

ID: W2950985821

TITLE: mixOmics: An R package for 'omics feature selection and multiple data integration

AUTHOR: ['Florian Rohart', 'Benoît Gautier', 'Amrit Singh', 'Kim-Anh Lê Cao']

ABSTRACT:

The advent of high throughput technologies has led to a wealth of publicly available 'omics data coming from different sources, such as transcriptomics, proteomics, metabolomics. Combining such large-scale biological data sets can lead to the discovery of important biological insights, provided that relevant information can be extracted in a holistic manner. Current statistical approaches have been focusing on identifying small subsets of molecules (a 'molecular signature') to explain or predict biological conditions, but mainly for a single type of 'omics. In addition, commonly used methods are univariate and consider each biological feature independently. We introduce mixOmics, an R package dedicated to the multivariate analysis of biological data sets with a specific focus on data exploration, dimension reduction and visualisation. By adopting a systems biology approach, the toolkit provides a wide range of methods that statistically integrate several data sets at once to probe relationships between heterogeneous 'omics data sets. Our recent methods extend Projection to Latent Structure (PLS) models for discriminant analysis, for data integration across multiple 'omics data or across independent studies, and for the identification of molecular signatures. We illustrate our latest mixOmics integrative frameworks for the multivariate analyses of 'omics data available from the package.

SOURCE: PLOS computational biology/PLoS computational biology

PDF URL: <https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1005752&type=printable>

CITED BY COUNT: 2254

PUBLICATION YEAR: 2017

TYPE: article

CONCEPTS: ['Univariate', 'Omics', 'Data integration', 'Computer science', 'Feature selection', 'Identification (biology)', 'Biological data', 'Multivariate statistics', 'Data mining', 'Data type', 'Dimensionality reduction', 'Data science', 'Computational biology', 'Bioinformatics', 'Biology', 'Machine learning', 'Botany', 'Programming language']