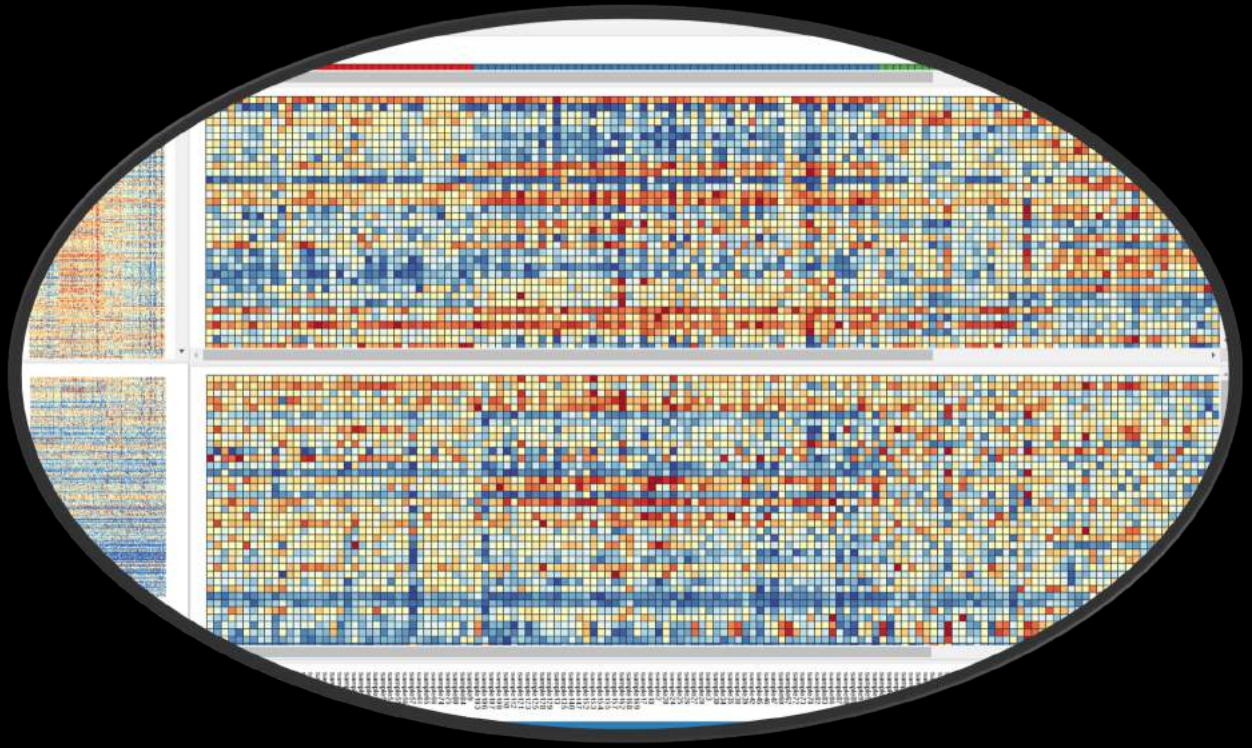
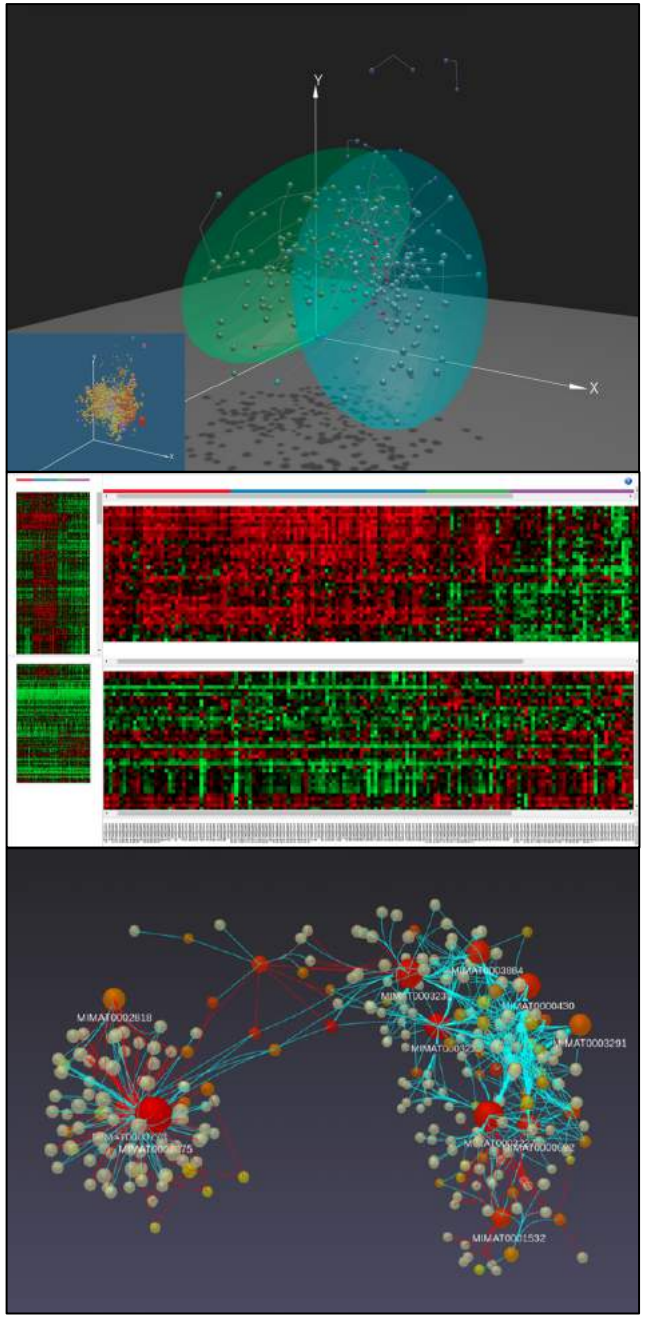
TUTORIAL 3:
Cluster analysis + dual heatmap



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

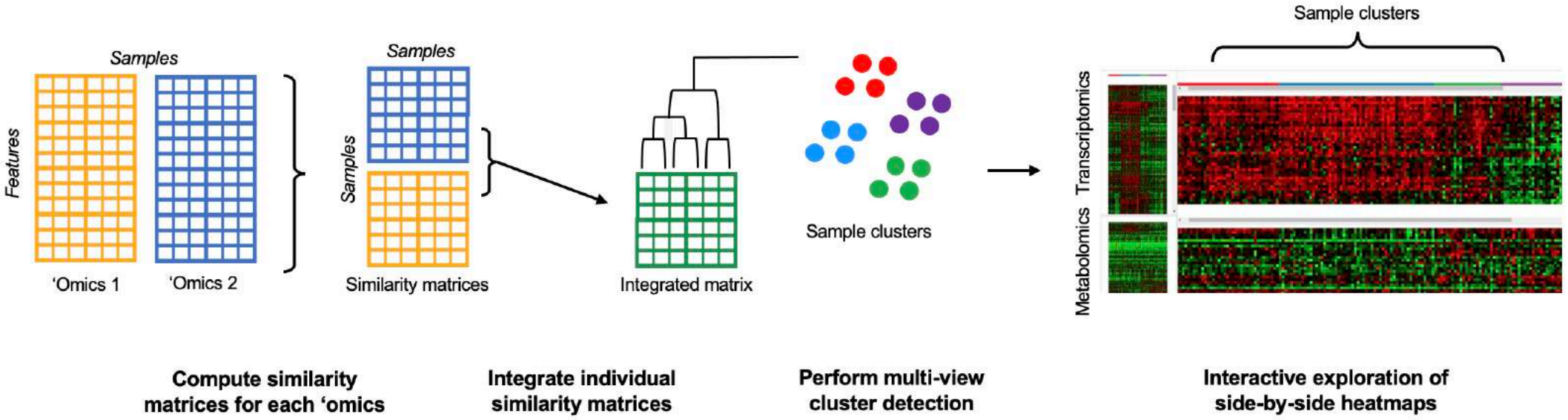
- Modern browser supporting WebGL.
- Ensure that WebGL is enabled in your browser!
  - Go to this page https://get.webgl.org/ to verify your browser supports WebGL.
  - Refer to the FAQs for instructions on how to enable WebGL.
- For the best performance and visualization, use the latest version of Google 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
- Dual heatmap viewer enable the simultaneous visualization of two feature layers to facilitate comparisons of global patterns
- The goal of this tutorial is to:
  - Introduce OmicsAnalyst's tools for heatmap visual analytics
  - Targeted analysis by performing functional enrichment analysis on selected regions from the heatmap

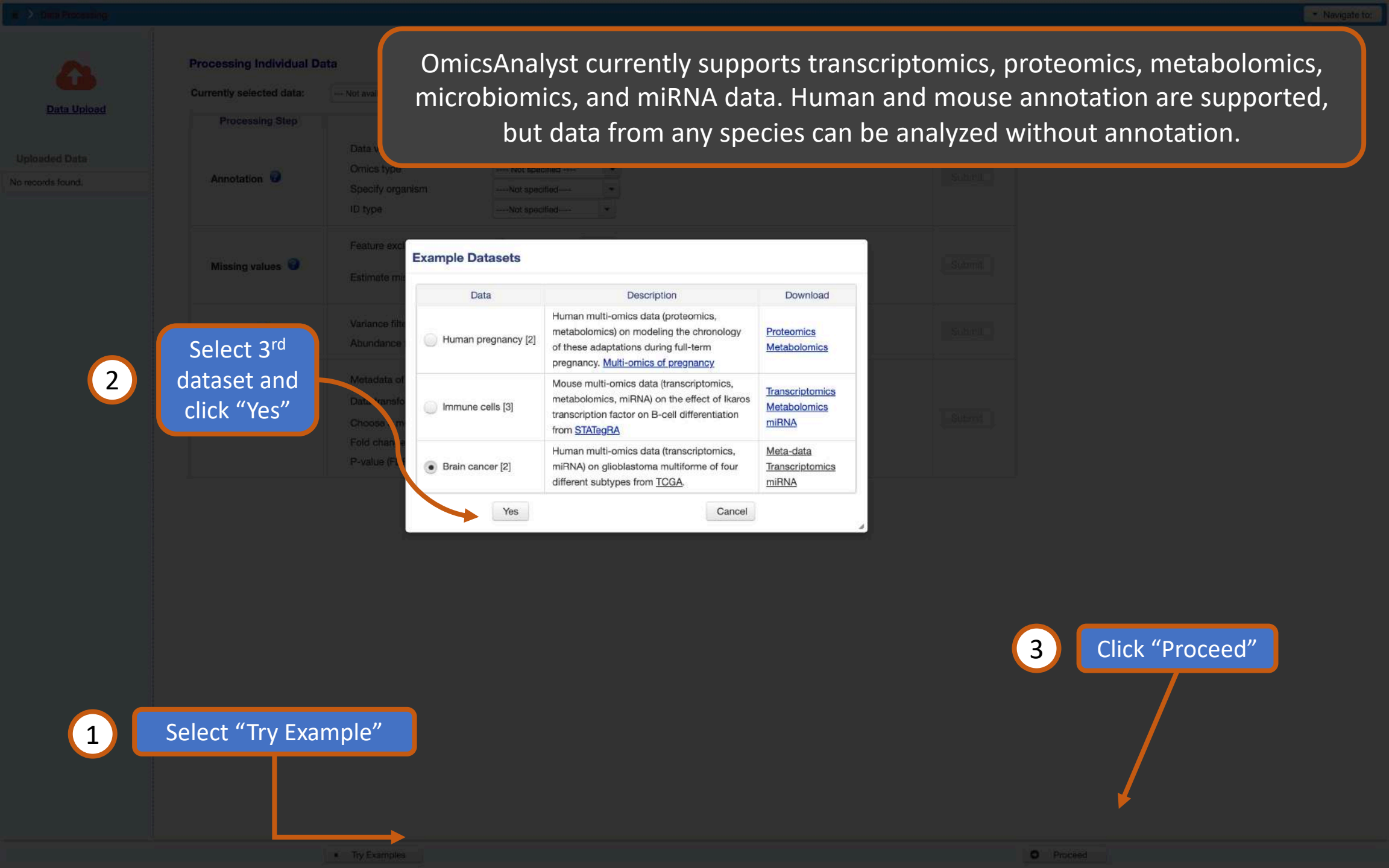
Sample workflow

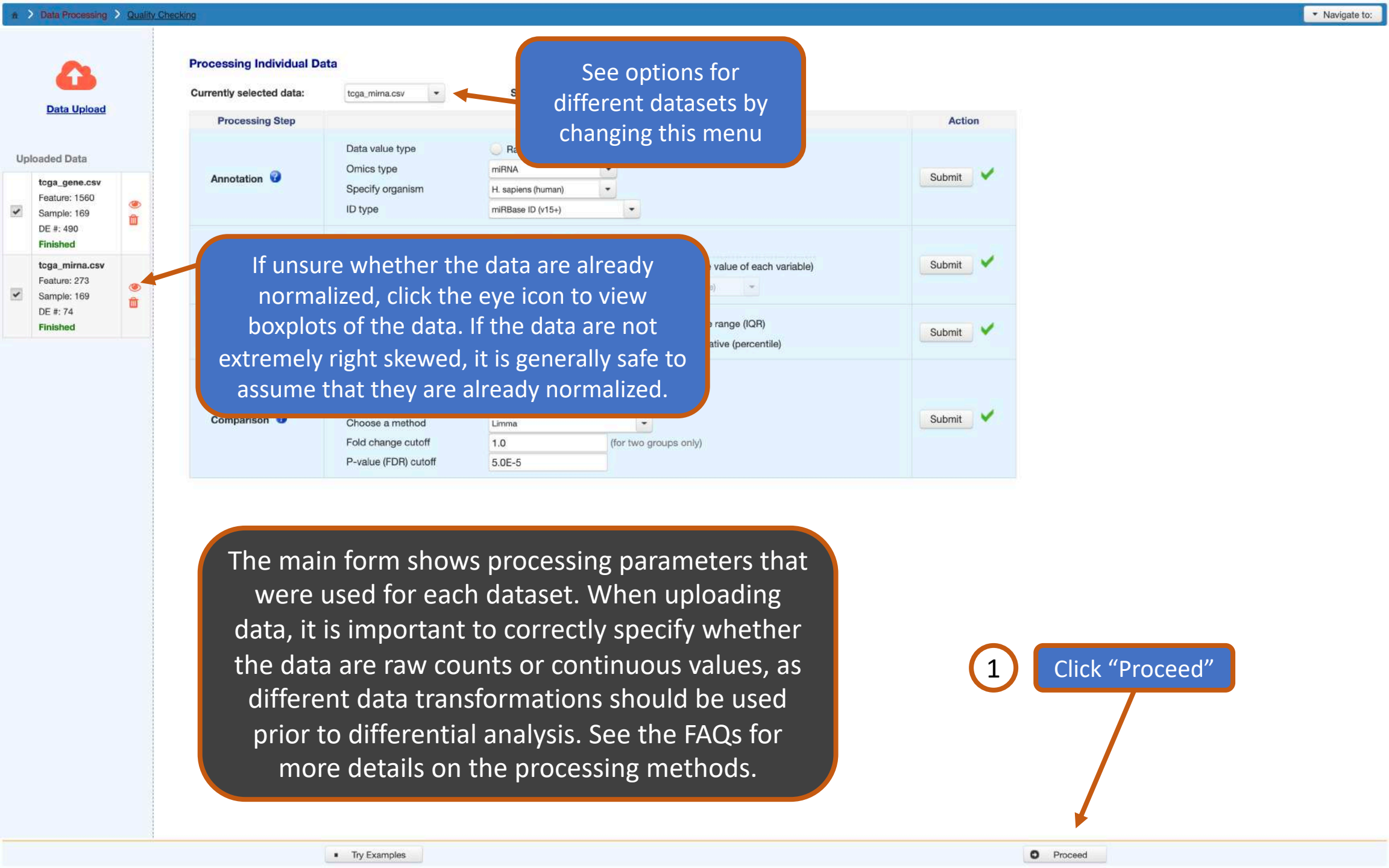
- An overview of this analysis track is:



- To start the tutorial, click **GET STARTED** on the homepage

OmicsAnalyst currently supports transcriptomics, proteomics, metabolomics, microbiomics, and miRNA data. Human and mouse annotation are supported, but data from any species can be analyzed without annotation.




Data Processing
Quality Checking
Navigate to:
Data Upload
Uploaded Data
tcga_gene.csv
Feature: 1560
Sample: 169
DE #: 490
Finished
tcga_mirna.csv
Feature: 273
Sample: 169
DE #: 74
Finished
Processing Individual Data
Currently selected data:
tcga_mirna.csv
See options for different datasets by changing this menu
Processing Step
Action
Annotation
Data value type
Omics type
miRNA
Specify organism
H. sapiens (human)
ID type
miRBase ID (v15+)
Submit
If unsure whether the data are already normalized, click the eye icon to view boxplots of the data. If the data are not extremely right skewed, it is generally safe to assume that they are already normalized.
Comparison
Choose a method
Limma
Fold change cutoff
1.0
(for two groups only)
P-value (FDR) cutoff
5.0E-5
The main form shows processing parameters that were used for each dataset. When uploading data, it is important to correctly specify whether the data are raw counts or continuous values, as different data transformations should be used prior to differential analysis. See the FAQs for more details on the processing methods.
1
Click “Proceed”
Try Examples
Proceed

The page provides graphics to ensure that the data has been properly normalized. Since different transformation may be preferred compared to differential analysis, you can update the method here. Scaling the datasets so that have comparable distributions across ‘omics types is recommended.

Uploaded Data

tcga_gene.csv
Feature: 1560
Sample: 169
DE #: 490

tcga_mirna.csv
Feature: 273
Sample: 169
DE #: 74

See the number of differential features from each dataset. Click the icon to update thresholds.

Data Quality Check

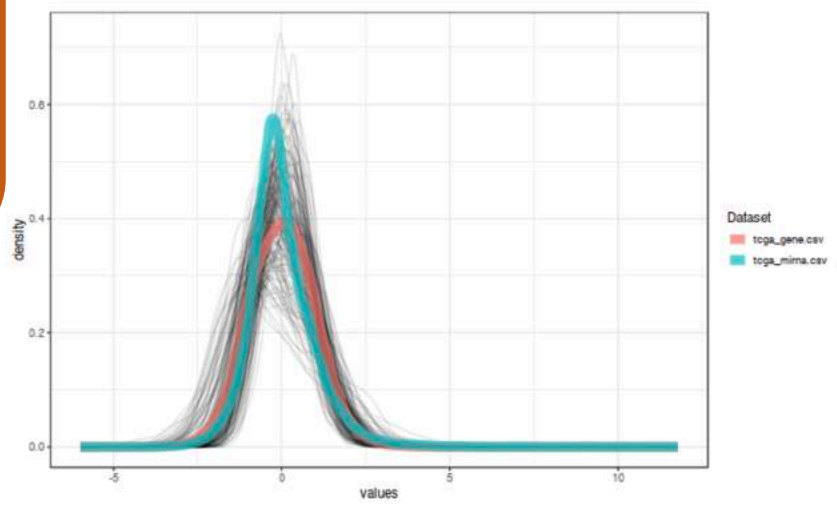
The uploaded omics datasets are summarized below:

Total number of samples:	169
Group names:	Classical; Mesenchymal; Neural; Proneural
Individual datasets:	tcga_gene.csv; tcga_mirna.csv
Corresponding feature number:	1560; 273

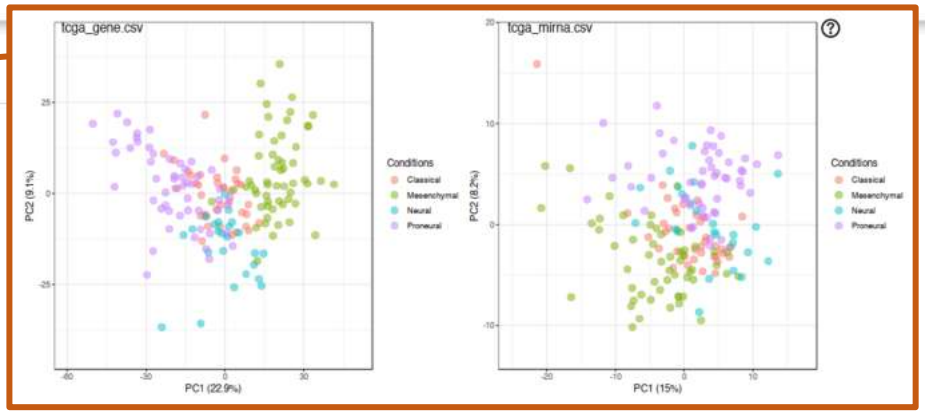
The density plot and PCA plots are generated to provide an overview of the omics datasets for qual
the distributions across different omics layers. If the overall distribution seems to be in very different ran
accounting for batch effect prior to uploading your data to OmicsAnalyst, for example with the R package COMBAT.

Current omics data: tcga_mirna.csv
Sample normalization: None
Data transformation: None
Data scaling: Auto scaling
Update

Density plot | PCA plot | t-SNE plot



1 Look at the “PCA plot”



2 Click “Proceed”

Previous | Proceed

Data Processing > Quality Checking > Method Selection

Navigate to:

Uploaded Data

tcga_gene.csv
Feature: 1560
Sample: 169
DE #: 490

tcga_mirna.csv
Feature:
Samp
DE #:

Please choose a method to proceed

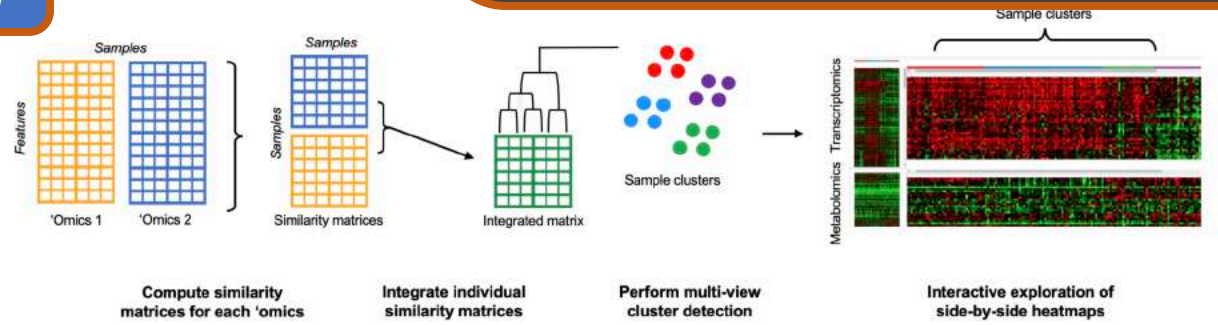
Correlation Network Analysis | Cluster Heatmap Analysis | Dimensior

The objective of this analysis is to understand relationships between samp
that integrate information from all 'omics datasets. Next, interactive heatm
d to either the detected clusters or the experime
le of the three visual analytics tools in OmicsAn

Go to the "Cluster Heatmap Analysis" tab

1

The clustering methods in this analysis track are "multi-view", in that they integrate information from multiple data views, or 'omics layers, prior to detecting clusters of samples. Select each method from the drop down menu to read a short description. To move on to the heatmap without detecting sample clusters, select the "Free Exploration" option.



Cluster analysis method: Similarity Network Fusion | Proceed

Click "Proceed"

3

Similarity Network Fusion (SNF) generates an integrated sample similarity matrix from multiple 'omics datasets by first computing similarity matrices for each dataset individually, and then fusing them together. Individual similarity matrices are computed using an exponential similarity kernel that scales the Euclidean distance between samples. These matrices are then fused together by an iterative approach that adjusts each matrix to make it more similar to the others. The SNF algorithm is iterated until the matrices converge. OmicsAnalyst then uses the clustering method from the Spectrum R package to define sample clusters in the SNF matrix (select the Spectrum option for more details). (more details ....)

2

Select "Similarity Network Fusion"

This page allows you to refine key tuneable parameters from the selected statistical method (if any). Diagnostic plots are provided to help determine whether the parameters should be changed.

1. Data Processing
2. Quality Checking
3. Method Selection
4. Parameter Tuning
4. Visual Analytics
6. Result Download

Analysis Overview & Refinement

Users can let the algorithm identify optimal cluster number or manually select a number of cluster in which the algorithm will partition the sa
underlying statistics (i.e. eigenvalue, AUC) behind identification of optimal cluster number is displayed. Normalized Mutual Information (NM

Cluster Number 2 Update

Datasets

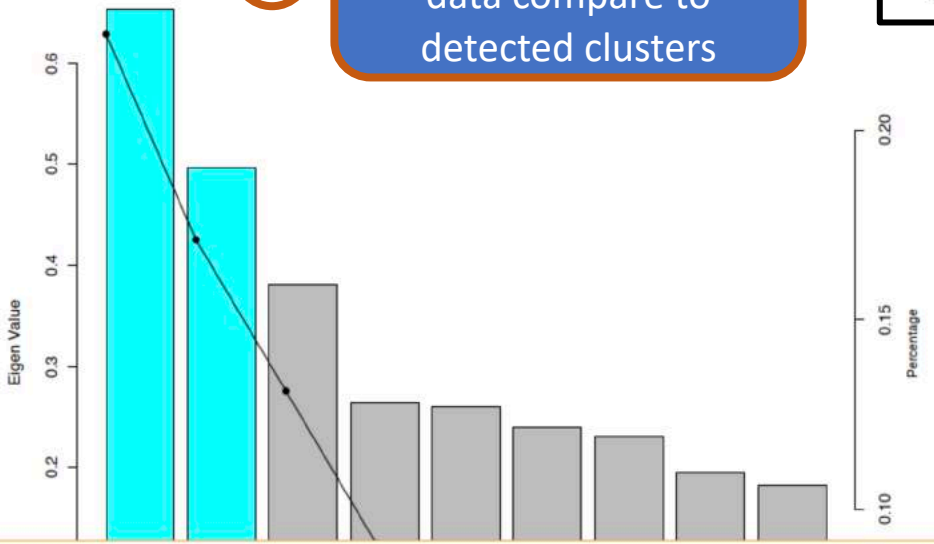
tcga_gene.csv
Feature: 1560
Sample: 169
DE #: 1494

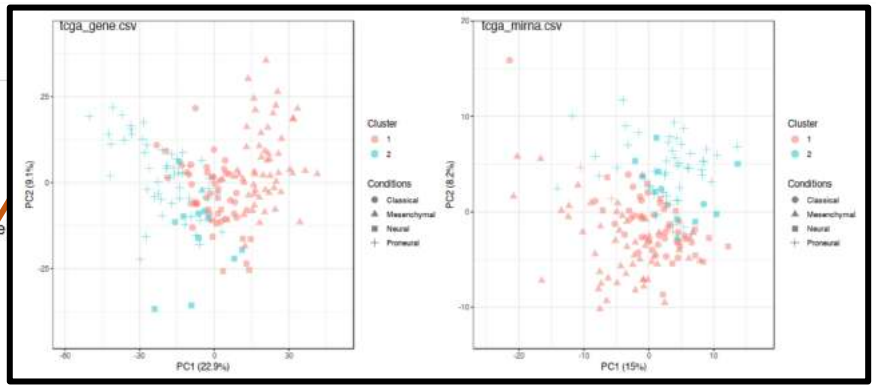
tcga_mirna.csv
Feature: 273
Sample: 169
DE #: 170

Diagnostic | PCA plot

Cluster number: 2.0
NMI: 0.445106

Diagnostic plot shows how eigenvalues relate to number of clus m is whe
eigenvalue happens.





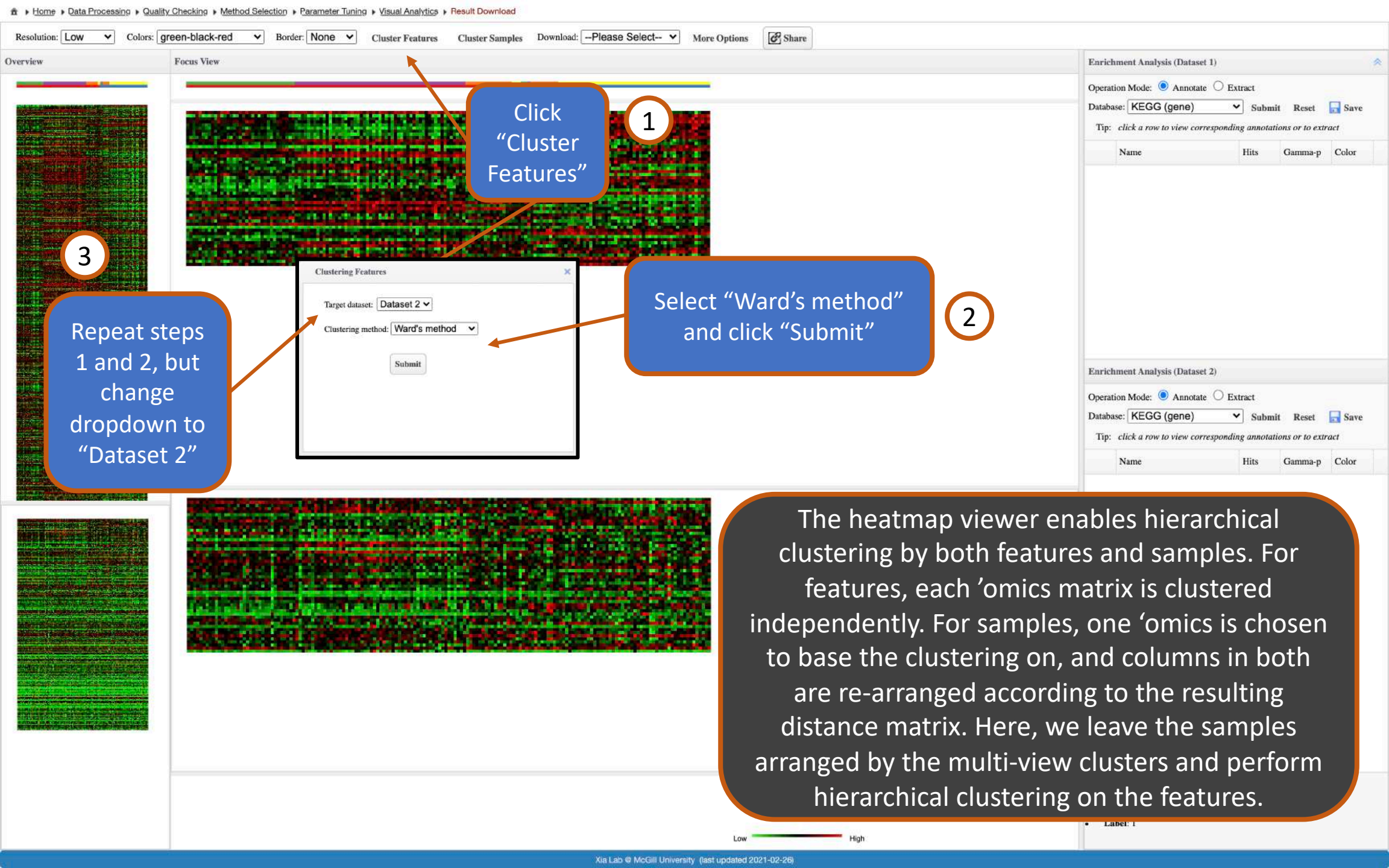
1

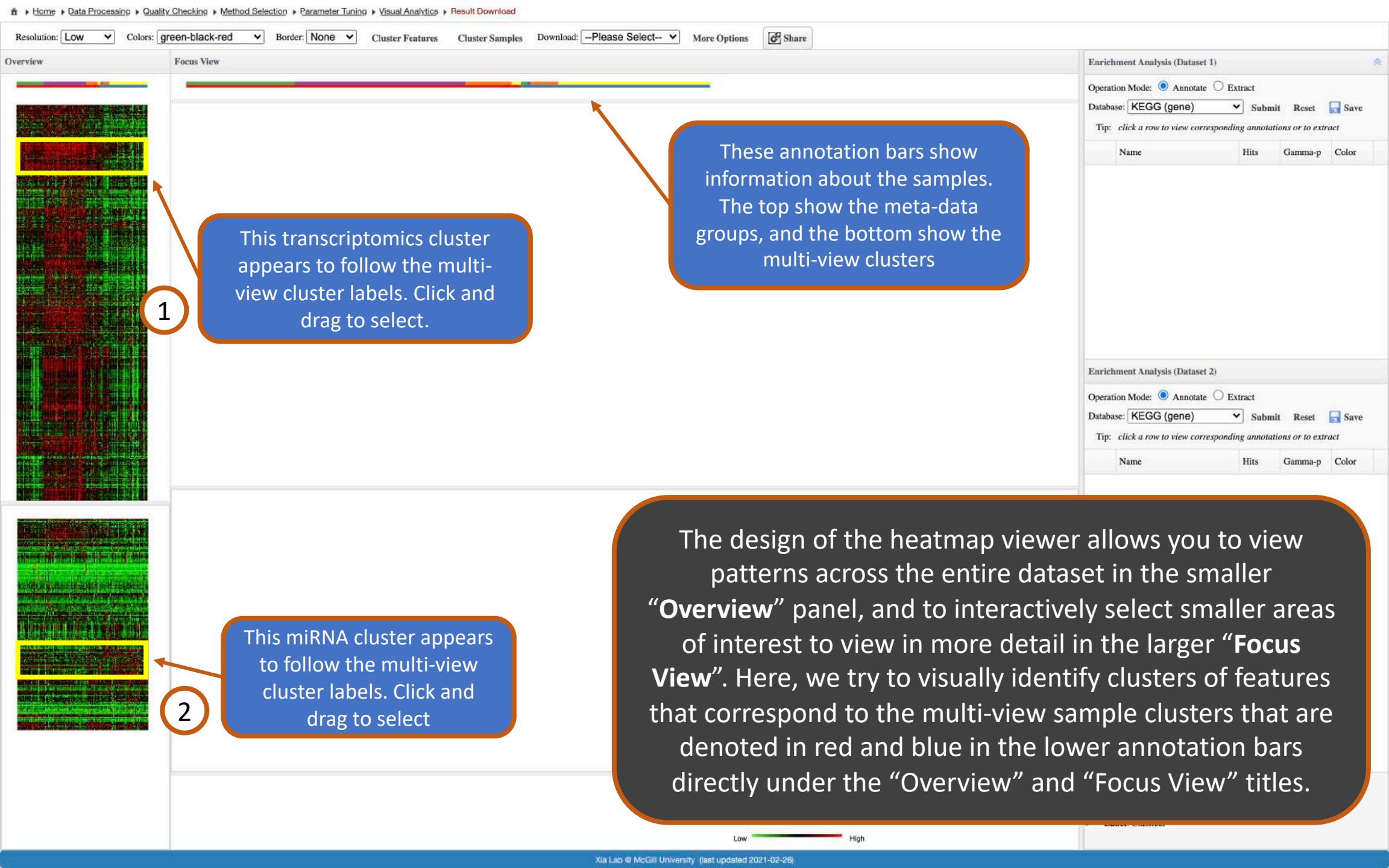
View the PCA plot to see how sample meta-data compare to detected clusters

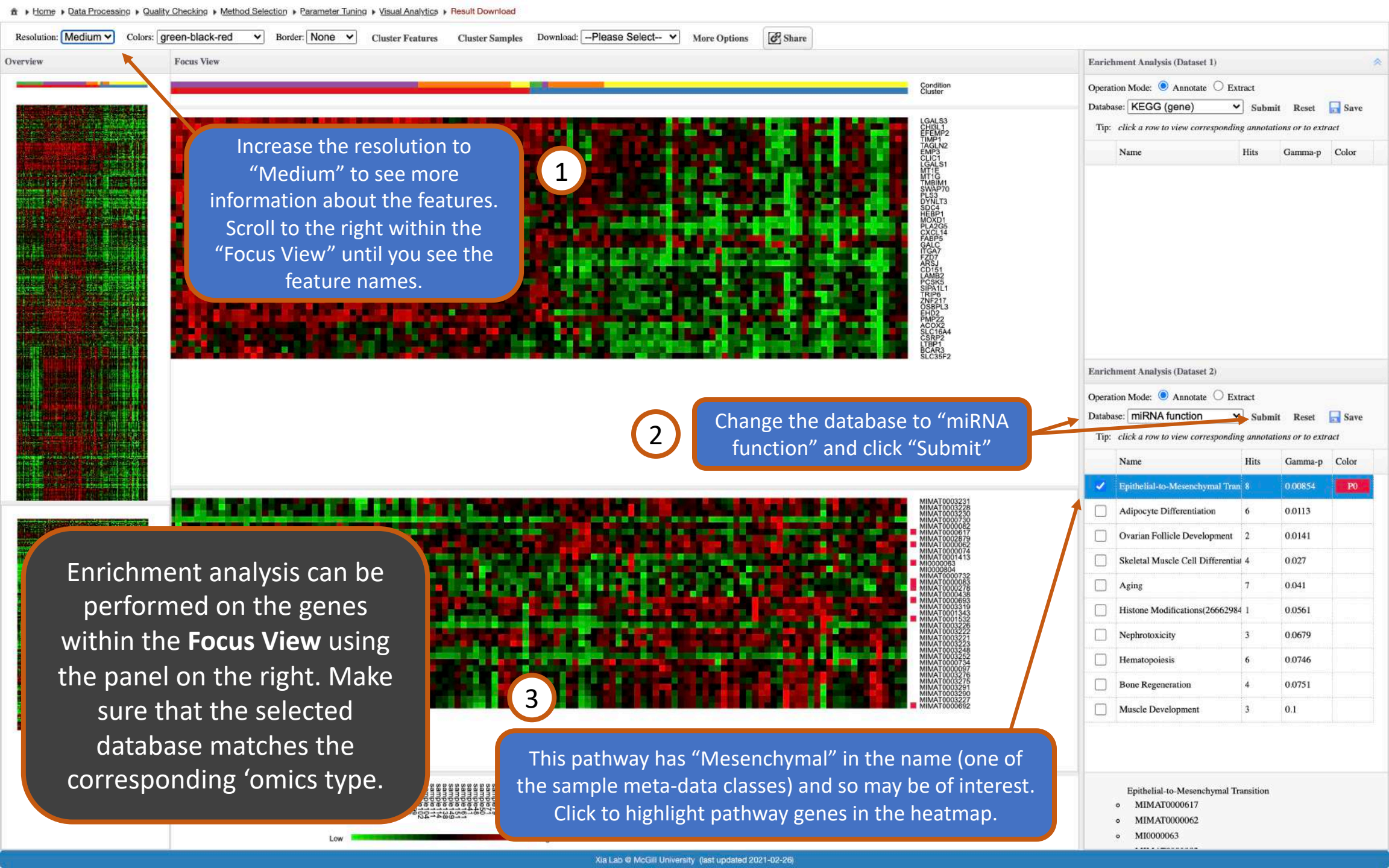
Click “Proceed” 2

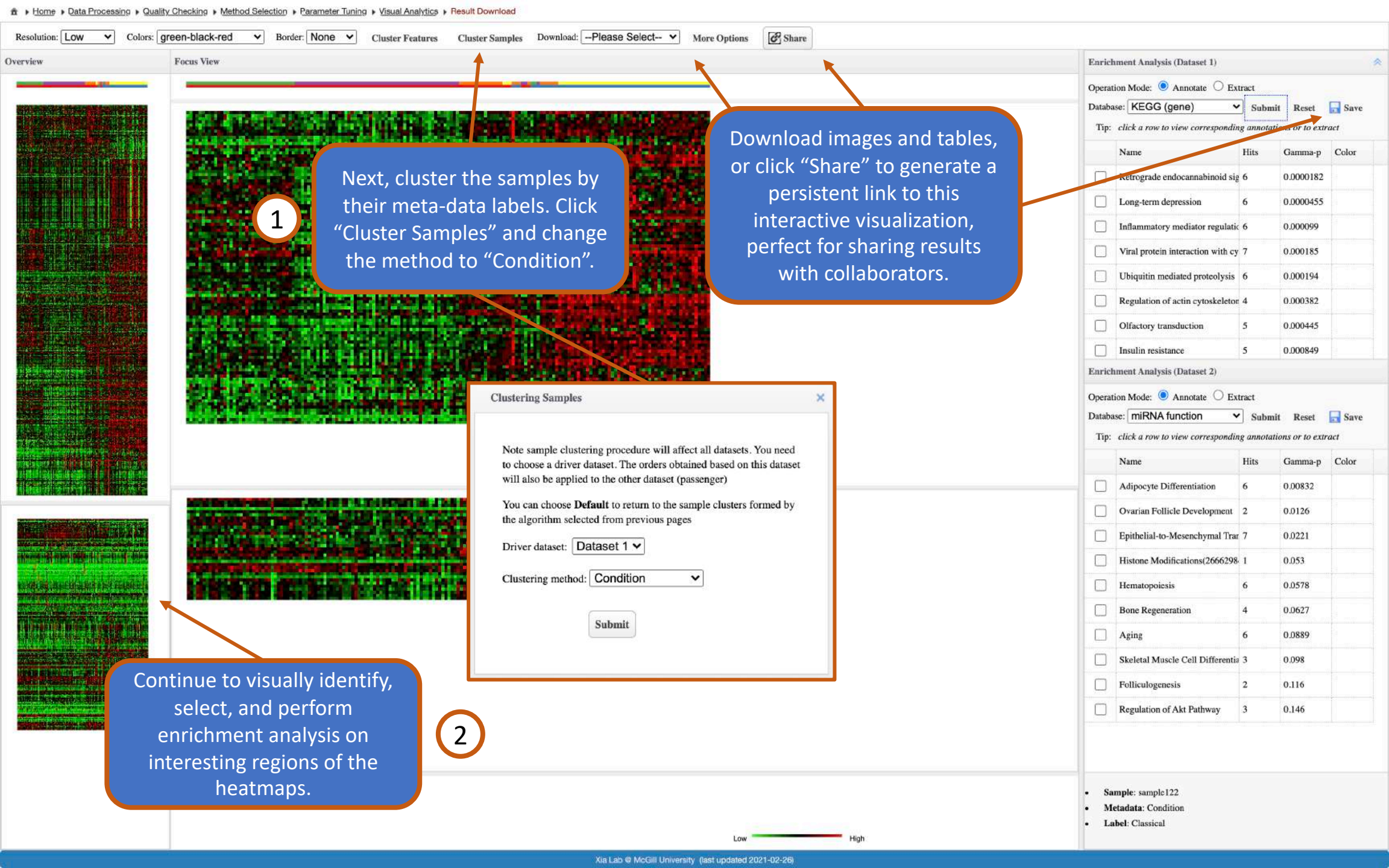
Previous Proceed

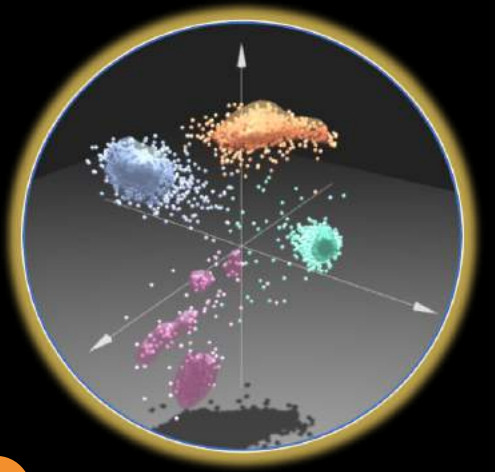

Home ▸ Data Processing ▸ Quality Checking ▸ Method Selection ▸ Parameter Tuning ▸ Visual Analytics ▸ Result Download
Resolution: Medium
Colors: navy-white-firebrick
Border: Default
Cluster Features
Cluster Samples
Download: --Please Select--
More Options
Share
Overview
Focus View
Change resolution to “Low”
1
Change colors to “green-black-red”
2
Remove border by selecting “None”
3
Enrichment Analysis (Dataset 1)
Operation Mode: Annotate Extract
Database: KEGG (gene)
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name
Hits
Gamma-p
Color
Enrichment Analysis (Dataset 2)
Operation Mode: Annotate Extract
Database: KEGG (gene)
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name
Hits
Gamma-p
Color
Low
High
Label:
Our heatmap viewer enables simultaneous visualization of two omics datasets, and allows targeted analysis on clusters of interest. There are numerous visualization options allowing users to customize the heatmap appearance.
Xia Lab @ McGill University (last updated 2021-02-26)


Home ▸ Data Processing ▸ Quality Checking ▸ Method Selection ▸ Parameter Tuning ▸ Visual Analytics ▸ Result Download
Resolution: Low
Colors: green-black-red
Border: None
Cluster Features
Cluster Samples
Download: --Please Select--
More Options
Share
Overview
Focus View
Enrichment Analysis (Dataset 1)
Operation Mode: Annotate Extract
Database: KEGG (gene)
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name
Hits
Gamma-p
Color
Click "Cluster Features"
1
Clustering Features
Target dataset: Dataset 2
Clustering method: Ward's method
Submit
Select "Ward's method" and click "Submit"
2
3
Repeat steps 1 and 2, but change dropdown to "Dataset 2"
Enrichment Analysis (Dataset 2)
The heatmap viewer enables hierarchical clustering by both features and samples. For features, each 'omics matrix is clustered independently. For samples, one 'omics is chosen to base the clustering on, and columns in both are re-arranged according to the resulting distance matrix. Here, we leave the samples arranged by the multi-view clusters and perform hierarchical clustering on the features.
Low
High
Xia Lab @ McGill University (last updated 2021-02-26)


Home ▸ Data Processing ▸ Quality Checking ▸ Method Selection ▸ Parameter Tuning ▸ Visual Analytics ▸ Result Download
Resolution: Low
Colors: green-black-red
Border: None
Cluster Features
Cluster Samples
Download: --Please Select--
More Options
Share
Overview
Focus View
Enrichment Analysis (Dataset 1)
Operation Mode: Annotate Extract
Database: KEGG (gene)
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name
Hits
Gamma-p
Color
Enrichment Analysis (Dataset 2)
Operation Mode: Annotate Extract
Database: KEGG (gene)
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name
Hits
Gamma-p
Color
These annotation bars show information about the samples. The top show the meta-data groups, and the bottom show the multi-view clusters
This transcriptomics cluster appears to follow the multi-view cluster labels. Click and drag to select.
1
This miRNA cluster appears to follow the multi-view cluster labels. Click and drag to select
2
The design of the heatmap viewer allows you to view patterns across the entire dataset in the smaller "Overview" panel, and to interactively select smaller areas of interest to view in more detail in the larger "Focus View". Here, we try to visually identify clusters of features that correspond to the multi-view sample clusters that are denoted in red and blue in the lower annotation bars directly under the "Overview" and "Focus View" titles.
Low
High
Xia Lab @ McGill University (last updated 2021-02-26)


Home ▸ Data Processing ▸ Quality Checking ▸ Method Selection ▸ Parameter Tuning ▸ Visual Analytics ▸ Result Download
Resolution: Medium
Colors: green-black-red
Border: None
Cluster Features
Cluster Samples
Download: --Please Select--
More Options
Share
Overview
Focus View
Condition
Cluster
Increase the resolution to "Medium" to see more information about the features. Scroll to the right within the "Focus View" until you see the feature names.
1
LGALS3
CHI3L1
EFEMP2
TIMP1
TAGLN2
EMP3
CLIC1
LGALS1
MT1E
MT1G
TMBIM1
SWAP70
PLS3
DYNLT3
SDC4
HEBP1
MOXD1
PLA2G5
CXCL14
FABP5
GALC
ITGA7
FZD7
ARSJ
CD151
LAMB2
PCSK5
SIPA1L1
TRIP6
ZNF217
OSBPL3
EHD2
PMP22
ACOX2
SLC16A4
CSRP2
LTBP1
BCAR3
SLC35F2
Enrichment Analysis (Dataset 1)
Operation Mode: Annotate Extract
Database: KEGG (gene)
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name
Hits
Gamma-p
Color
2
Change the database to "miRNA function" and click "Submit"
Enrichment Analysis (Dataset 2)
Operation Mode: Annotate Extract
Database: miRNA function
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name
Hits
Gamma-p
Color
Epithelial-to-Mesenchymal Tran 8 0.00854 P0
Adipocyte Differentiation 6 0.0113
Ovarian Follicle Development 2 0.0141
Skeletal Muscle Cell Differentiat 4 0.027
Aging 7 0.041
Histone Modifications(26662984 1 0.0561
Nephrotoxicity 3 0.0679
Hematopoiesis 6 0.0746
Bone Regeneration 4 0.0751
Muscle Development 3 0.1
Enrichment analysis can be performed on the genes within the Focus View using the panel on the right. Make sure that the selected database matches the corresponding 'omics type.
3
This pathway has "Mesenchymal" in the name (one of the sample meta-data classes) and so may be of interest. Click to highlight pathway genes in the heatmap.
Epithelial-to-Mesenchymal Transition
MIMAT0000617
MIMAT0000062
MI0000063
Low
Xia Lab @ McGill University (last updated 2021-02-26)


Home ▸ Data Processing ▸ Quality Checking ▸ Method Selection ▸ Parameter Tuning ▸ Visual Analytics ▸ Result Download
Resolution: Low
Colors: green-black-red
Border: None
Cluster Features
Cluster Samples
Download: --Please Select--
More Options
Share
Overview
Focus View
1
Next, cluster the samples by their meta-data labels. Click “Cluster Samples” and change the method to “Condition”.
Download images and tables, or click “Share” to generate a persistent link to this interactive visualization, perfect for sharing results with collaborators.
Clustering Samples
Note sample clustering procedure will affect all datasets. You need to choose a driver dataset. The orders obtained based on this dataset will also be applied to the other dataset (passenger)
You can choose Default to return to the sample clusters formed by the algorithm selected from previous pages
Driver dataset: Dataset 1
Clustering method: Condition
Submit
Continue to visually identify, select, and perform enrichment analysis on interesting regions of the heatmaps.
2
Low
High
Enrichment Analysis (Dataset 1)
Operation Mode: Annotate Extract
Database: KEGG (gene)
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name Hits Gamma-p Color
Retrograde endocannabinoid sig 6 0.0000182
Long-term depression 6 0.0000455
Inflammatory mediator regulatic 6 0.000099
Viral protein interaction with cy 7 0.000185
Ubiquitin mediated proteolysis 6 0.000194
Regulation of actin cytoskeletor 4 0.000382
Olfactory transduction 5 0.000445
Insulin resistance 5 0.000849
Enrichment Analysis (Dataset 2)
Operation Mode: Annotate Extract
Database: miRNA function
Submit
Reset
Save
Tip: click a row to view corresponding annotations or to extract
Name Hits Gamma-p Color
Adipocyte Differentiation 6 0.00832
Ovarian Follicle Development 2 0.0126
Epithelial-to-Mesenchymal Trar 7 0.0221
Histone Modifications(2666298 1 0.053
Hematopoiesis 6 0.0578
Bone Regeneration 4 0.0627
Aging 6 0.0889
Skeletal Muscle Cell Differentia 3 0.098
Folliculogenesis 2 0.116
Regulation of Akt Pathway 3 0.146
Sample: sample122
Metadata: Condition
Label: Classical
Xia Lab @ McGill University (last updated 2021-02-26)



THE END

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

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