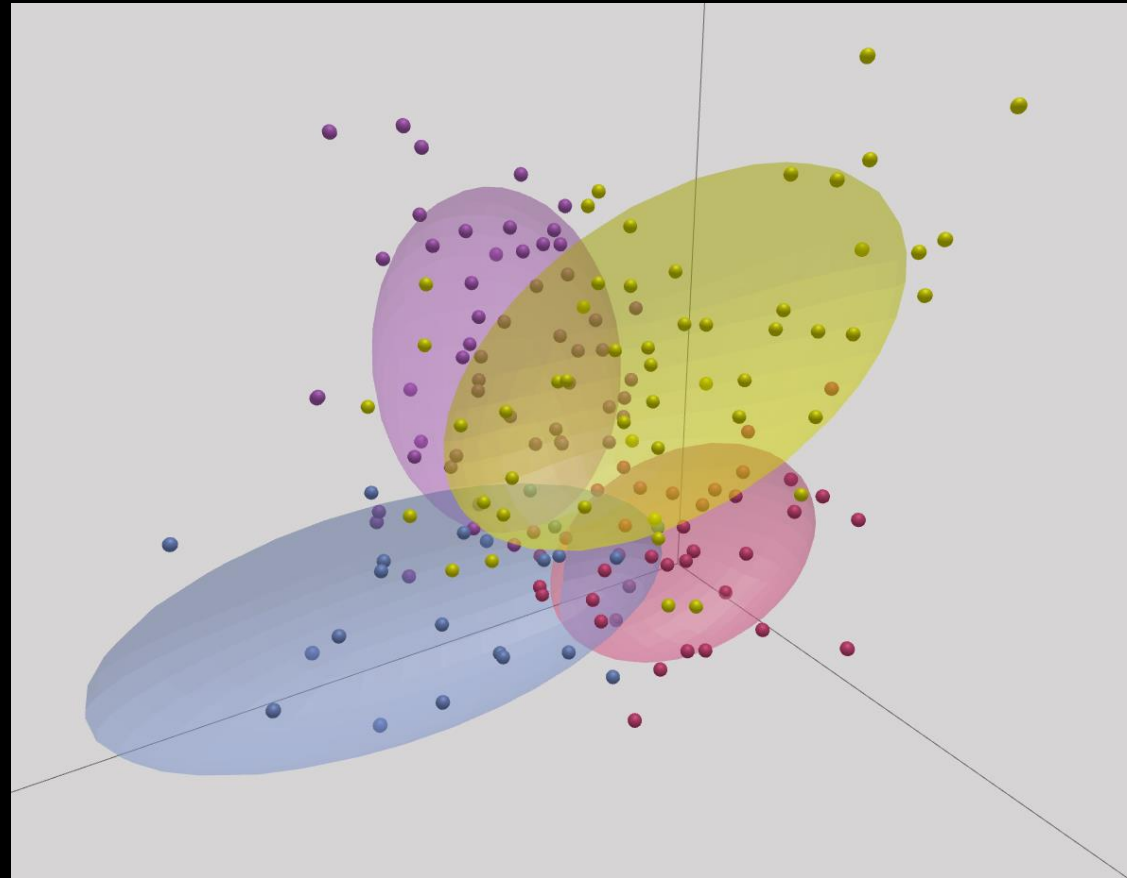
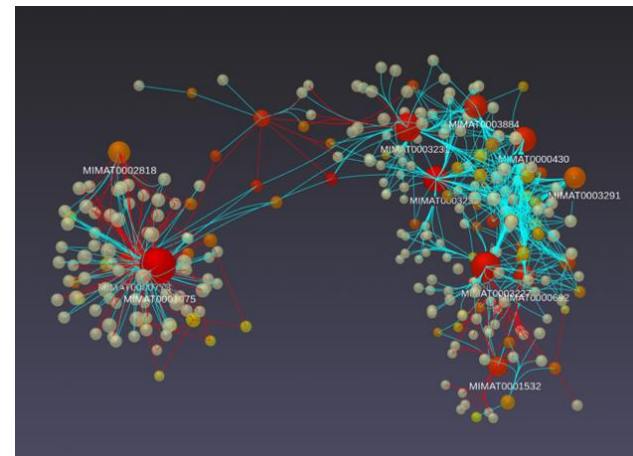
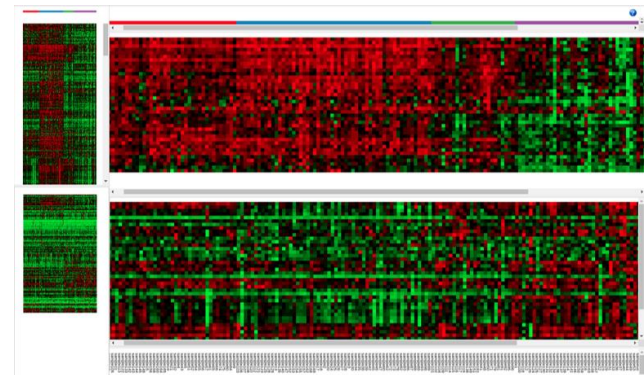
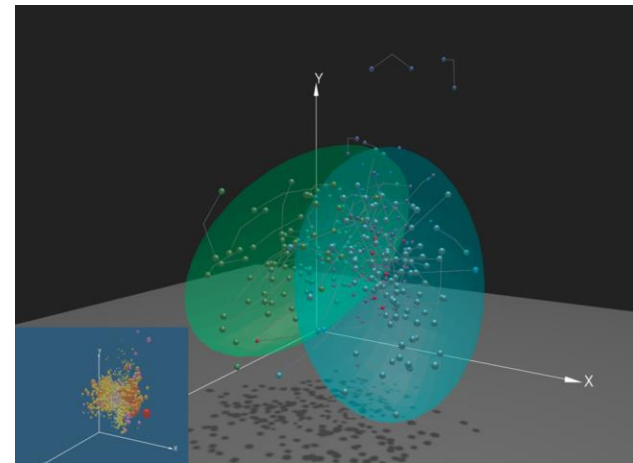


TUTORIAL 1: Interpreting multi-omics using interactive scatter plot



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

🏠 > Data Upload 📄 Navigate to:

Upload your multi-omics data

- A single metadata and at least two omics data tables (.csv) are required;
- The metadata table should describe the same sample IDs shared across all omics data; a small percentage of missing values are OK.
- The omics data should already be normalized using methods appropriate for the corresponding omics type; a maximum of 5 omics data can be uploaded;

Metadata table
✓

normalized_lipids.c
✓

OmicsData #4 OmicsData #5
Reset ↻

Example Datasets

Data	Description	Download
<input checked="" type="radio"/> Diabetes	Human multi-omics data from islet tissue (lipidomics, RNA-seq): no diabetes (ND), impaired glucose tolerance (IGT), type 3c diabetes (T3D), and type 2 diabetes (T2D). See Wigger L. et al. for more details.	Metadata RNA-seq Lipidomics
<input type="radio"/> Human pregnancy	Human multi-omics data (proteomics, metabolomics) on modeling the chronology of these adaptations during full-term pregnancy. Multi-omics of pregnancy .	Metadata Proteomics Metabolomics
<input type="radio"/> Brain cancer	Human multi-omics data (transcriptomics, miRNA) on glioblastoma multiforme of four different subtypes from TCGA .	Meta-data Transcriptomics miRNA

YesCancel

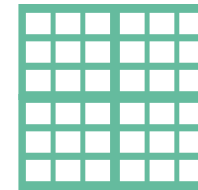
☆ Try Examples» Proceed

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

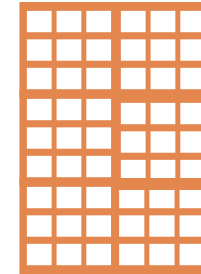
Select "Try Example"

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



Lipidomics
43 samples



RNA-seq
133 samples

Multi-omics harmonization

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

Home > 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

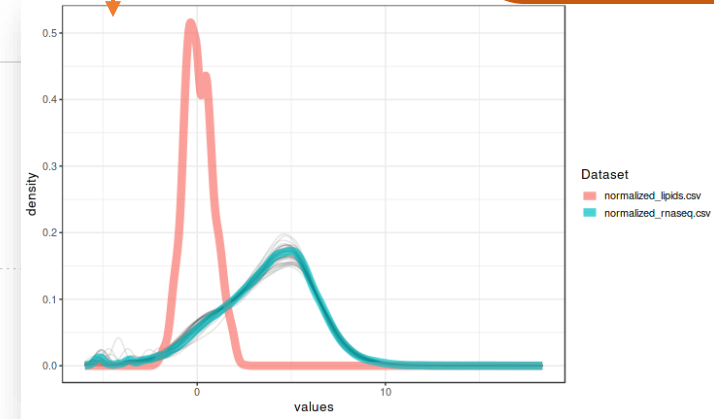
Matched sample number: 43

OmicsData Editor

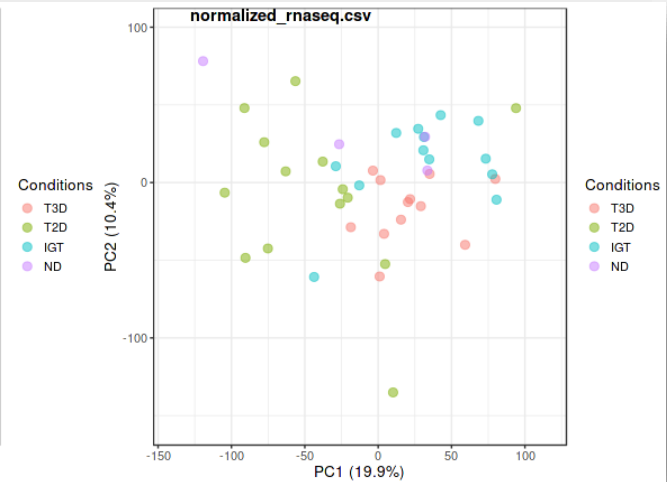
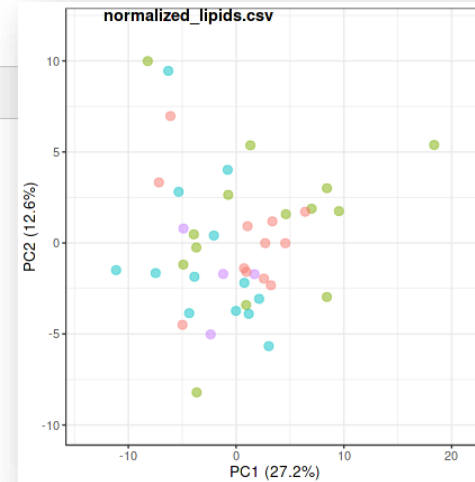
Use the graphical summaries below - Density plot and PCA plot for more information

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

Home > Data Upload > Quality Checking > Linear Modeling Navigate to:

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

Comparison analysis using linear models with covariate adjustments

The underlying method is based on [limma](#) for its high-performance implementation. Some data may include some form of blocking in the study design, which can be modeled as either fixed or random effects. Please note that **keeping this option unspecified** (the default). Using fixed effect model not only is computationally more efficient, but also gives results that are more consistent with the interpretation of differences. Please refer to the excellent book by [Paul D. Allison \(2009\)](#) for more technical discussions.

Primary metadata: HbA1c

Covariates (control for): Age x BMI

Blocking factor: -- Unspecified --

P-value cutoff: 0.05

View detailed result table

Click a point to view; drag to zoom; reset zoom at bottom

normalized_lipids.csv

Significant (44) Sig. when adjusted (1) Non-sig. (65) Non-sig. when adjusted (6)

Abundance

TAG.51.3.0

Abundance

Diagnosis

name

- T3D
- T2D
- IGT
- ND

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

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.

OK
Comparison analysis succeeded for normalized_lipids.csv.

OK
Comparison analysis succeeded for normalized_rnaseq.csv.

Detailed result table

Download the table as a csv file

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

Feature Details Table

Click a feature name to edit its name and then click the next column to save the change. Click the view link to visualize a graphical summary of the distribution. The bar plots on the left show the original values (mean +/- SD). The box and whisker plots on the right summarize the normalized values. Note, positive infinite numbers are represented as 999999, and negative infinite numbers as -999999.

To update a name suitable for graphical display, **click the name** to edit and then click the next column to save [Download](#)

Name ↑↓	coefficient ↑↓	AveExpr ↑↓	t ↑↓	P.Value ↑↓	adj.P.Val ↑↓	B ↑↓	
<input type="text"/>							
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	View
TAG.56.7.0	0.074941	0.74675	2.8065	0.0072203	0.049502	-2.6501	View
TAG.54.7.0	0.067703	0.2726	2.7916	0.0075112	0.049502	-2.685	View
PC.O.18.1.0.18.2.0	-0.053364	-0.25662	-2.7459	0.0084741	0.049502	-2.7914	View

Click on view to visualize expression/abundance pattern of a feature

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 - MCI and MOFA one supervised method - DIABLO are available

Home > 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

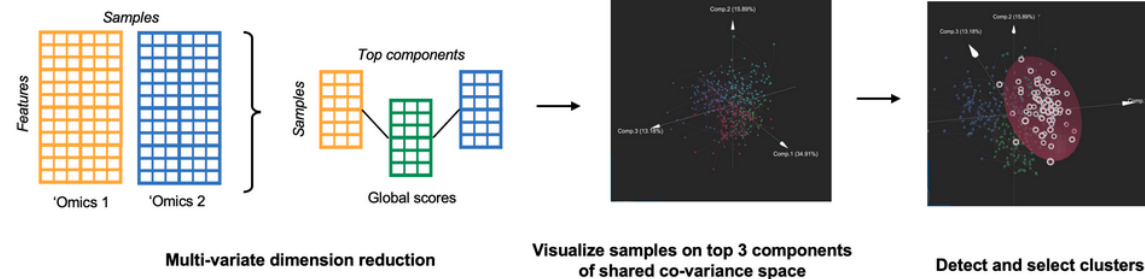
normalized_rnaseq.csv
Feature: 14399
Sample: 43
Sig. #: 2845

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.



Dimension reduction method MCI

Multiple co-inertia analysis (MCI) is a robust method of finding related multi-dimensional components across multiple datasets. MCI 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. MCI 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.

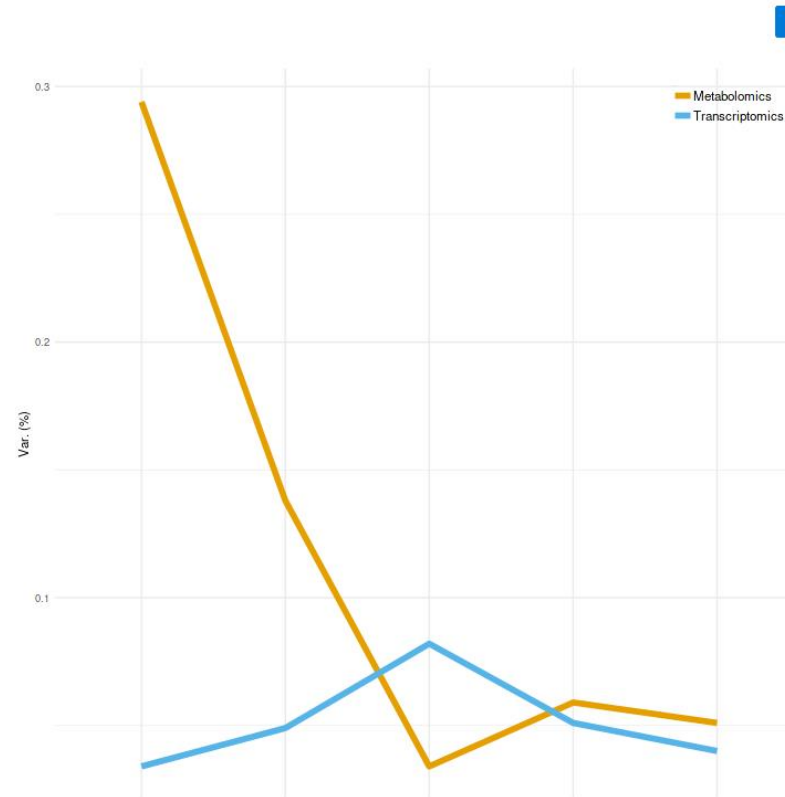
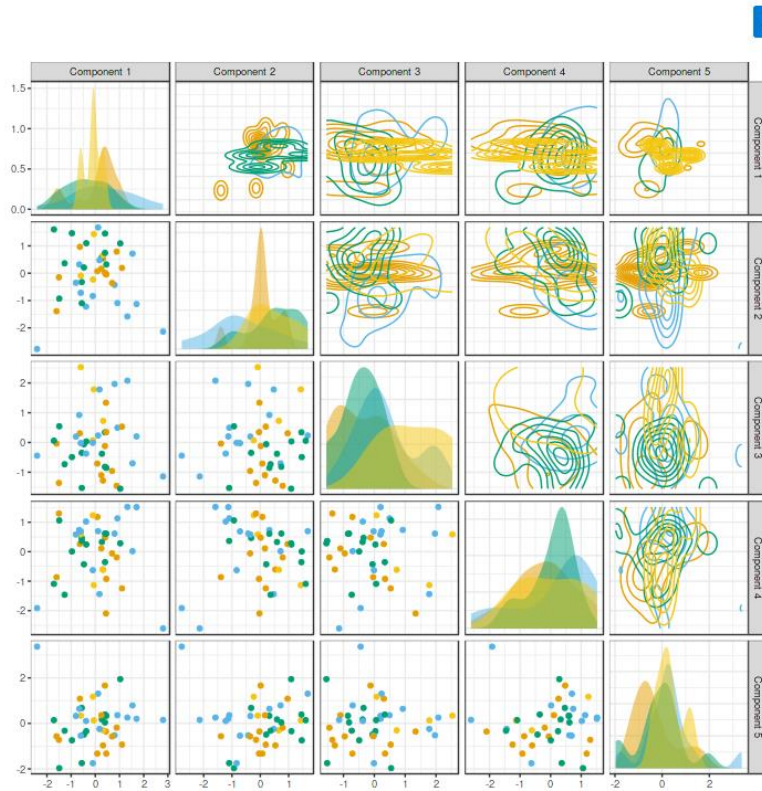
[Home](#) > [Data Upload](#) > [Quality Checking](#) > [Linear Modeling](#) > [Method Selection](#) > [Parameter Tuning](#)

Dimension Reduction Results

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

MCIA factors simultaneously 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](#)



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

Scatter plot viewer

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

View Type: Score plot (samples) | Node Style: -- Specify -- | Node display: Default | Biplot Arrow: Show Label | Download: -- Specify -- | Advanced Options

Settings

Background

Floor

Wall

Shadow

Axis

Overview

Select: | Colors: | Submit

<input type="checkbox"/>	Name	Size	Color	Edi
<input type="checkbox"/>	IGT	13		
<input type="checkbox"/>	ND	4		
<input type="checkbox"/>	T3D	12		
<input type="checkbox"/>	T2D	14		

Ranked Features

Omics: | View: | Update

Feature	Stat	P-row	P-adj	Color
TAG.53.4.0	0.070	0.000373	0.0212	
TAG.51.3.0	0.065	0.000539	0.0212	
PC.O.16.1.0.18.:	-0.074	0.000548	0.0212	
PC.O.18.2.0.18.:	-0.069	0.00103	0.0260	
PC.O.16.0.0.20.:	-0.070	0.00143	0.0260	
TAG.52.2.0	0.072	0.00153	0.0260	
TAG.53.3.0	0.069	0.00157	0.0260	
TAG.51.2.0	0.065	0.00197	0.0286	
TAG.54.3.0	0.064	0.00403	0.0495	

Page 1 of 3

Enrichment Analysis

Method: | Database: | Submit

Name	Hits	P-val	P-val(ad
------	------	-------	----------

Current Selection

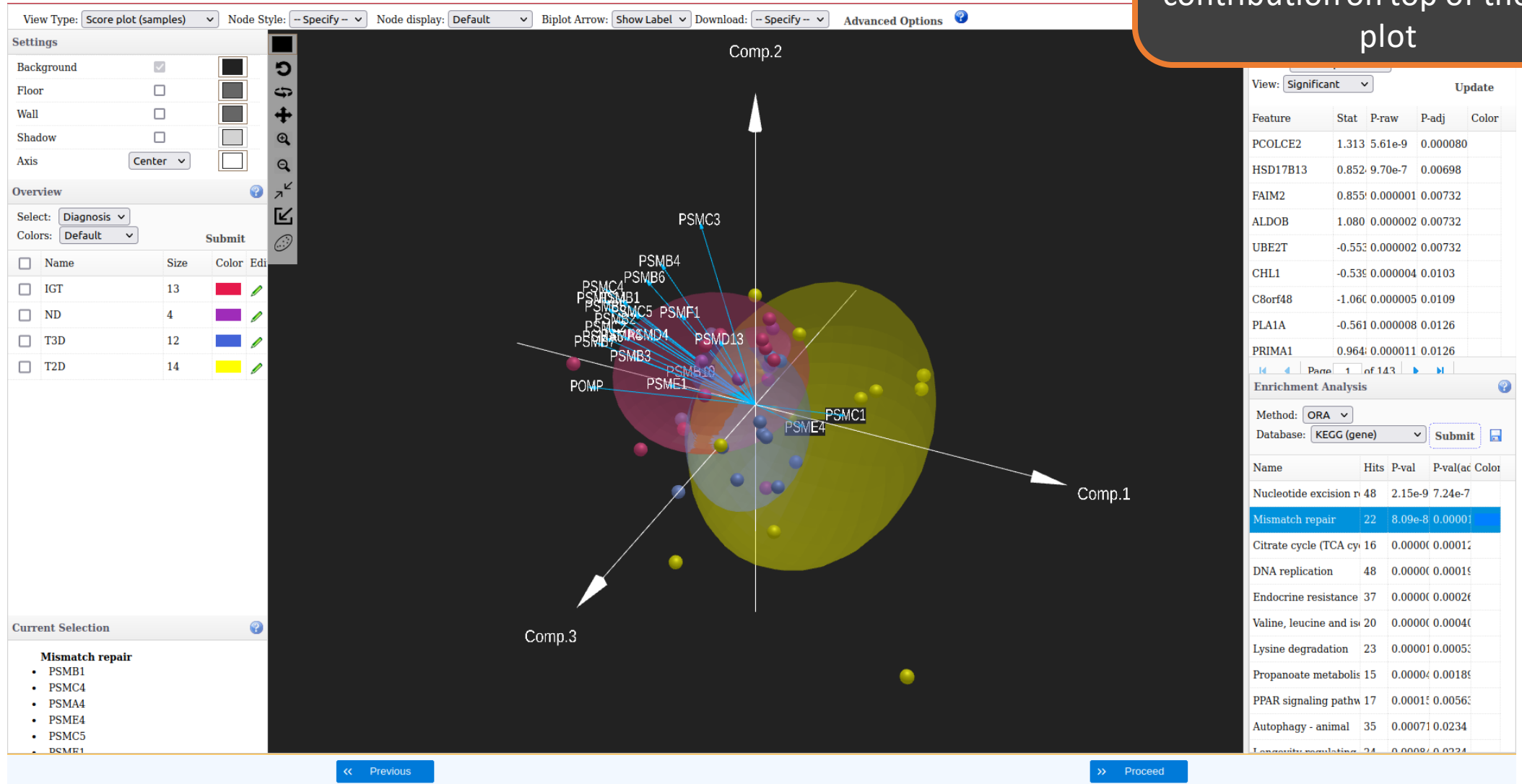
Previous | Proceed

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 current view by changing the settings here

Customize view

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.

View Type: **Score plot (samples)** Node Style: **--Specify--** Node display: **Default** Biplot Arrow: **Show Label** Download: **--Specify--** **Advanced Options**

Settings

Background

Floor

Wall

Shadow

Axis **Center**

Overview

Select: **Diagnosis** Colors: **Default** **Submit**

<input type="checkbox"/>	Name	Size	Color	Edi
<input type="checkbox"/>	IGT	13		
<input type="checkbox"/>	ND	4		

Show loading plot as inset

Ranked Features

Omics: **Transcriptomics** View: **Significant** **Update**

Feature	Stat	P-row	P-adj	Color
PCOLCE2	1.313	5.61e-9	0.000080	
HSD17B13	0.852	9.70e-7	0.00698	
FAIM2	0.855	0.000001	0.00732	
ALDOB	1.080	0.000002	0.00732	
UBE2T	-0.553	0.000002	0.00732	
CHL1	-0.539	0.000004	0.0103	
C8orf48	-1.060	0.000005	0.0109	
PLA1A	-0.561	0.000008	0.0126	
PRIMA1	0.964	0.000011	0.0126	

Enrichment Analysis

Method: **ORA** Database: **KEGG (gene)** **Submit**

Name	Hits	P-val	P-val(ac Color)
Nucleotide excision r	48	2.15e-9	7.24e-7
Mismatch repair	22	8.09e-8	0.00001
Citrate cycle (TCA cy	16	0.00000	0.00017
DNA replication	48	0.00000	0.00019
Endocrine resistance	37	0.00000	0.00026
Valine, leucine and is	20	0.00000	0.00040
Lysine degradation	23	0.00001	0.00050
Propanoate metabolis	15	0.00004	0.00189
PPAR signaling pathw	17	0.00015	0.00560
Autophagy - animal	35	0.00071	0.0234
Longevity regulatin	24	0.00084	0.0234

Current Selection

Mismatch repair

- PSMB1
- PSMC4
- PSMA4
- PSME4
- PSMC5
- PSME1

Navigation: **<< Previous** **Proceed >>**

Loading plot in the main view

View Type: **Score plot (samples)** Node Style: **-- Specify --** Node display: **Default** Biplot Arrow: **Show Label** Download: **-- Specify --** **Advanced Options**

Settings

Background

Floor

Wall

Shadow

Axis **Center**

Overview

Select: **Diagnosis** Colors: **Default** **Submit**

<input type="checkbox"/>	Name	Size	Color	Edi
<input type="checkbox"/>	IGT	13	Red	
<input type="checkbox"/>	ND	4	Purple	
<input type="checkbox"/>	T3D	12	Blue	
<input type="checkbox"/>	T2D	14	Yellow	

Current Selection

Mismatch repair

- PSMB1
- PSMC4
- PSMA4
- PSME4
- PSMC5
- PSME1

Ranked Features

Omics: **Transcriptomics** View: **Significant** **Update**

Feature	Stat	P-raw	P-adj	Color
PCOLCE2	1.313	5.61e-9	0.000080	
HSD17B13	0.852	9.70e-7	0.00698	
FAIM2	0.855	0.000001	0.00732	
ALDOB	1.080	0.000002	0.00732	
UBE2T	-0.552	0.000002	0.00732	
CHL1	-0.532	0.000004	0.0103	
C8orf48	-1.060	0.000005	0.0109	
PLA1A	-0.561	0.000008	0.0126	
PRIMA1	0.964	0.000011	0.0126	

Page 1 of 143

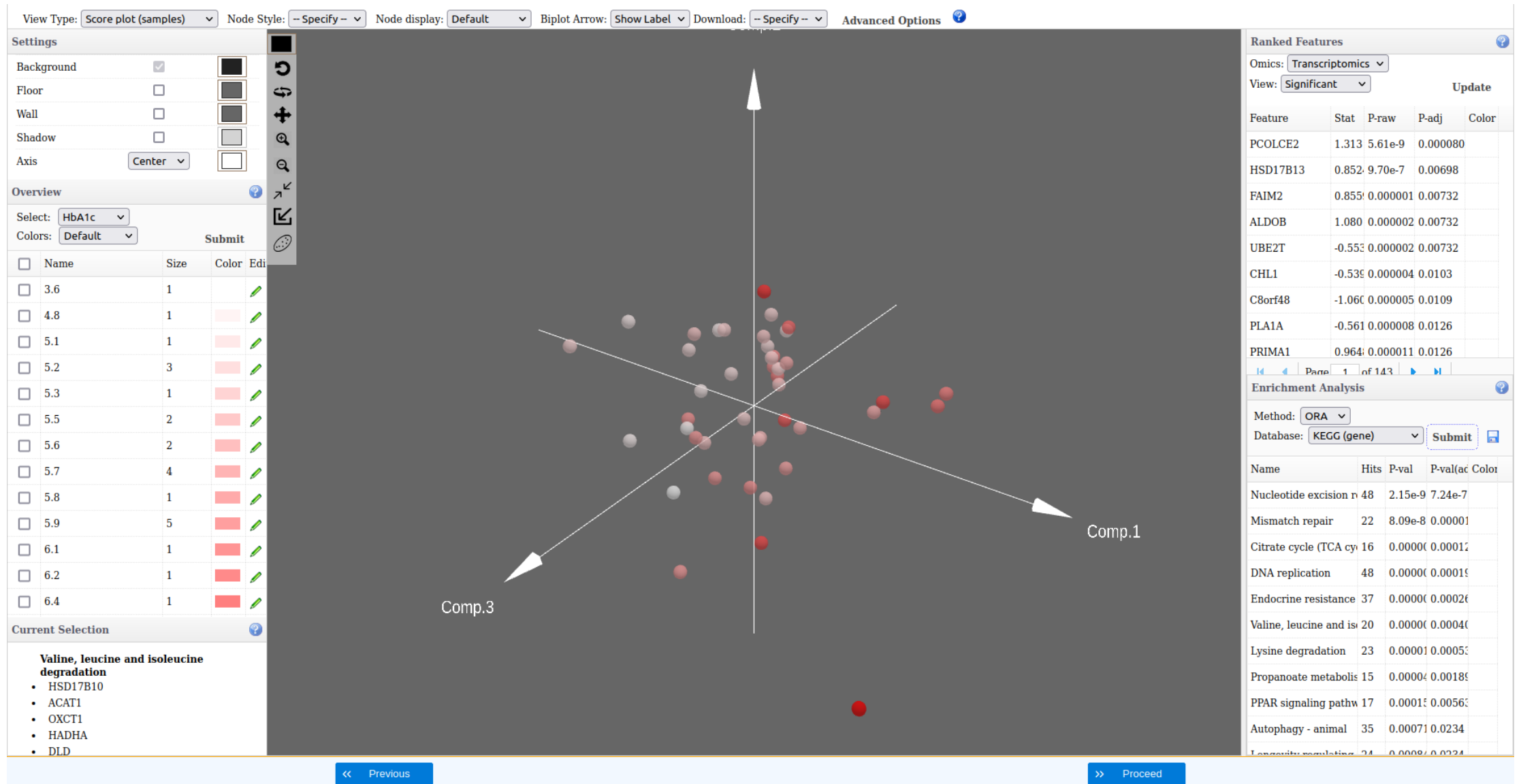
Enrichment Analysis

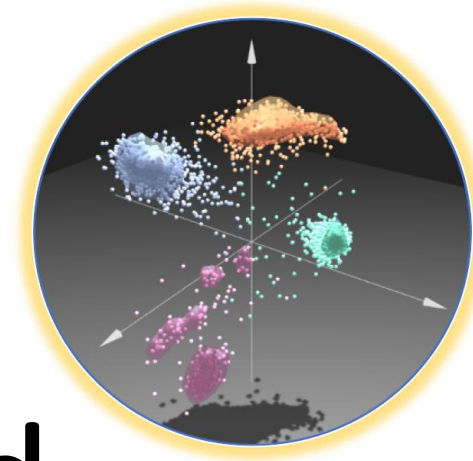
Method: **ORA** Database: **KEGG (gene)** **Submit**

Name	Hits	P-val	P-val(ac Color)
Nucleotide excision r	48	2.15e-9	7.24e-7
Mismatch repair	22	8.09e-8	0.00001
Citrate cycle (TCA cy	16	0.00000	0.00012
DNA replication	48	0.00000	0.00019
Endocrine resistance	37	0.00000	0.00026
Valine, leucine and is	20	0.00000	0.00040
Lysine degradation	23	0.00001	0.00052
Propanoate metabolis	15	0.00004	0.00189
PPAR signaling pathw	17	0.00012	0.00562
Autophagy - animal	35	0.00071	0.0234
Longevity regulatin	24	0.00084	0.0234

Previous **Proceed**

Color nodes by continuous metadata





The End

For more information, visit the **FAQs, Tutorials, Resources**
and **Contact** pages on www.omicsanalyst.ca