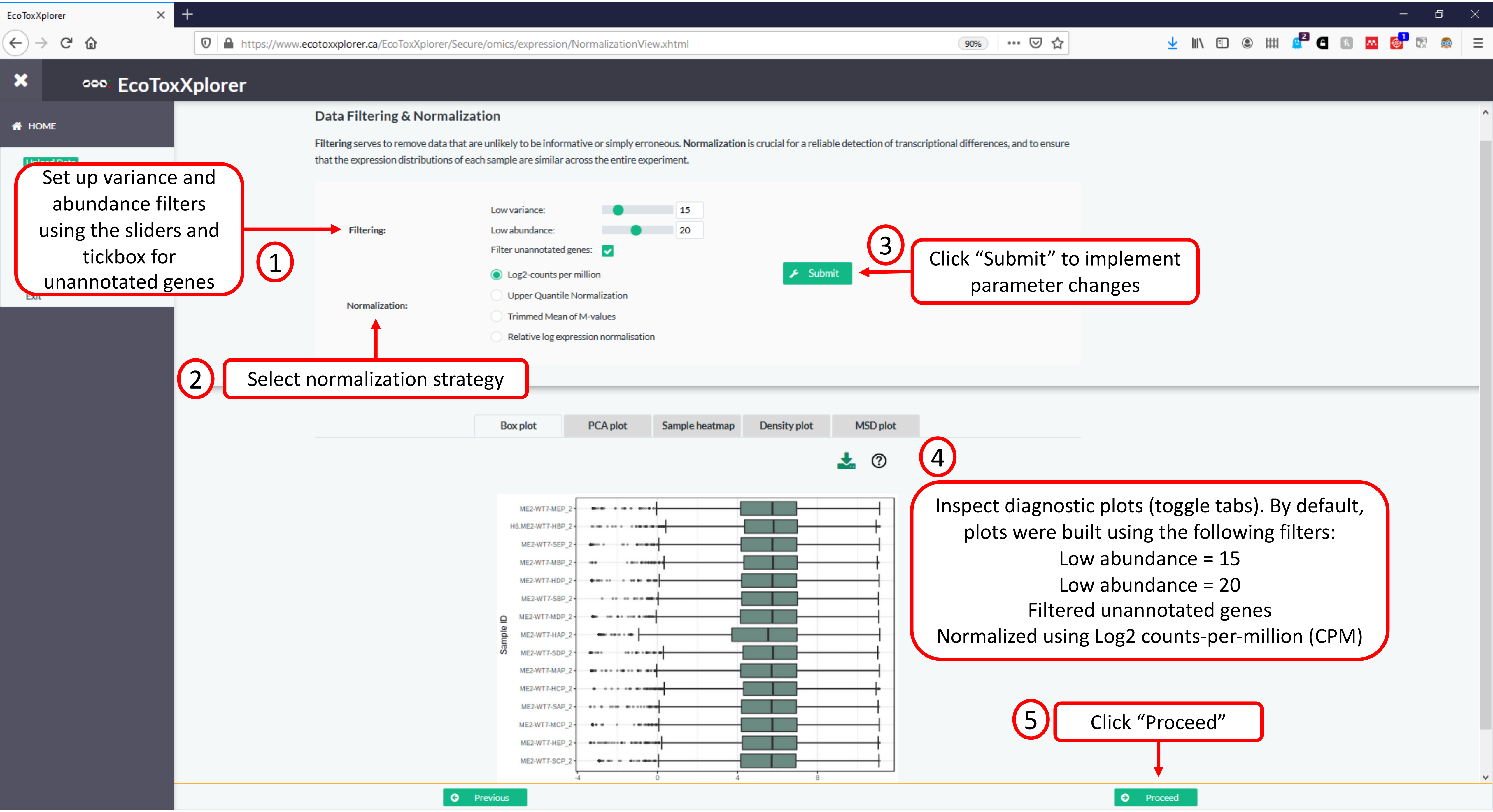
Quality Check

The uploaded expression samples are summarized below in text as well as four diagnostic plots commonly employed in quality check.

Data type:	RNA count table
Total feature number:	14343
Matched gene number:	5671 (39%)
Sample number:	15
Number of experimental factors:	1
Total read counts:	7.99e+07
Average counts per sample:	5.33e+06
Maximum counts per sample:	7.50e+06
Minimum counts per sample:	3.65e+06
Group names:	high; medium; solvent Edit group

This screen displays the summary statistics of raw data input. It is supported by four(4) diagnostic figures to visually inspect raw data. Users may configure samples to be included for further downstream analysis (See Appendix).

Click “Proceed”



Data Filtering & Normalization

Filtering serves to remove data that are unlikely to be informative or simply erroneous. **Normalization** is crucial for a reliable detection of transcriptional differences, and to ensure that the expression distributions of each sample are similar across the entire experiment.

Filtering:

Low variance: 15

Low abundance: 20

Filter unannotated genes: ☑

Normalization:

● Log2-counts per million

○ Upper Quantile Normalization

○ Trimmed Mean of M-values

○ Relative log expression normalisation

Submit

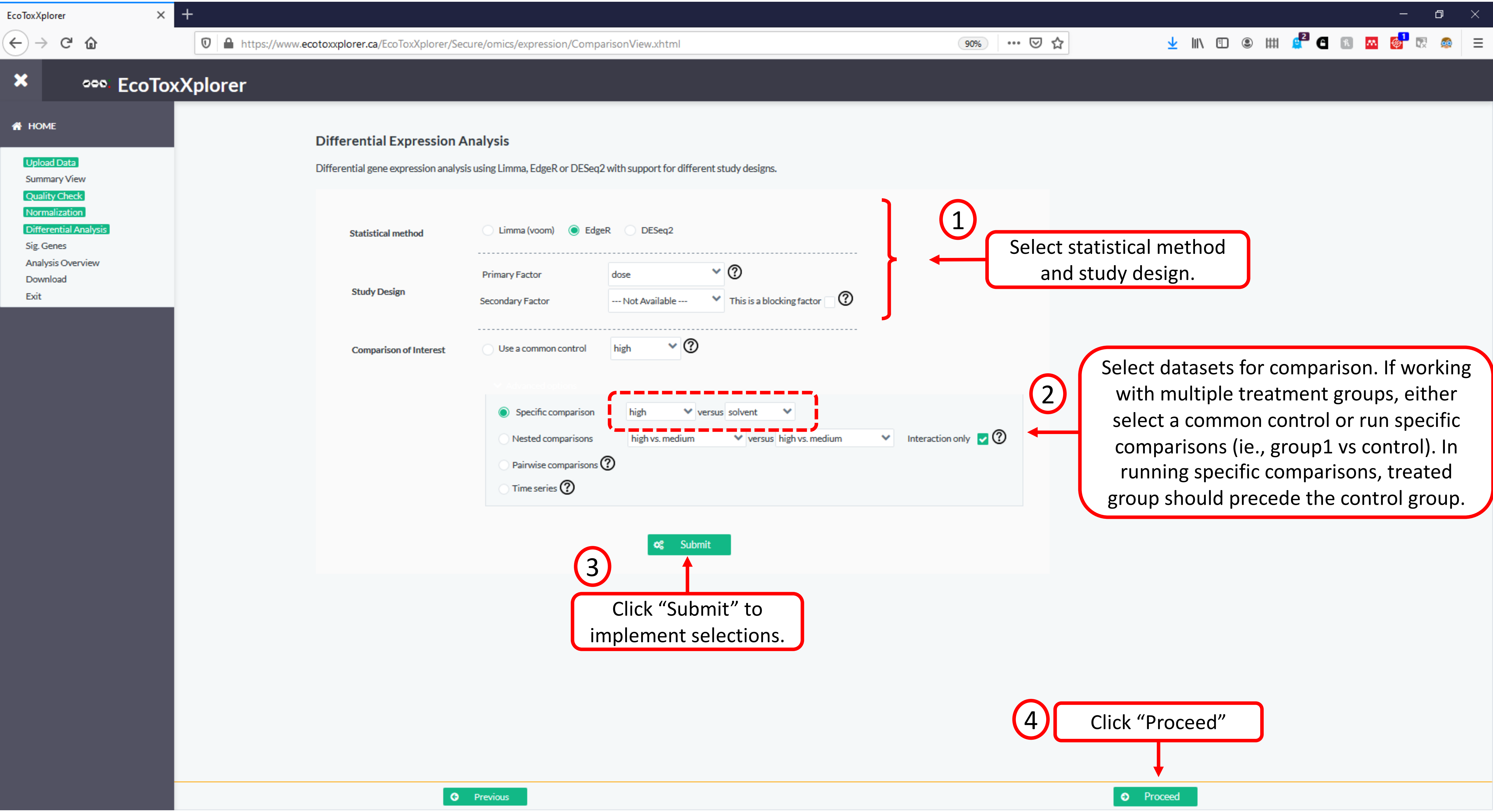
① Set up variance and abundance filters using the sliders and tickbox for unannotated genes

② Select normalization strategy

③ Click “Submit” to implement parameter changes

④ Inspect diagnostic plots (toggle tabs). By default, plots were built using the following filters:
Low abundance = 15
Low abundance = 20
Filtered unannotated genes
Normalized using Log2 counts-per-million (CPM)

⑤ Click “Proceed”



Differential Expression Analysis

Differential gene expression analysis using Limma, EdgeR or DESeq2 with support for different study designs.

Statistical method ○ Limma (voom) ● EdgeR ○ DESeq2

Study Design
Primary Factor: dose
Secondary Factor: --- Not Available --- This is a blocking factor ☐

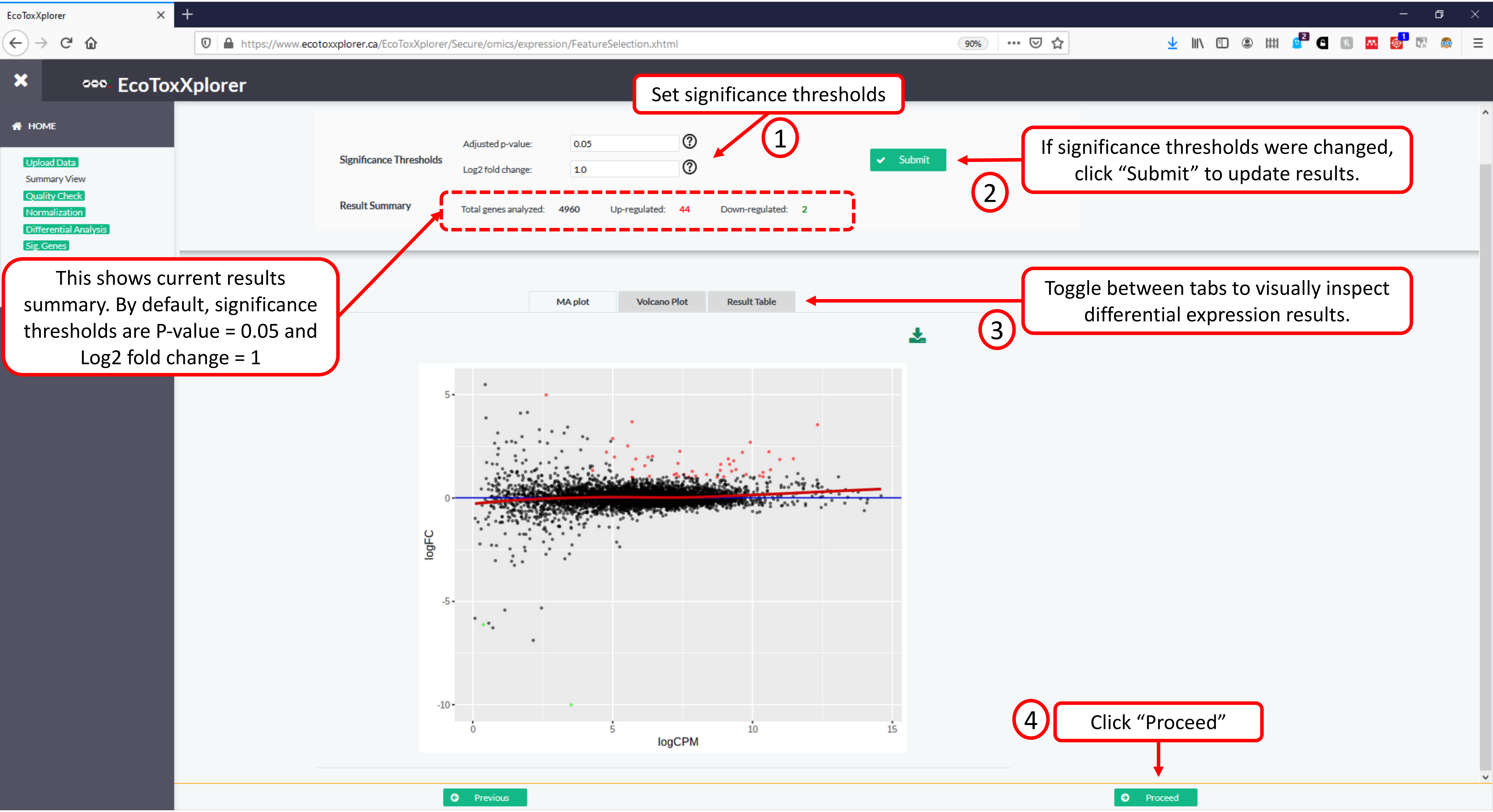
Comparison of Interest
○ Use a common control: high
● Specific comparison: high versus solvent
○ Nested comparisons: high vs. medium versus high vs. medium Interaction only ☑
○ Pairwise comparisons
○ Time series

1. Select statistical method and study design.

2. Select datasets for comparison. If working with multiple treatment groups, either select a common control or run specific comparisons (ie., group1 vs control). In running specific comparisons, treated group should precede the control group.

3. Click “Submit” to implement selections.

4. Click “Proceed”



Significance Thresholds

Adjusted p-value: 0.05

Log2 fold change: 1.0

Result Summary

Total genes analyzed: 4960 Up-regulated: 44 Down-regulated: 2

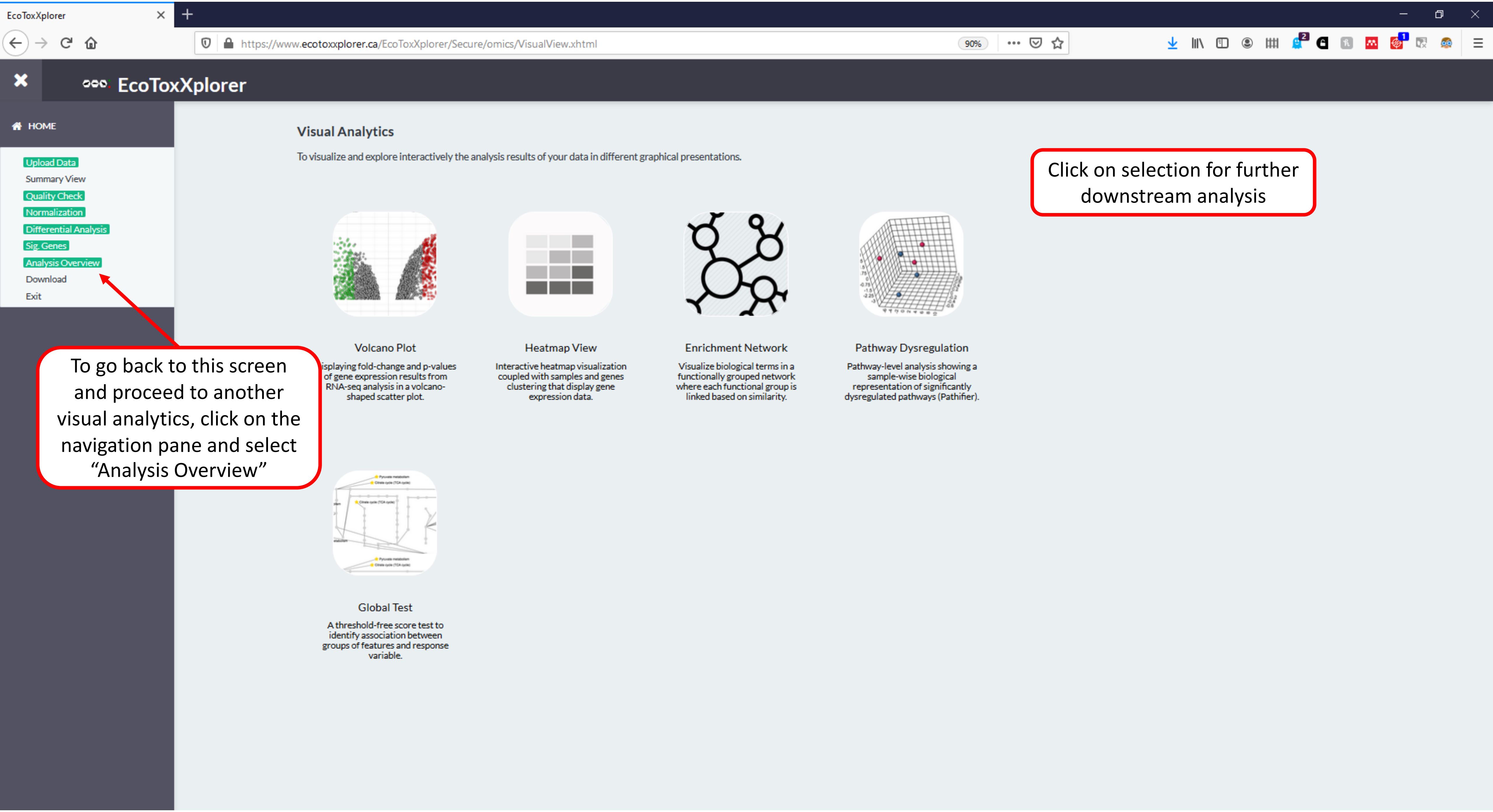
① Set significance thresholds

② If significance thresholds were changed, click “Submit” to update results.

This shows current results summary. By default, significance thresholds are P-value = 0.05 and Log2 fold change = 1

③ Toggle between tabs to visually inspect differential expression results.

④ Click “Proceed”



Visual Analytics

To visualize and explore interactively the analysis results of your data in different graphical presentations.

Volcano Plot

isplaying fold-change and p-values of gene expression results from RNA-seq analysis in a volcano-shaped scatter plot.

Heatmap View

Interactive heatmap visualization coupled with samples and genes clustering that display gene expression data.

Enrichment Network

Visualize biological terms in a functionally grouped network where each functional group is linked based on similarity.

Pathway Dysregulation

Pathway-level analysis showing a sample-wise biological representation of significantly dysregulated pathways (Pathifier).

Global Test

A threshold-free score test to identify association between groups of features and response variable.

Click on selection for further downstream analysis

To go back to this screen and proceed to another visual analytics, click on the navigation pane and select “Analysis Overview”

Click navigation pane and select "Analysis Overview" to go back to Visual Analytics main screen.

Gene Set Enrichment

Query: Sig. All

Database: GO:BP

Submit

☐	Pathway	Hits	P-value
☑	response to oxidative stress	3	0.0000366
☐	sphingolipid biosynthetic proce	1	0.00534
☐	superoxide metabolic process	1	0.0186
☐	protein complex assembly	1	0.0212
☐	response to UV	1	0.0238
☐	proteolysis	4	0.0308
☐	response to xenobiotic stimulus	1	0.0317
☐	negative regulation of angiogen	1	0.0394
☐	heme biosynthetic process	1	0.042
☐	response to stress	1	0.042

Page 1 of 1

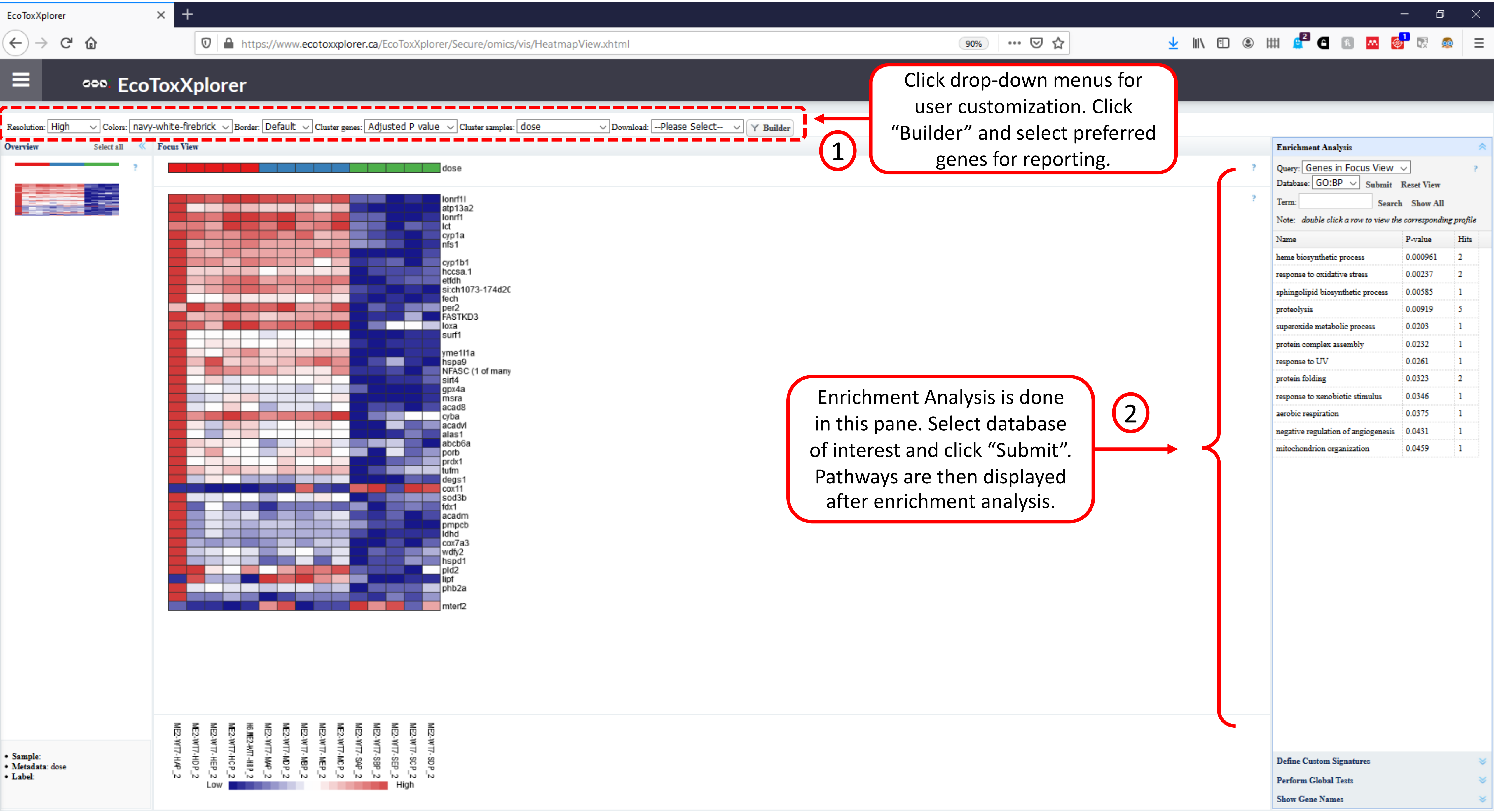
① Select a pathway to highlight (blue) associated gene(s) in the volcano plot.

Gene View

Interactive Volcano Plot

Double-click on a point to view, or select an area to explore (mouse-drag).

② Select a gene to display response across treatment groups



Resolution: High
Colors: navy-white-firebrick
Border: Default
Cluster genes: Adjusted P value
Cluster samples: dose
Download: --Please Select--

- **Sample:**
- **Metadata:** dose
- **Label:**

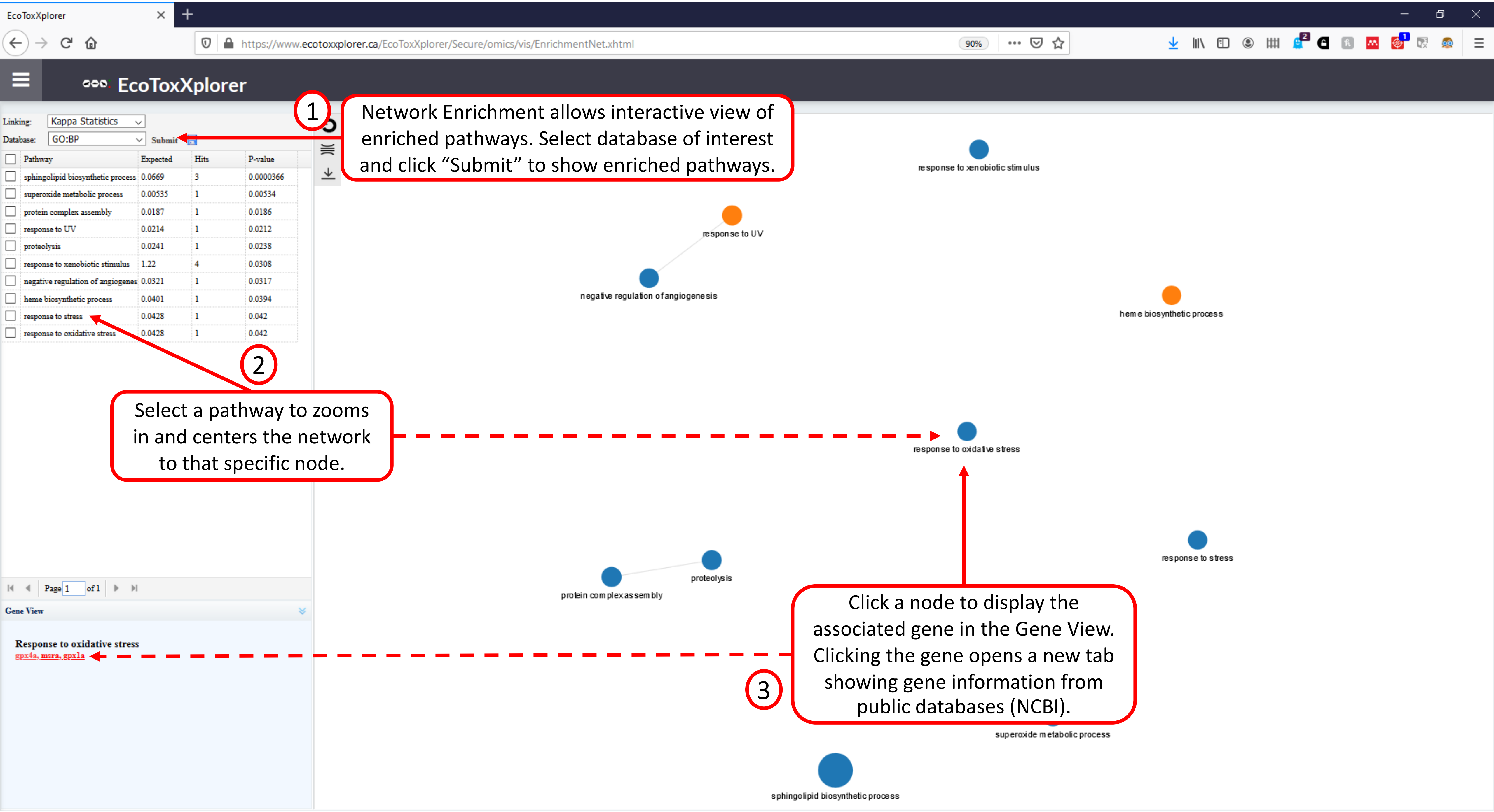
① Click drop-down menus for user customization. Click "Builder" and select preferred genes for reporting.

Enrichment Analysis

Query: Genes in Focus View
Database: GO:BP
Term:
Note: *double click a row to view the corresponding profile*

Name	P-value	Hits
heme biosynthetic process	0.000961	2
response to oxidative stress	0.00237	2
sphingolipid biosynthetic process	0.00585	1
proteolysis	0.00919	5
superoxide metabolic process	0.0203	1
protein complex assembly	0.0232	1
response to UV	0.0261	1
protein folding	0.0323	2
response to xenobiotic stimulus	0.0346	1
aerobic respiration	0.0375	1
negative regulation of angiogenesis	0.0431	1
mitochondrion organization	0.0459	1

② Enrichment Analysis is done in this pane. Select database of interest and click "Submit". Pathways are then displayed after enrichment analysis.



Linking: Kappa Statistics

Database: GO:BP

Submit

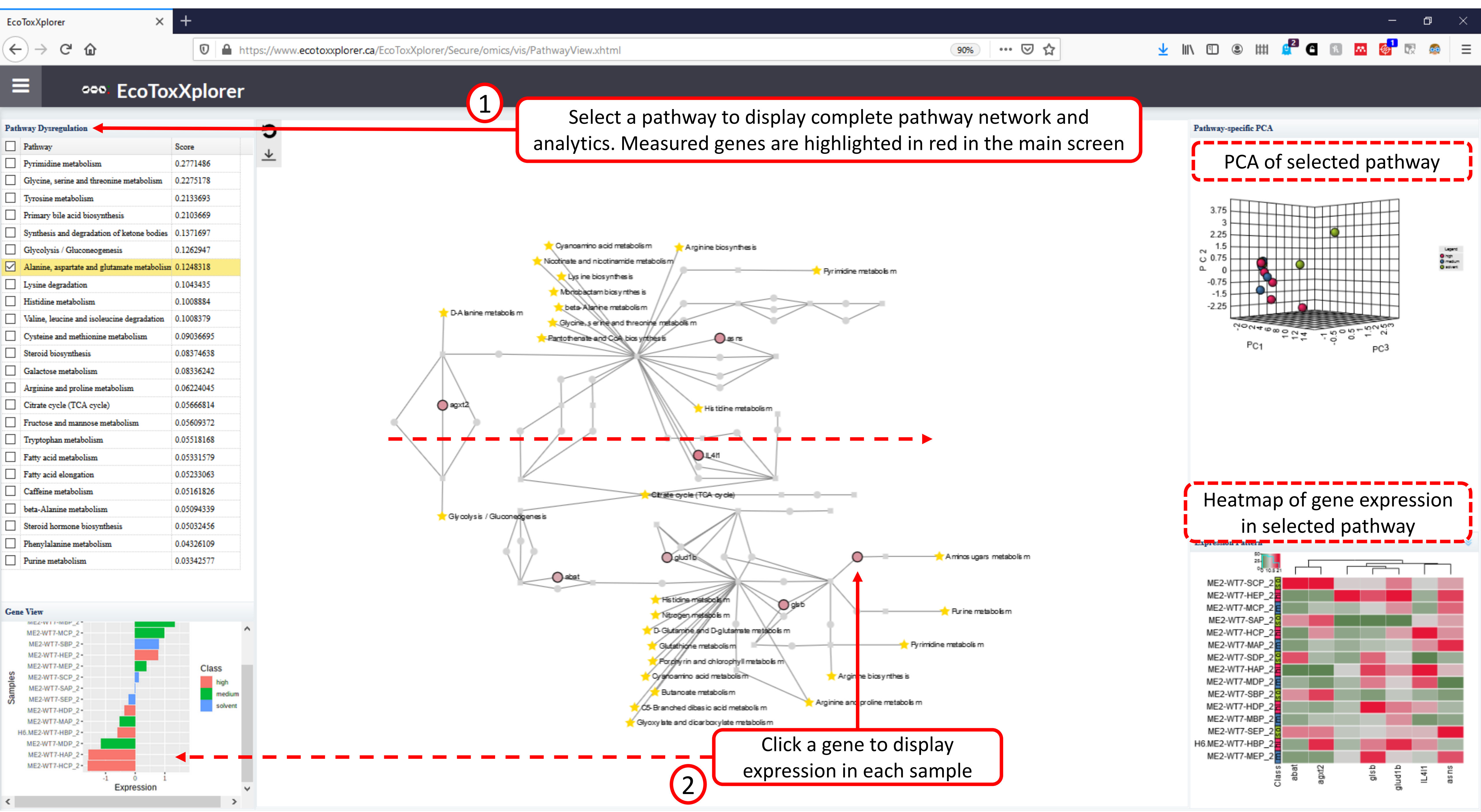
☐	Pathway	Expected	Hits	P-value
☐	sphingolipid biosynthetic process	0.0669	3	0.0000366
☐	superoxide metabolic process	0.00535	1	0.00534
☐	protein complex assembly	0.0187	1	0.0186
☐	response to UV	0.0214	1	0.0212
☐	proteolysis	0.0241	1	0.0238
☐	response to xenobiotic stimulus	1.22	4	0.0308
☐	negative regulation of angiogenes	0.0321	1	0.0317
☐	heme biosynthetic process	0.0401	1	0.0394
☐	response to stress	0.0428	1	0.042
☐	response to oxidative stress	0.0428	1	0.042

Response to oxidative stress
gpx4a, msra, gpx1a

① Network Enrichment allows interactive view of enriched pathways. Select database of interest and click "Submit" to show enriched pathways.

② Select a pathway to zooms in and centers the network to that specific node.

③ Click a node to display the associated gene in the Gene View. Clicking the gene opens a new tab showing gene information from public databases (NCBI).



① Select a pathway to display complete pathway network and analytics. Measured genes are highlighted in red in the main screen

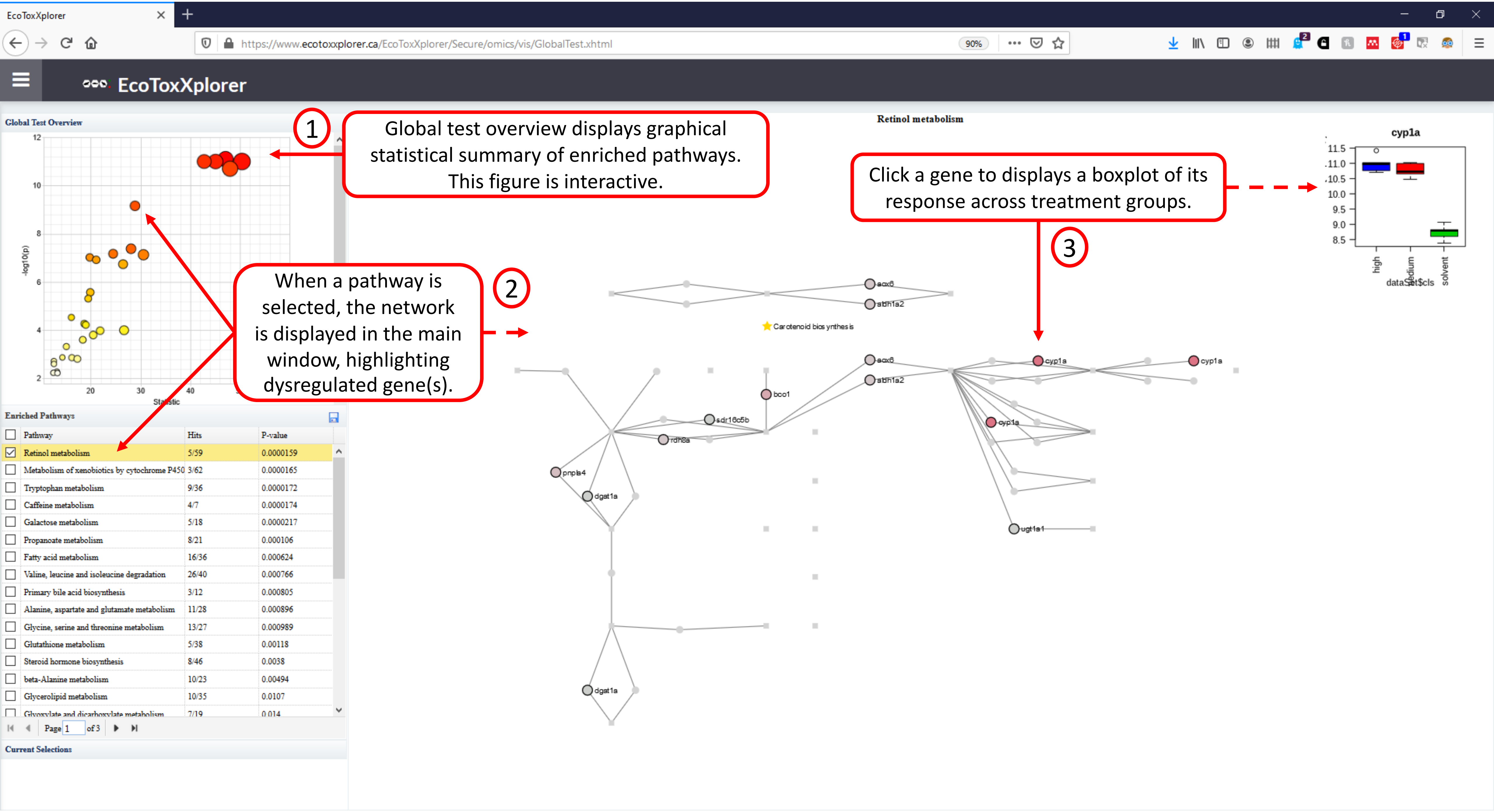
PCA of selected pathway

Heatmap of gene expression in selected pathway

② Click a gene to display expression in each sample

Pathway Dysregulation

☐	Pathway	Score
☐	Pyrimidine metabolism	0.2771486
☐	Glycine, serine and threonine metabolism	0.2275178
☐	Tyrosine metabolism	0.2133693
☐	Primary bile acid biosynthesis	0.2103669
☐	Synthesis and degradation of ketone bodies	0.1371697
☐	Glycolysis / Gluconeogenesis	0.1262947
☑	Alanine, aspartate and glutamate metabolism	0.1248318
☐	Lysine degradation	0.1043435
☐	Histidine metabolism	0.1008884
☐	Valine, leucine and isoleucine degradation	0.1008379
☐	Cysteine and methionine metabolism	0.09036695
☐	Steroid biosynthesis	0.08374638
☐	Galactose metabolism	0.08336242
☐	Arginine and proline metabolism	0.06224045
☐	Citrate cycle (TCA cycle)	0.05666814
☐	Fructose and mannose metabolism	0.05609372
☐	Tryptophan metabolism	0.05518168
☐	Fatty acid metabolism	0.05331579
☐	Fatty acid elongation	0.05233063
☐	Caffeine metabolism	0.05161826
☐	beta-Alanine metabolism	0.05094339
☐	Steroid hormone biosynthesis	0.05032456
☐	Phenylalanine metabolism	0.04326109
☐	Purine metabolism	0.03342577



①

Global test overview displays graphical statistical summary of enriched pathways. This figure is interactive.

②

When a pathway is selected, the network is displayed in the main window, highlighting dysregulated gene(s).

③

Click a gene to displays a boxplot of its response across treatment groups.

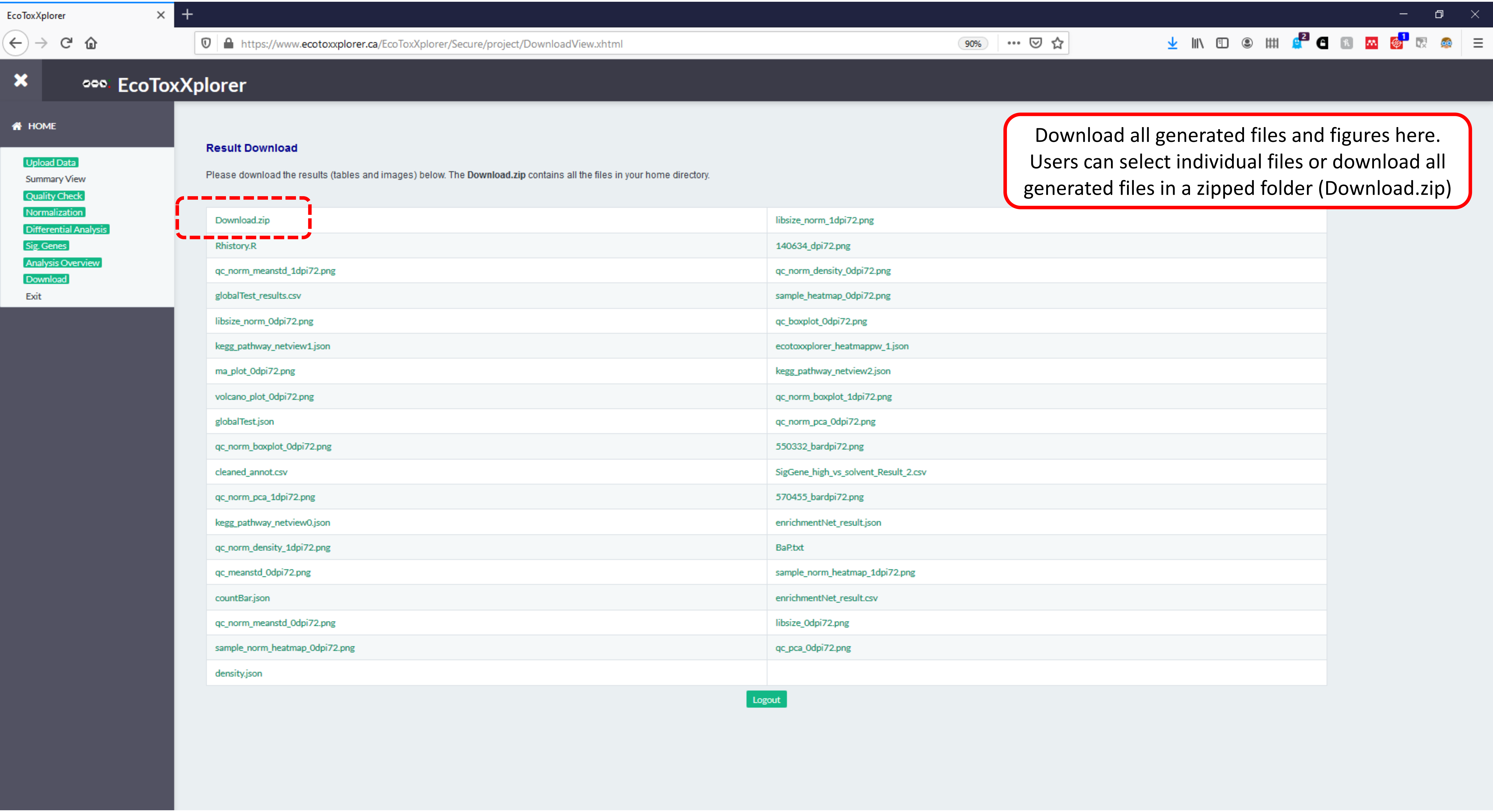
Enriched Pathways

☐	Pathway	Hits	P-value
☑	Retinol metabolism	5/59	0.0000159
☐	Metabolism of xenobiotics by cytochrome P450	3/62	0.0000165
☐	Tryptophan metabolism	9/36	0.0000172
☐	Caffeine metabolism	4/7	0.0000174
☐	Galactose metabolism	5/18	0.0000217
☐	Propanoate metabolism	8/21	0.000106
☐	Fatty acid metabolism	16/36	0.000624
☐	Valine, leucine and isoleucine degradation	26/40	0.000766
☐	Primary bile acid biosynthesis	3/12	0.000805
☐	Alanine, aspartate and glutamate metabolism	11/28	0.000896
☐	Glycine, serine and threonine metabolism	13/27	0.000989
☐	Glutathione metabolism	5/38	0.00118
☐	Steroid hormone biosynthesis	8/46	0.0038
☐	beta-Alanine metabolism	10/23	0.00494
☐	Glycerolipid metabolism	10/35	0.0107
☐	Glyoxylate and dicarboxylate metabolism	7/19	0.014

Page 1 of 3

Current Selections

Download all generated files and figures here. Users can select individual files or download all generated files in a zipped folder (Download.zip)

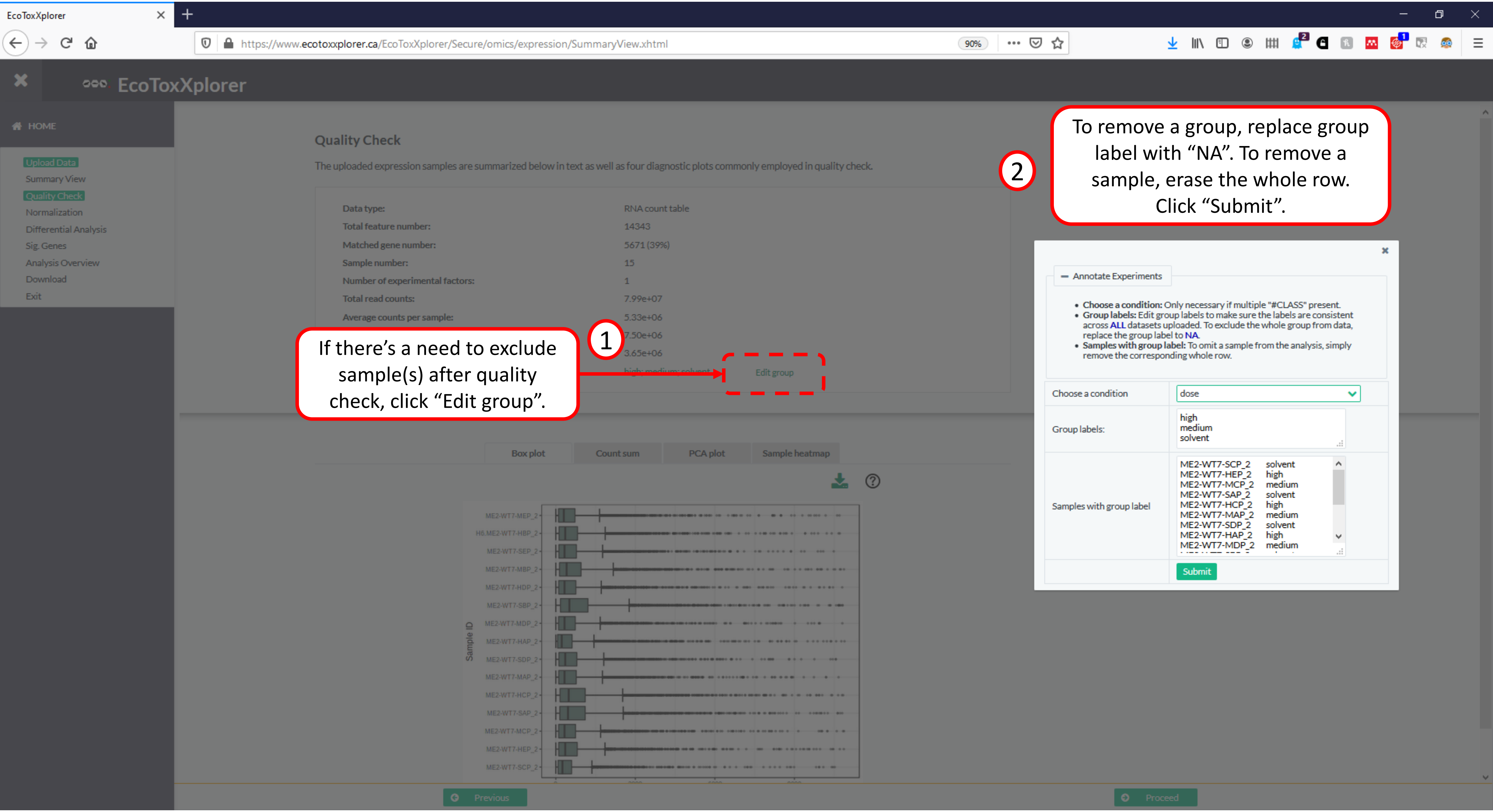


Result Download

Please download the results (tables and images) below. The **Download.zip** contains all the files in your home directory.

Download.zip	libsize_norm_1dpi72.png
Rhistory.R	140634_dpi72.png
qc_norm_meanstd_1dpi72.png	qc_norm_density_0dpi72.png
globalTest_results.csv	sample_heatmap_0dpi72.png
libsize_norm_0dpi72.png	qc_boxplot_0dpi72.png
kegg_pathway_netview1.json	ecotoxxplorer_heatmappw_1.json
ma_plot_0dpi72.png	kegg_pathway_netview2.json
volcano_plot_0dpi72.png	qc_norm_boxplot_1dpi72.png
globalTest.json	qc_norm_pca_0dpi72.png
qc_norm_boxplot_0dpi72.png	550332_bardpi72.png
cleaned_annot.csv	SigGene_high_vs_solvent_Result_2.csv
qc_norm_pca_1dpi72.png	570455_bardpi72.png
kegg_pathway_netview0.json	enrichmentNet_result.json
qc_norm_density_1dpi72.png	BaP.txt
qc_meanstd_0dpi72.png	sample_norm_heatmap_1dpi72.png
countBar.json	enrichmentNet_result.csv
qc_norm_meanstd_0dpi72.png	libsize_0dpi72.png
sample_norm_heatmap_0dpi72.png	qc_pca_0dpi72.png
density.json	

Appendix



Quality Check

The uploaded expression samples are summarized below in text as well as four diagnostic plots commonly employed in quality check.

Data type:	RNA count table
Total feature number:	14343
Matched gene number:	5671 (39%)
Sample number:	15
Number of experimental factors:	1
Total read counts:	7.99e+07
Average counts per sample:	5.33e+06
	7.50e+06
	3.65e+06
	high; medium; solvent

Annotate Experiments

- **Choose a condition:** Only necessary if multiple "#CLASS" present.
- **Group labels:** Edit group labels to make sure the labels are consistent across **ALL** datasets uploaded. To exclude the whole group from data, replace the group label to **NA**.
- **Samples with group label:** To omit a sample from the analysis, simply remove the corresponding whole row.

Choose a condition	dose
Group labels:	high medium solvent
Samples with group label	ME2-WT7-SCP_2 solvent ME2-WT7-HEP_2 high ME2-WT7-MCP_2 medium ME2-WT7-SAP_2 solvent ME2-WT7-HCP_2 high ME2-WT7-MAP_2 medium ME2-WT7-SDP_2 solvent ME2-WT7-HAP_2 high ME2-WT7-MDP_2 medium
	Submit

① If there's a need to exclude sample(s) after quality check, click "Edit group".

② To remove a group, replace group label with "NA". To remove a sample, erase the whole row. Click "Submit".

Questions?

Contact us at

https://www.ecotoxxplorer.ca/doc/ContactView.xhtml