BOX 1 | Glossary

Metabolomics: The study of chemical processes involving the small molecules that are the direct and indirect products of metabolic pathways. Metabolites are often classified into primary metabolites involved in growth, development, and reproduction, and secondary metabolites that are more important for ecological function.

Lipidomics: The study of complete fatty acid and lipid profiles in a sample. Lipid molecules are crucial for short- and long-term energy storage.

Plasticity: Ability of one genotype to produce more than one phenotype.

Metabolic plasticity: An organism's capacity to modulate energy production, allocation, and use.

Compound: A distinct chemical substance that is measured and analyzed within a biological sample such as a metabolite or lipid.

Steady state: Absolute or relative concentrations of a compound at a particular point in time.

Metabolic flux: Rate at which compounds pass through a metabolic pathway.

Liquid chromatography mass spectrometry (LC-MS): Combines physical separation capability of liquid chromatography with mass analysis capability of mass spectrometry. Can characterize a wide range of molecule types, and is considered more robust for lipidomic applications.

Gas chromatography mass spectrometry (GC-MS): Combines features of gas chromatography with mass spectrometry. Best suited for volatile molecules, and is considered more robust for metabolomic applications.

Nuclear magnetic resonance (NMR): Quantify compounds by placing a sample in a magnetic field and using the inherent magnetic properties to identify the compounds. 1H-NMR is most commonly used with metabolomics data.

Targeted: Predefined list of compounds quantified using standards.

Semi-targeted: Sscreening for a broad set of known compounds without necessarily having all standards.

Untargeted: Profiling as many compounds as possible including unknown compounds.

Permutational analysis of variance (PERMANOVA): Unsupervised multivariate approach used to test if centroids of groups are significantly different from each other. Often paired with permutational analyses of dispersion (PERMDISP).

Weighted gene co-expression network analysis (WGCNA): Correlation-based approach originally developed to identify groups of genes that shared expression patterns. Can be applied to identify metabolites or lipids with shared abundance patterns.

ANOVA-simultaneous components analysis (ASCA): Decomposes multivariate data according to variables of interest. Useful for examining multivariate responses across time or multiple additional factors.

Principal components analysis (PCA): Unsupervised molecular approach generally used for exploratory analysis. Reduces the number of dimensions in large datasets to principal components that retain most of the original information.

Partial least squares discriminant analysis (PLS-DA): Supervised approach that uses group labels to reduce dimensionality by identifying variables that explain relationships between predictors and responses. Orthogonal PLS-DA (OPLS-DA) is a variation commonly used with metabolomic and lipidomic data.

Variable Importance in Projection (VIP): Quantifies the discriminatory power of a compound from a PLS-DA model.

Significance Analysis of Microarray (SAM): Feature-specific modified t-tests combined with permutation analysis used to determine if a compound has differential abundance between experimental conditions.

Machine learning (ML): Computer systems that learn or adapt using statistical models to mimic human behavior and recognize patterns. Useful when datasets are complex, large, or require automation.

KEGG: Kyoto Encyclopedia of Genes and Genomes. Publicly-available database for pathway annotation.

HMDB: Human Metabolome Database. Publicly-available database for small molecule metabolites present in humans.

Overrepresentation-based enrichment: Determines if specific pathways or functions are observed in a target dataset more than expected by chance in comparison to a background dataset.

Network topology-based enrichment: Incorporates additional factors that impact pathway activity, such as feature position in a pathway or feature-feature interactions, into an enrichment analysis.