
awesome-multi-omics

A community-maintained list of software packages for multi-omics data analysis.

While many of the packages here are marketed for “omics” data (transcriptomics, proteomics, etc.), other more general terms for this type of data analysis are:

- multi-modal
- multi-table
- multi-way

The common thread among the methods listed here is that the same samples are measured across different assays. The data can be described as multiple matrices/tables with the same number of samples and varying number of features.

The repo is in the style of Sean Davis’ awesome-single-cell repo for single-cell analysis methods.

Contributions welcome...

For brevity, below lists only the first author of multi-omics methods.

Software packages and methods

Multi-omics correlation or factor analysis

- 2007 - **SCCA** - Parkhomenko - sparse CCA - paper 1, paper 2
- 2008 - **PCCA** - Waaijenborg - penalized CCA / CCA-EN - paper
- 2009 - PMA - Witten - Sparse Multi CCA - paper 1, paper 2
- 2009 - **sPLS** - Lê Cao - sparse PLS - paper
- 2009 - gesca - Hwang - RGSCA regularized generalized structured component analysis - paper
- 2010 - **Regularized dual CCA** - Soneson - paper
- 2011 - RGCCA - Tenenhaus - Regularized Generalized CCA and Sparse Generalized CCA - paper 1, paper 2
- 2011 - **SNMNMf** - Zhang - Sparse Network-regularized Multiple Non-negative Matrix Factorization - paper
- 2011 - scca - Lee - Sparse Canonical Covariance Analysis for High-throughput Data - paper
- 2012 - STATIS/DiSTATIS - Abdi - structuring three-way statistical tables - paper
- 2012 - **joint NMF** - Zhang - extension of NMF to multiple datasets - paper
- 2012 - **sMBPLS** - Li - sparse MultiBlock Partial Least Squares - paper
- 2012 - **Bayesian group factor analysis** - Virtanen - paper
- 2012 - RIMBANET - Zhu - Reconstructing Integrative Molecular Bayesian Networks - paper

-
- 2013 - FactoMineR - Abdi - MFA: multiple factor analysis - paper
 - 2013 - JIVE - Lock - joint & individual variance explained - paper
 - 2013 - pandaR - Schlauch - Passing Attributes between Networks for Data Assimilation - paper
 - 2014 - omicade4 - Meng - MCIA: multiple co-interia analysis - paper
 - 2014 - STATegRa - Planell - DISCO, JIVE, & O2PLS - paper
 - 2014 - **Joint factor model** - Ray - paper
 - 2014 - GFAsparse - Khan - group factor analysis sparse paper 1, paper 2
 - 2015 - **Sparse CCA** - Gao (3rd paper first author is Chen) - paper 1, paper 2, paper 3
 - 2015 - CCAGFA - Klami - Bayesian Canonical Correlation Analysis and Group Factor Analysis - paper 1, paper 2
 - 2016 - CMF - Klami - collective matrix factorization - paper
 - 2016 - moGSA - Meng - multi-omics gene set analysis - paper
 - 2016 - iNMF - Yang - integrative NMF - paper
 - 2016 - BASS - Zhao - Bayesian group factor analysis - paper
 - 2016 - **imputeMFA** in missMDA - Voillet - multiple imputation for multiple factor analysis (MI-MFA) - paper
 - 2016 - PLSCA - Beaton - Partial Least Square Correspondence Analysis - paper
 - 2017 - mixOmics - Rohart - various methods - paper1, paper2
 - 2017 - mixedCCA - Yoon - sparse CCA for data of mixed types - paper
 - 2017 - SLIDE - Gaynanova - Structural Learning and Integrative Decomposition of Multi-View Data - paper
 - 2017 - fCCAC - Madrigal - functional canonical correlation analysis to evaluate covariance - paper
 - 2017 - TSKCCA - Yoshida - Sparse kernel canonical correlation analysis - paper
 - 2017 - **SMSMA** - Kawaguchi - Supervised multiblock sparse multivariable analysis - paper
 - 2018 - AJIVE - Feng - angle-based JIVE - paper
 - 2018 - MOFA - Argelaguet - multi-omics factor analysis - paper 1, paper 2, application
 - 2018 - PCA+CCA - Brown - paper
 - 2018 - JACA - Zhang - Joint Association and Classification Analysis - paper
 - 2018 - **iPCA** - Tang - Integrated Principal Components Analysis - paper
 - 2018 - pCIA - Min - penalized COI - paper
 - 2018 - **sSCCA** - Safo - structured sparse CCA - paper
 - 2018 - **SWCCA** - Min - Sparse Weighted CCA - paper
 - 2018 - OmicsPLS - Bouhaddani - O2PLS implemented in R, with an alternative cross-validation scheme - paper
 - 2018 - SCCA-BC - Pimentel - Biclustering by sparse canonical correlation analysis - paper
 - 2018 - mixKernel - Mariette - kernel method for unsupervised multi-omics integration - paper 1, paper 2
 - 2019 - WON-PARAFAC - Kim - weighted orthogonal nonnegative parallel factor analysis - paper

-
- 2019 - BIDIFAC - Park - bidimensional integrative factorization - paper 1, paper 2
 - 2019 - SmCCNet - Shi - sparse multiple canonical correlation network analysis - paper
 - 2020 - msPLS - Csala - multiset sparse partial least squares path modeling - paper
 - 2020 - **MOTA** - Fan - network-based multi-omic data integration for biomarker discovery - paper
 - 2020 - D-CCA - Shu - Decomposition-based Canonical Correlation Analysis - paper
 - 2020 - COMBI - Hawinkel - Compositional Omics Model-Based Integration - paper
 - 2020 - DPCCA - Gundersen - Deep Probabilistic CCA - paper
 - 2020 - MEFISTO - Velten - spatial or temporal relationships - preprint
 - 2020 - MultiPower - Tarazona - Sample size in multi-omic experiments - paper
 - 2020 - mixedCCA - Yoon - Sparse semiparametric CCA for data of mixed types - paper
 - 2020 - smCIA/ssmCIA - Min - Sparse (structured sparse) multiple co-Inertia analysis - paper
 - 2023 - MuVI - Qoku - Integrate noisy feature sets - paper

Ecology multi-table literature

- 1994 - **COI** - Doleddec - Co-inertia analysis - paper
- 2007 - ade4 - Dray - Implementing the Duality Diagram for Ecologists - paper

Chemometrics multi-table literature

- 1987 - - Wold - Multi-way principal components-and PLS-analysis - paper
- 1996 - - Wold - Hierarchical multiblock PLS - paper
- 2003 - - Trygg - O2-PLS, a two-block (X-Y) latent variable regression (LVR) - paper
- 2011 - - Hanafi - Connections between multiple COI and consensus PCA - paper
- 2015 - THEME - Verron - THEmatic Model Exploration - paper

Behavioral research multi-table literature

- 2013 - DISCO SCA - Schouteden - distinctive and common components with simultaneous-component analysis - paper 1, paper 2

Multi-omics clustering / classification / prediction

Note: I think that prediction of genomic tracks, e.g. ChIP-seq, from other genomic tracks is a large area of research that may deserve a separate repository. Below are methods for clustering / classification of samples into sub-types or prediction of outcomes.

-
- 2009 - iCluster - Shen - paper
 - 2011 - PSDF - Yuan - Data fusion by Bayesian nonparametric Dirichlet modeling - website, publication
 - 2012 - MDI - Kirk - paper1, paper2
 - 2013 - iClusterPlus - Mo - paper
 - 2013 - BCC - Lock - Bayesian consensus clustering - paper
 - 2013 - iBAG - Wang - Integrative Bayesian Analysis of Genomics - paper
 - 2014 - SNF - Wang - paper
 - 2015 - moCluster - Meng - Derivative of iClusterPlus - paper
 - 2017 - clusternomics - Gabasova - paper
 - 2017 - PFA - Shi - Pattern Fusion analysis - publication
 - 2019 - IBOOST - Wong - paper
 - 2019 - Spectrum - John - paper
 - 2019 - NEMO - Rappoport - Similarity-based Clustering - paper
 - 2020 - INF - Chierici and Bussola - paper
 - 2021 - ClustOmics - Brière - Consensus clustering - paper
 - 2021 - MOGONET - Tongxin Wang - Multi-Omics Graph cOnvolutional NETworks - paper

Multi-omics autoencoders

- 2019 - maui - Ronen - Stacked VAE + clustering predictive of survival - paper
- 2019 - IntegrativeVAEs - Simidjievski - Variational autoencoders + classification - paper
- 2019 - OmiVAE - Xiaoyu Zhang - Integrated Multi-omics Analysis Using Variational Autoencoders - paper
- 2021 - DeepProg - Poirion - DL and ML ensemble + survival prediction - paper
- 2021 - SHAE - Wissel - Supervised Hierarchical Autoencoder + survival prediction - preprint

Multi-omics networks

- 2018 - MolTi-DREAM - Didier - identifying communities from multiplex networks, and annotated the obtained clusters article
- 2018 NetICS - Christos Dimitrakopoulos - Network-based integration of multi-omics data for prioritizing cancer genes - paper
- 2019 - RWR-MH - Valdeolivas - Random walk with restart on multiplex and heterogeneous biological networks article
- 2020 - MOGAMUN - Novoa-del-toro - A multi-objective genetic algorithm to find active modules in multiplex biological networks preprint

-
- 2021 - RWRF - Wen - Random Walk with Restart for multi-dimensional data Fusion paper

Single cell multi-omics

- 2018 - cardelino - - gene expression states to clones (SNVs from scRNA-seq + bulk exome data) -
- 2018 - clonealign - Campbell - gene expression states to clones (scRNA-seq + scDNA-seq (CNV)) - paper
- 2020 - CiteFuse - Kim - CITE-seq data analysis paper
- 2021 - CoSpar - Wang - infer dynamics by integrating state and lineage information - paper

Multi-study correlation or factor analysis

- 2016 - MSFA - De Vito - multi-study factor analysis: same features, different samples - paper

Multi-omics simulation

- 2016 - InterSIM - Chalise - methylation, gene expression and protein expression - paper
- 2019 - MOSim - Martinez-Mira - RNA-seq, ATAC-seq (DNase-seq), ChIP-seq, small RNA-seq and Methyl-seq. - preprint
- 2019 - OmicsSIMLA - Chung - DNA, CNV, WGBS, RNAseq, Protein expression - paper

Multi-omics reviews / evaluations

- 2008 - Holmes - Multivariate data analysis: The French way
- 2014 - Kohl - A practical data processing workflow for multi-OMICS projects
- 2016 - Josse - Measuring multivariate association and beyond
- 2016 - Ebrahim - Multi-omic data integration enables discovery of hidden biological regularities
- 2016 - Meng - Dimension reduction techniques for the integrative analysis of multi-omics data
- 2016 - Li - A review on machine learning principles for multi-view biological data integration
- 2017 - Huang - More Is Better: Recent Progress in Multi-Omics Data Integration Methods
- 2017 - Hasin - Multi-omics approaches to disease
- 2017 - Allen - Statistical data integration: Challenges and opportunities
- 2018 - Rappoport - Multi-omic and multi-view clustering algorithms: review and cancer benchmark
- 2018 - Bougeard - Current multiblock methods: Competition or complementarity? A comparative study in a unified framework
- 2018 - Karczewski - Integrative omics for health and disease

-
- 2018 - Yan - Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data
 - 2019 - Misra - Integrated omics: tools, advances and future approaches
 - 2019 - Chauvel - Evaluation of integrative clustering methods for the analysis of multi-omics data
 - 2019 - McCabe - Consistency and overfitting of multi-omics methods on experimental data - code
 - 2019 - Pierre-Jean - Clustering and variable selection evaluation of 13 unsupervised methods for multi-omics data integration
 - 2019 - Pinu - Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community
 - 2019 - Wu - A Selective Review of Multi-Level Omics Data Integration Using Variable Selection
 - 2019 - Sankaran - Multitable methods for microbiome data integration - code
 - 2020 - Lee - Heterogeneous Multi-Layered Network Model for Omics Data Integration and Analysis
 - 2020 - Herrmann - Large-scale benchmark study of survival prediction methods using multi-omics data - code
 - 2020 - Nguyen - Multiview learning for understanding functional multiomics
 - 2020 - Eicher - Metabolomics and multi-omics integration: a survey of computational methods and resources
 - 2020 - Cantini - Benchmarking joint multi-omics dimensionality reduction approaches for cancer study
 - 2020 - Subramanian - Multi-omics Data Integration, Interpretation, and Its Application
 - 2020 - Krassowski - State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing - code
 - 2021 - Espinosa - Data-Driven Modeling of Pregnancy-Related Complications
 - 2022 - Jiang - Uncovering Cross-Cohort Molecular Features with Multi-Omics Integration Analysis
 - 2022 - Cai - Machine learning for multi-omics data integration in cancer

Multi-omics application papers

- 2007 - Fagan - A multivariate analysis approach to the integration of proteomic and gene expression data
- 2011 - De la Cruz - The duality diagram in data analysis: Examples of modern applications - R notebook
- 2014 - Tomescu - Integrative omics analysis. A study based on Plasmodium falciparum mRNA and protein data

-
- 2014 - Costello (NCI/DREAM) - A community effort to assess and improve drug sensitivity prediction algorithms
 - 2015 - Wang - Inferring gene–gene interactions and functional modules using sparse canonical correlation analysis
 - 2016 - Wan - TCGA2STAT: simple TCGA data access for integrated statistical analysis in R - R notebook
 - 2017 - Butler - Integrating single-cell transcriptomic data across different conditions, technologies, and species.
 - 2018 - Skelly - Reference trait analysis reveals correlations between gene expression and quantitative traits in disjoint samples - R notebook
 - 2018 - Stuart - Comprehensive integration of single cell data
 - 2018 - Ash - Joint analysis of gene expression levels and histological images identifies genes associated with tissue morphology
 - 2019 - Xu - Identifying subpathway signatures for individualized anticancer drug response by integrating multi-omics data
 - 2019 - Ghaemi - Multiomics modeling of the immunome, transcriptome, microbiome, proteome and metabolome adaptations during human pregnancy Multi-omics in pregnancy using stacked generalization

Multi-omics data management

- 2017 - MultiAssayExperiment - Ramos - Software for the integration of multi-omics experiments in Bioconductor - paper.
- 2021 - muon - Bredikhin - Multimodal omics analysis framework

Batch effect correction

- 2020 - MultiBaC - Ugidos - MultiBaC: A strategy to remove batch effects between different omic data types and R package publication
- A multivariate method to correct for batch effects in microbiome data

Meetings and workshops

- 2020 - Mathematical Frameworks for Integrative Analysis of Emerging Biological Data Types - Hackathon details - June 14-19, 2020 in Banff, Canada