XiaLab Analytics

Empowering researchers through trainings, tools and AI

https://www.xialab.ca • contact@xialab.ca

Omics Data Science Training Course

Winter 2024

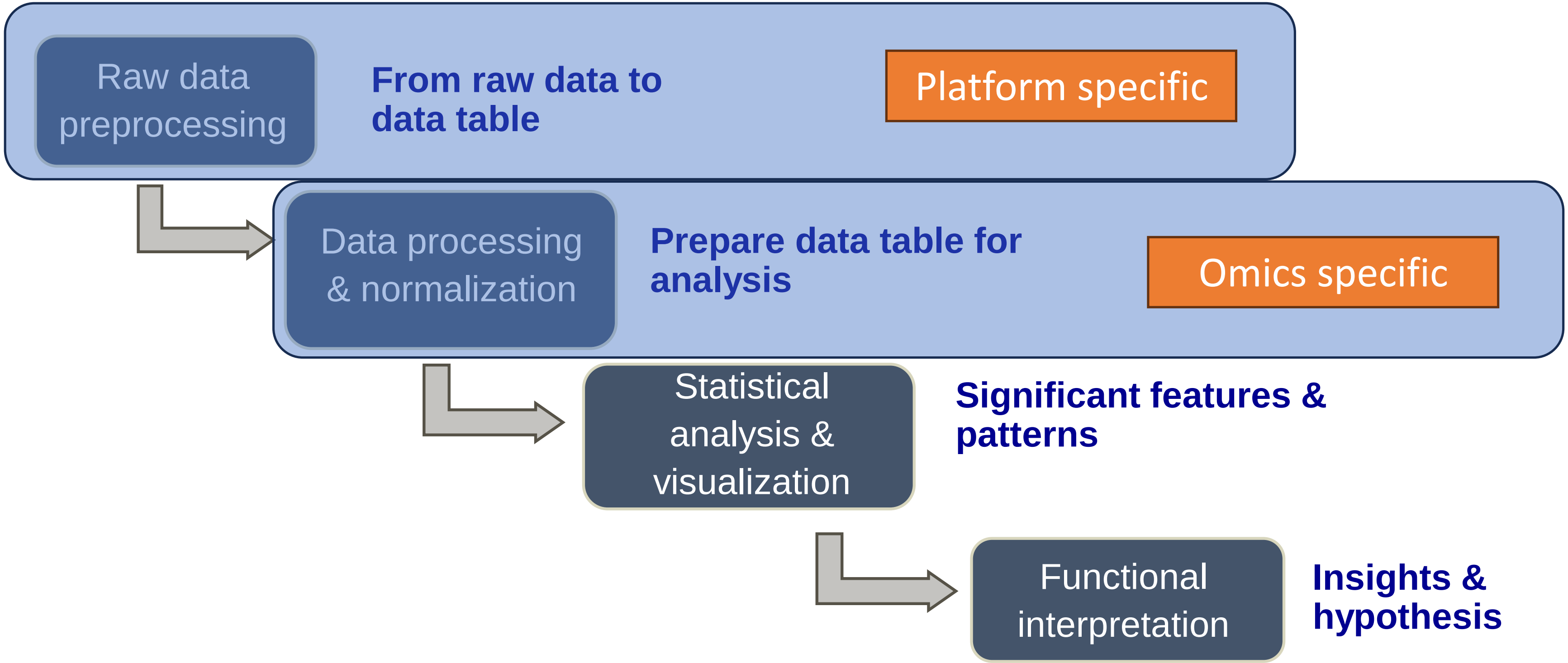
Our Syllabus

Topic	Date	Lecture	Lab
Omics Data Science Foundations	Jan. 6	Omics data processing, statistics and visualization	--
	Jan. 13	From raw data to functional insights	--
Transcriptomics	Jan. 20	Gene expression data analysis (part I)	ExpressAnalyst & NetworkAnalyst
	Jan. 27	Gene expression data analysis (part II)	ExpressAnalyst & Seq2Fun
Proteomics, Networks, & Biomarkers	Feb. 3	Biological network analysis & gene regulatory networks	NetworkAnalyst & miRNet
	Feb. 10	Proteomics & biomarker analysis	ExpressAnalyst & MetaboAnalyst
Metabolomics	Feb. 17	Targeted metabolomics data analysis	MetaboAnalyst
	Feb. 24	LC-MS untargeted metabolomics data analysis	MetaboAnalyst
Microbiomics	Mar. 2	Marker gene data analysis	MicrobiomeAnalyst
	Mar. 9	Functional microbiome data analysis	MicrobiomeAnalyst
Multi-omics	Mar. 16	Knowledge-driven multi-omics integration	OmicsNet
	Mar. 23	Data-driven multi-omics integration	OmicsAnalyst

Schedule for today

Time	**Topics**
9:00 – 9:10	Overview of data-driven multi-omics
9:10 – 9:50	Dimensionality reduction
9:50 – 10:10	Live demo & hands-on
10:15 – 10:40	Correlation analysis
10:40 – 10:55	Live demo & hands-on
10:55 – 11:10	Clustering analysis
11:10 – 11:25	Live demo & hands-on
Summary & Discussion	

Omics general workflow



Raw data preprocessing

From raw data to data table

Platform specific

Data processing & normalization

Prepare data table for analysis

Omics specific

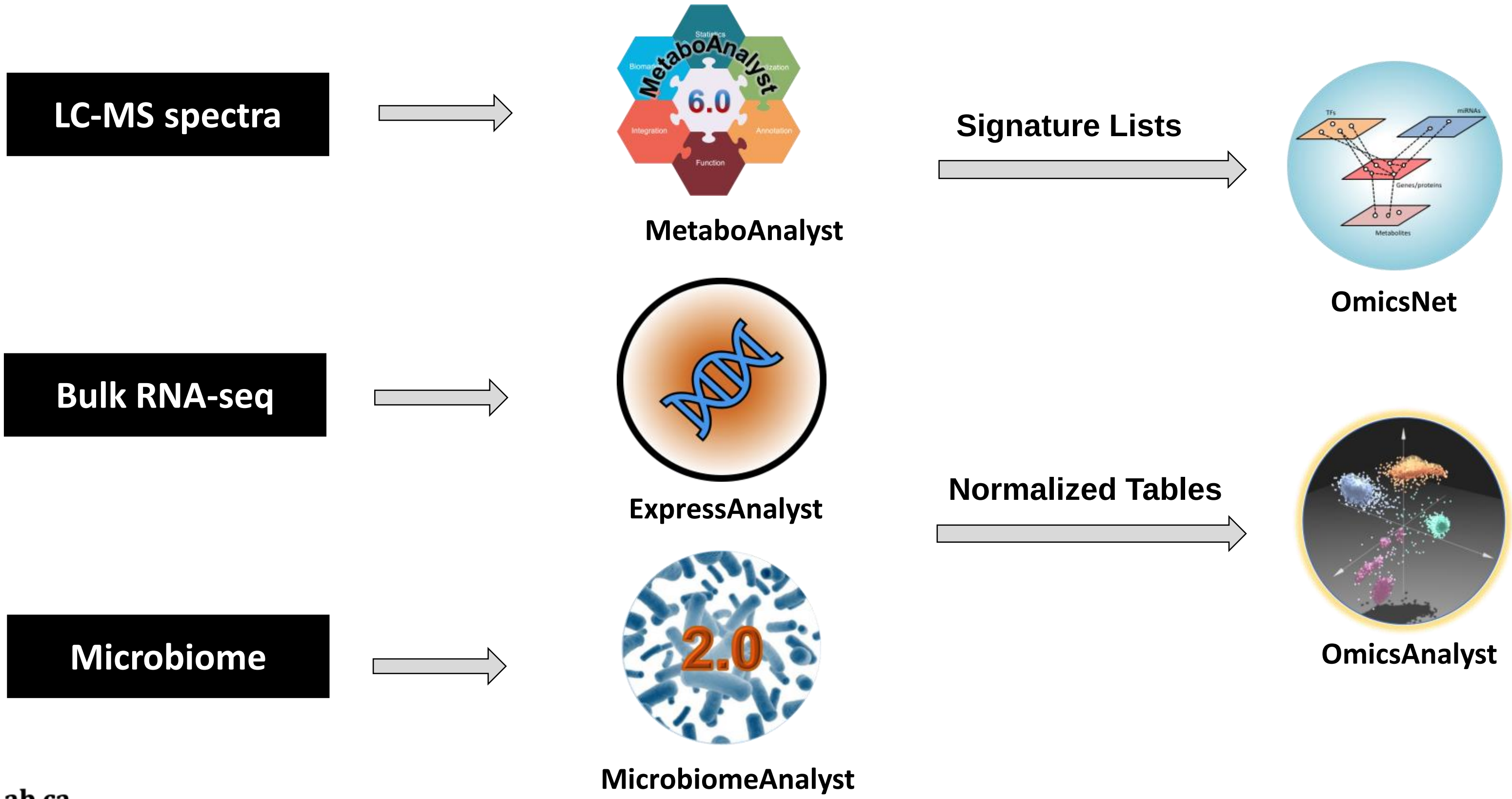
Statistical analysis & visualization

Significant features & patterns

Functional interpretation

Insights & hypothesis

Single omics to multi-omics



LC-MS spectra

Bulk RNA-seq

Microbiome

MetaboAnalyst

ExpressAnalyst

MicrobiomeAnalyst

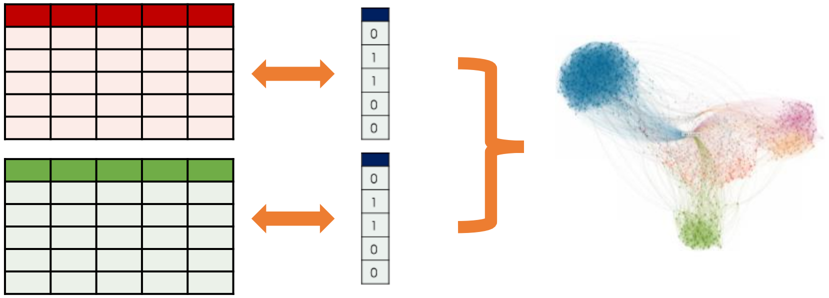
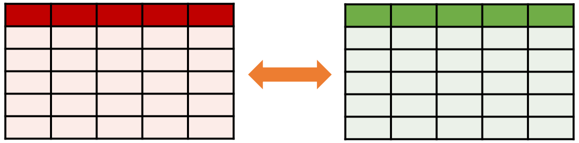
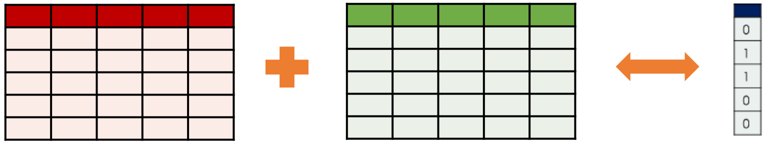
Signature Lists

Normalized Tables

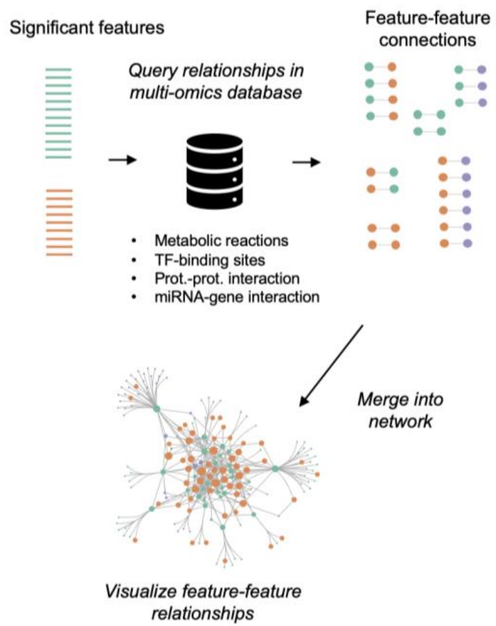
OmicsNet

OmicsAnalyst

Common workflows in multi-omics analysis

Aim	Strategy
Interpretation	
Correlation	
Prediction	

Multi-omics integration via knowledge graph



1. Perform comprehensive analysis on individual omics data to identify key signatures
2. Project the signatures from each omics layer to a knowledge graph
3. Customize the networks to suitable form
4. Visualize and apply different algorithms for network analysis & interpretation

Connect the dots

Knowledge-driven networks

MetaboAnalyst

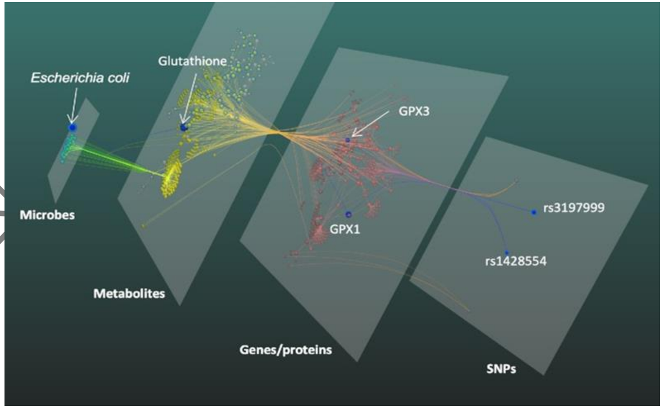
ExpressAnalyst

MicrobiomeAnalyst

A list of metabolites or LC-MS peaks

A list of genes or proteins

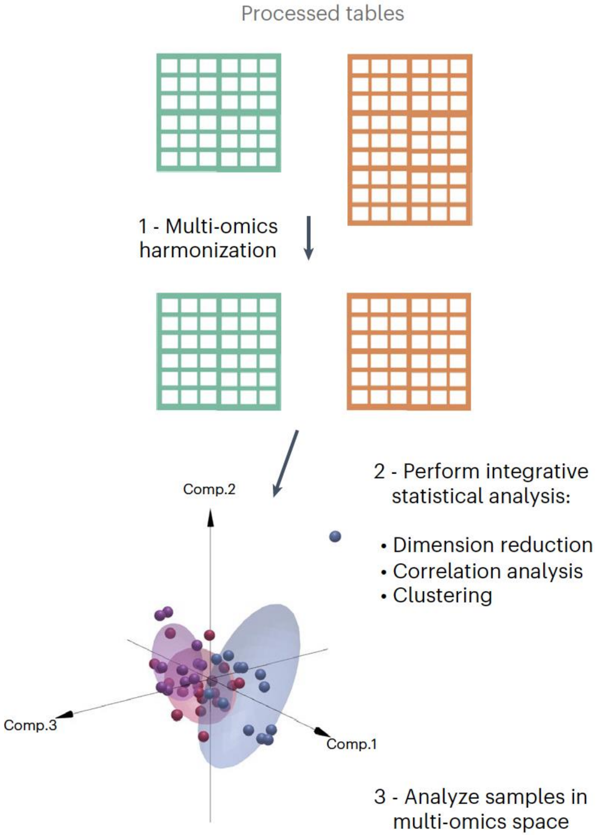
A list of microbes / taxa or genes



Qualitative Analysis

Data & model driven integration

1. Require large sample sizes (> 20 per group) and strictly matching samples
2. Perform *de novo* identification of shared patterns and correlations across different omics layers
3. Examine the main contributing features to infer their functional implications.
4. Visualization and analysis (i.e. enrichment analysis) to interpret results



2 - Perform integrative statistical analysis:

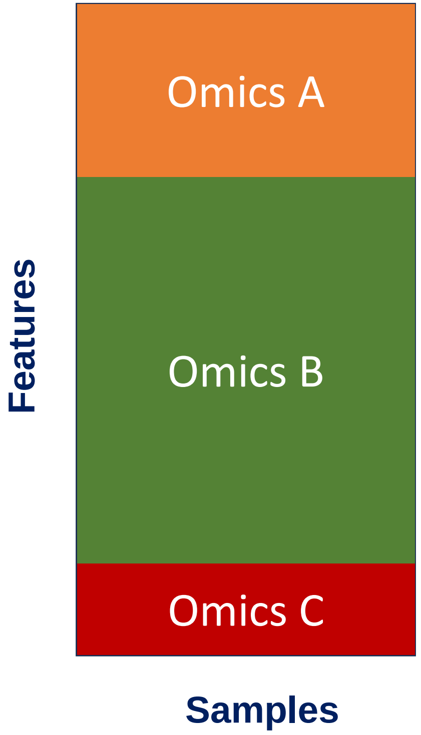
- Dimension reduction
- Correlation analysis
- Clustering

3 - Analyze samples in multi-omics space

Quantitative Analysis

Main challenges in quantitative analysis

- **High dimensional**
  - Omics data is high-dimensional
  - Multi-omics is ultra high-dimensional
- **Size difference when integration**
  - Transcriptomics: 10,000s
  - Proteomics: 1000s
  - Metabolomics: 100s ~ 1000s
  - Microbiome: 100s ~ 1000s
- **Scale difference when integration**
  - Can be of very different scale (order of magnitude)
  - Raw intensity values / counts can be ~1,000,000
  - Normalized values can be -1 ~ 1



Omics A

Omics B

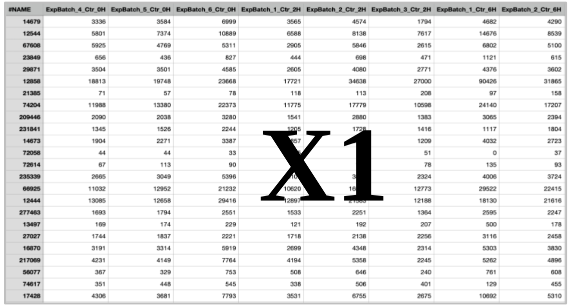
Omics C

Features

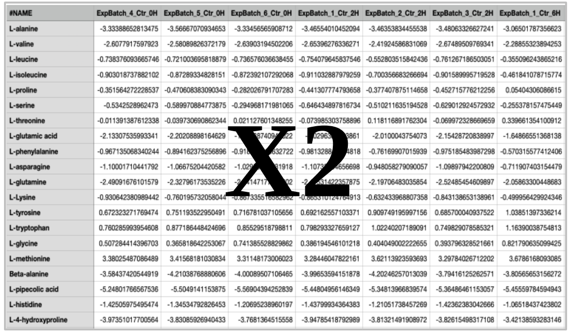
Samples

Key Strategies

- ➢Data filtering
  - o Make data comparable in size
- ➢Data scaling
  - o Make data similar in scale
- ➢Correlation analysis
  - o Focus on correlated features
- ➢Clustering analysis
  - o Reveal shared patterns (univariate)
- ➢Dimensionality reduction
  - o Reveal global patterns (multivariate)

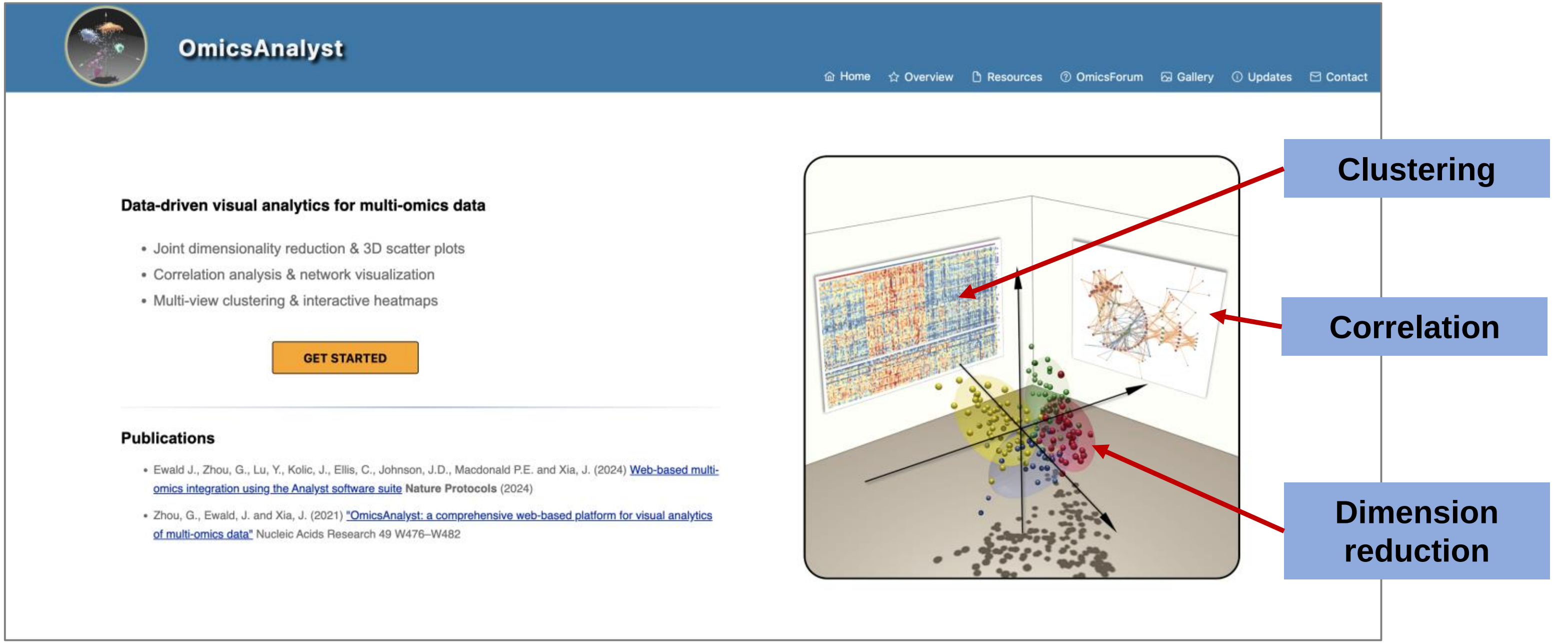


#NAME	ExpBatch_4_Ctr_0H	ExpBatch_5_Ctr_0H	ExpBatch_6_Ctr_0H	ExpBatch_1_Ctr_2H	ExpBatch_2_Ctr_2H	ExpBatch_3_Ctr_2H	ExpBatch_1_Ctr_6H	ExpBatch_2_Ctr_6H
14679	3336	3584	6999	3565	4574	1794	4682	4290
12544	5801	7374	10889	6588	8138	7617	14676	8539
67608	5925	4769	5311	2905	5846	2615	6802	5100
23849	656	436	827	444	698	471	1121	615
29871	3504	3501	4585	2605	4080	2771	4376	3602
12858	18813	19748	23668	17721	34638	27000	90426	31865
21385	71	57	78	118	113	208	97	158
74204	11988	13380	22373	11775	17779	10598	24140	17207
209446	2090	2038	3280	1541	2880	1383	3065	2394
231841	1345	1526	2244	1205	1728	1416	1117	1804
14673	1904	2271	3387	[illegible]	[illegible]	1209	4032	2723
72058	44	44	33	[illegible]	[illegible]	51	0	37
72614	67	113	90	[illegible]	[illegible]	78	135	93
235339	2665	3049	5396	[illegible]	[illegible]	2324	4006	3724
66925	11032	12952	21232	10620	[illegible]	12773	29522	22415
12444	13085	12658	29416	12897	[illegible]	12188	18130	21616
277463	1693	1794	2551	1533	2251	1364	2595	2247
13497	169	174	229	121	192	207	500	178
27027	1744	1837	2221	1718	2138	2256	3116	2458
16870	3191	3314	5919	2699	4348	2314	5303	3830
217069	4231	4149	7764	4194	5358	2245	5262	4896
56077	367	329	753	508	646	240	761	608
74617	351	448	545	338	506	401	129	455
17428	4306	3681	7793	3531	6755	2675	10692	5310



#NAME	ExpBatch_4_Ctr_0H	ExpBatch_5_Ctr_0H	ExpBatch_6_Ctr_0H	ExpBatch_1_Ctr_2H	ExpBatch_2_Ctr_2H	ExpBatch_3_Ctr_2H	ExpBatch_1_Ctr_6H
L-alanine	-3.33388652813475	-3.56667070934653	-3.33456565908712	-3.46554010452094	-3.46353834455538	-3.48063326627241	-3.06501787356623
L-valine	-2.6077917597923	-2.58089826372179	-2.63903194502206	-2.65396276336271	-2.41924586831069	-2.67489509769341	-2.28855323894253
L-leucine	-0.738376093665746	-0.721003695818879	-0.736576036638455	-0.754079645837546	-0.552803515842436	-0.761267186503051	-0.355096243865216
L-isoleucine	-0.903018737882102	-0.87289334828151	-0.872392107292068	-0.911032887979259	-0.700356683266694	-0.901589995719528	-0.461841078715774
L-proline	-0.351564272228537	-0.470608383090343	-0.282026791707283	-0.441307774793658	-0.377407875114658	-0.452715776212256	0.05404306086615
L-serine	-0.5342528962473	-0.589970884773875	-0.294968171981065	-0.646434897816734	-0.510211635194528	-0.629012924572932	-0.255378157475449
L-threonine	-0.011391387612338	-0.039730690862344	0.021127601348255	-0.073985303758896	0.118116891762304	-0.069972328669659	0.339661354100912
L-glutamic acid	-2.13307535993341	-2.20208898164629	[illegible]	[illegible]	-2.0100043754073	-2.15428720838997	-1.64866551368138
L-phenylalanine	-0.967135068340244	-0.894162375256896	[illegible]	[illegible]	-0.76169907015939	-0.975185483987298	-0.570315577412406
L-asparagine	-1.10001710441792	-1.06675204420582	[illegible]	[illegible]	-0.948058279090057	-1.09897942200809	-0.711907403154479
L-glutamine	-2.49091676101579	-2.32796173535226	[illegible]	[illegible]	-2.19706483035854	-2.52485454609897	-2.05863300448683
L-Lysine	-0.930642380989442	-0.760195732058044	-0.867335516582962	-0.865310124764913	-0.632433968807358	-0.843138653138961	-0.499956429924346
L-tyrosine	0.672323271769474	0.751193522950491	0.716781037105656	0.692162557103371	0.909749195997156	0.685700040937522	1.03851397336214
L-tryptophan	0.760285993954608	0.877186448424696	0.85529518798811	0.798293327659127	1.02240207189091	0.749829078585321	1.16390038754813
L-glycine	0.507284414396703	0.365818642253067	0.741385528829862	0.386194546101218	0.404049002222655	0.393796328521661	0.821790635099425
L-methionine	3.38025487086489	3.41568181030834	3.31148173006023	3.28446047822161	3.62113923593693	3.29784026712202	3.6786168093085
Beta-alanine	-3.58437420544919	-4.21038768880606	-4.00089507106465	-3.99653594151878	-4.20246257013039	-3.79416125262571	-3.80565653156272
L-pipecolic acid	-5.24801766567536	-5.5049141153875	-5.56904394252839	-5.44804956146349	-5.34813966839574	-5.36486461153057	-5.45559784594943
L-histidine	-1.42505975495474	-1.34534792826453	-1.20695238960197	-1.43799934364383	-1.21051738457269	-1.42362383042666	-1.06518437423802
L-4-hydroxyproline	-3.97351017700564	-3.83085926940433	-3.7681364515558	-3.94785418792989	-3.81321491908972	-3.82615498317108	-3.42138593283146

Overall design of OmicsAnalyst



Data-driven visual analytics for multi-omics data

- Joint dimensionality reduction & 3D scatter plots
- Correlation analysis & network visualization
- Multi-view clustering & interactive heatmaps

Publications

- Ewald J., Zhou, G., Lu, Y., Kolic, J., Ellis, C., Johnson, J.D., Macdonald P.E. and Xia, J. (2024) Web-based multi-omics integration using the Analyst software suite **Nature Protocols** (2024)
- Zhou, G., Ewald, J. and Xia, J. (2021) "OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data" Nucleic Acids Research 49 W476–W482

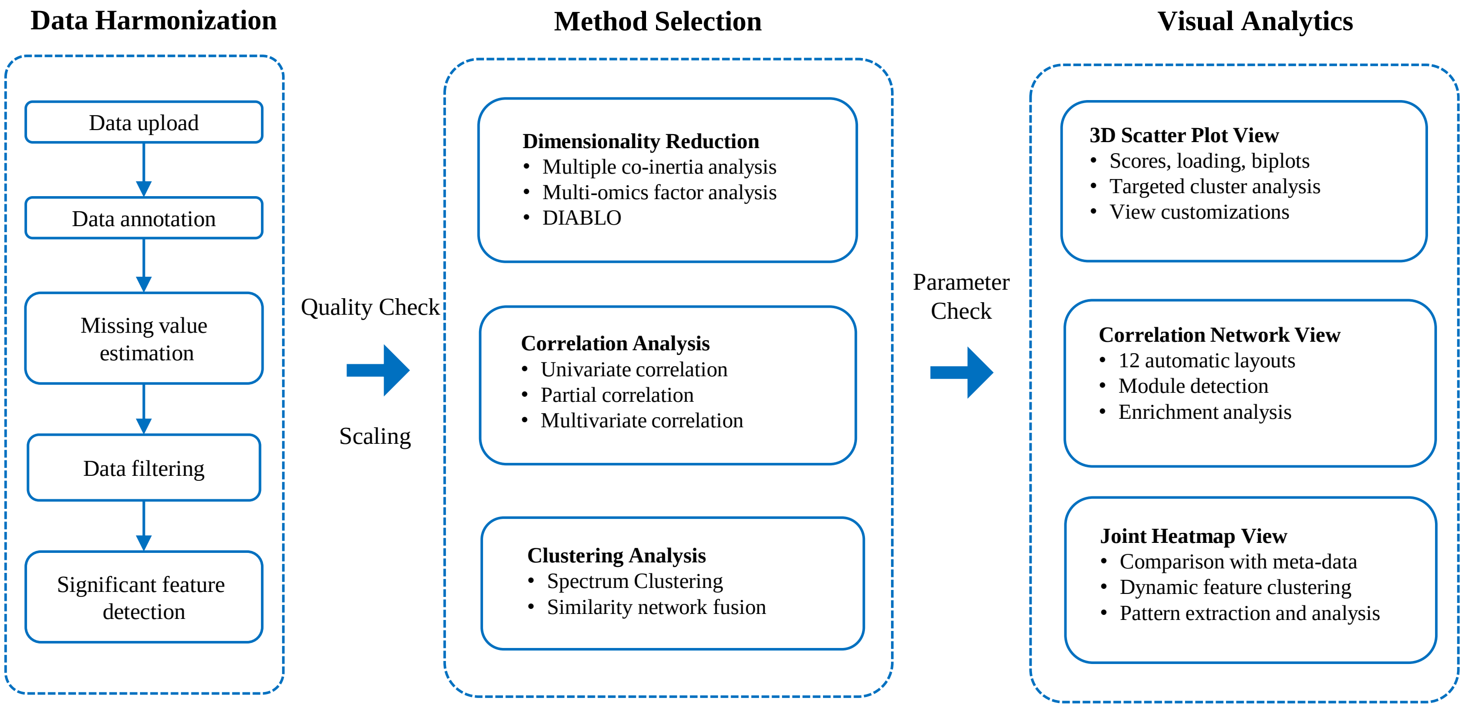
Clustering

Correlation

Dimension reduction

Built-in filtering & scaling

Under the hood



Data Harmonization

- Data upload
- Data annotation
- Missing value estimation
- Data filtering
- Significant feature detection

Quality Check

Scaling

Method Selection

Dimensionality Reduction
- Multiple co-inertia analysis
- Multi-omics factor analysis
- DIABLO

Correlation Analysis
- Univariate correlation
- Partial correlation
- Multivariate correlation

Clustering Analysis
- Spectrum Clustering
- Similarity network fusion

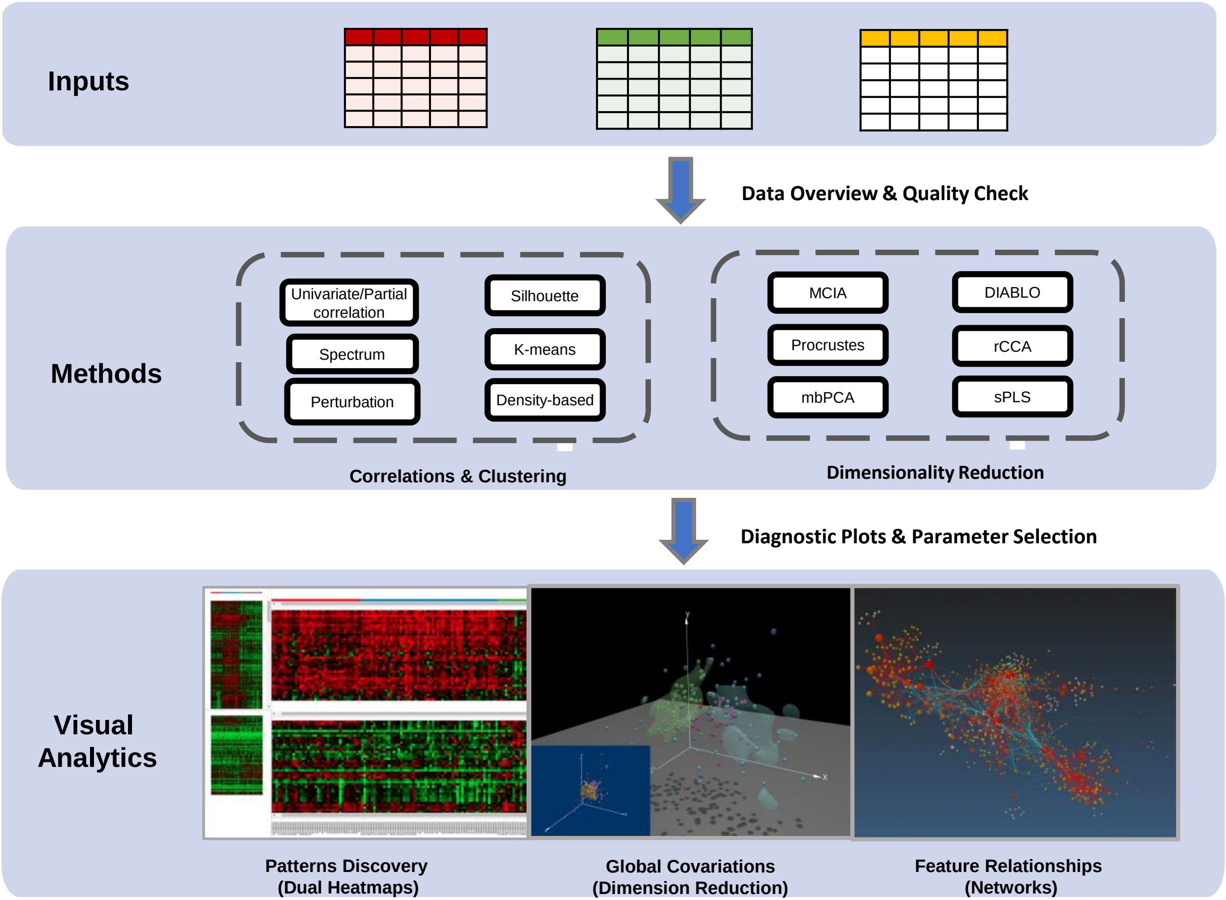
Parameter Check

Visual Analytics

3D Scatter Plot View
- Scores, loading, biplots
- Targeted cluster analysis
- View customizations

Correlation Network View
- 12 automatic layouts
- Module detection
- Enrichment analysis

Joint Heatmap View
- Comparison with meta-data
- Dynamic feature clustering
- Pattern extraction and analysis



Inputs

Data Overview & Quality Check

Methods

Univariate/Partial correlation

Silhouette

Spectrum

K-means

Perturbation

Density-based

Correlations & Clustering

MCIA

DIABLO

Procrustes

rCCA

mbPCA

sPLS

Dimensionality Reduction

Diagnostic Plots & Parameter Selection

Visual Analytics

Patterns Discovery (Dual Heatmaps)

Global Covariations (Dimension Reduction)

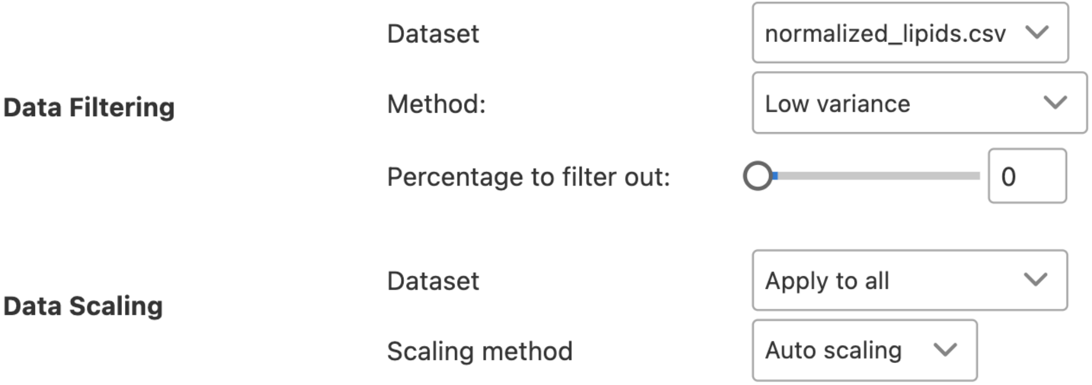
Feature Relationships (Networks)

Data Input

- Max five omics data
  - Recommended
    - Processed & normalized data table follow the best practices of individual omics fields
- A metadata table
  - Must all share the same sample IDs
  - No missing values are allowed for metadata
- When some samples are missing, only the overlap samples will be used in joint analysis

Data Filtering & Scaling

- Different omics data types often have very different number of features and variances
- Many multi-omics integration methods are sensitive to imbalanced dimensionality or variance (i.e. omics layer containing many more features or large variance could dominate the analysis)
- It is advisable to perform data processing to make them more comparable



Data Filtering

Dataset: normalized_lipids.csv

Method: Low variance

Percentage to filter out: 0

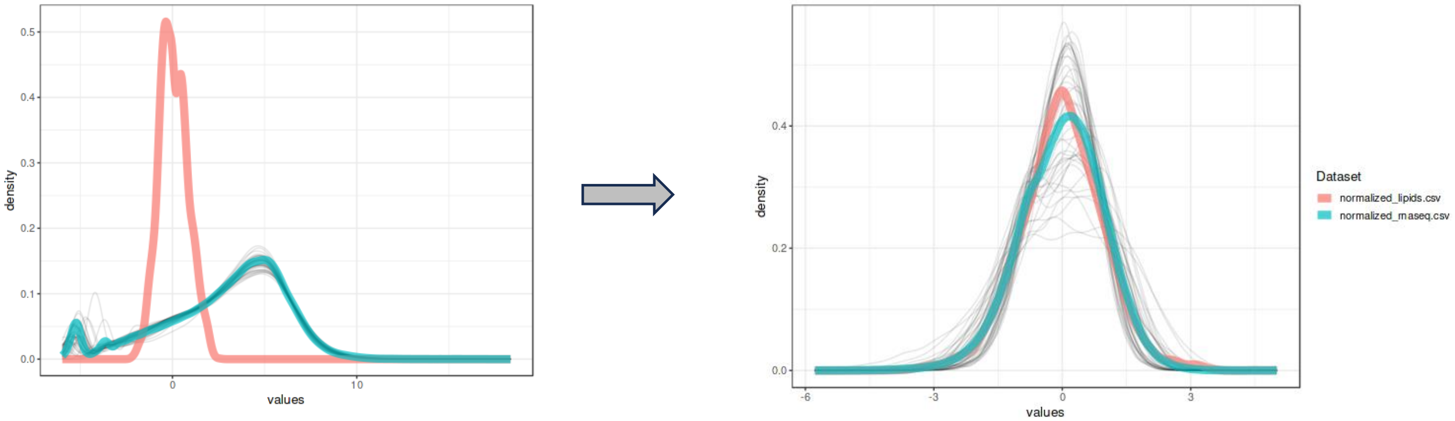
Data Scaling

Dataset: Apply to all

Scaling method: Auto scaling

General Considerations

- ➢ Number of features
  - ➢ Stronger filtering for larger omics data
- ➢ Feature abundance values are at similar scale
  - ➢ Unit scaling (auto-scaling), pareto, mean-centering, range



Schedule for today

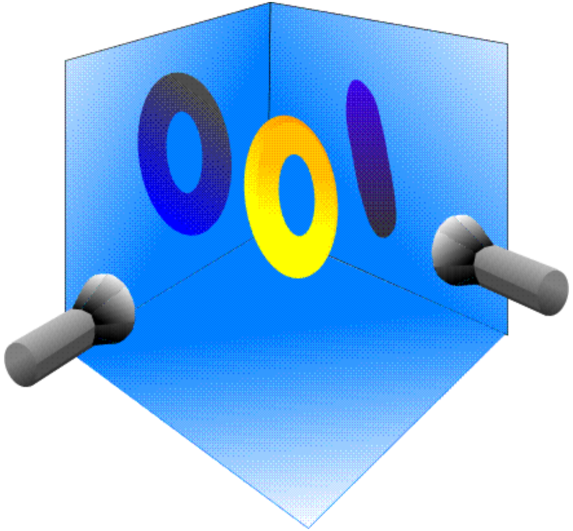
Time	**Topics**
9:00 – 9:10	Overview of data-driven multi-omics
9:10 – 9:50	Dimensionality reduction
9:50 – 10:10	Live Demo
10:15 – 10:40	Feature correlation analysis
10:40 – 10:55	Live Demo
10:55 – 11:10	Clustering analysis
11:10 – 11:25	Live Demo
Summary & Discussion	

Dimensionality Reduction

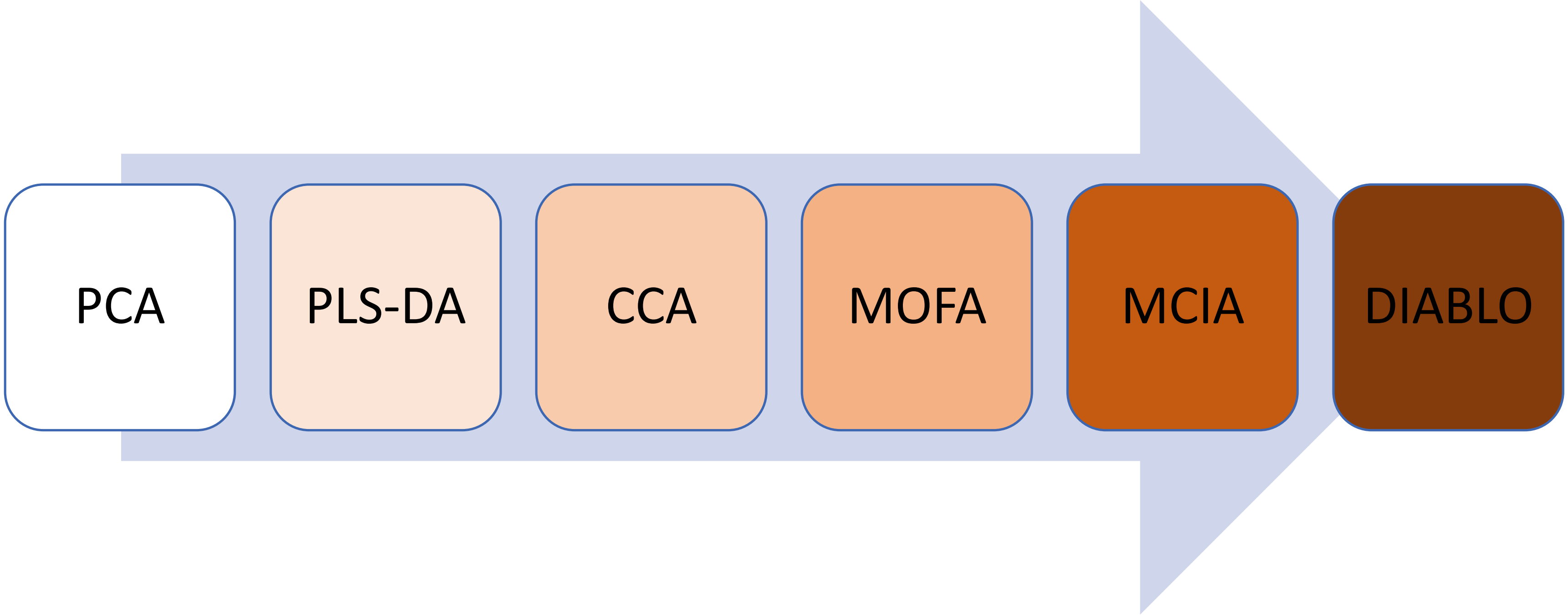
From single omics to multi-omics

About Dimensionality Reduction

- To compute a low-dimensional representation that captures the main characteristics of the high-dimensional data
- Main assumptions
  1. There are redundancies in omics data
     - ✓ Molecules involved in the same biological processes that are often correlated
  2. Typical summary statistics (variance, covariance) can capture the main characteristics of the data



Evolution of Dimensionality Reduction



PCA

PLS-DA

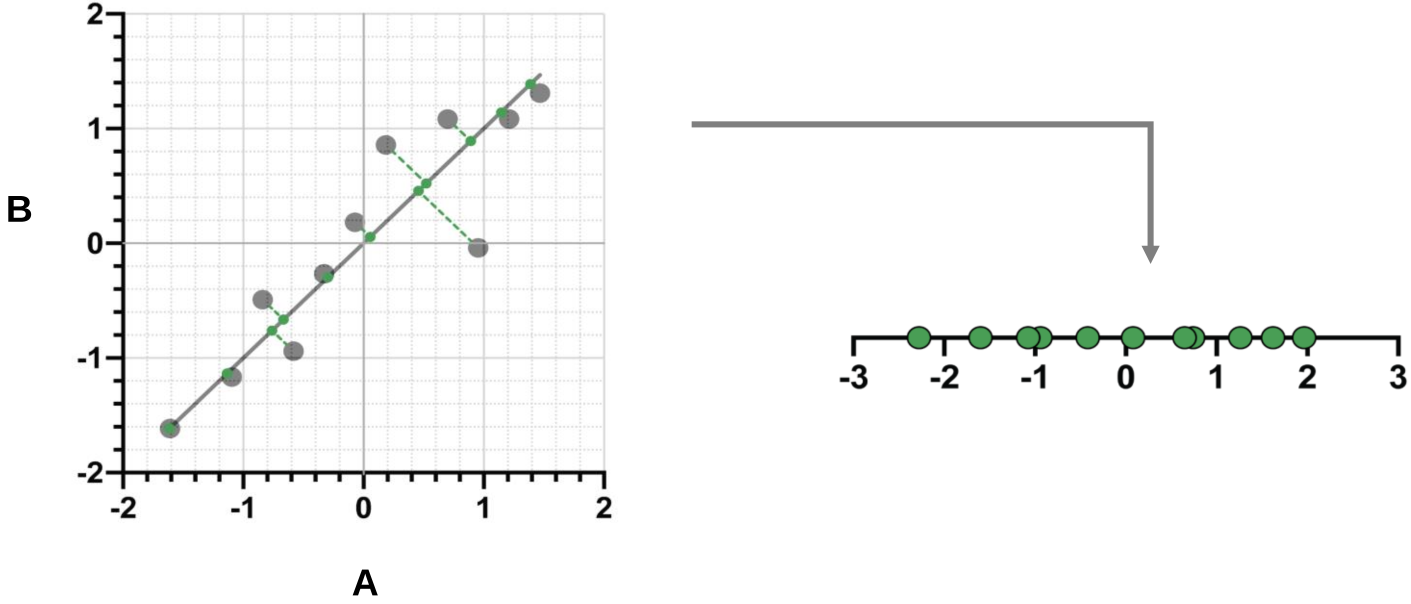
CCA

MOFA

MCIA

DIABLO

Principal Component Analysis (PCA)



PCA details

From *k* original variables: $x_1, x_2, \ldots, x_k$:

Produce *k* new variables: $t_1, t_2, \ldots, t_k$

$t_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k$

$t_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2k}x_k$

...

$t_k = a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kk}x_k$

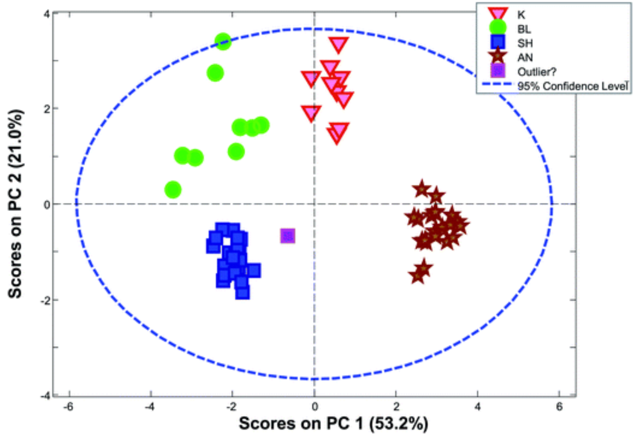
Linear combinations

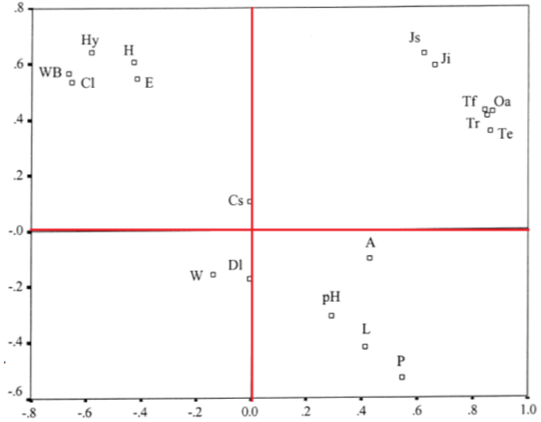
Such that:

- t_k's are uncorrelated (orthogonal)
- t_1 explains as much as possible of original variance in data set
- t_2 explains as much as possible of remaining variance, etc.

Scores & Loadings

- **Scores:** samples in the low-dimensional space
  - Can be used to view patterns
- **Loadings:** feature coefficients –covariances/correlations between the original variables and the samples in new
  - Can be used to view main feature contributors to the patterns of interest



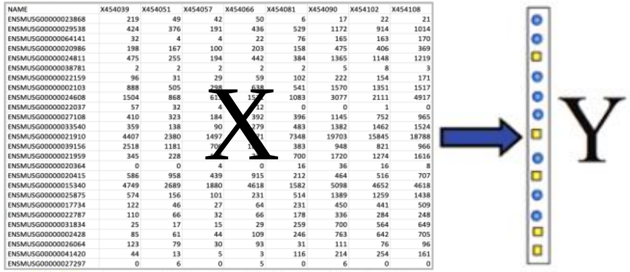


Loadings

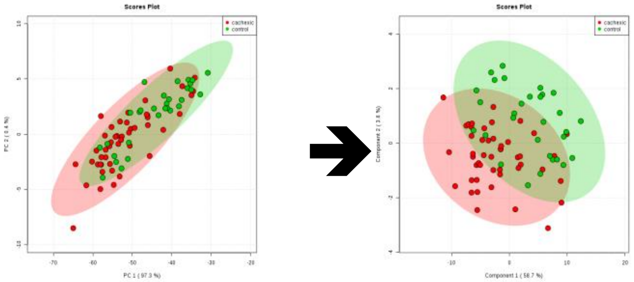
Partial least squares- discriminant analysis (PLS-DA)

- When the experimental effects are subtle or moderate, PCA will not show good separation patterns

- PLS-DA is a supervised method that uses multiple linear regression technique to find the direction of maximum covariance between a data set (X) and the class membership (Y)



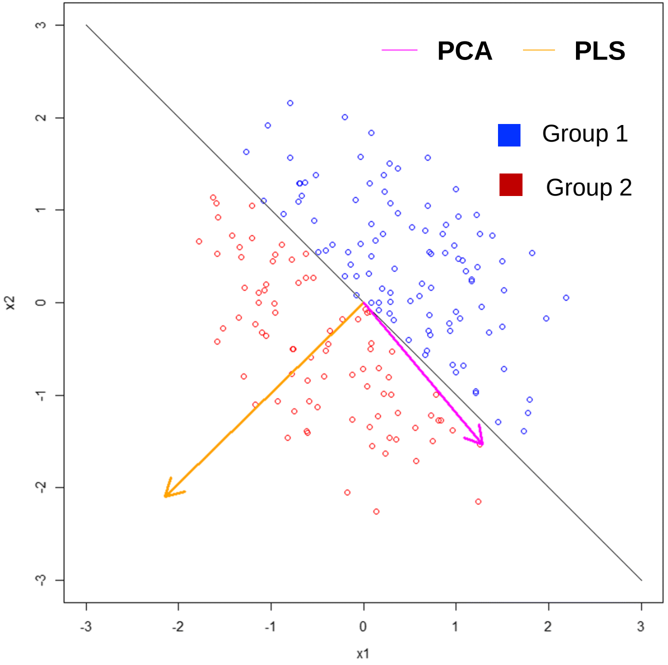
NAME	X454039	X454051	X454057	X454066	X454081	X454090	X454102	X454108
ENSMUSG00000023868	219	49	42	50	6	17	22	21
ENSMUSG00000029538	424	376	191	436	529	1172	914	1014
ENSMUSG00000064141	32	4	4	22	76	165	163	170
ENSMUSG00000020986	198	167	100	203	158	475	406	369
ENSMUSG00000024811	475	255	194	442	384	1365	1148	1219
ENSMUSG00000038781	2	2	2	2	2	5	8	3
ENSMUSG00000022159	96	31	29	59	102	222	154	171
ENSMUSG00000002103	888	505	298	638	541	1570	1351	1517
ENSMUSG00000024608	1504	868	[illegible]	[illegible]	1083	3077	2111	4917
ENSMUSG00000022037	57	32	4	12	0	0	1	0
ENSMUSG00000027108	410	323	184	392	396	1145	752	965
ENSMUSG00000033540	359	138	90	279	483	1382	1462	1524
ENSMUSG00000021910	4407	2380	1497	[illegible]	7348	19703	15845	18788
ENSMUSG00000039156	2518	1181	[illegible]	[illegible]	383	948	821	966
ENSMUSG00000021959	345	228	[illegible]	[illegible]	700	1720	1274	1616
ENSMUSG00000020364	0	0	4	0	16	36	16	8
ENSMUSG00000020415	586	958	439	915	212	464	516	707
ENSMUSG00000015340	4749	2689	1880	4618	1582	5098	4652	4618
ENSMUSG00000025875	574	156	101	231	514	1389	1259	1438
ENSMUSG00000017734	122	46	27	64	231	450	441	509
ENSMUSG00000022787	110	66	32	66	178	336	284	248
ENSMUSG00000031834	25	17	15	29	259	700	564	649
ENSMUSG00000002428	85	61	44	109	246	763	642	705
ENSMUSG00000026064	123	79	30	93	31	111	76	96
ENSMUSG00000041420	44	13	5	3	116	214	254	161
ENSMUSG00000027297	0	6	0	5	0	6	0	0



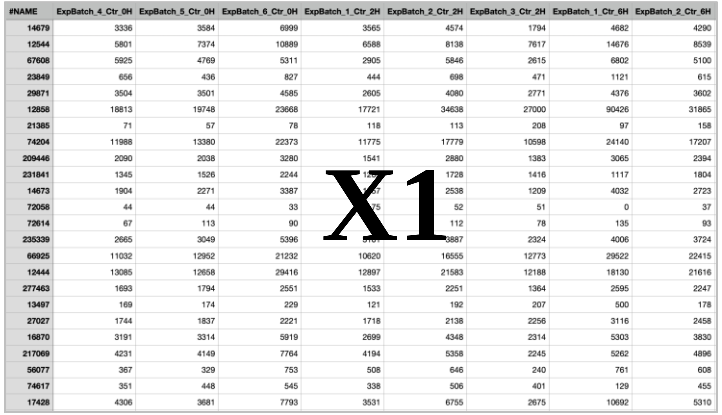
PCA ➔ PLS-DA

PCA vs PLS (variance vs co-variance)

Directions identified by PCA vs PLS-DA can be different

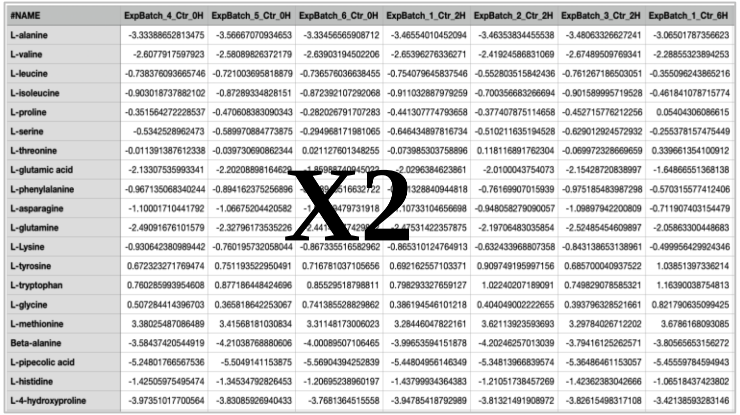


Extend to multiple omics datasets



X1

#NAME	ExpBatch_4_Ctr_0H	ExpBatch_5_Ctr_0H	ExpBatch_6_Ctr_0H	ExpBatch_1_Ctr_2H	ExpBatch_2_Ctr_2H	ExpBatch_3_Ctr_2H	ExpBatch_1_Ctr_6H	ExpBatch_2_Ctr_6H
14679	3336	3584	6999	3565	4574	1794	4682	4290
12544	5801	7374	10889	6588	8138	7617	14676	8539
67608	5925	4769	5311	2905	5846	2615	6802	5100
23849	656	436	827	444	698	471	1121	615
29871	3504	3501	4585	2605	4080	2771	4376	3602
12858	18813	19748	23668	17721	34638	27000	90426	31865
21385	71	57	78	118	113	208	97	158
74204	11988	13380	22373	11775	17779	10598	24140	17207
209446	2090	2038	3280	1541	2880	1383	3065	2394
231841	1345	1526	2244	[illegible]	1728	1416	1117	1804
14673	1904	2271	3387	[illegible]	2538	1209	4032	2723
72058	44	44	33	75	52	51	0	37
72614	67	113	90	[illegible]	112	78	135	93
235339	2665	3049	5396	[illegible]	3887	2324	4006	3724
66925	11032	12952	21232	10620	16555	12773	29522	22415
12444	13085	12658	29416	12897	21583	12188	18130	21616
277463	1693	1794	2551	1533	2251	1364	2595	2247
13497	169	174	229	121	192	207	500	178
27027	1744	1837	2221	1718	2138	2256	3116	2458
16870	3191	3314	5919	2699	4348	2314	5303	3830
217069	4231	4149	7764	4194	5358	2245	5262	4896
56077	367	329	753	508	646	240	761	608
74617	351	448	545	338	506	401	129	455
17428	4306	3681	7793	3531	6755	2675	10692	5310



X2

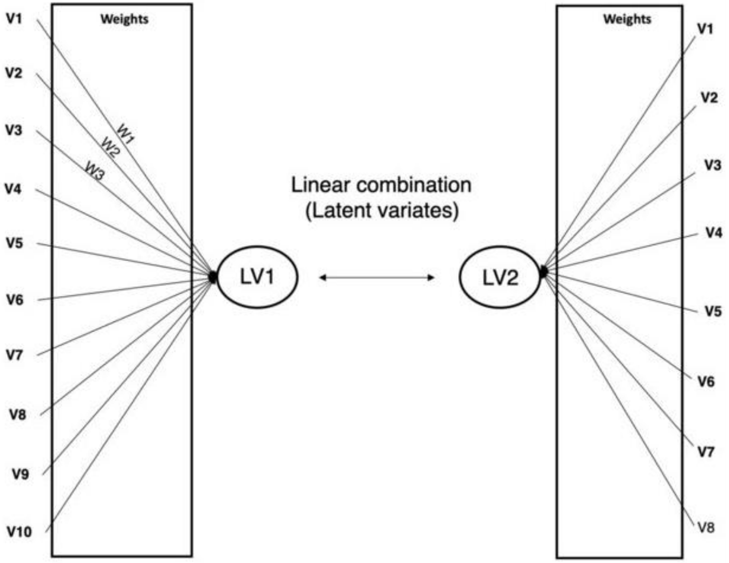
#NAME	ExpBatch_4_Ctr_0H	ExpBatch_5_Ctr_0H	ExpBatch_6_Ctr_0H	ExpBatch_1_Ctr_2H	ExpBatch_2_Ctr_2H	ExpBatch_3_Ctr_2H	ExpBatch_1_Ctr_6H
L-alanine	-3.33388652813475	-3.56667070934653	-3.33456565908712	-3.46554010452094	-3.46353834455538	-3.48063326627241	-3.06501787356623
L-valine	-2.6077917597923	-2.58089826372179	-2.63903194502206	-2.65396276336271	-2.41924586831069	-2.67489509769341	-2.28855323894253
L-leucine	-0.738376093665746	-0.721003695818879	-0.736576036638455	-0.754079645837546	-0.552803515842436	-0.761267186503051	-0.355096243865216
L-isoleucine	-0.903018737882102	-0.87289334828151	-0.872392107292068	-0.911032887979259	-0.700356683266694	-0.901589995719528	-0.461841078715774
L-proline	-0.351564272228537	-0.470608383090343	-0.282026791707283	-0.441307774793658	-0.377407875114658	-0.452715776212256	0.05404306086615
L-serine	-0.5342528962473	-0.589970884773875	-0.294968171981065	-0.646434897816734	-0.510211635194528	-0.629012924572932	-0.255378157475449
L-threonine	-0.011391387612338	-0.039730690862344	0.021127601348255	-0.073985303758896	0.118116891762304	-0.069972328669659	0.339661354100912
L-glutamic acid	-2.13307535993341	-2.20208898164629	-1.85988740945022	-2.0296384623861	-2.0100043754073	-2.15428720838997	-1.64866551368138
L-phenylalanine	-0.967135068340244	-0.894162375256896	[illegible]	[illegible]	-0.76169907015939	-0.975185483987298	-0.570315577412406
L-asparagine	-1.10001710441792	-1.06675204420582	[illegible]	-1.10733104656698	-0.948058279090057	-1.09897942200809	-0.711907403154479
L-glutamine	-2.49091676101579	-2.32796173535226	[illegible]	-2.47531422357875	-2.19706483035854	-2.52485454609897	-2.05863300448683
L-Lysine	-0.930642380989442	-0.760195732058044	-0.867335516582962	-0.865310124764913	-0.632433968807358	-0.843138653138961	-0.499956429924346
L-tyrosine	0.672323271769474	0.751193522950491	0.716781037105656	0.692162557103371	0.909749195997156	0.685700040937522	1.03851397336214
L-tryptophan	0.760285993954608	0.877186448424696	0.85529518798811	0.798293327659127	1.02240207189091	0.749829078585321	1.16390038754813
L-glycine	0.507284414396703	0.365818642253067	0.741385528829862	0.386194546101218	0.404049002222655	0.393796328521661	0.821790635099425
L-methionine	3.38025487086489	3.41568181030834	3.31148173006023	3.28446047822161	3.62113923593693	3.29784026712202	3.6786168093085
Beta-alanine	-3.58437420544919	-4.21038768880606	-4.00089507106465	-3.99653594151878	-4.20246257013039	-3.79416125262571	-3.80565653156272
L-pipecolic acid	-5.24801766567536	-5.5049141153875	-5.56904394252839	-5.44804956146349	-5.34813966839574	-5.36486461153057	-5.45559784594943
L-histidine	-1.42505975495474	-1.34534792826453	-1.20695238960197	-1.43799934364383	-1.21051738457269	-1.42362383042666	-1.06518437423802
L-4-hydroxyproline	-3.97351017700564	-3.83085926940433	-3.7681364515558	-3.94785418792989	-3.81321491908972	-3.82615498317108	-3.42138593283146

Joint Dimensionality Reduction (jDR)

- Mainly extensions of PCA and PLS-DA
- Simultaneously project multiple data tables to a **shared** low-dimensional space with or without consideration of class labels.
- Each method computes components that maximize some **statistical terms**.
- The maximized term can integrate multiple statistics, which is the key concept in jDR

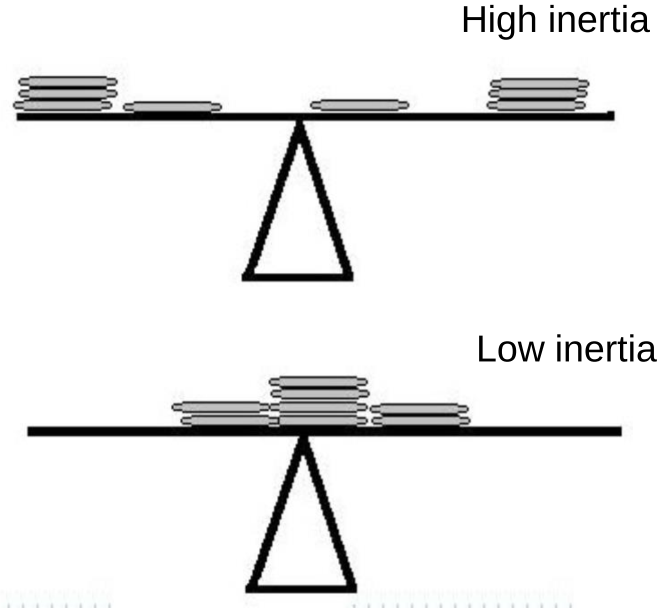
Canonical Correlation Analysis (CCA)

- To extract latent features shared between multiple data by finding the linear combinations of features-referred to as canonical variables (CVs)-within each data that achieve maximal <u>cross-matrix correlation</u>
- Assume linear model
- Limited to n > p (i.e. more samples than features)
  - Not suitable for omics



Multiple co-inertia analysis (MCIA)

- Inertia is a measure for the variability of the data
    - The inertia of an object is the tendency of an object at rest to stay at rest. The inertia of an object suspended from its centroid is directly related to how widely dispersed the mass is away from its centroid
- The inertia of a set of points relative to one point P is defined by the weighted sum of the squared distances between each considered point and the point P.
    - The inertia of a centered matrix (mean is equal to zero) is simply the sum of the squared matrix elements.



Multiple co-inertia analysis (MCIA)

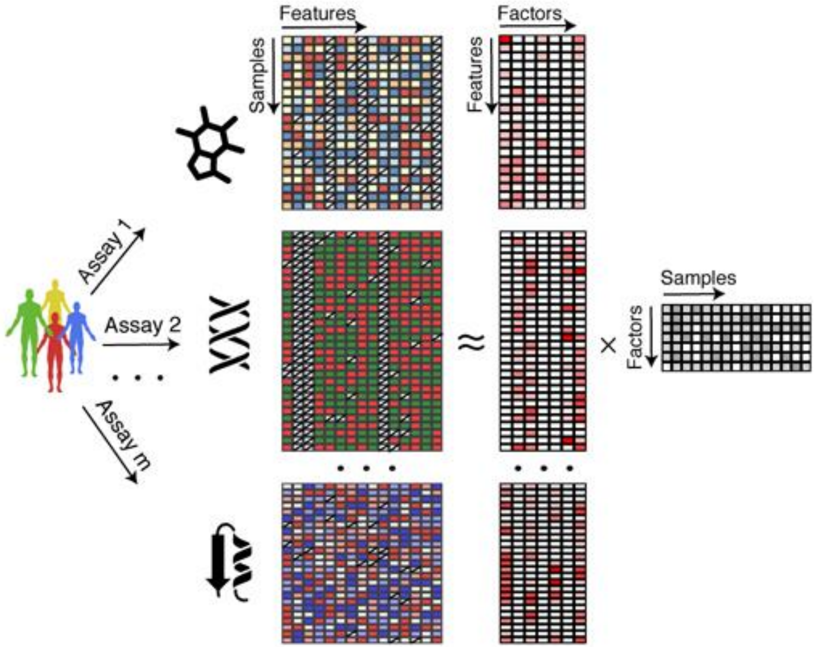
- Co-inertia is a global measure for the co-variability of two data sets (for example, two high-dimensional random variables). If the data sets are centered, the co-inertia is the sum of squared covariances
- MCIA is very similar to CCA, performed in a two steps.
    1. Dimension reduction method is performed on each individual data.
    2. Project the two dimensionally reduced matrices into a same hyperspace while imposing the constraint of maximizing covariance between each matrix.

MCIA vs CCA

- MCIA performs well when the number of features are much greater than the number of samples (i.e. omics data)
- MCIA finds components that simultaneously maximize sources of variability within each dataset, and correlation of the components across datasets. This means that MCIA components capture variability trends that are shared across all omics datasets.
- It is more robust to outliers and has fewer tuneable parameters
- MCIA is symmetric, therefore the order that the 'omics datasets are uploaded will not impact the results

Multi-Omics Factor Analysis (MOFA)

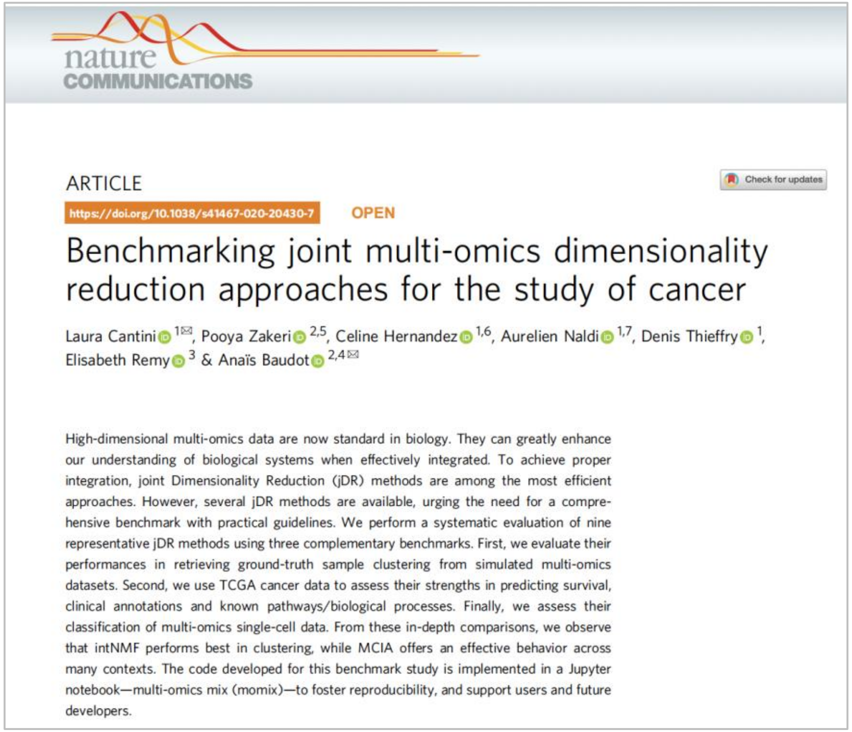
- Generalization of PCA to multi-omics data
- MOFA identifies latent factors that capture the main sources of variation across the different omics datasets. These factors are derived from multiple data types simultaneously.
- Each factor represents a biological or technical signal that is shared across the datasets to varying degrees.



MCIA vs. MOFA

- MCIA identify components simultaneously but separately in each layer, by maximizing a term that includes variance of each data and correlation across data.
    - This finds a balance between components that both explain a substantial proportion of the variability within each layer and are shared across layers.
- MOFA first performs an additional normalization step to correct for systematic differences in ‘shape’. Then all omics features are directly merged into the same matrix, and subject to PCA.
    - There is no stipulation that components should be correlated across layers, it is possible for some components to be almost 100% driven by one omics layer. This allows us to find both shared and complementary factors across omics layers

How should I choose?



nature COMMUNICATIONS

ARTICLE

Check for updates

https://doi.org/10.1038/s41467-020-20430-7 OPEN

Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer

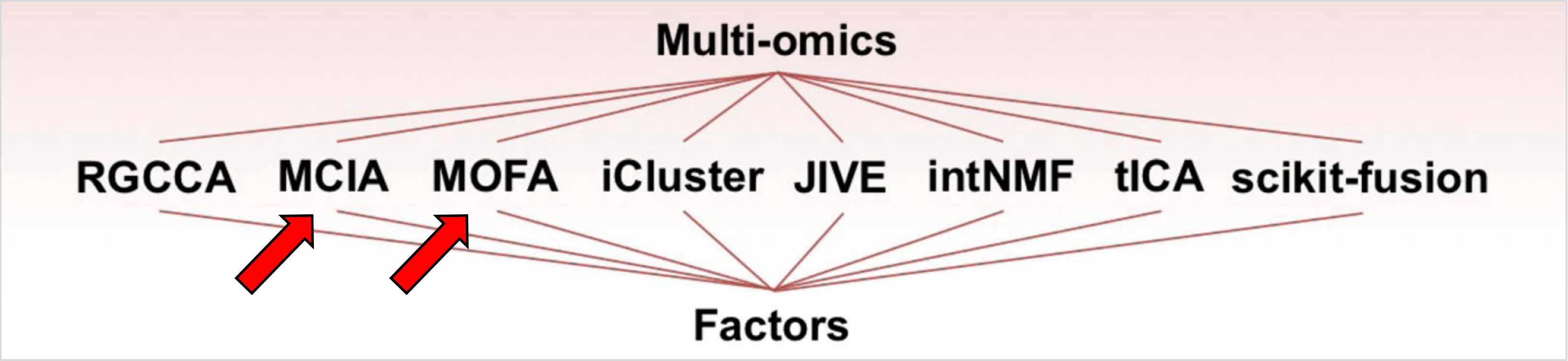
Laura Cantini [1✉], Pooya Zakeri [2,5], Celine Hernandez [1,6], Aurelien Naldi [1,7], Denis Thieffry [1], Elisabeth Remy [3] & Anaïs Baudot [2,4✉]

High-dimensional multi-omics data are now standard in biology. They can greatly enhance our understanding of biological systems when effectively integrated. To achieve proper integration, joint Dimensionality Reduction (jDR) methods are among the most efficient approaches. However, several jDR methods are available, urging the need for a comprehensive benchmark with practical guidelines. We perform a systematic evaluation of nine representative jDR methods using three complementary benchmarks. First, we evaluate their performances in retrieving ground-truth sample clustering from simulated multi-omics datasets. Second, we use TCGA cancer data to assess their strengths in predicting survival, clinical annotations and known pathways/biological processes. Finally, we assess their classification of multi-omics single-cell data. From these in-depth comparisons, we observe that intNMF performs best in clustering, while MCIA offers an effective behavior across many contexts. The code developed for this benchmark study is implemented in a Jupyter notebook—multi-omics mix (momix)—to foster reproducibility, and support users and future developers.

https://www.nature.com/articles/s41467-020-20430-7

Results

MCIA, MOFA, and RGCCA showed the best performance among the set of methods not intrinsically designed for clustering. In the cancer data benchmark, when we evaluated the associations of the factors with survival or clinical annotations, **MCIA,** JIVE, **MOFA,** and RGCCA were the most efficient methods.

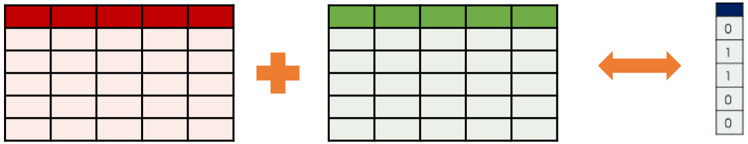


https://www.nature.com/articles/s41467-020-20430-7

Prediction & Classifcation

Issue: unsupervised jDR methods identify features that are highly correlated but led to poor discriminative ability.

Naïve approach on classification & prediction



Same sample size, feature # drastically increase

➢ Most models will become worse (i.e. overfitting)

Mix different scales (OTU counts & concentrations)

➢ Most models cannot accommodate

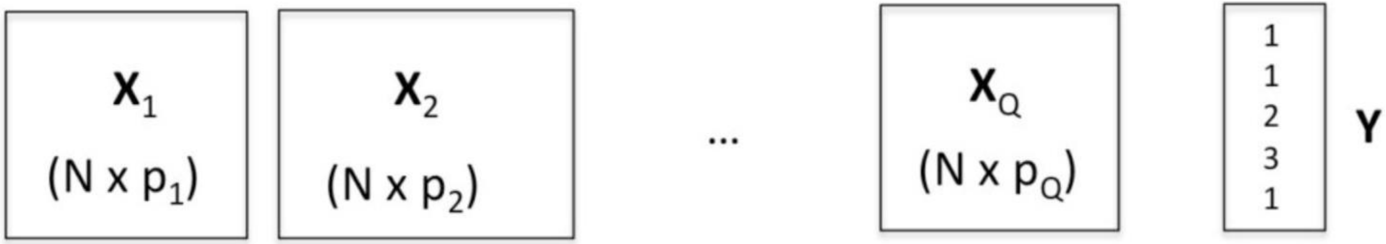
Different sizes (100s metabolites ~1000s OTUs ~ 10,000s genes)

➢ Larger data will dominate the analysis

➔ Integrating dimensionality reduction into classification

DIABLO

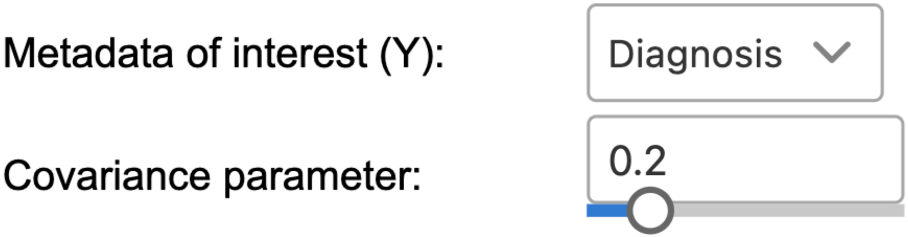
- Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO)
- Aims to identify coherent patterns between datasets that change with respect to different phenotypes.
- DIABLO is supervised as it also considers the variance of a single metadata variable (Y).



http://mixomics.org/mixdiablo/

Balance between covariance and predictivity

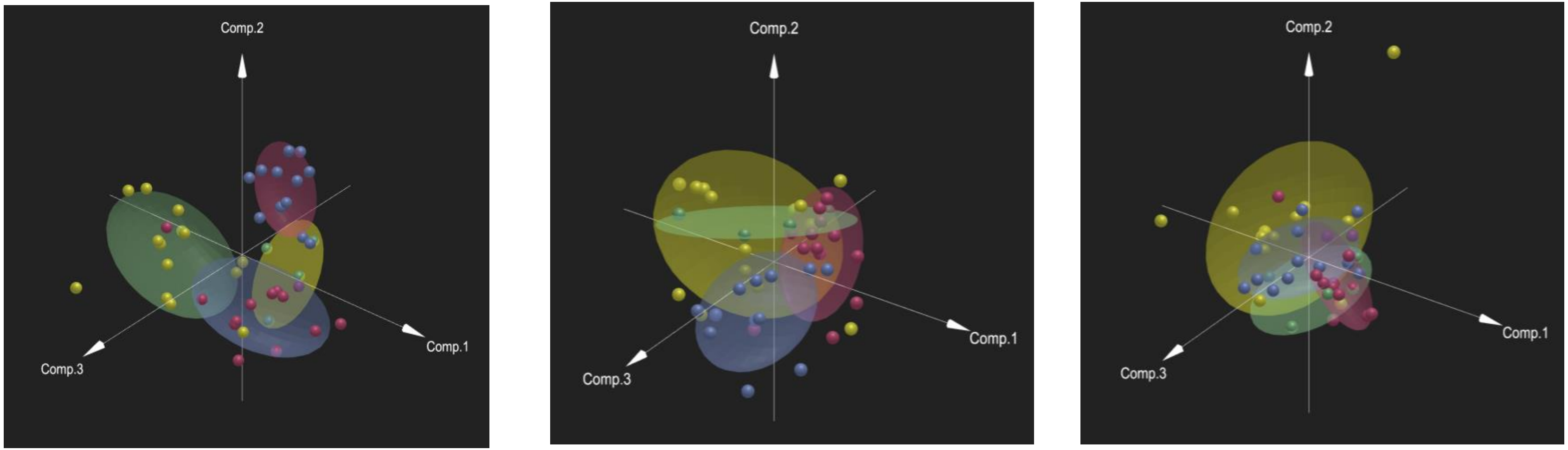
- DIABLO maximizes the ability of the components to explain metadata of interest and the covariance across omics data.
- The **covariance** parameter adjusts the weight of these two goals. A value of 0 does not consider covariance at all (i.e. maximizing separation w.r.t metadata of interest). A value of 1 does not consider the metadata at all (i.e. maximizing covariance across omics layers), making it very similar to MCIA.



Metadata of interest (Y): Diagnosis

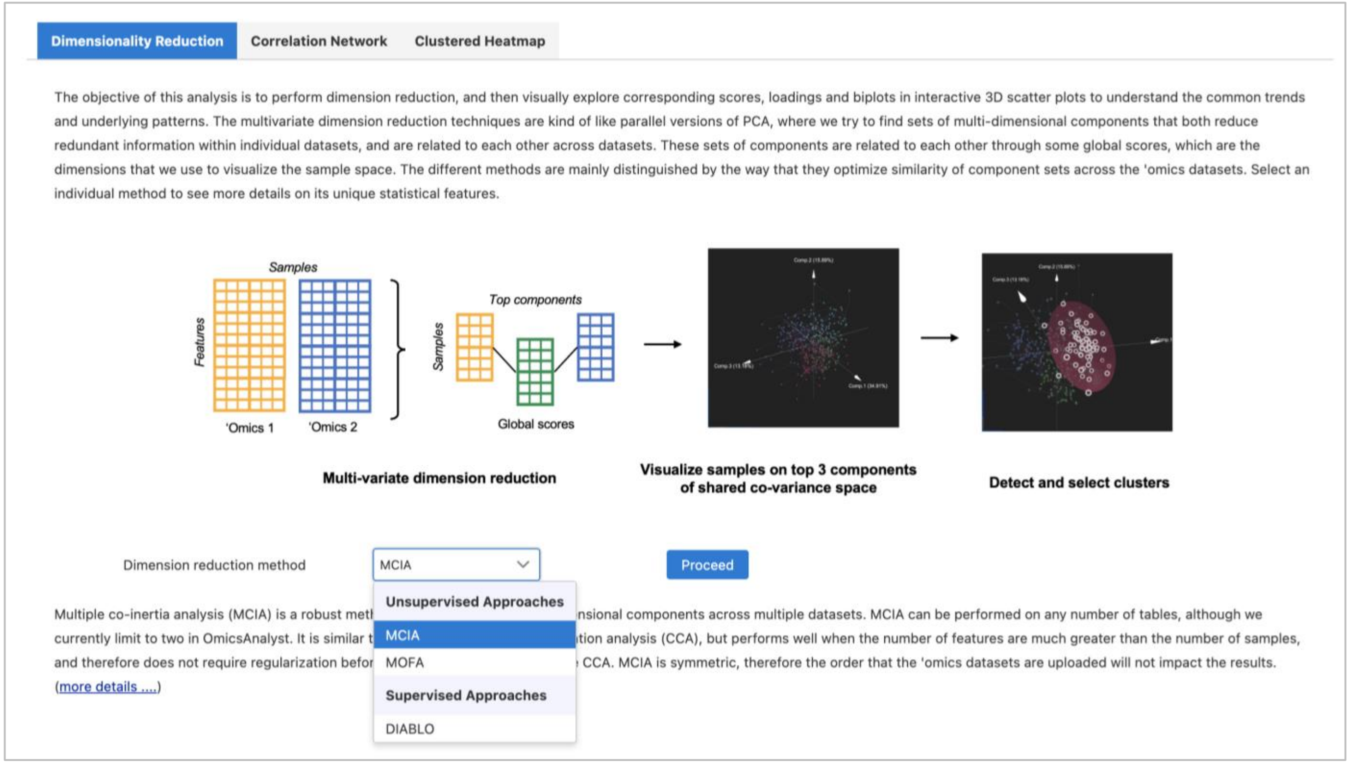
Covariance parameter: 0.2

Balance between covariance vs predictivity



Covariance: 0 0.2 0.4

Dimension Reduction track in OmicsAnalyst



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.

Samples
Features
'Omics 1
'Omics 2
Samples
Top components
Global scores
Multi-variate dimension reduction
Visualize samples on top 3 components of shared co-variance space
Detect and select clusters

Dimension reduction method: MCIA

Multiple co-inertia analysis (MCIA) is a robust meth nsional components across multiple datasets. MCIA can be performed on any number of tables, although we currently limit to two in OmicsAnalyst. It is similar t tion analysis (CCA), but performs well when the number of features are much greater than the number of samples, and therefore does not require regularization befor CCA. MCIA is symmetric, therefore the order that the 'omics datasets are uploaded will not impact the results. (more details ....)

XiaLab.ca

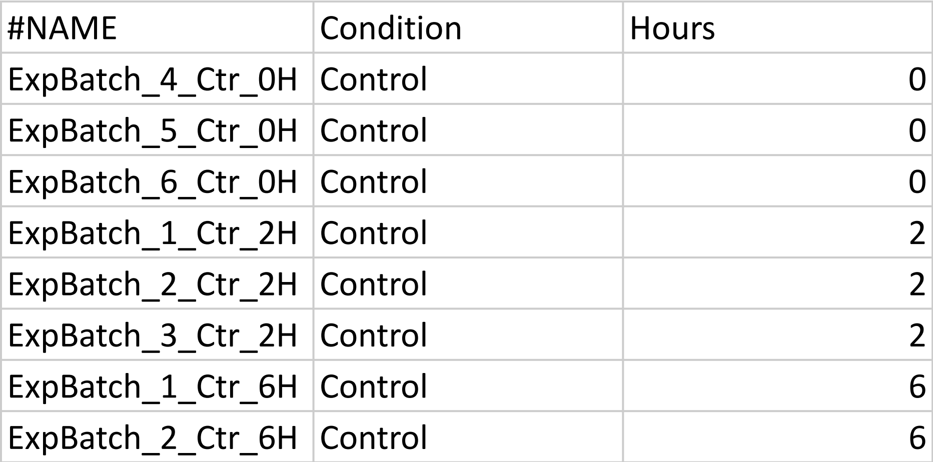
Empowering researchers through trainings, tools, and AI

Live Demo

Background

- Mouse multi-omics data on the effect of Ikaros transcription factor on B-cell differentiation
- Transcriptomics, Metabolomics, miRNA
- Metadata
    - Condition: Control/Ikaros
    - Hours: 6 time points

Meta-data table

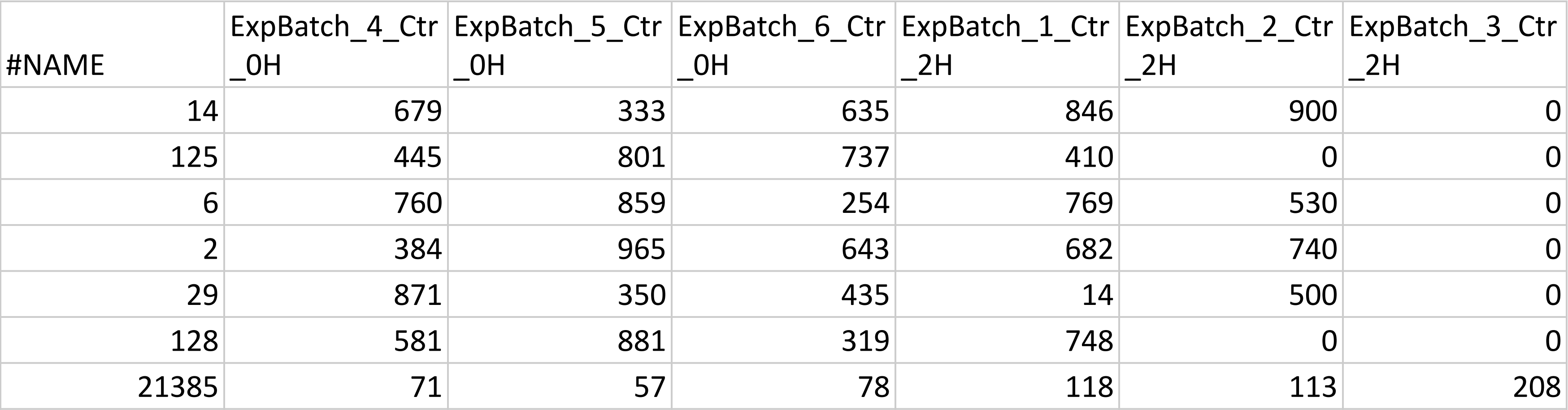


#NAME	Condition	Hours
ExpBatch_4_Ctr_0H	Control	0
ExpBatch_5_Ctr_0H	Control	0
ExpBatch_6_Ctr_0H	Control	0
ExpBatch_1_Ctr_2H	Control	2
ExpBatch_2_Ctr_2H	Control	2
ExpBatch_3_Ctr_2H	Control	2
ExpBatch_1_Ctr_6H	Control	6
ExpBatch_2_Ctr_6H	Control	6

⋮

- ➢Samples in rows, metadata group in columns
- ➢Make sure to exclude metadata group that only contains a single group.

Omics data 1 - transcriptomics



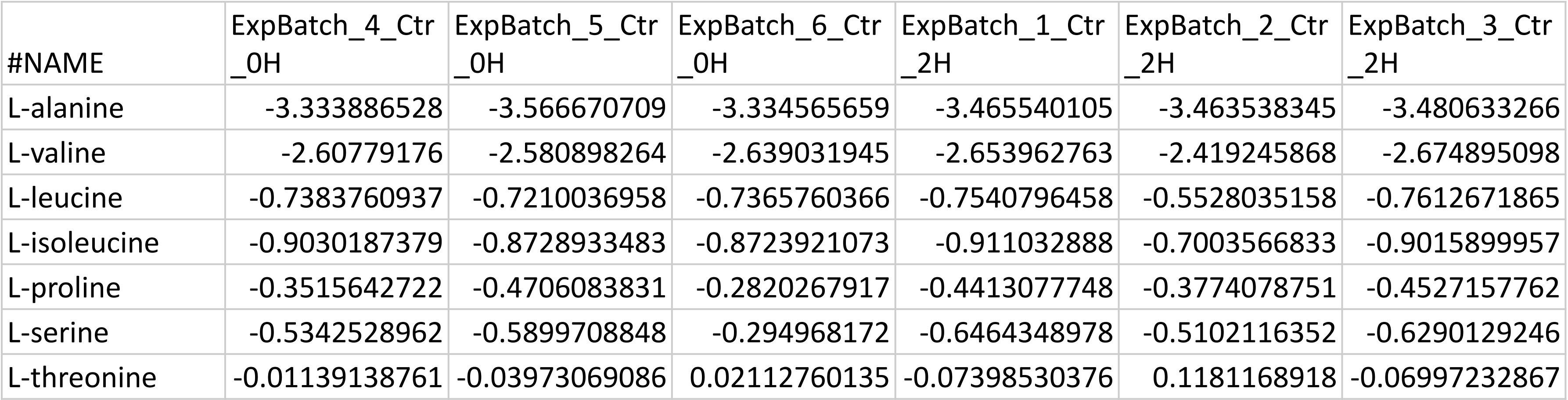
#NAME	ExpBatch_4_Ctr_0H	ExpBatch_5_Ctr_0H	ExpBatch_6_Ctr_0H	ExpBatch_1_Ctr_2H	ExpBatch_2_Ctr_2H	ExpBatch_3_Ctr_2H
14	679	333	635	846	900	0
125	445	801	737	410	0	0
6	760	859	254	769	530	0
2	384	965	643	682	740	0
29	871	350	435	14	500	0
128	581	881	319	748	0	0
21385	71	57	78	118	113	208

…

⋮

- ➢Processed data matrix
- ➢Samples in columns, features in rows (Entrez id).

Omics data #2 - Metabolomics



#NAME	ExpBatch_4_Ctr_0H	ExpBatch_5_Ctr_0H	ExpBatch_6_Ctr_0H	ExpBatch_1_Ctr_2H	ExpBatch_2_Ctr_2H	ExpBatch_3_Ctr_2H
L-alanine	-3.333886528	-3.566670709	-3.334565659	-3.465540105	-3.463538345	-3.480633266
L-valine	-2.60779176	-2.580898264	-2.639031945	-2.653962763	-2.419245868	-2.674895098
L-leucine	-0.7383760937	-0.7210036958	-0.7365760366	-0.7540796458	-0.5528035158	-0.7612671865
L-isoleucine	-0.9030187379	-0.8728933483	-0.8723921073	-0.911032888	-0.7003566833	-0.9015899957
L-proline	-0.3515642722	-0.4706083831	-0.2820267917	-0.4413077748	-0.3774078751	-0.4527157762
L-serine	-0.5342528962	-0.5899708848	-0.294968172	-0.6464348978	-0.5102116352	-0.6290129246
L-threonine	-0.01139138761	-0.03973069086	0.02112760135	-0.07398530376	0.1181168918	-0.06997232867

...

⋮

XiaLab.ca

Empowering researchers through trainings, tools, and AI

Omics data #3 - Metabolomics

#NAME	ExpBatch_4_Ctr_0H	ExpBatch_5_Ctr_0H	ExpBatch_6_Ctr_0H	ExpBatch_1_Ctr_2H	ExpBatch_2_Ctr_2H	ExpBatch_3_Ctr_2H
mmu-let-7g-3p	99	112	185	85	60	9
mmu-miR-1a-3p	0.2	0.2	0.2	0.2	0.2	0.2
mmu-miR-15b-5p	25915	28316	25890	36141	21100	5348
mmu-miR-15b-3p	4135	3687	5975	4817	2407	645
mmu-miR-23b-5p	11	20	24	2	2	1
mmu-miR-23b-3p	2824	2812	4048	3319	1905	406
mmu-miR-27b-5p	460	596	747	1161	919	120

⋮

Schedule for today

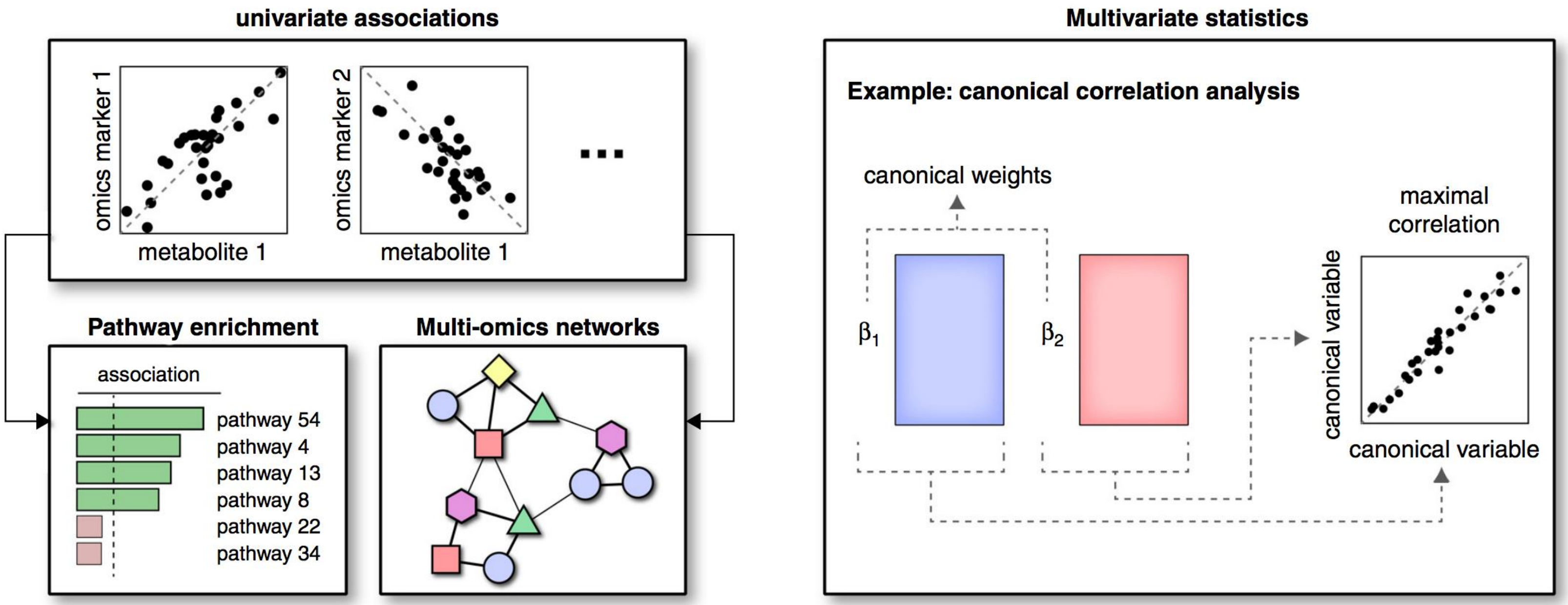
Time	Topics
9:00 – 9:10	Overview of data-driven multi-omics
9:10 – 9:50	Dimensionality reduction
9:50 – 10:10	Live Demo
10:15 – 10:40	Feature correlation analysis
10:40 – 10:55	Live Demo
10:55 – 11:10	Clustering analysis
11:10 – 11:25	Live Demo
Summary & Discussion	

Feature Correlation Analysis

What are informative features?

1. Features that are significant associated with phenotype
    - Differential expression analysis (univariate)
        - T-tests, ANOVA, limma, Fold change
    - Already discussed in the previous lectures
2. Features that are highly correlated across omics layers
    - Univariate correlation analysis
        - Parametric
        - Non-parametric
        - Partial correlation
    - Multivariate correlation analysis
        - Features that have large loadings in the jDR methods

Within-omics and between-omics correlation



Current Opinion in Biotechnology 2016, 39:198–206

Pearson’s covariance & correlation

Measures the relative strength of the linear relationship between two variables

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}(x_i-\overline{X})(y_i-\overline{Y})}{n-1}$$

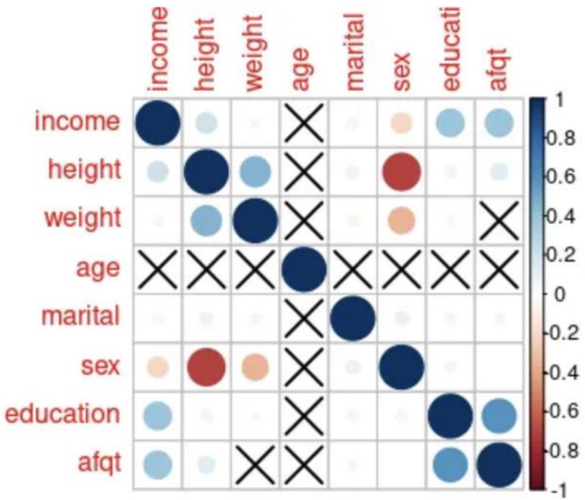
$$r=\frac{\operatorname{cov} ariance(x,y)}{\sqrt{\operatorname{var} x}\sqrt{\operatorname{var} y}}$$

cov(X,Y) > 0	X and Y are positively correlated
cov(X,Y) < 0	X and Y are inversely correlated
cov(X,Y) = 0	X and Y are independent

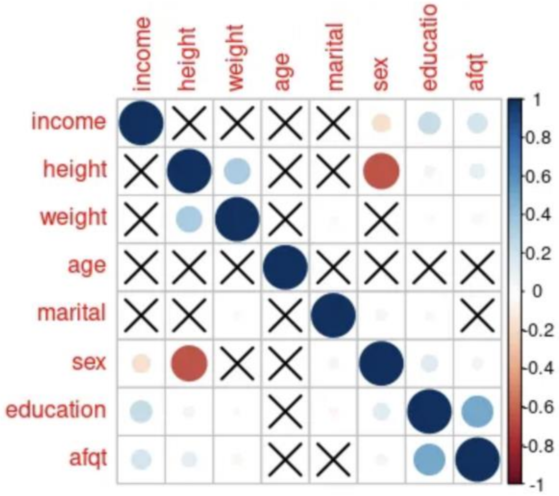
Pearson’s Correlation Coefficient is standardized covariance (unit-less)

Partial correlation

- ➢ The partial correlation coefficient is a measure of the strength of the linear relationship between two variables after entirely controlling for the effects of other variables



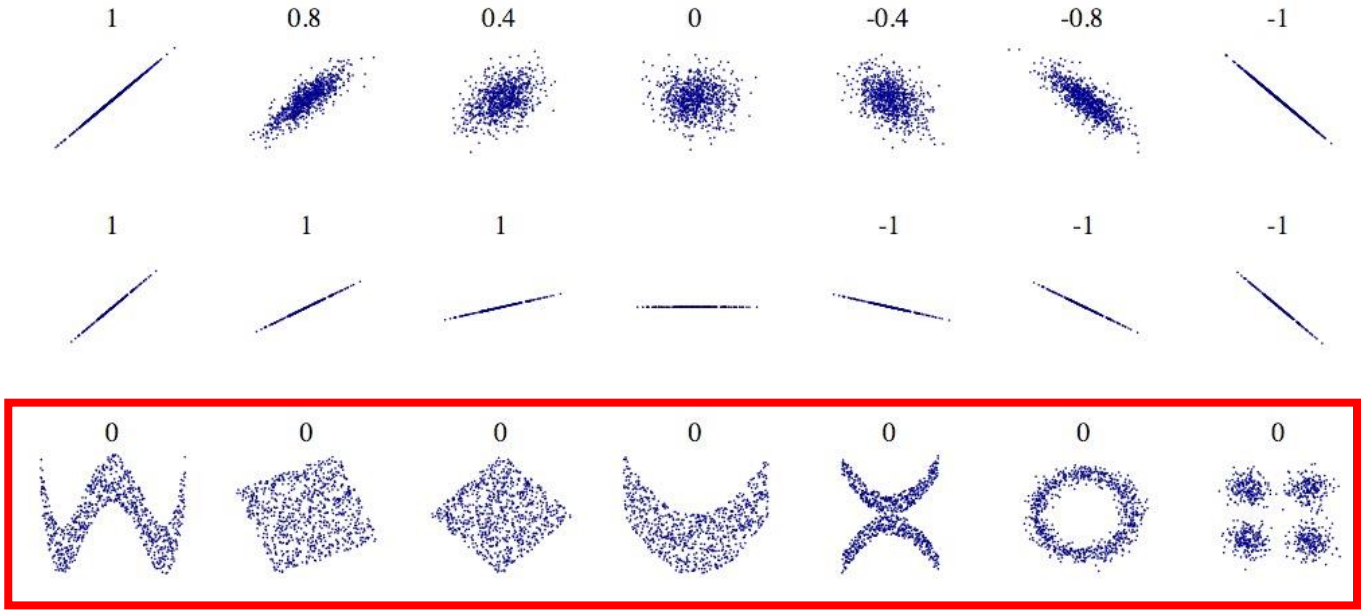
Correlation Matrix



Partial Correlation Matrix

https://towardsdatascience.com/partial-correlation-508353cd8b5

Pearson Correlation Coefficient Limitations



Source: wikimedia commons

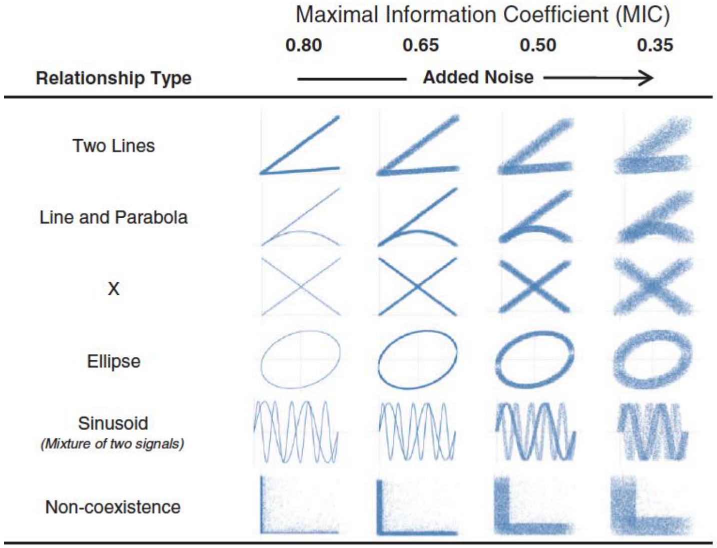
They are correlated!!

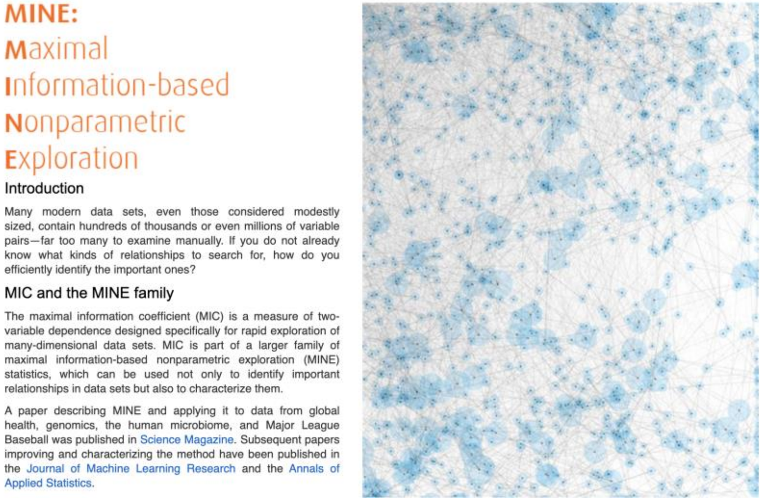
XiaLab.ca

Empowering researchers through trainings, tools, and AI

Detecting nonlinear correlations - Mutual Information

- Maximal Information Coefficient (MIC)





MINE: Maximal Information-based Nonparametric Exploration

Introduction

Many modern data sets, even those considered modestly sized, contain hundreds of thousands or even millions of variable pairs—far too many to examine manually. If you do not already know what kinds of relationships to search for, how do you efficiently identify the important ones?

MIC and the MINE family

The maximal information coefficient (MIC) is a measure of two-variable dependence designed specifically for rapid exploration of many-dimensional data sets. MIC is part of a larger family of maximal information-based nonparametric exploration (MINE) statistics, which can be used not only to identify important relationships in data sets but also to characterize them.

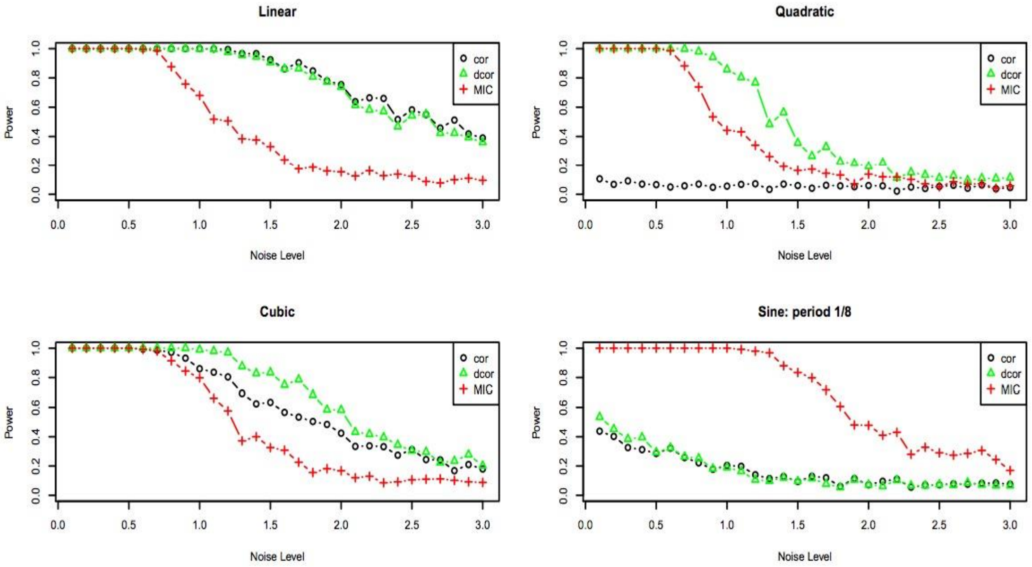
A paper describing MINE and applying it to data from global health, genomics, the human microbiome, and Major League Baseball was published in Science Magazine. Subsequent papers improving and characterizing the method have been published in the Journal of Machine Learning Research and the Annals of Applied Statistics.

http://www.exploredata.net

XiaLab.ca
Empowering researchers through trainings, tools, and AI

Nonlinear correlations - distance correlation

- Computes distance covariance and distance correlation statistics, which are multivariate measures of dependence.

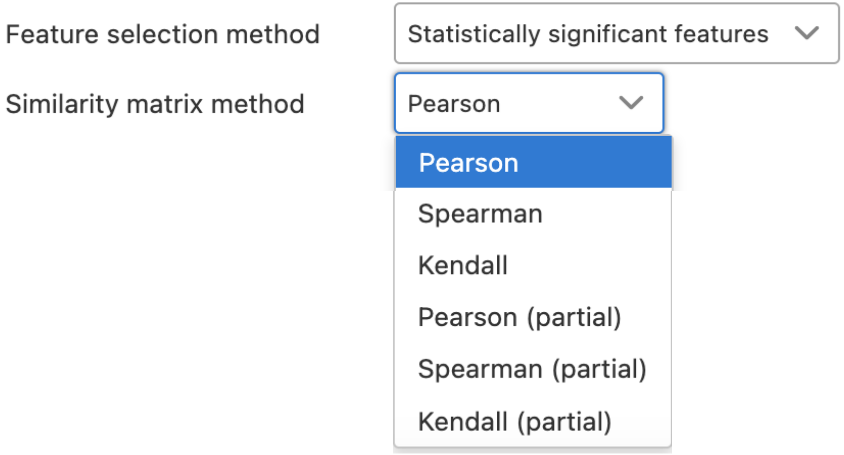


http://www-stat.stanford.edu/~tibs/reshef/comment.pdf

Correlation analysis in OmicsAnalyst

Feature level correlation analysis comparing 1000s vs 1000s features (> 1 million comparisons!)

- ➢ Significant features only
- ➢ Use linear methods



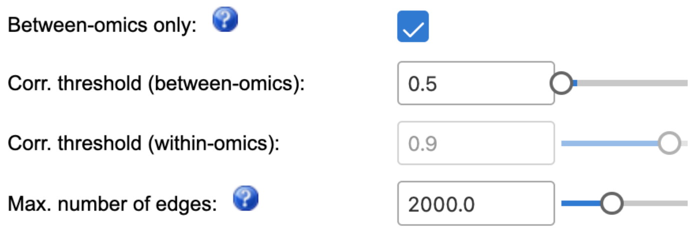
Feature selection method: Statistically significant features

Similarity matrix method: Pearson

Correlation network

There are systematic difference in between-omics and within-omics correlations

➔ Apply different cut-offs

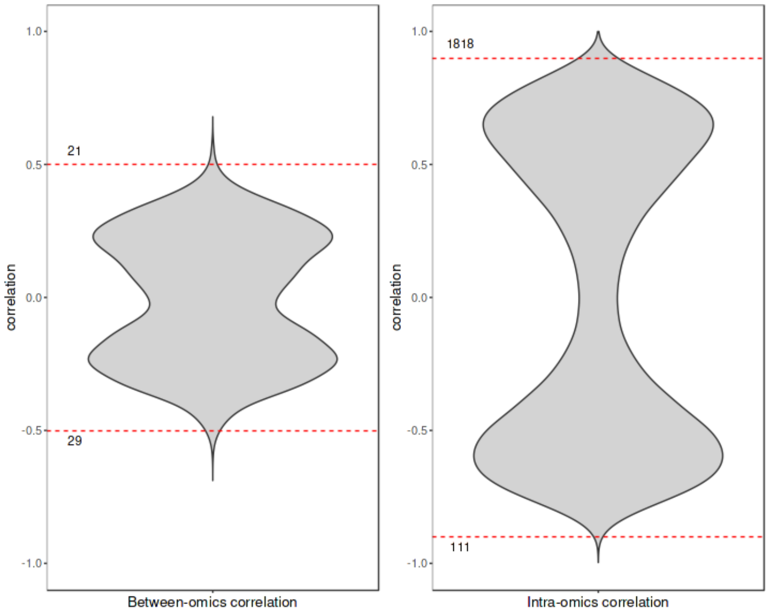


Between-omics only: ☑

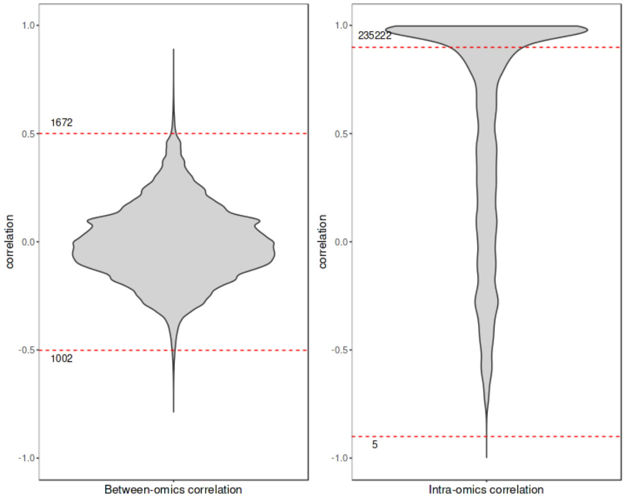
Corr. threshold (between-omics): 0.5

Corr. threshold (within-omics): 0.9

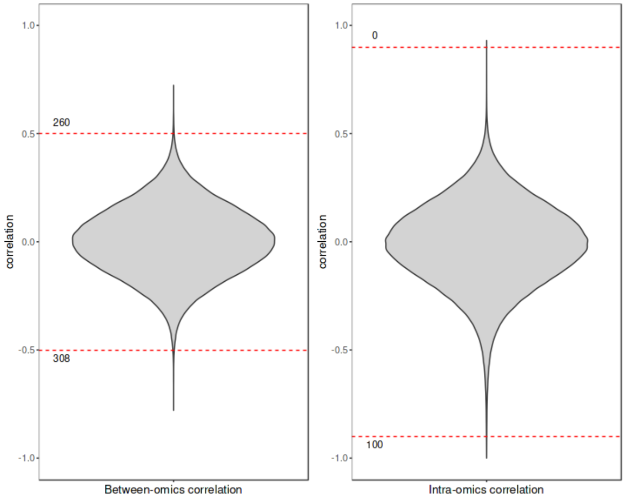
Max. number of edges: 2000.0



Correlation vs. partial correlation

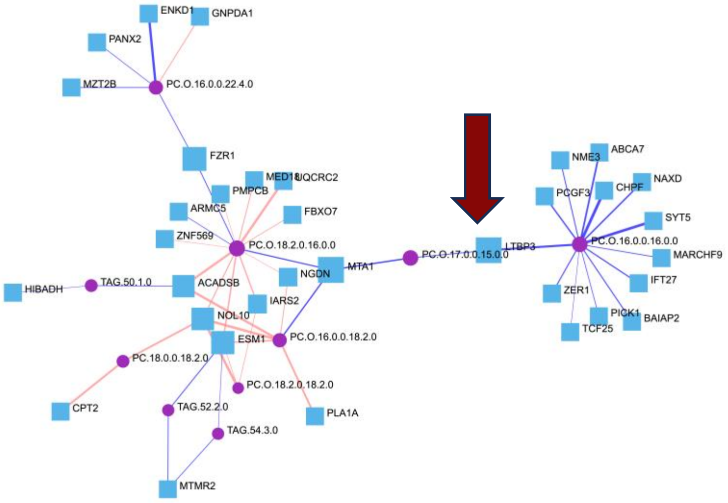


Pearson correlation

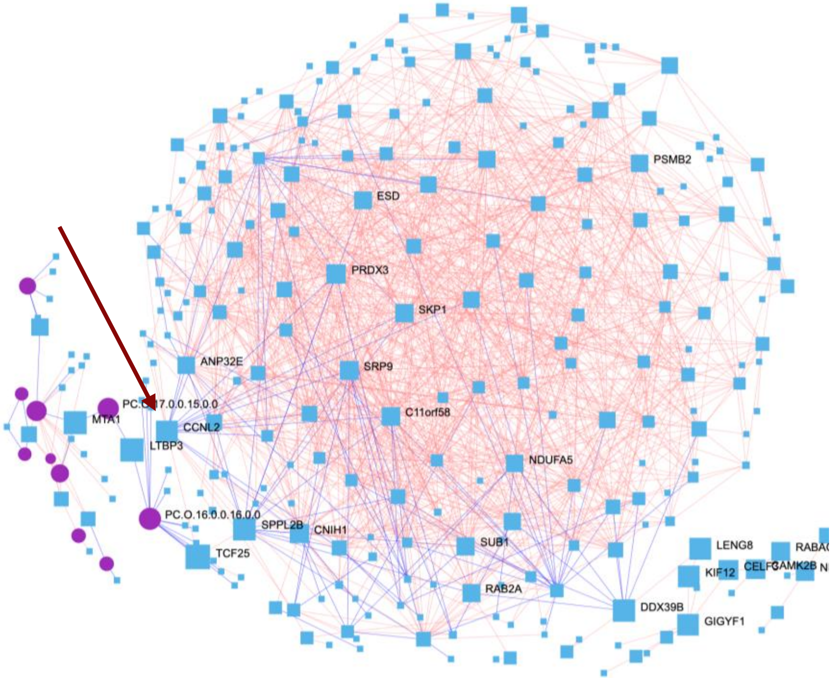


Pearson partial correlation

Multi-omics correlation network



Between omics only

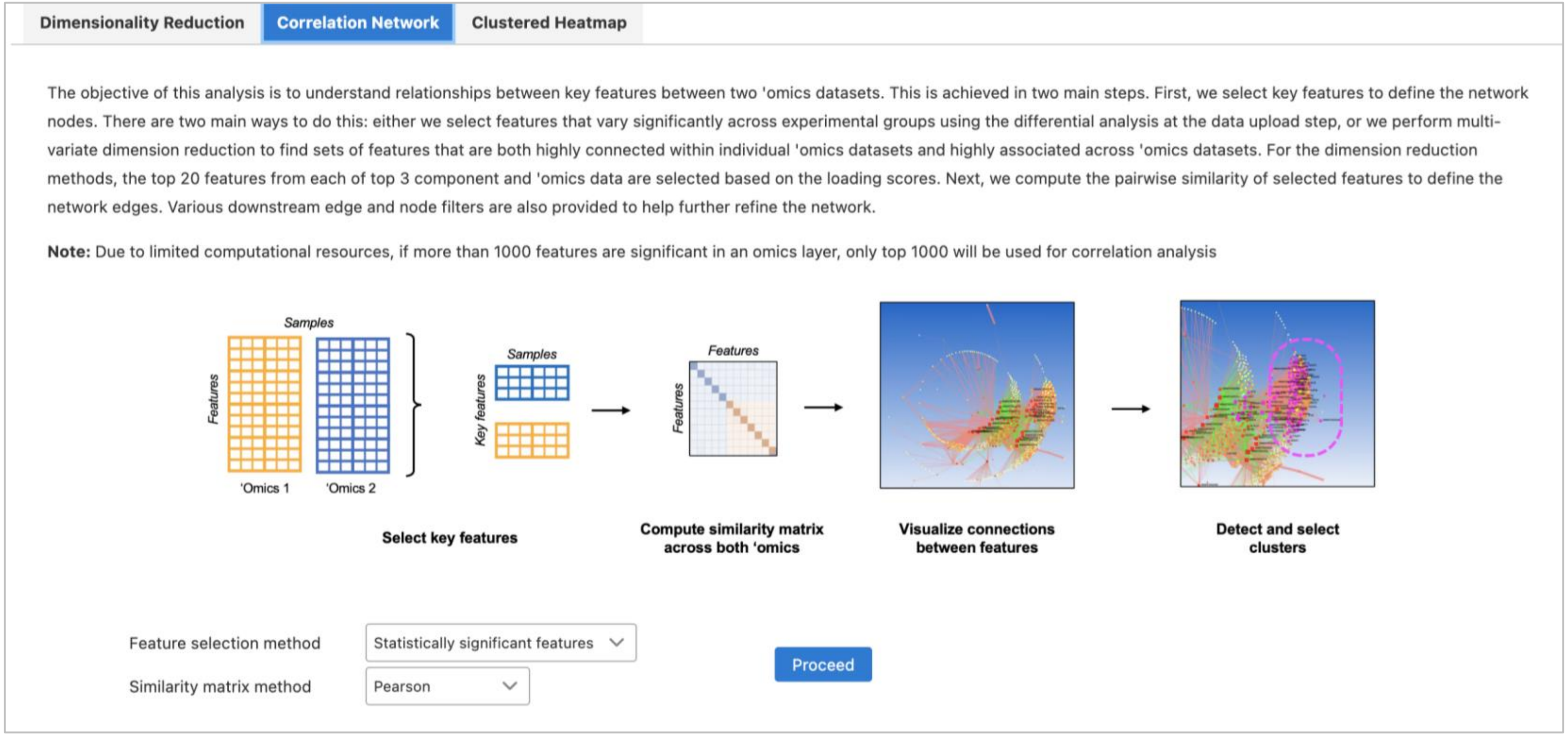


Within and between omics

XiaLab.ca

Empowering researchers through trainings, tools, and AI

Correlation Analysis track in OmicsAnalyst



The objective of this analysis is to understand relationships between key features between two 'omics datasets. This is achieved in two main steps. First, we select key features to define the network nodes. There are two main ways to do this: either we select features that vary significantly across experimental groups using the differential analysis at the data upload step, or we perform multi-variate dimension reduction to find sets of features that are both highly connected within individual 'omics datasets and highly associated across 'omics datasets. For the dimension reduction methods, the top 20 features from each of top 3 component and 'omics data are selected based on the loading scores. Next, we compute the pairwise similarity of selected features to define the network edges. Various downstream edge and node filters are also provided to help further refine the network.

Note: Due to limited computational resources, if more than 1000 features are significant in an omics layer, only top 1000 will be used for correlation analysis

Select key features → **Compute similarity matrix across both 'omics** → **Visualize connections between features** → **Detect and select clusters**

Feature selection method: Statistically significant features

Similarity matrix method: Pearson

Live Demo

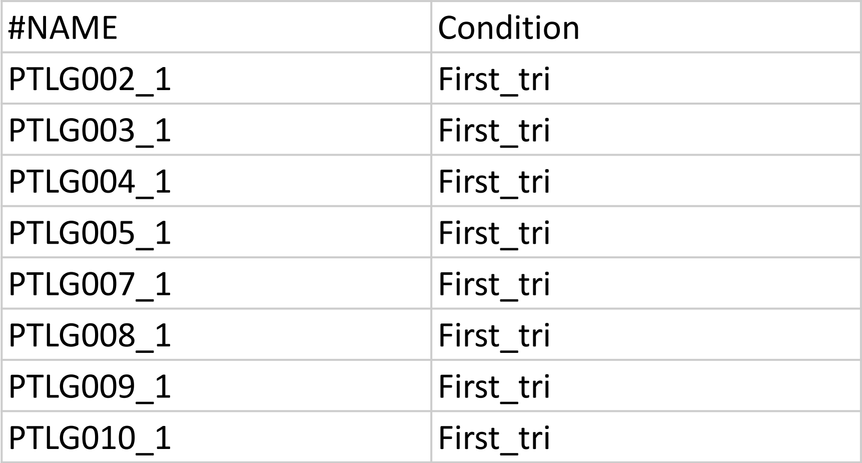
Summary

1. Feature selection
   - Significant features from DE analysis
   - Top features identified from dimensional reduction methods
2. Correlation analysis among selected features
   - Correlation or partial correlation
3. Network building
   - Inter-omics/Intra-omics relationships
   - Control size through correlation threshold
4. Network visualization and analysis
   - 2D/3D
   - Topological and enrichment analysis (if applicable)

Background

- Human multi-omics data on pregnancy progress
- Lipidomics, RNA-seq
- Metadata (4 groups)
    - First trimester
    - Second trimester
    - Third trimester
    - Baseline

Meta-data table

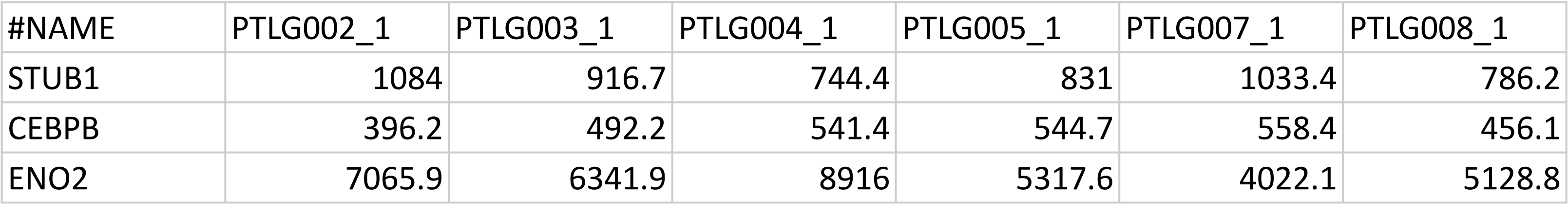


#NAME	Condition
PTLG002_1	First_tri
PTLG003_1	First_tri
PTLG004_1	First_tri
PTLG005_1	First_tri
PTLG007_1	First_tri
PTLG008_1	First_tri
PTLG009_1	First_tri
PTLG010_1	First_tri

⋮

- ➢Samples in rows, metadata group in columns
- ➢Make sure to exclude metadata group that only contains a single group.

Omics data 1 - proteomics



#NAME	PTLG002_1	PTLG003_1	PTLG004_1	PTLG005_1	PTLG007_1	PTLG008_1
STUB1	1084	916.7	744.4	831	1033.4	786.2
CEBPB	396.2	492.2	541.4	544.7	558.4	456.1
ENO2	7065.9	6341.9	8916	5317.6	4022.1	5128.8

…

⋮

- ➢Processed data matrix
- ➢Samples in columns, features in rows.

Omics data #2 - Metabolomics

#NAME	PTLG002_1	PTLG003_1	PTLG004_1	PTLG005_1	PTLG007_1	PTLG008_1
N1-Methyl-2-pyridone-5-carboxamide	0.0550991485	0.06030218125	0.0745380945	0.04959586675	0.0507334805	0.133186114
Barringtogenol C	0.05766222871	0.07476740471	0.1102784697	0.08434155943	0.2133026861	0.1387500359
3beta-Acetoxy-11alpha-methoxy-12-ursen-28-oic acid	0.057382463	0.056692533	0.067014164	0.0516164	0.091558005	0.134262229
Basilimoside	0.04493315375	0.0590499555	0.0632840915	0.05531613575	0.117367835	0.13811012
2s-Pyrrolidin-2-Ylmethylamine	0.040983833	0.049911143	0.053143402	0.045142262	0.045710933	0.132127336

...

⋮

Schedule for today

Time	**Topics**
9:00 – 9:10	Overview of data-driven multi-omics
9:10 – 9:50	Dimensionality reduction
9:50 – 10:10	Live Demo
10:15 – 10:40	Feature selection and correlation
10:40 – 10:55	Live Demo
➡ **10:55 – 11:10**	Clustering analysis
11:10 – 11:25	Live Demo
Summary & Discussion	

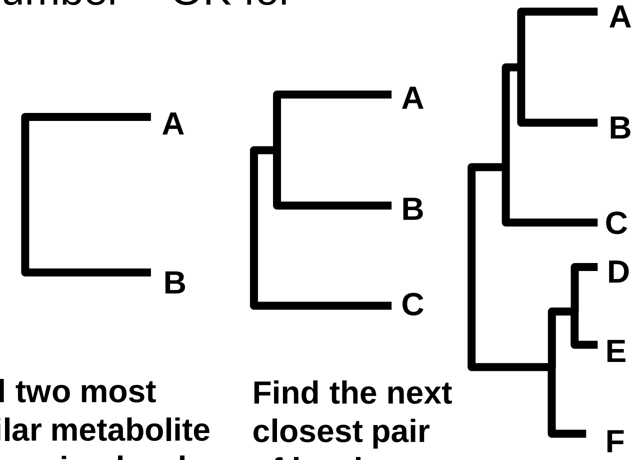
Clustering Analysis

Common clustering algorithms for omics data

- Hierarchical clustering
- K-means clustering
- Spectrum clustering
- Similar Network Fusion

Hierarchical clustering & heatmap

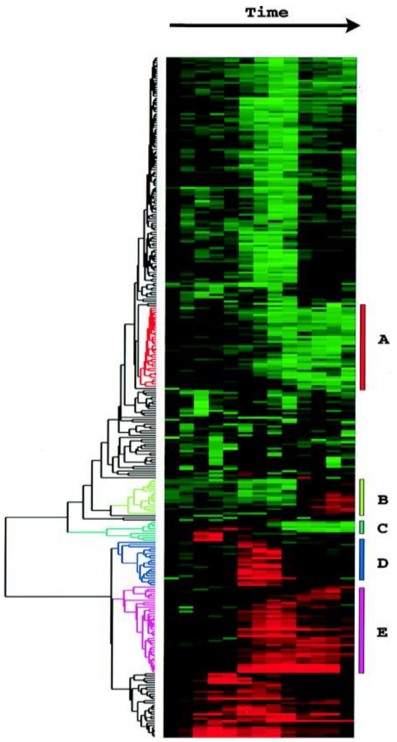
- Produces a set of nested clusters in which each pair of objects is progressively nested into a larger cluster until only one cluster remains
- No explicit set for cluster number – OK for human, hard for computer



Find two most similar metabolite expression levels or curves

Find the next closest pair of levels or curves

Iterate

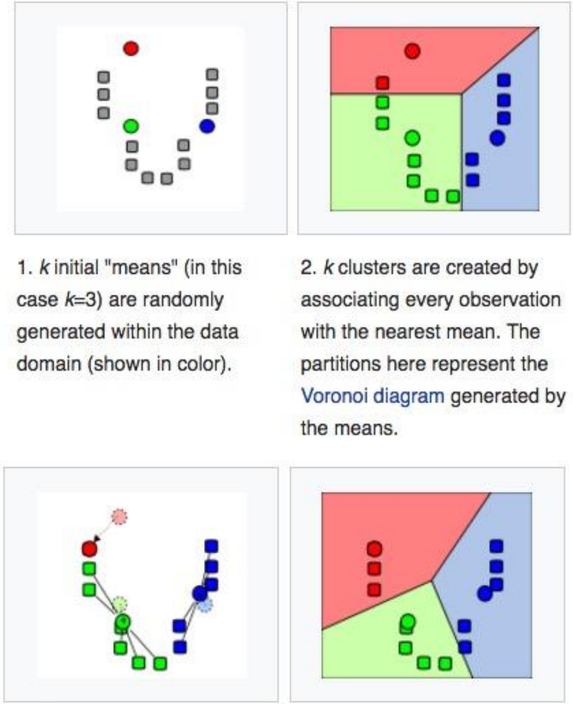


K-means

Goal: minimize the cost which is defined as the sum of squared distances between all data points and their cluster centers.

Initialisation: set seed points (randomly)

1) Assign each object to the cluster of the nearest seed point measured with a specific distance metric
2) Compute new seed points as the centroids of the clusters of the current partition (the centroid is the centre, i.e., *mean point*, of the cluster)
3) Go back to Step 1), stop when no more new assignment (i.e., membership in each cluster no longer changes)



1. *k* initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color).

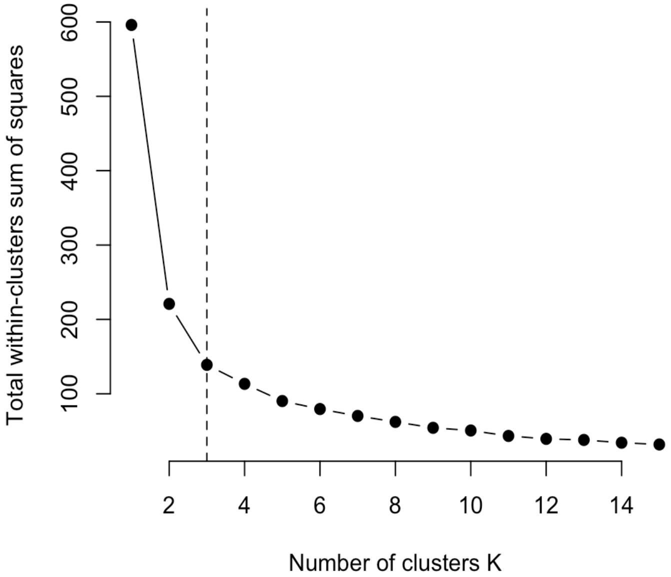
2. *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

3. The centroid of each of the *k* clusters becomes the new mean.

4. Steps 2 and 3 are repeated until convergence has been reached.

https://en.wikipedia.org/wiki/K-means_clustering

K-means – the value of K

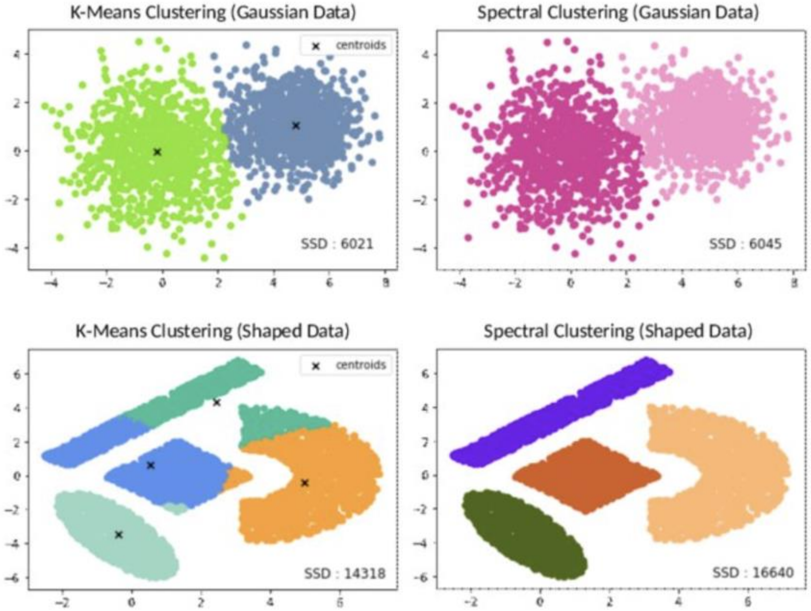


Testing with different K values.

Tries to minimize the within-cluster sum of squares error (WCSS)

Spectrum clustering

- K-means algorithm generally assumes that the clusters are spherical or round i.e. within k-radius from the cluster centroid
- Spectral clustering helps us overcome two major problems in clustering: one being the shape of the cluster and the other is determining the cluster centroid
- Spectrum combines the strengths of several other methods: an adaptive density-aware kernel is used to strengthen connections in the graph based on common nearest neighbors.



Spectrum clustering

- Key steps in Spectrum clustering
  1. Compute a weighted adjacent matrix is derived from the input dataset.
  2. Compute eigenvalues and eigenvectors of this matrix to partition the data.
  3. Apply K-means on the “embedding” space to derive clustering

- OmicsAnalyst uses the eigengap mode in the Spectrum R package, which is suited for Gaussian distributed data

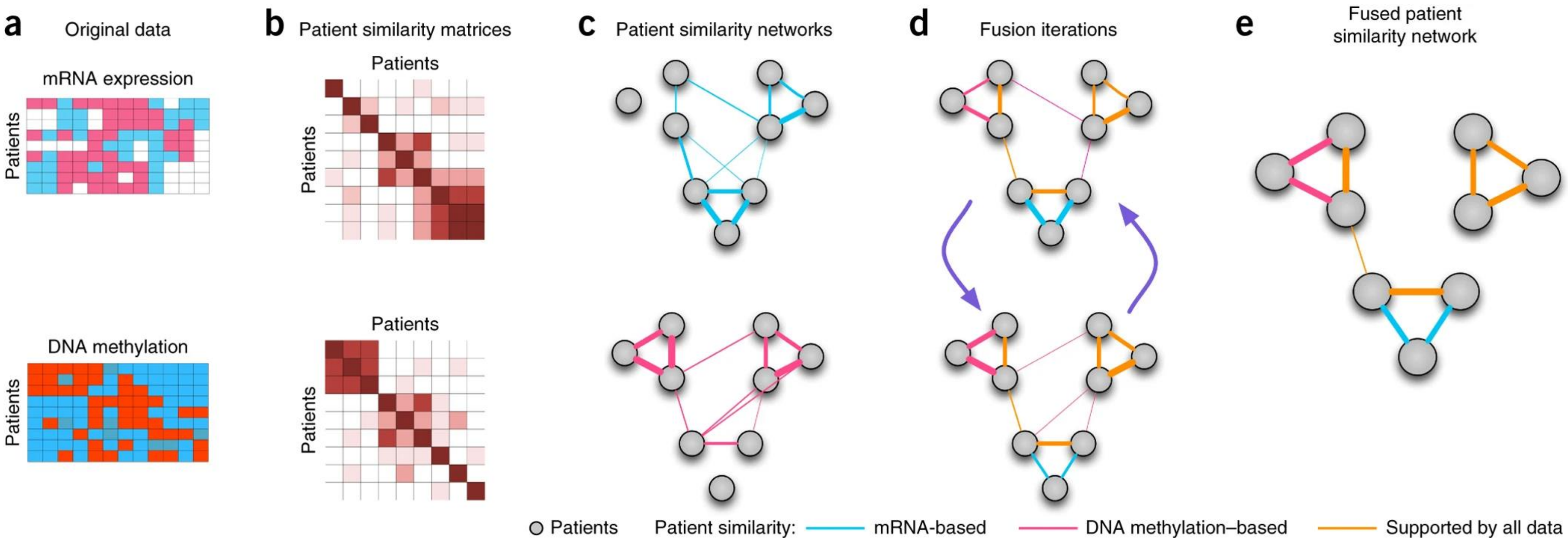
Similar Network Fusion (SNF)

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.

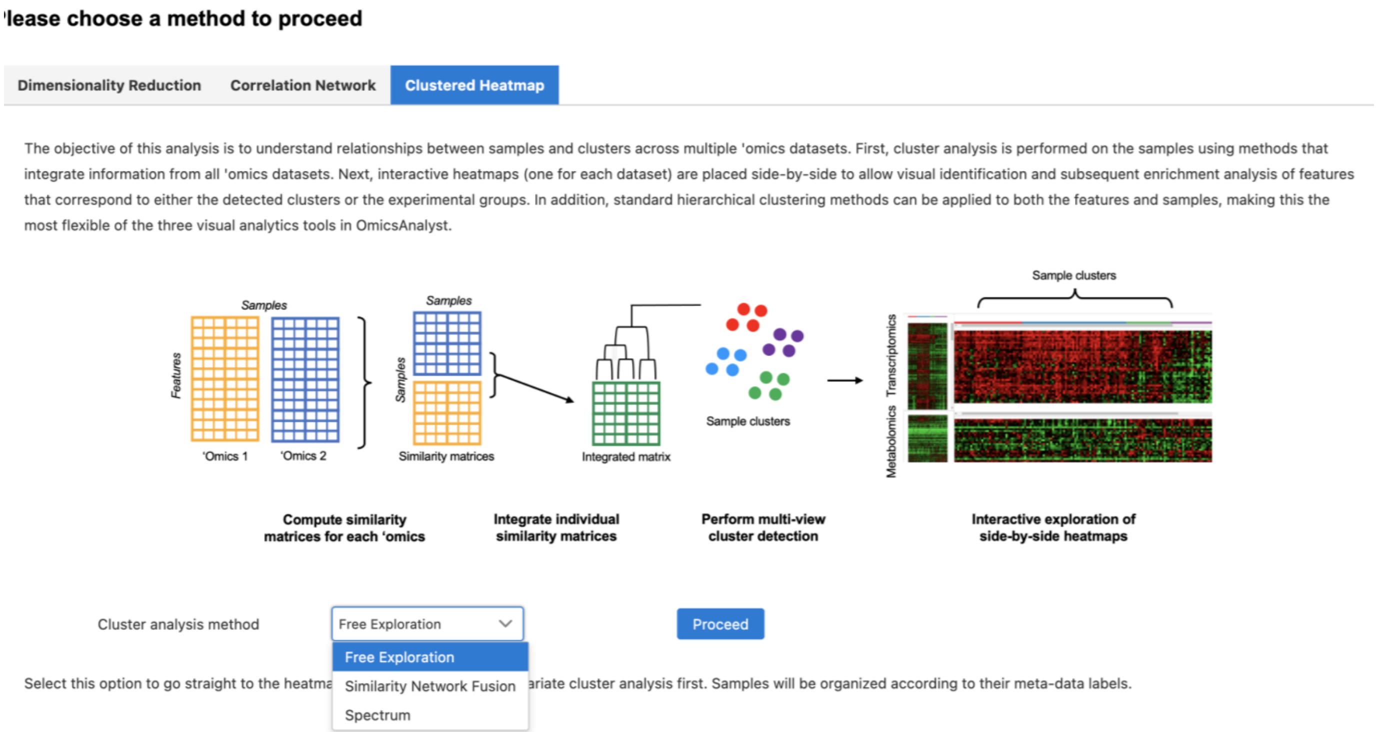
1. Individual similarity matrices are computed using an exponential similarity kernel that scales the Euclidean distance between samples.
2. These matrices are then fused together by an iterative approach that adjusts each matrix to make it more similar to the others.
3. The SNF algorithm is iterated until the matrices converge.

The fused network captures both shared and complementary information from different data sources

Similar Network Fusion Workflow



Clustering Analysis in OmicsAnalyst



'lease choose a method to proceed

The objective of this analysis is to understand relationships between samples and clusters across multiple 'omics datasets. First, cluster analysis is performed on the samples using methods that integrate information from all 'omics datasets. Next, interactive heatmaps (one for each dataset) are placed side-by-side to allow visual identification and subsequent enrichment analysis of features that correspond to either the detected clusters or the experimental groups. In addition, standard hierarchical clustering methods can be applied to both the features and samples, making this the most flexible of the three visual analytics tools in OmicsAnalyst.

Samples
Features
'Omics 1
'Omics 2
Samples
Samples
Similarity matrices
Integrated matrix
Sample clusters
Sample clusters
Transcriptomics
Metabolomics
Compute similarity matrices for each 'omics
Integrate individual similarity matrices
Perform multi-view cluster detection
Interactive exploration of side-by-side heatmaps

Cluster analysis method: Free Exploration

Select this option to go straight to the heatma ariate cluster analysis first. Samples will be organized according to their meta-data labels.

Clustering analysis track in OmicsAnalyst

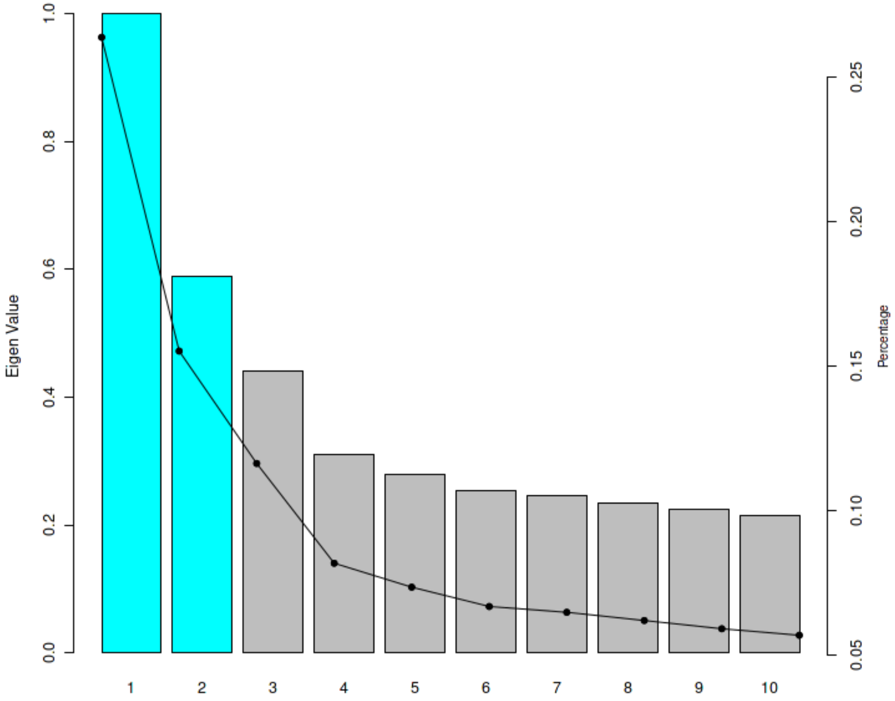
To understand relationships between samples and clusters across two 'omics datasets.

1. First, cluster analysis is performed on the samples using methods that integrate information from all 'omics datasets.
2. Interactive heatmaps (one for each dataset) are placed side-by-side to allow visual identification and subsequent enrichment analysis of features that correspond to either the detected clusters or the experimental groups.

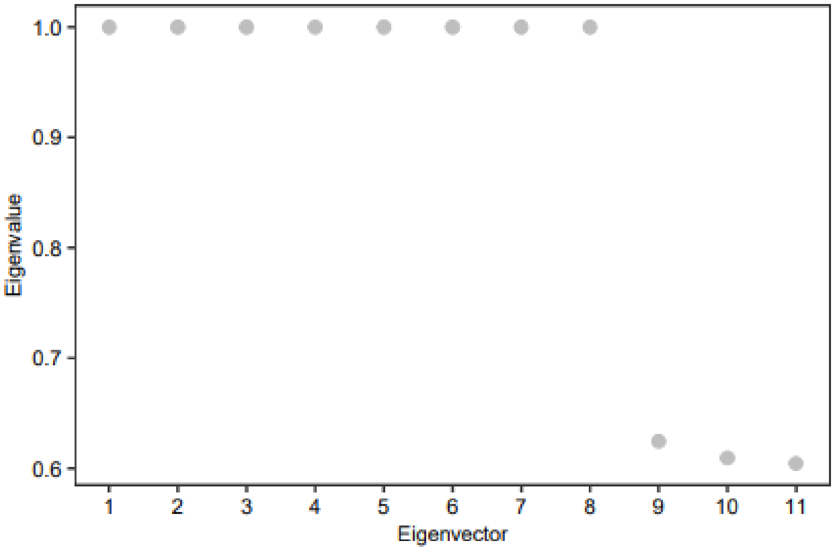
Live Demo

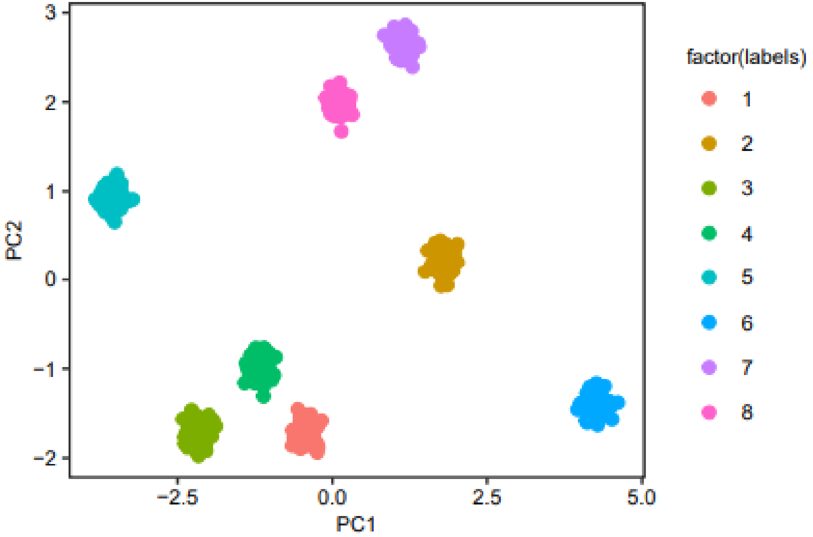
Eigenvalue per cluster number

- "Eigengap" method to identify the optimal number of clusters
- **Assumption**:
  - The eigenvalues can reveal the intrinsic clustering structure of the dataset by indicating points of significant change
- The cluster number where **largest drop in eigenvalue** happens.
- **Not a *hard* rule.** It's a heuristic to help you choose.



Eigengap example





Schedule for today

Time	Topics
9:00 – 9:10	Overview of data-driven multi-omics
9:10 – 9:50	Dimensionality reduction
9:50 – 10:10	Live Demo
10:15 – 10:40	Feature selection and correlation
10:40 – 10:55	Live Demo
10:55 – 11:10	Clustering analysis
11:10 – 11:25	Live Demo
Summary & Discussion	

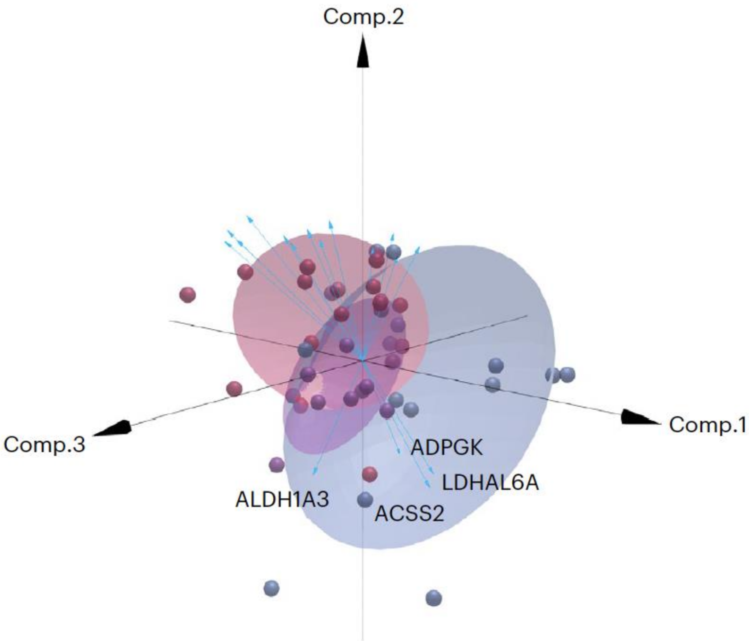
General comments on DR

- Mainly for exploratory analysis
- Very complex
- Common performance evaluation methods cannot be readily applied
  - P values
  - Permutations
  - Cross validations
- How to make choices?

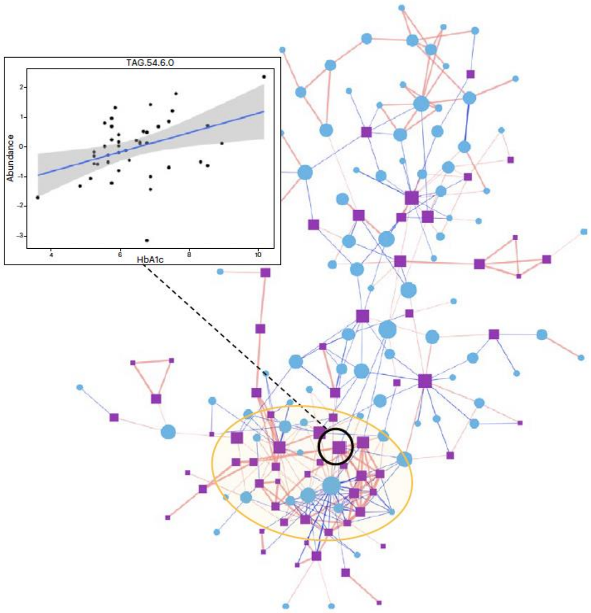
Some rules of thumb to reduce false patterns

- **Occam's razor,** assuming that the simplest consistent hypothesis about the target function is actually the best.
  - Minimal number of features and simple algorithms
- **Minimum cross-validation error:** when trying to choose among hypotheses, select the hypothesis with the lowest cross-validation error.
- **Maximum separation distance:** when drawing a boundary between two classes, attempt to maximize the width of the boundary. The assumption is that distinct classes tend to be separated by wide boundaries.
- **Nearest neighbors:** assume that most of the cases in a small neighborhood in feature space belong to the same class. The assumption is that cases that are near each other tend to belong to the same class.

Combining multiple independent methods

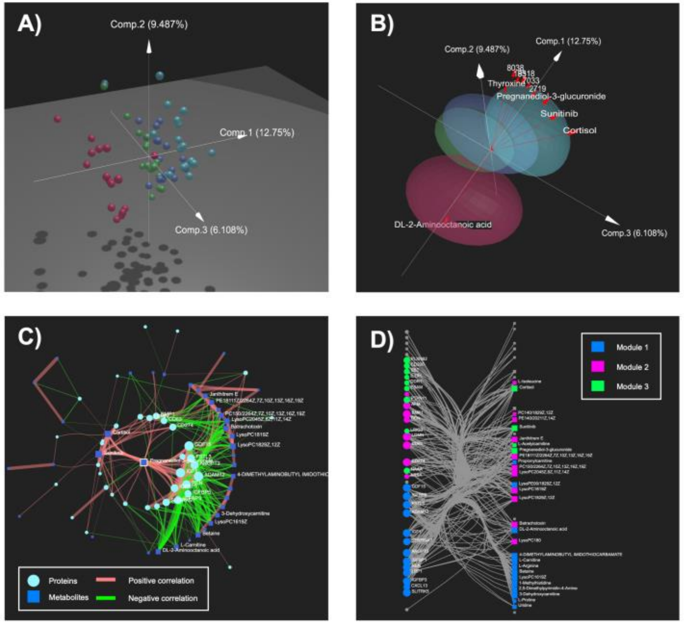


View features from dimension reductions

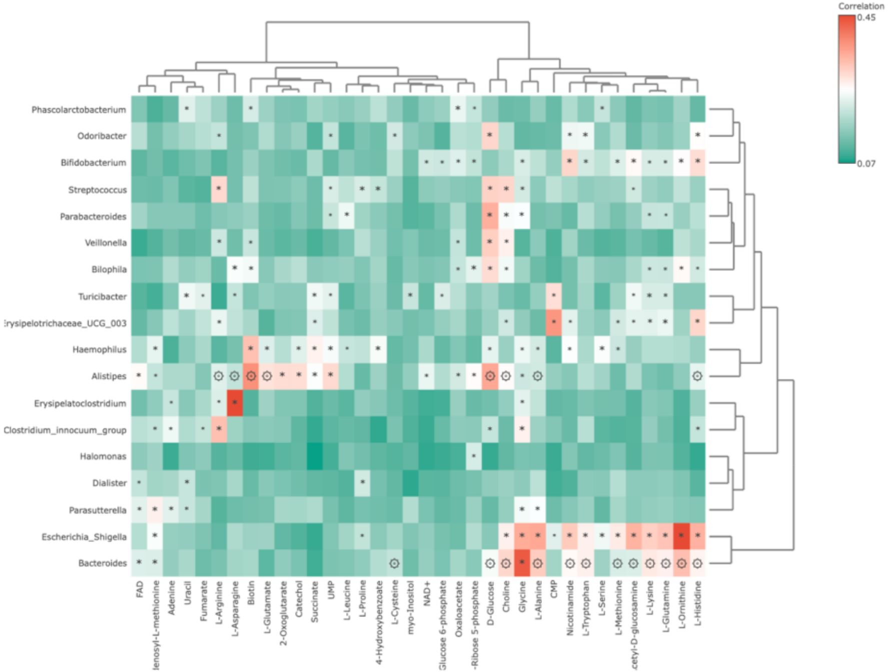


Correlation network of high-magnitude loadings from the top 3 components.

Visualization & biology



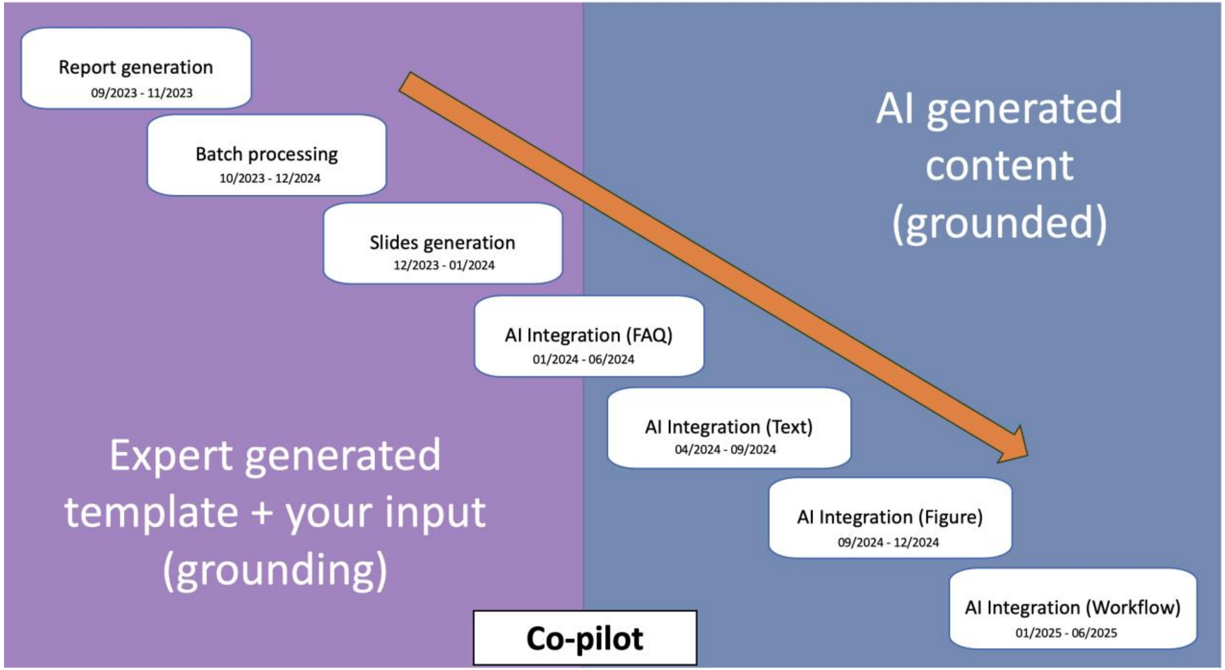
Different perspectives



Overlay statistics and knowledge annotation

Conclusions

Towards AI co-pilot



Report generation
09/2023 - 11/2023

Batch processing
10/2023 - 12/2024

Slides generation
12/2023 - 01/2024

AI Integration (FAQ)
01/2024 - 06/2024

AI Integration (Text)
04/2024 - 09/2024

AI Integration (Figure)
09/2024 - 12/2024

AI Integration (Workflow)
01/2025 - 06/2025

Expert generated template + your input (grounding)

AI generated content (grounded)

Co-pilot

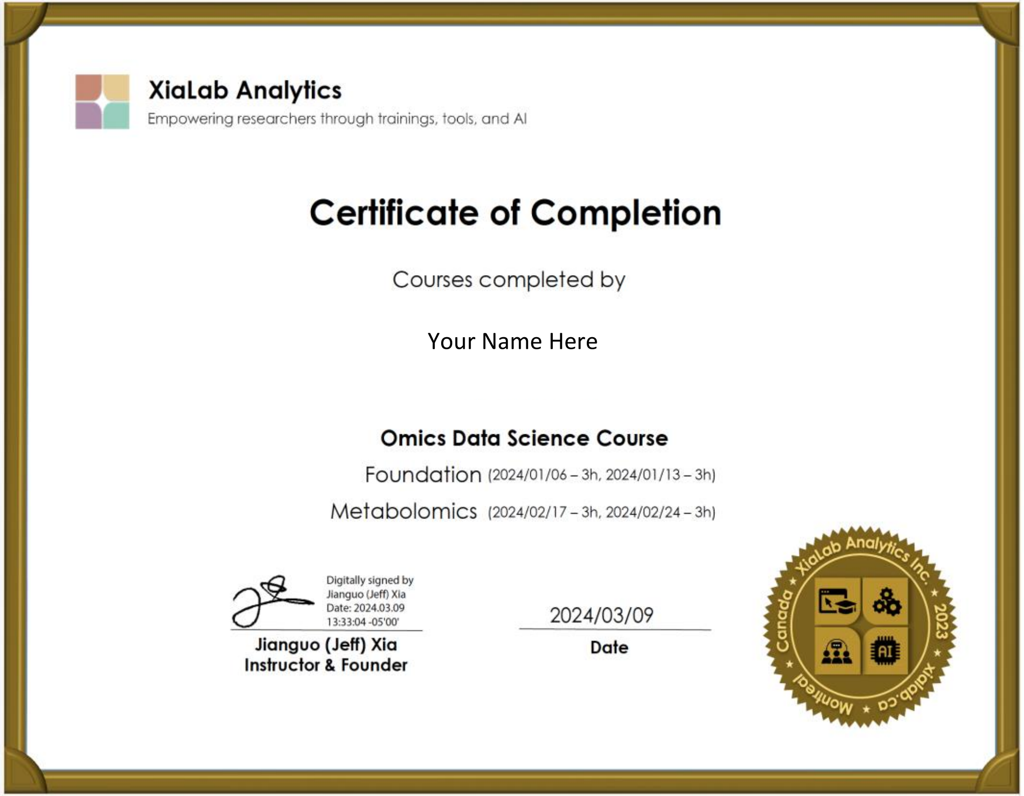
Empowering research

- AI can do straightforward, time-consuming tasks

Tasks	Grounding	Status
Run data analysis workflow	Our validated pipelines	✓
Generate analysis report	Report / slide templates	✓
Provide FAQs & writing summary (manuscript draft)	Searching literature & forum	Ongoing
Manuscript 1.0	Your turn	--

We will send you the Zoom invitation in May

Certificate



XiaLab Analytics

Empowering researchers through trainings, tools, and AI

Certificate of Completion

Courses completed by

Your Name Here

Omics Data Science Course

Foundation (2024/01/06 – 3h, 2024/01/13 – 3h)

Metabolomics (2024/02/17 – 3h, 2024/02/24 – 3h)

Digitally signed by Jianguo (Jeff) Xia
Date: 2024.03.09
13:33:04 -05'00'

Jianguo (Jeff) Xia
Instructor & Founder

2024/03/09

Date

XiaLab Analytics Inc. ★ 2023 ★ xialab.ca ★ Montreal ★ Canada

contact@xialab.ca with your name

Help pass the word

- Three times per year
  - Winter session: Saturday morning, Jan. - March.
  - Fall session: Saturday morning, Sept. - Nov.
  - Summer bootcamp: tentatively Aug. 5 - 9

➔ You are welcome to attend (need to register)

- Active membership (annual)
- Same selections

Thank you & see you in May!