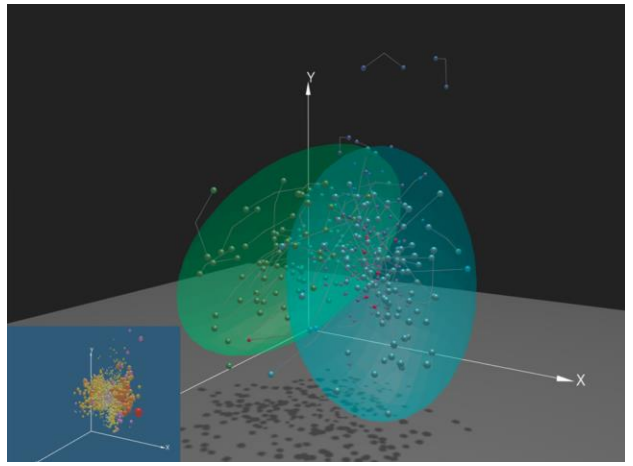
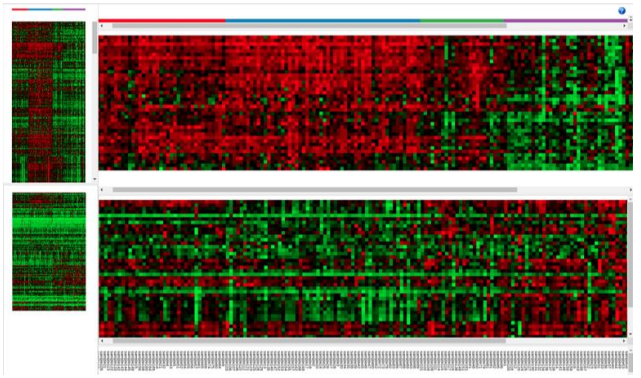
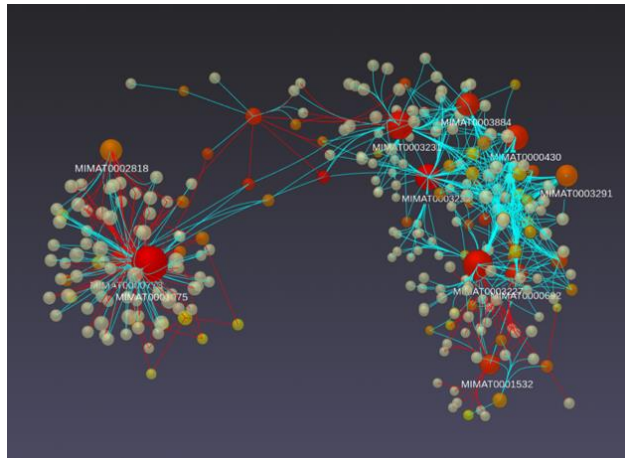
TUTORIAL 1: Interpreting multi-omics using interactive scatter plot

Intro to OmicsAnalyst

- Web-based platform designed for data-driven multi-omics integration and visualization
- Designed to be accessible to bench scientists rather than bioinformatician
- Integrates well-established multivariate and univariate statistics with innovative visual analytics to support:
  - Integrative multi-omics analysis
  - Clustering and pattern discovery
  - Correlation analysis







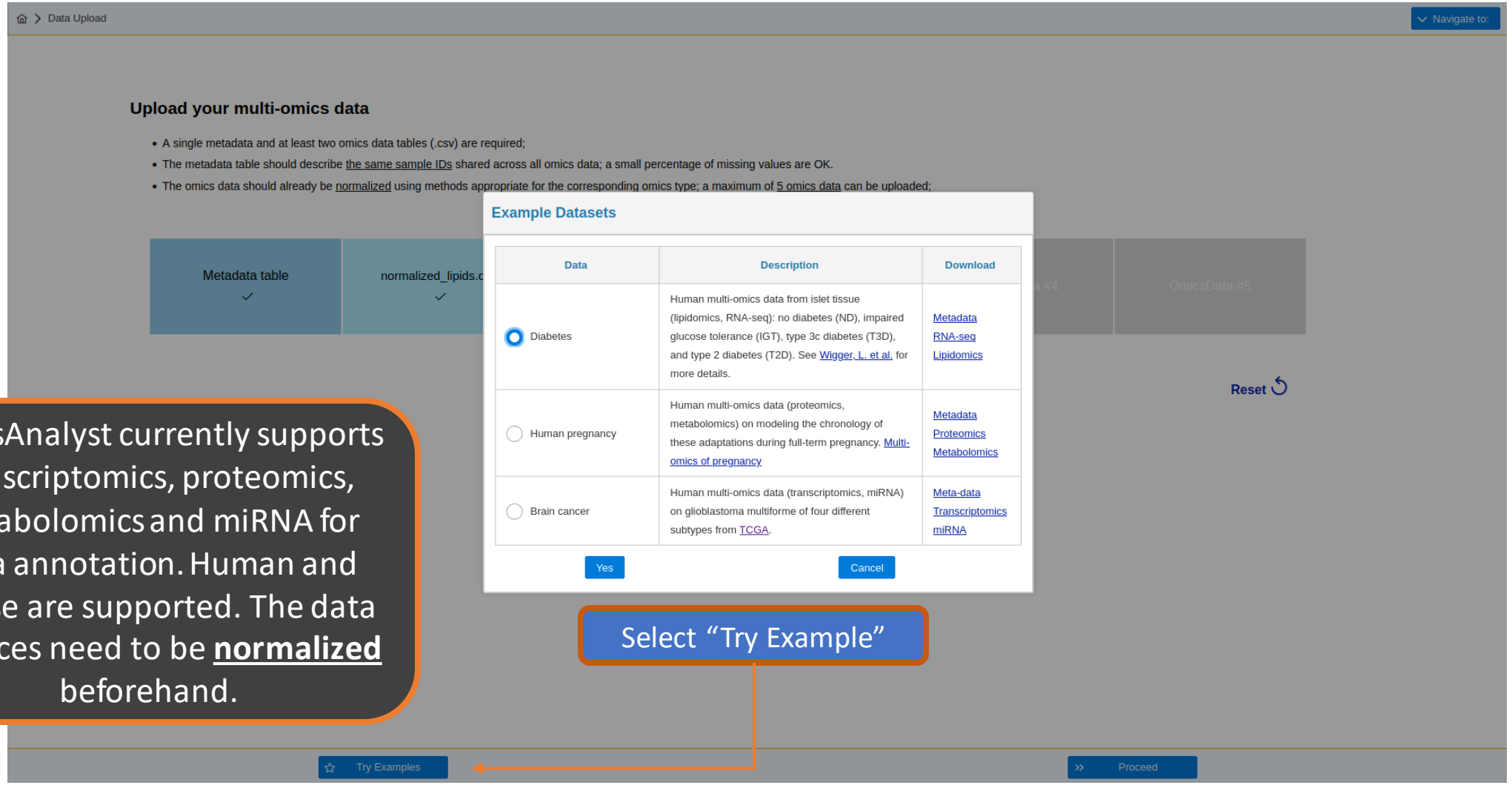
Requirements

- Modern browser supporting WebGL.
- Ensure that WebGL is enabled in your browser!
  - Go to this page https://get.webgl.org/ to verify your browser supports WebGL.
  - Refer to the FAQs for instructions on how to enable WebGL.
- For the best performance and visualization, use the latest version of Google Chrome.

Goal for this tutorial

- A challenge for multi-omics analysis is to interpret highly complex and heterogeneous data without being overwhelmed by it.
- Visualizing global patterns and clusters facilitate data interpretation
- Multivariate statistical methods integrates and transform omics data into dimensionally reduced data points in both feature space and sample space.
- The goal of this tutorial is to:
  - Introduce OmicsAnalyst's tools for visualizing feature and sample space
  - Perform clustering and functional analysis to explore its overall pattern and biological functions.

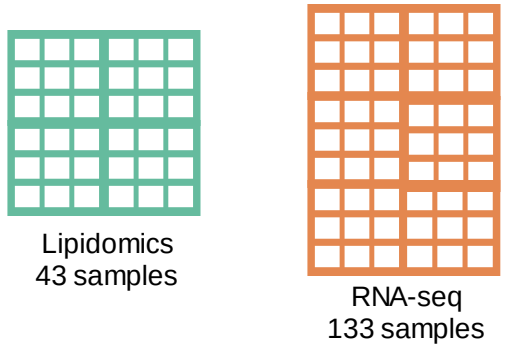
Select and process example dataset



OmicsAnalyst currently supports transcriptomics, proteomics, metabolomics and miRNA for data annotation. Human and mouse are supported. The data matrices need to be **normalized** beforehand.

Selected example dataset

- Lipidomics and RNA-seq data from healthy individuals and patients with different diabetes diagnosis.
- Metadata file containing different variables that have been recorded
  - i.e Age, BMI, Diagnosis, etc.
- OmicsAnalyst will trim samples to retain those that have both lipidomics and RNA-seq data



Multi-omics harmonization

Notice the two datasets do not have same data distribution. This can be addressed using data scaling.

Data Upload > Quality Checking

Uploaded Data

Metadata

t2d_metadata.csv
Primary: Diagnosis
Factors #: 5
Sample: 135

Omics Data

normalized_lipids.csv
Feature: 116
Sample: 43
Sig. #: 0

normalized_rnaseq.csv
Feature: 14406
Sample: 133
Sig. #: 0

R Command History

Multi-omics data harmonization

Omics Data Overview | Metadata Overview

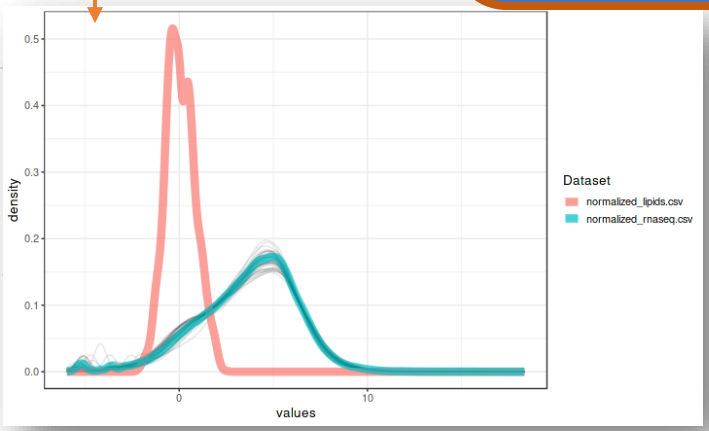
Omics data:	normalized_lipids.csv; normalized_rnaseq.csv
Annotated feature number:	116; 14399
Filtered feature number:	116; 14399
Matched sample number:	43

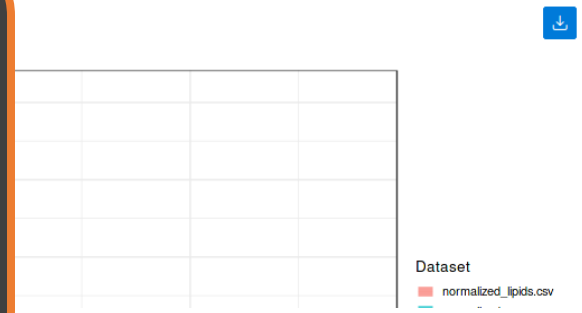
OmicsData Editor

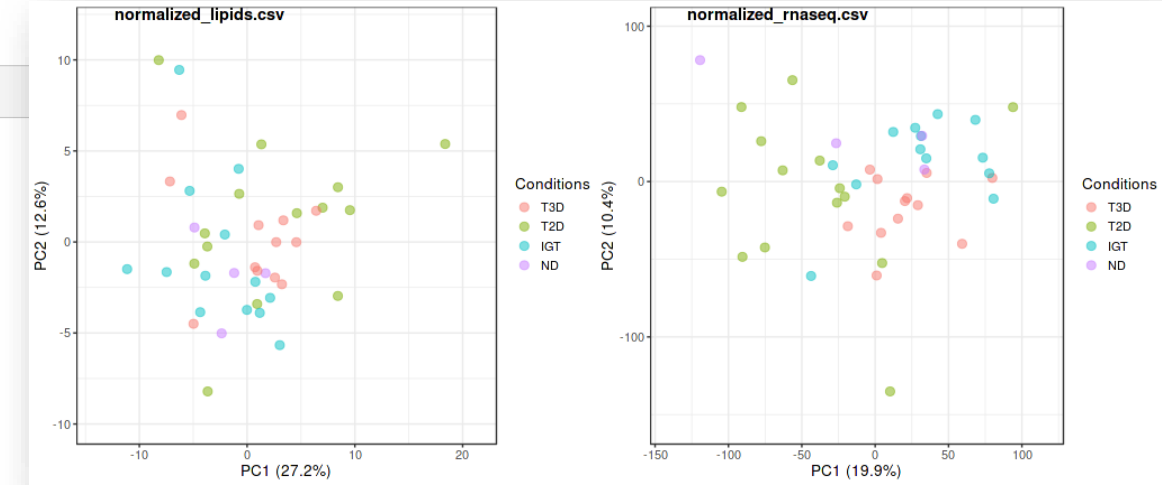
Use the graphical summaries below - **Density plot** and **PCA plot** for more information

Data Filtering
Dataset: Apply to all
Remove % by variance: 0
Update

Data Scaling
Dataset: Apply to all
Scaling method: Auto scaling
Update

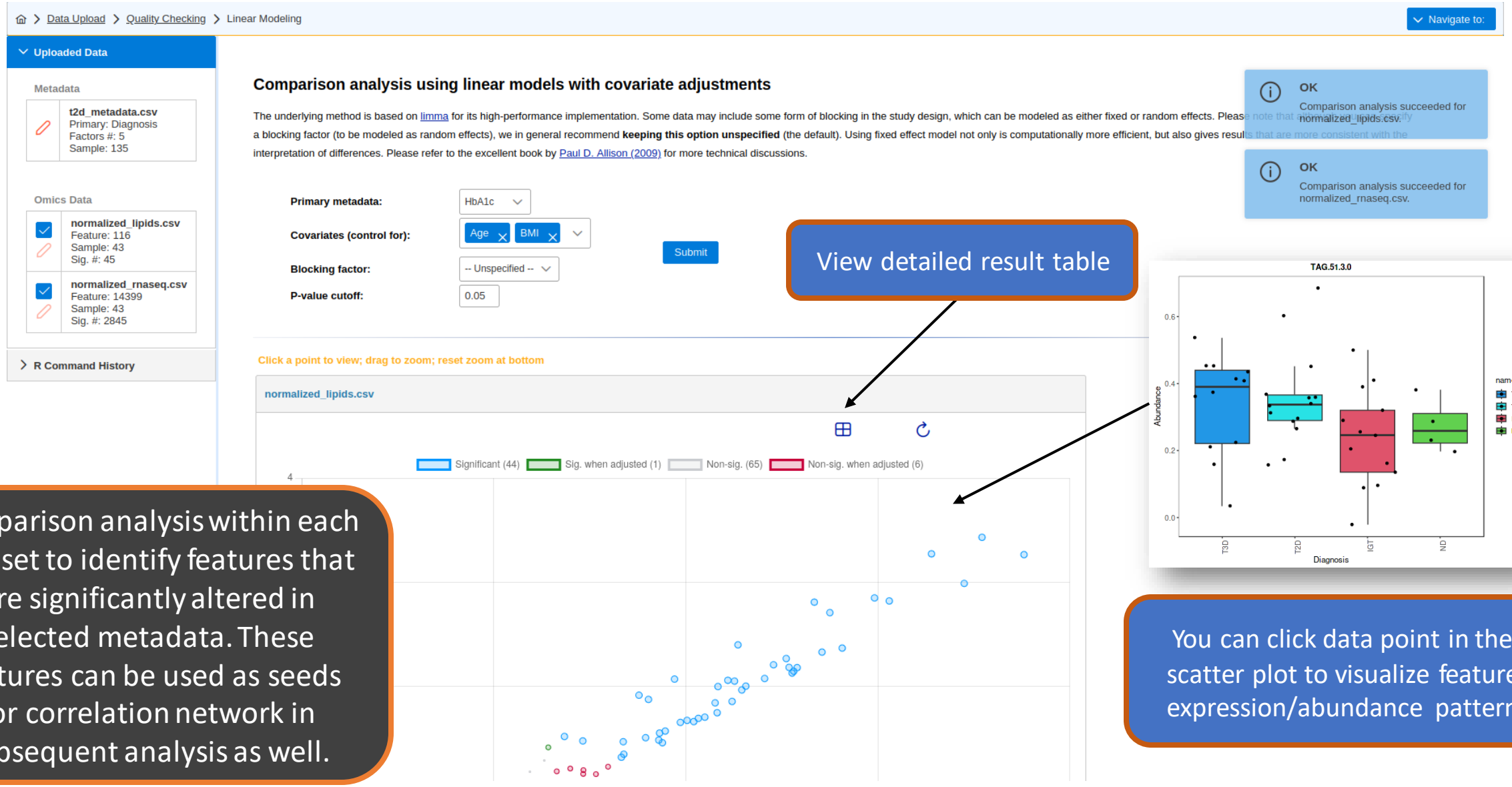






Harmonization page allows you to further process your datasets by making them more comparable. You can use density plot to have an overview of the data distribution and PCA for sample clustering

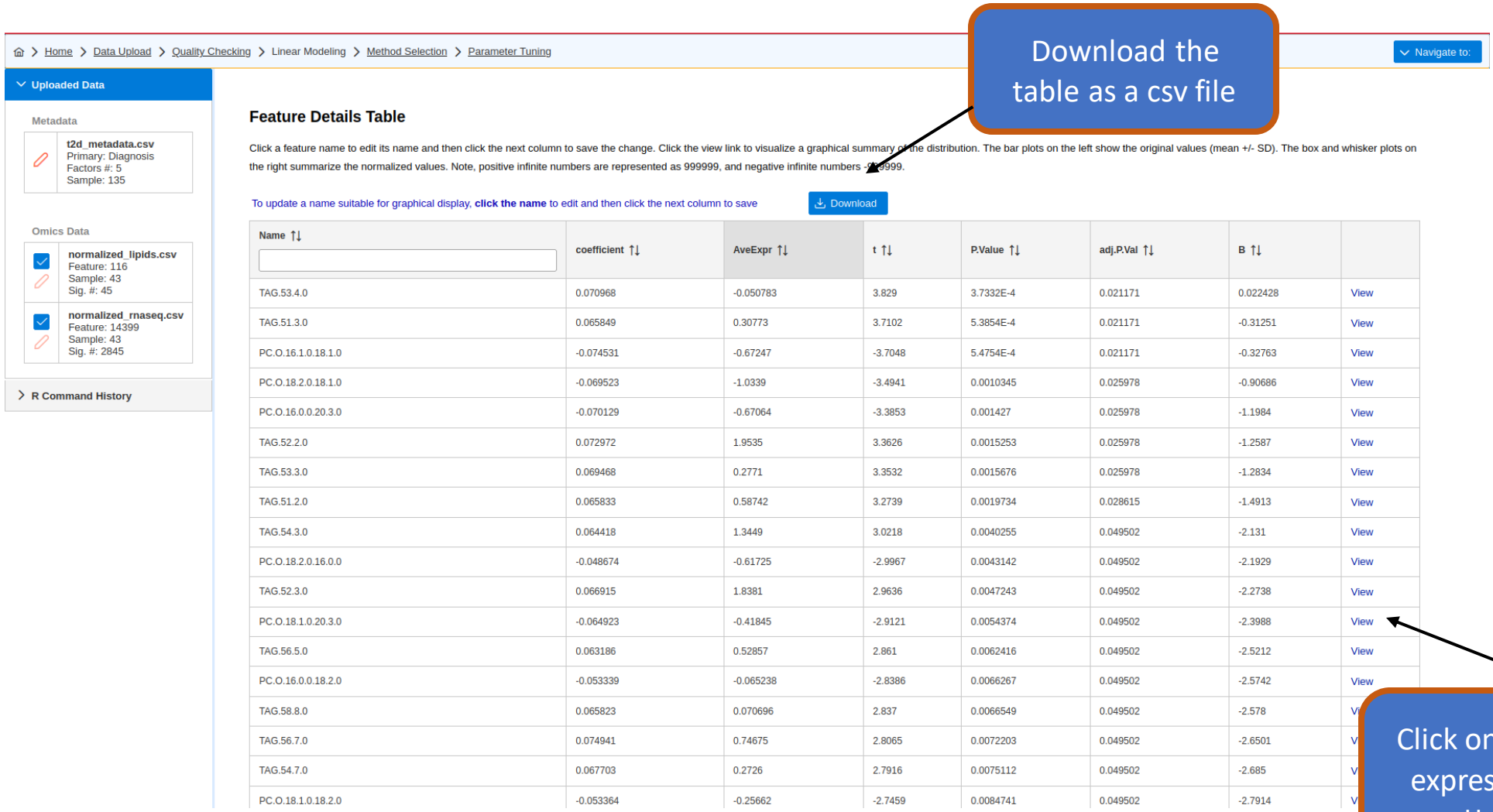
Comparison analysis



Comparison analysis within each dataset to identify features that are significantly altered in selected metadata. These features can be used as seeds for correlation network in subsequent analysis as well.

You can click data point in the scatter plot to visualize feature expression/abundance pattern

Detailed result table



Name	coefficient	AveExpr	t	P.Value	adj.P.Val	B	
TAG.53.4.0	0.070968	-0.050783	3.829	3.7332E-4	0.021171	0.022428	View
TAG.51.3.0	0.065849	0.30773	3.7102	5.3854E-4	0.021171	-0.31251	View
PC.O.16.1.0.18.1.0	-0.074531	-0.67247	-3.7048	5.4754E-4	0.021171	-0.32763	View
PC.O.18.2.0.18.1.0	-0.069523	-1.0339	-3.4941	0.0010345	0.025978	-0.90686	View
PC.O.16.0.0.20.3.0	-0.070129	-0.67064	-3.3853	0.001427	0.025978	-1.1984	View
TAG.52.2.0	0.072972	1.9535	3.3626	0.0015253	0.025978	-1.2587	View
TAG.53.3.0	0.069468	0.2771	3.3532	0.0015676	0.025978	-1.2834	View
TAG.51.2.0	0.065833	0.58742	3.2739	0.0019734	0.028615	-1.4913	View
TAG.54.3.0	0.064418	1.3449	3.0218	0.0040255	0.049502	-2.131	View
PC.O.18.2.0.16.0.0	-0.048674	-0.61725	-2.9967	0.0043142	0.049502	-2.1929	View
TAG.52.3.0	0.066915	1.8381	2.9636	0.0047243	0.049502	-2.2738	View
PC.O.18.1.0.20.3.0	-0.064923	-0.41845	-2.9121	0.0054374	0.049502	-2.3988	View
TAG.56.5.0	0.063186	0.52857	2.861	0.0062416	0.049502	-2.5212	View
PC.O.16.0.0.18.2.0	-0.053339	-0.065238	-2.8386	0.0066267	0.049502	-2.5742	View
TAG.58.8.0	0.065823	0.070696	2.837	0.0066549	0.049502	-2.578	V
TAG.56.7.0	0.074941	0.74675	2.8065	0.0072203	0.049502	-2.6501	V
TAG.54.7.0	0.067703	0.2726	2.7916	0.0075112	0.049502	-2.685	V
PC.O.18.1.0.18.2.0	-0.053364	-0.25662	-2.7459	0.0084741	0.049502	-2.7914	V

Covariate adjustments in comparison analysis

- Study design of multi-omics datasets may have different variables (metadata) that affect the experimental outcome.
  - i.e clinical data from heterogeneous population

- We use linear modeling method based from *limma* R package for covariate adjustment. The objective is to control for the effects of potential confounding variables when performing comparison on the primary metadata.

Method Selection

Three analysis tracks are available. Under dimension reduction, two unsupervised methods - MCIA and MOFA one supervised method - DIABLO are availabe

Data Upload > Quality Checking > Linear Modeling > Method Selection

Uploaded Data

Metadata

t2d_metadata.csv
Primary: Diagnosis
Factors #: 5
Sample: 135

Omics Data

normalized_lipids.csv
Feature: 116
Sample: 43
Sig. #: 45

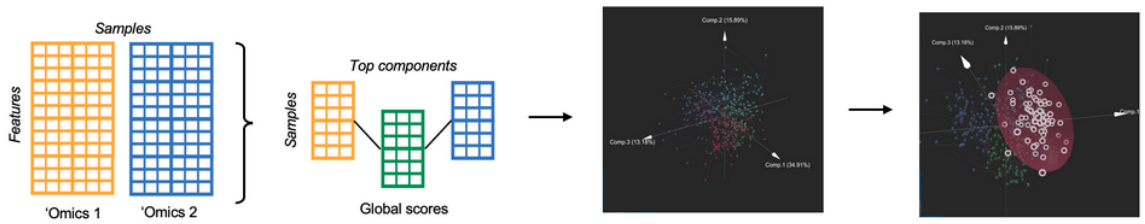
normalized_rnaseq.csv
Feature: 14399
Sample: 43
Sig. #: 2845

R Command History

Please choose a method to proceed

Dimensionality Reduction | **Correlation Network** | **Clustered Heatmap**

The objective of this analysis is to perform dimension reduction, and then visually explore corresponding scores, loadings and biplots in interactive 3D scatter plots to understand the common trends and underlying patterns. The multivariate dimension reduction techniques are kind of like parallel versions of PCA, where we try to find sets of multi-dimensional components that both reduce redundant information within individual datasets, and are related to each other across datasets. These sets of components are related to each other through some global scores, which are the dimensions that we use to visualize the sample space. The different methods are mainly distinguished by the way that they optimize similarity of component sets across the 'omics datasets. Select an individual method to see more details on its unique statistical features.



Multi-variate dimension reduction | **Visualize samples on top 3 components of shared co-variance space** | **Detect and select clusters**

Dimension reduction method: MCIA | Proceed

Multiple co-inertia analysis (MCIA) is a robust method of finding related multi-dimensional components across multiple datasets. MCIA can be performed on any number of tables, although we currently limit to two in OmicsAnalyst. It is similar to more familiar canonical correlation analysis (CCA), but performs well when the number of features are much greater than the number of samples, and therefore does not require regularization before being used for 'omics data like CCA. MCIA is symmetric, therefore the order that the 'omics datasets are uploaded will not impact the results. (more details ....)

Results summary

The sample separation in top 5 components are shown in the PCA. The line chart displays variance percentage per omics type for the top five components.

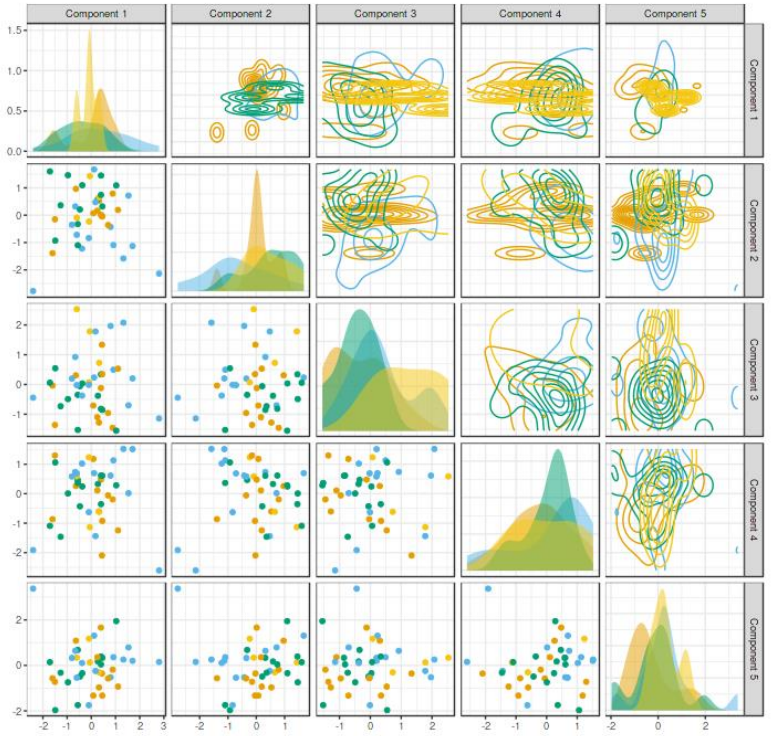
Home > Data Upload > Quality Checking > Linear Modeling > Method Selection > Parameter Tuning

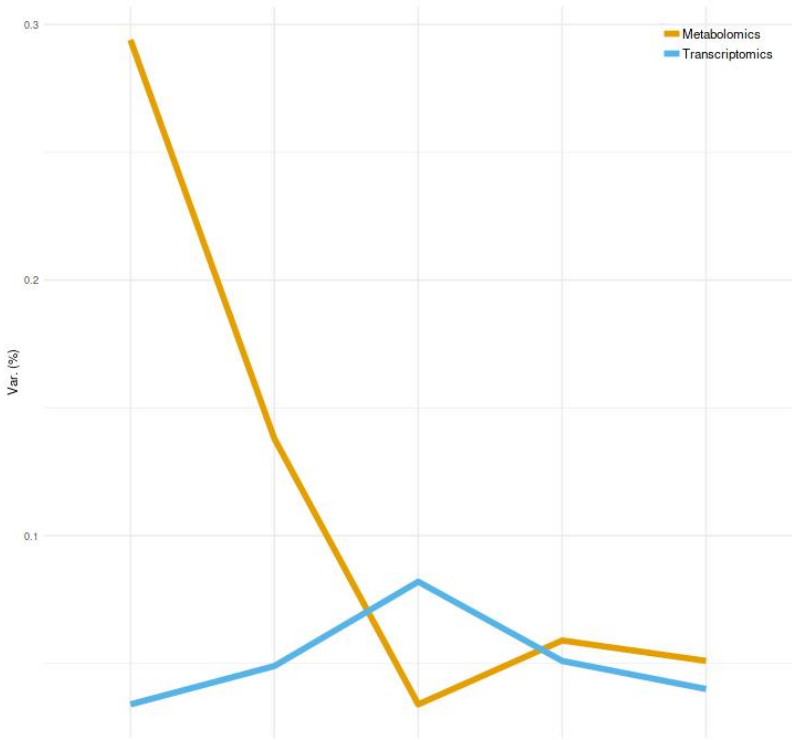
Dimension Reduction Results

Each multivariate method calculates **integrated factor scores** based on **feature weights** from each omics layer.

MCIA factors simulataneously maximize the variability in each individual omics layer AND correlation between layers. This means that each factor is shared across omics layers.

View Graphical Summary

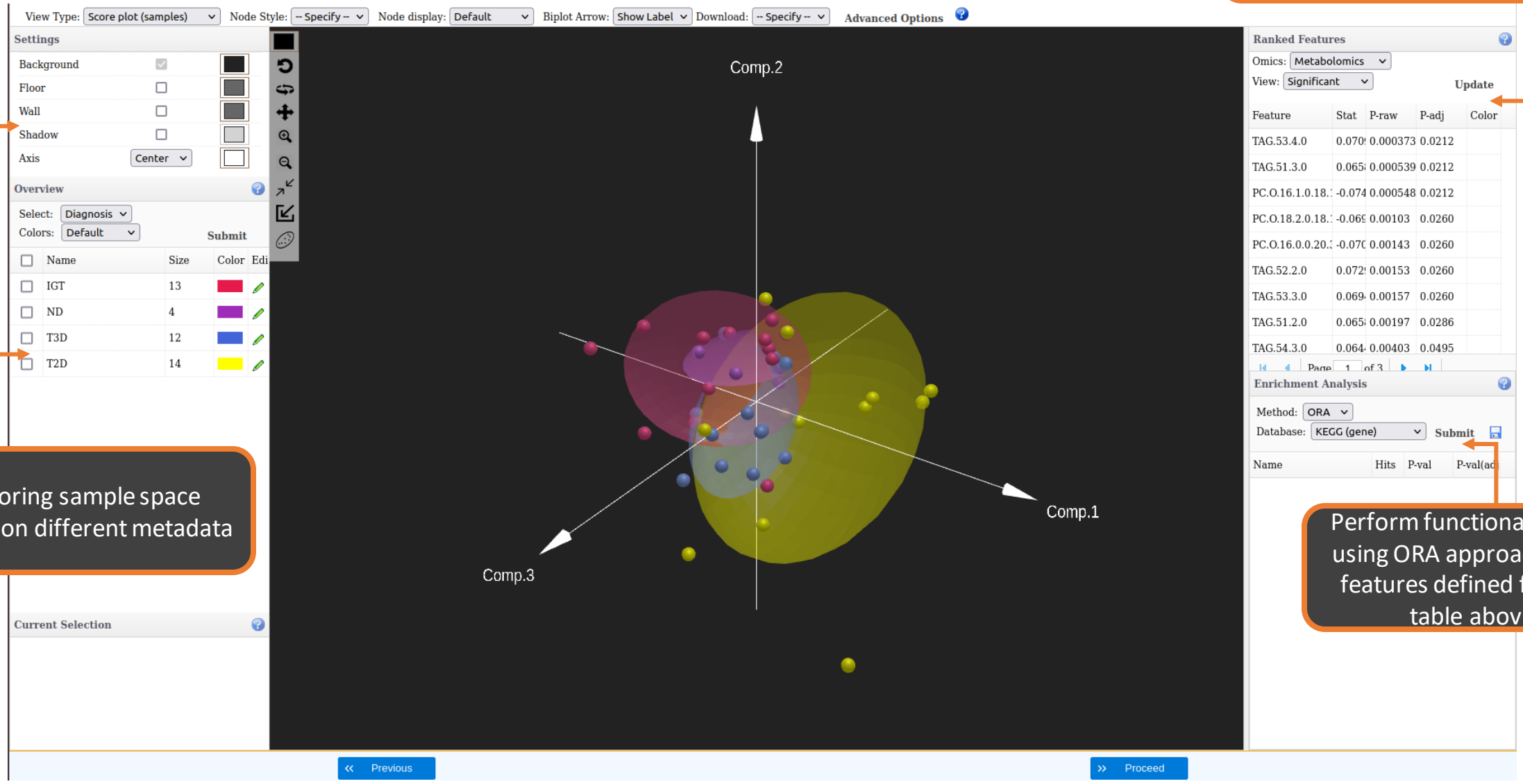




Scatter plot viewer

Overall visual settings that modify the background environment of the plot.

Show list of significant features from comparison analysis or based on loading score of component 1, 2 or 3.

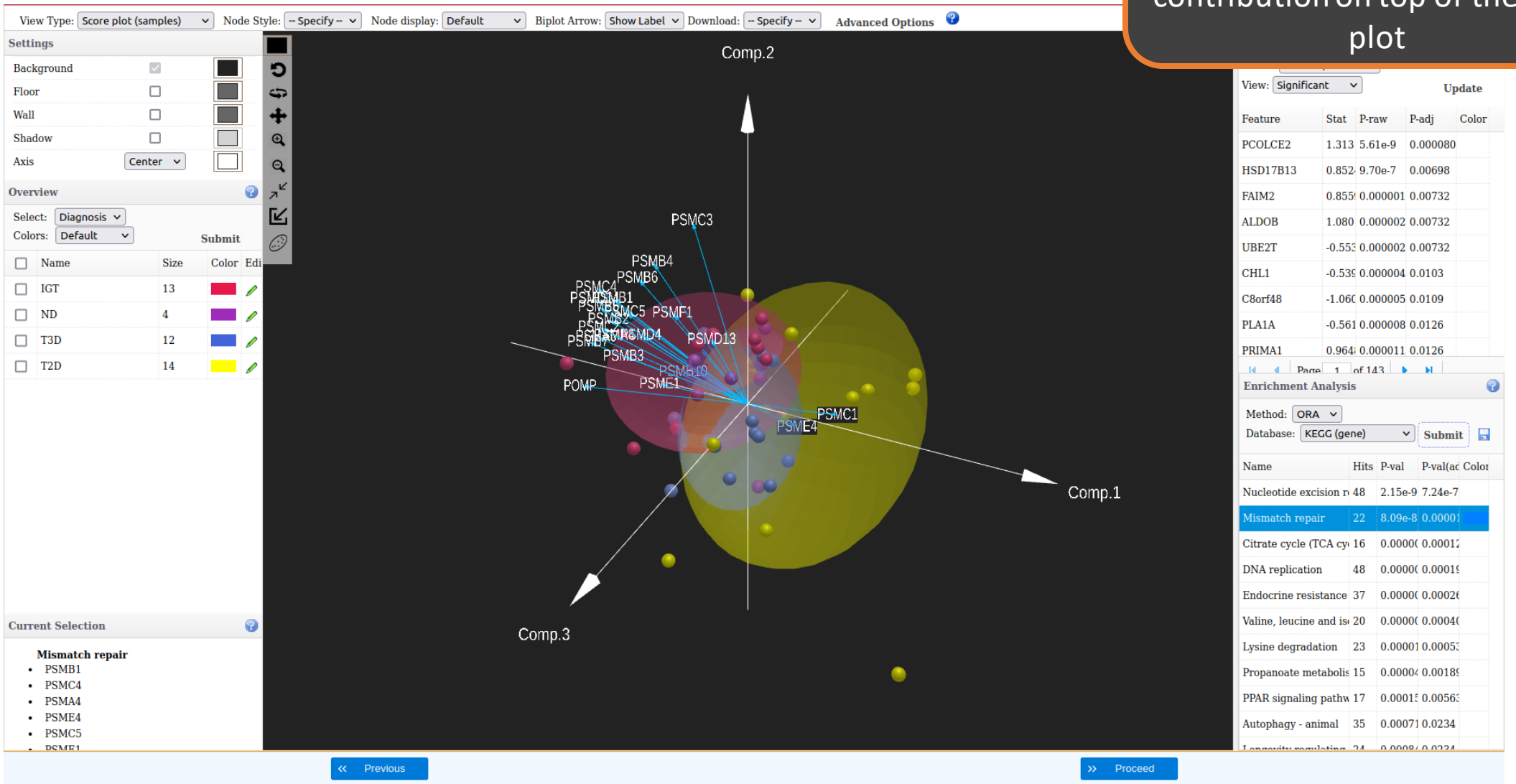


Coloring sample space based on different metadata

Perform functional analysis using ORA approach on the features defined from the table above.

Enrichment analysis

Selecting individual feature or enriched pathway/set will project features loading contribution on top of the score plot

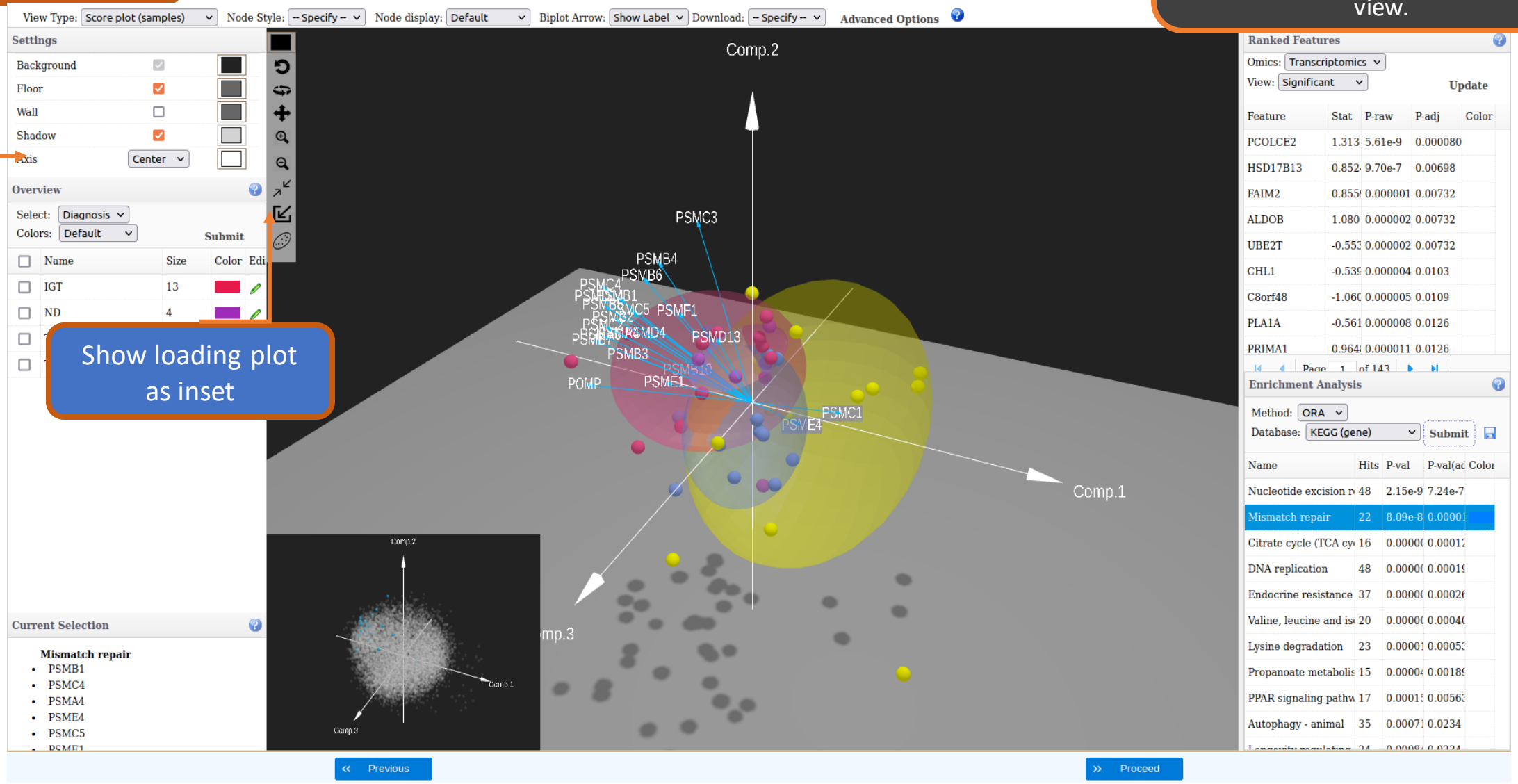


Customize view

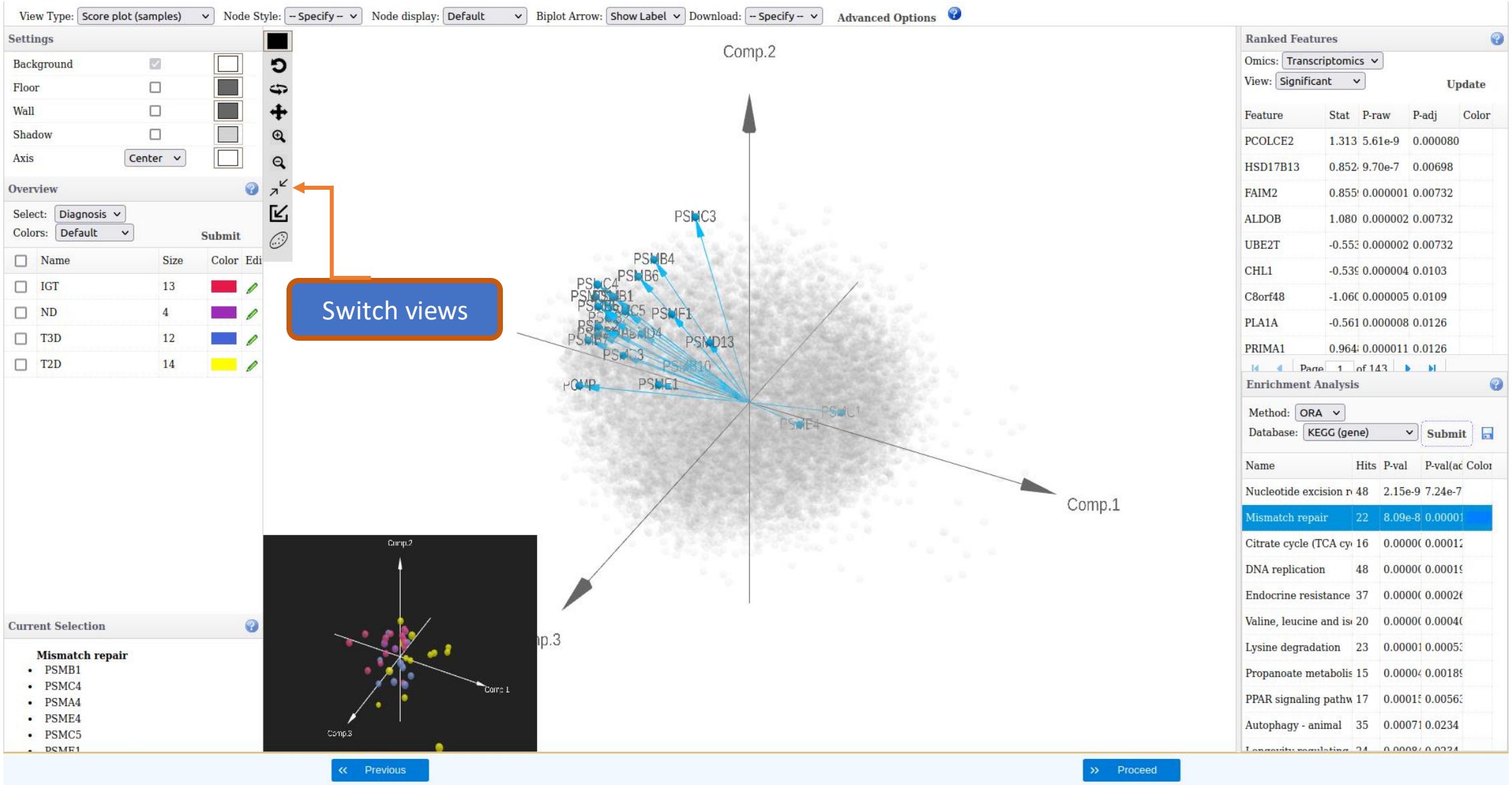
Customize current view by changing the settings here

Customize the current view space by adding floor, wall, shadow, changing axis position, background color. It can be applied to both main view or inset view.

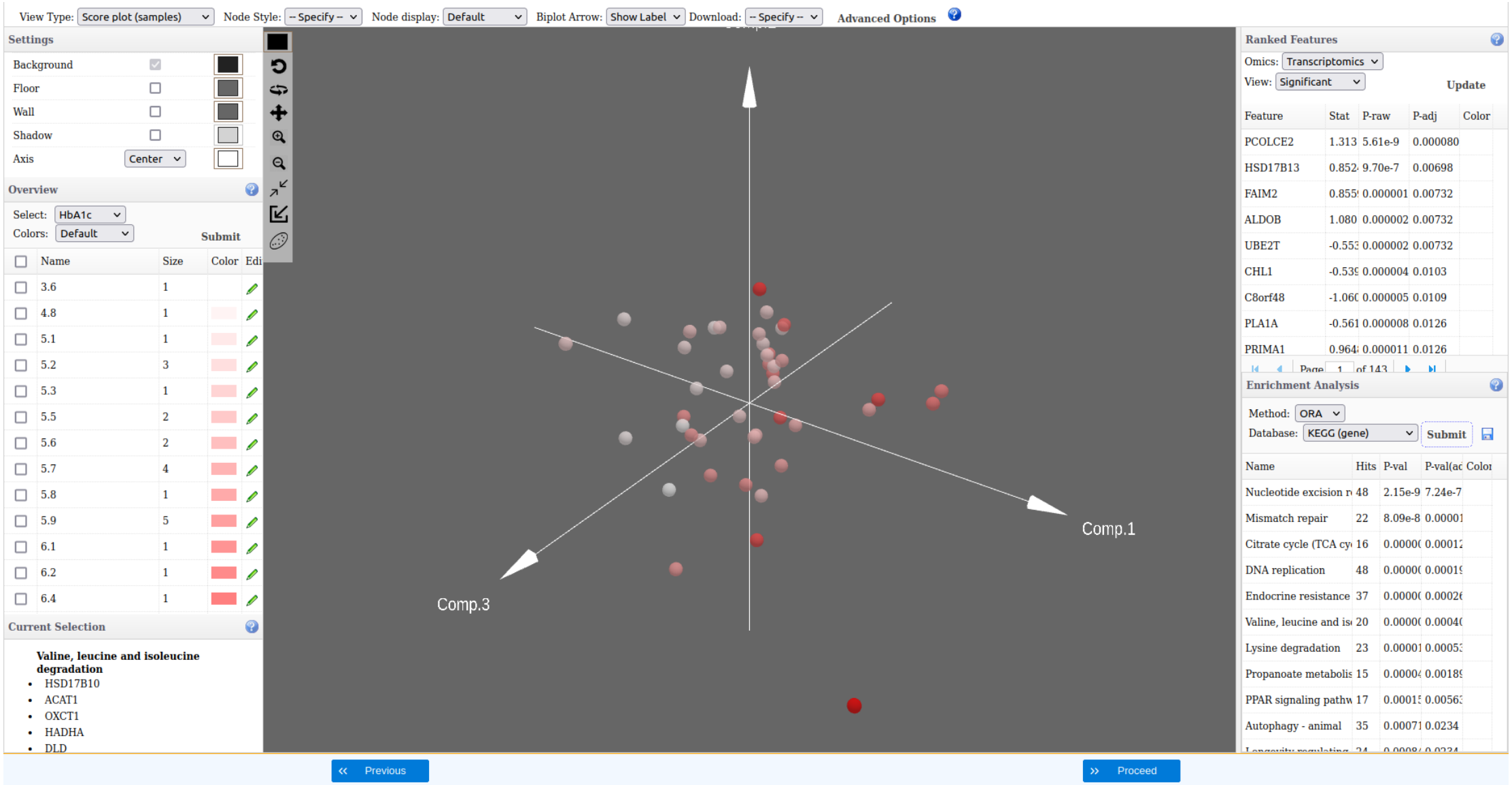
Show loading plot as inset

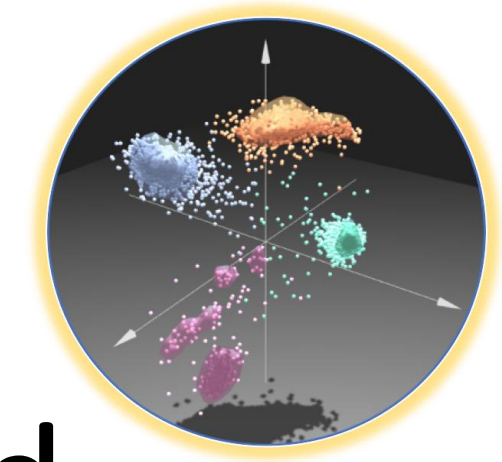


Loading plot in the main view



Color nodes by continuous metadata





The End

For more information, visit the **FAQs, Tutorials, Resources**

and **Contact** pages on www.omicsanalyst.ca