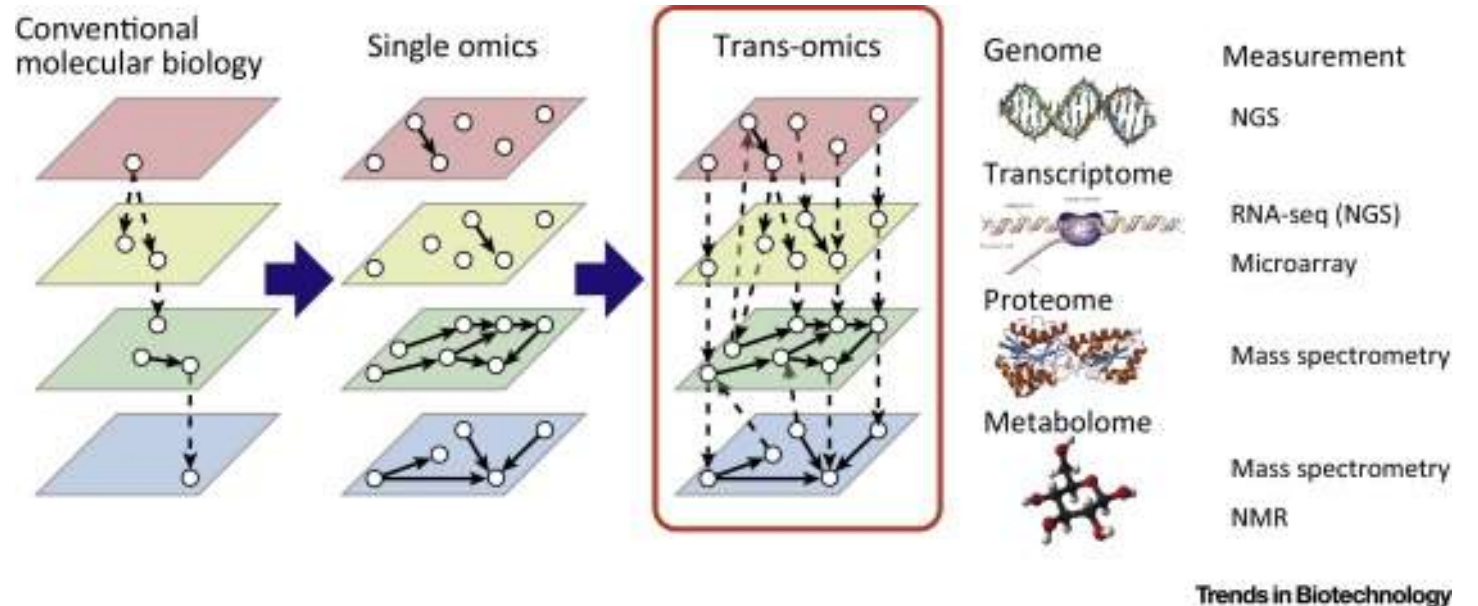


# OmicAnalyst

- a comprehensive platform for visual analytics of multi-omics data

# Multi-omics integration

- Disease subtyping, patient stratification and classification based on multi-omics signatures
- Biomarker prediction for diagnostics and identifying driver features for diseases
- Gain insights into disease biology through understanding interplay of molecules



Y, Katsuyuki, et al. *Trends in biotechnology*, 2016

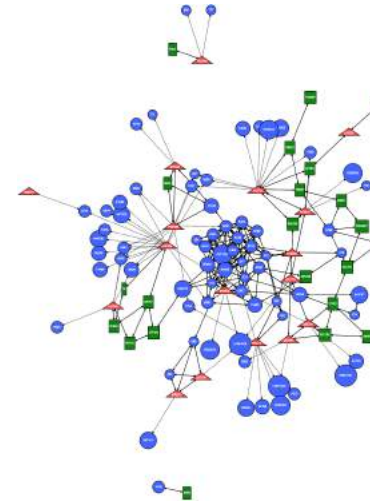
# Integration strategies

- Knowledge-driven:
  - Perform differential analysis separately on each 'omics dataset
  - Integrate lists of significant features using existing knowledge (typical pathway libraries)
  - Disadvantage: limits novel discovery because integration is constrained to relationships within the knowledge base
- Data-driven:
  - Use multi-variate statistical approaches to directly integrate and find relationships between multiple 'omics datasets
  - Less biased and more systematic
  - Disadvantage: methods are more difficult to implement and results are harder to visualize and interpret

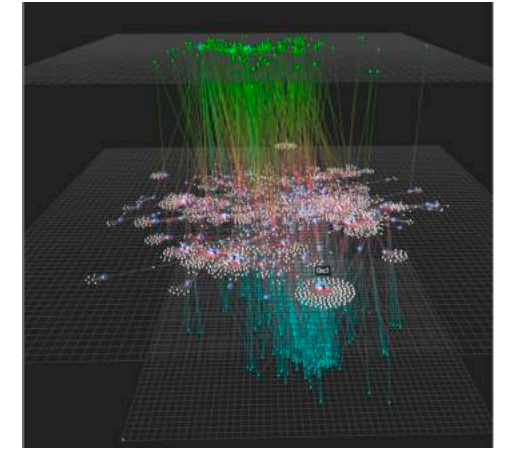
# Existing web-based tools

## Existing tools are largely knowledge-driven:

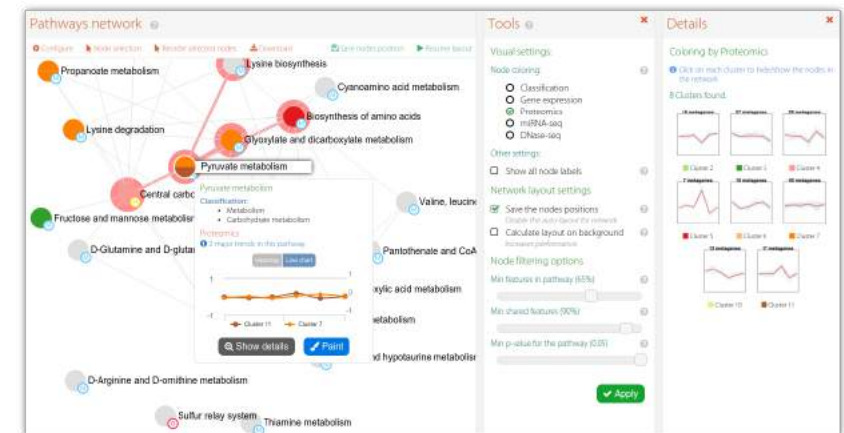
- **3Omics**: Correlation and co-expression analysis coupled with functional enrichment and pathway analysis
- **Paintomics**: Mapping to KEGG pathways coupled with visualization and functional enrichment
- **OmicsNet**: Mapping to molecular interactions for 3D network visual analytics coupled with functional enrichment analysis
- **GeneTrail2**: Multiple-omics integration coupled with enrichment analysis and identification of dysregulated subnetwork



3Omics



OmicsNet



PaintOmics

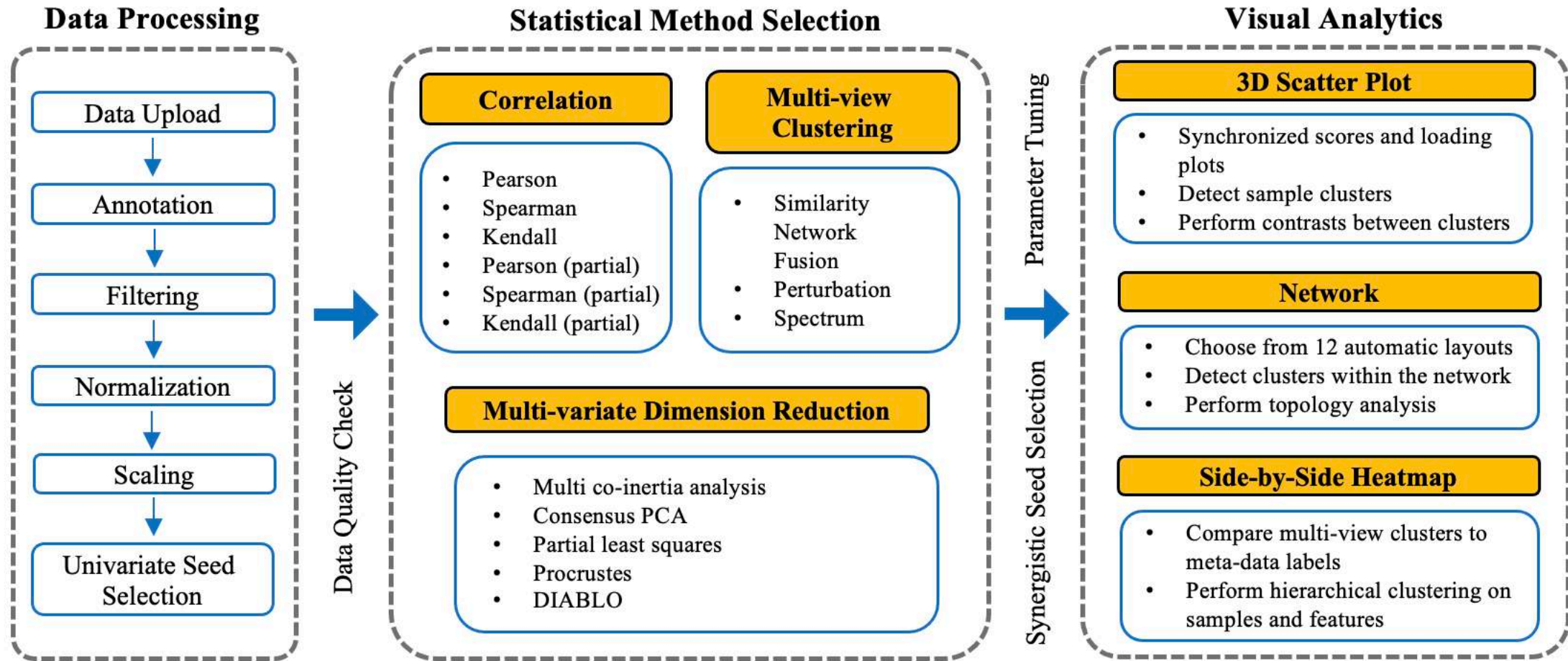
# Overview

- **Gap:**
  - Existing tools for data-driven multi-omics integration are standalone and require programming knowledge
  - Difficult to visualize and interpret results from multivariate methods
- **Goal:**
  - Provide an intuitive platform for performing multi-omics integration and visualization in a data-driven and unbiased manner
    - Input: 'omics abundance tables and associated meta-data
    - Support transcriptomics, proteomics, metabolomics, microbiomics, and miRNA data.
- **Approach:**
  - Couple multivariate and univariate statistical analysis with visual analytics

# Projected applications

- Explore inherent **trends** and **patterns** in multi-omics data and whether samples cluster according to biological condition
  - Heatmap viewer, 3D scatterplot
- Identify **correlated** features between two datasets
  - PLS, DIABLO, CPCA, MCIA, univariate, and partial correlation networks
- Identify potential **biomarker** features
  - DIABLO, differential analysis, dimension reduction loading scores
- Identify multi-view **sample clusters** from multiple 'omics data
  - SNF, Spectrum, and Perturbation methods

# General workflow



# Requirements

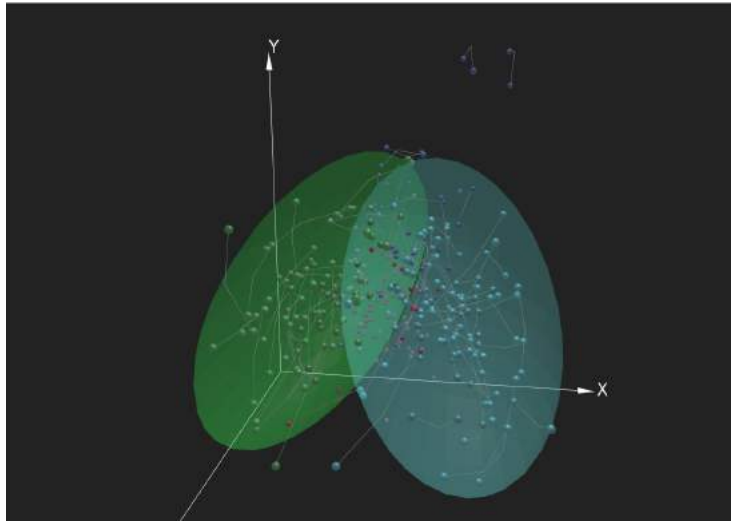
- Modern browser supporting WebGL.
- Ensure that WebGL is enabled in your browser!
  - Go to this page <https://get.webgl.org/> to verify your browser supports WebGL.
  - Refer to the FAQs for instructions on how to enable WebGL.
- For the best performance and visualization, use the latest version of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



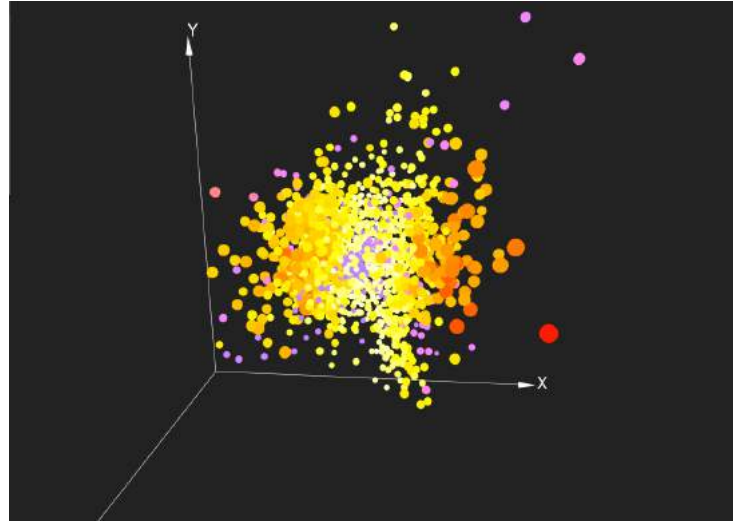
# Multi-variate dimension reduction

- OmicsAnalyst supports several well-established algorithms for integrating multiple-omics data
  - Dimensionality reduction to visualize hidden patterns and clusters
  - Identify sets of features that maximize correlation between omics datasets
- Both supervised and unsupervised methods
  - **Supervised:** DIABLO
  - **Unsupervised:** PLS, MCIA, CPCA
- Visualize overall sample and feature space using synchronized 3D scatter plot
- Visualize correlation structure of top dimensions using network viewer

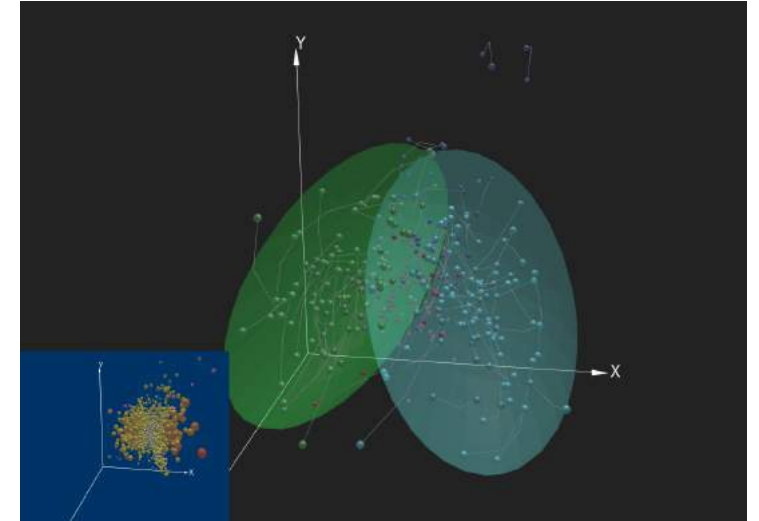
# Synchronized 3D scatterplot



**Sample space** (score plot)  
to identify global patterns  
and clusterings



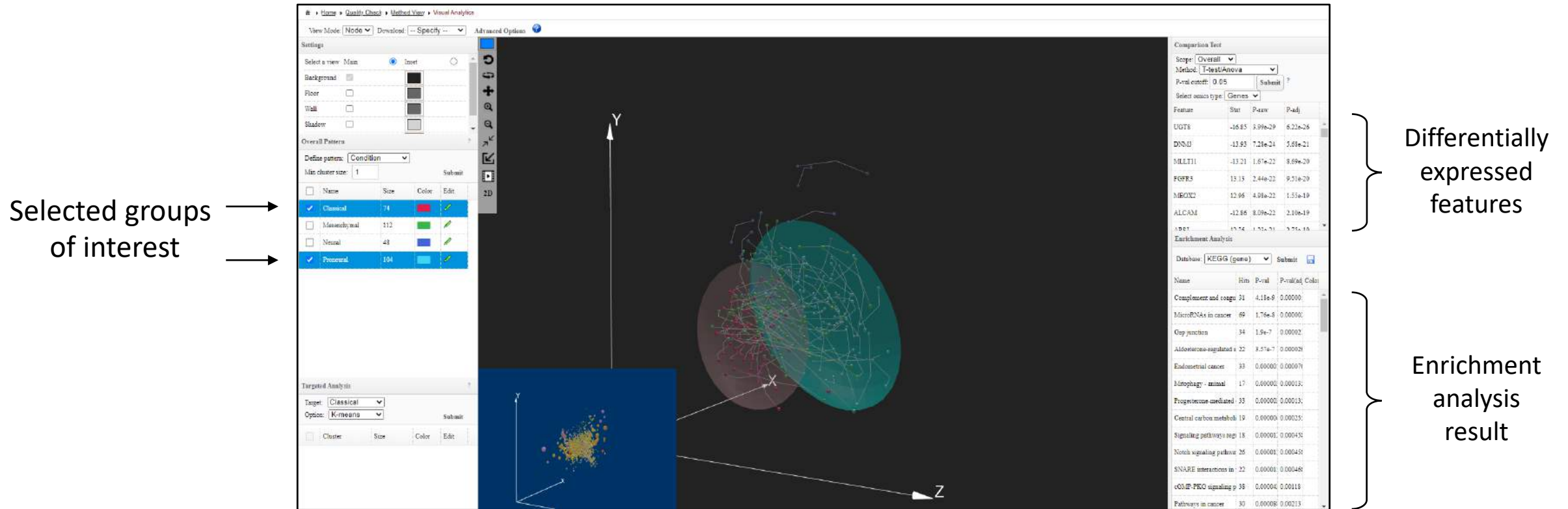
**Feature space** (loadings plot)  
to enable targeted analytics of  
important features



**Joint space** coupled with  
synchronized score and loadings  
plot to identify features driving  
sample separation

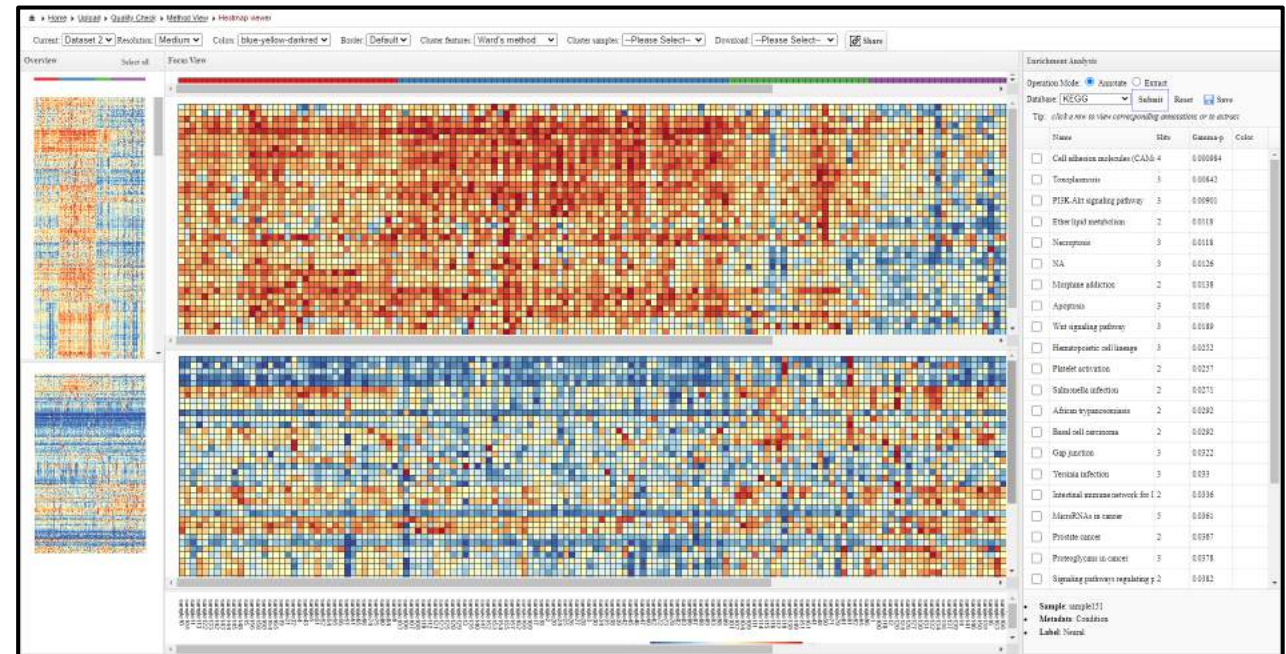
# Targeted analysis

- Flexible comparison tests to identify differentially abundant features between different experimental groups and clusters.
  - Kruskal-Wallis, Student T-test, ANOVA, Limma, EdgeR



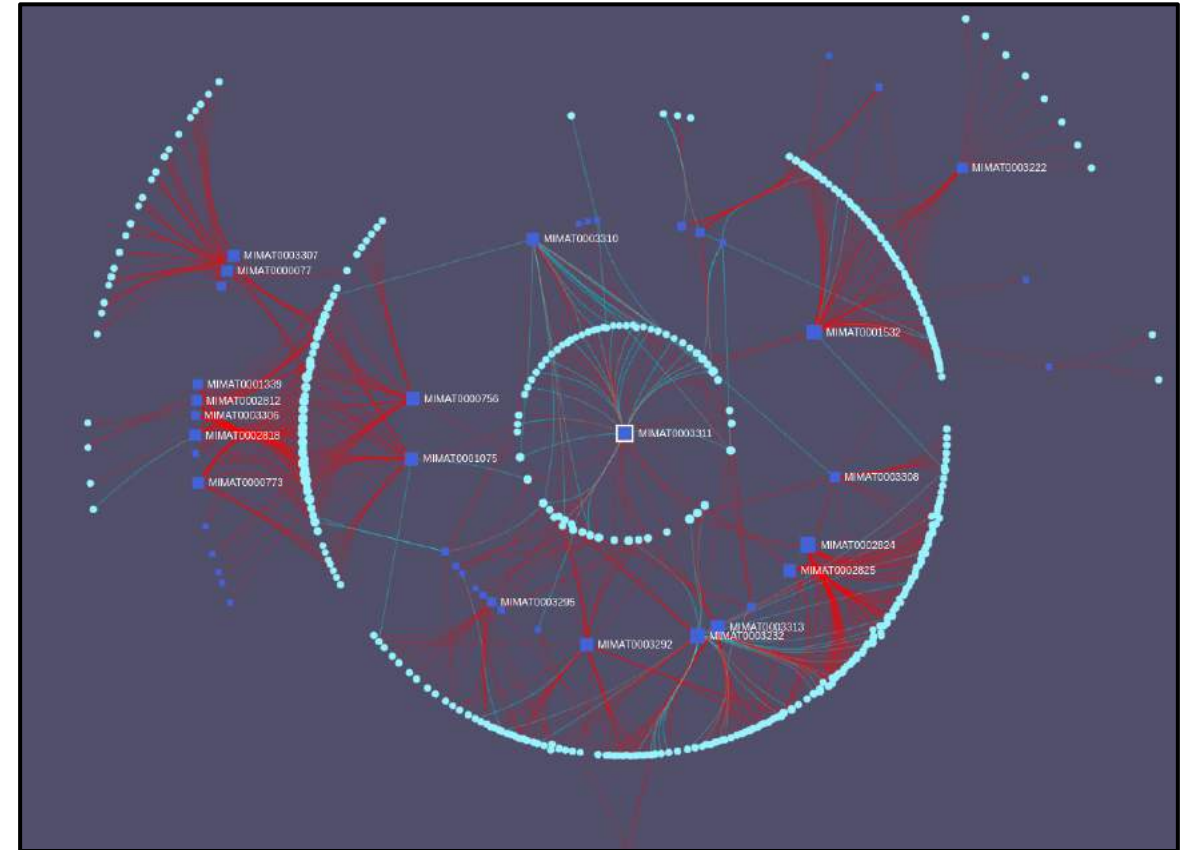
# Multi-view clustering and heatmap

- Multi-view clustering groups samples using multiple data “views” of the data, for example multiple ‘omics datasets
- Combine detected sample clusters with flexible hierarchical clustering in side-by-side heatmaps
- Visually identify discrete clusters and patterns that would be difficult to find using statistics alone

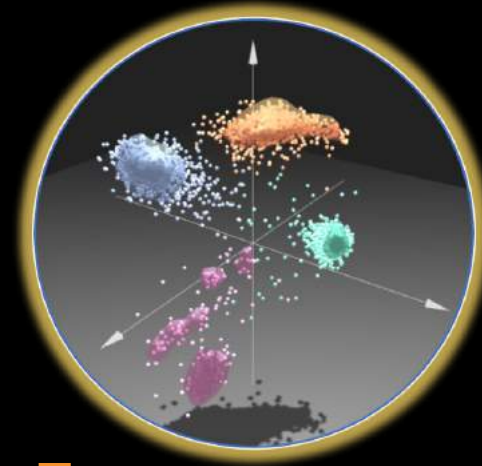


# Correlation network

- Start with seeds selected using multi-variate or univariate methods
- Well-established univariate and multi-variate (partial) correlation methods to identify pairs of highly related features across omics layers
  - **Parametric:** Pearson
  - **Non-parametric:** Spearman, Kendall
- Visualize the most significant connections using the network viewer



Correlation network displayed using concentric circle layout



# The End

For more information, visit the **FAQs, Tutorials, Resources**  
and **Contact** pages on [www.omicsanalyst.ca](http://www.omicsanalyst.ca)