Une image contenant bâtiment, fenêtre, plein air, ville

Description générée automatiquement

# Une image contenant Graphique, clipart, graphisme, conception Description générée automatiquementMulti-Omics datas

Multi-omics, short for multiple omics, refers to an integrated approach in biological research that combines data from multiple "omics" disciplines. Omics disciplines involve the comprehensive study of various biological molecules or components within an organism, such as genomics (study of genes), transcriptomics (study of RNA transcripts), proteomics (study of proteins), metabolomics (study of metabolites), and others.

By integrating data from multiple omics disciplines, researchers aim to gain a holistic understanding of complex biological systems. This approach enables scientists to explore relationships, interactions, and regulatory mechanisms across different molecular levels, leading to insights into biological processes, disease mechanisms, and potential therapeutic targets.

For example, in cancer research, multi-omics analyses can help identify genetic mutations (genomics), gene expression patterns (transcriptomics), protein abundance (proteomics), and metabolic changes (metabolomics) associated with tumor development and progression. Integrating these diverse datasets can provide a more comprehensive understanding of the molecular basis of cancer and aid in the development of personalized treatments.

In the realm of data science, there isn't a one-size-fits-all solution for handling multi-omics data. Instead, there are two primary strategies: early fusion and late fusion. With early fusion, data from different omics layers are combined at the outset of analysis, while late fusion involves integrating data after separate analyses. In the case of Stabl, we've adopted a hybrid approach, called halfway fusion, combining features selected on each omic separately to predict the outcome.

## Early Fusion

Early fusion is a strategy used in multi-omics data analysis where features from different omics are combined at the beginning of the analysis pipeline, typically before any specific analysis or modeling takes place. In early fusion, data from various omics are integrated into a single dataset, by merging the data based on common identifiers (e.g., gene names, patient IDs).

# Une image contenant texte, capture d’écran, Police, ligne Description générée automatiquement

Early fusion is a commonly used approach in multi-omics studies to aggregate them into one dataset for further analysis.

However, early fusion may also face challenges such as data heterogeneity, batch effects, and scalability issues, which need to be carefully addressed during data preprocessing and integration.

## Late Fusion (or stacked generalization)

Late fusion is a strategy used in multi-omics data analysis where information from different omics layers is integrated after separate analyses or processing of each omics dataset. In late fusion, data from individual omics platforms (such as genomics, transcriptomics, proteomics, metabolomics) are analyzed independently, and the results are then combined or integrated during subsequent stages of analysis.

# Une image contenant texte, capture d’écran, Police, conception Description générée automatiquement

Late fusion approaches are commonly used in multi-omics studies to leverage the strengths of individual omics and to address specific research questions. However, late fusion may also encounter challenges such as loss of information due to separate analyses, potential inconsistencies between different omics layers, and the need for robust integration methods.

## Halfway fusion

Halfway fusion, also referred to as intermediate fusion or mid-level fusion, is a hybrid approach that combines elements of both early and late fusion strategies for integrating multi-omics data. In halfway fusion, data from different omics layers are partially integrated at an intermediate stage of the analysis pipeline, striking a balance between the comprehensive integration of early fusion and the modular analysis of late fusion. In our case, the intermediate stage is the feature selection step. The feature selection is performed on each omic independently, then the selected features are concatenated into one dataset used for further analysis.

# Une image contenant texte, capture d’écran, Police, diagramme Description générée automatiquement

When employing an univariate selection method, the process resembles early fusion. However, with multivariate selection, such as the approach utilized in Stabl, there's a noticeable effect. The objective is to assess features within each omics dataset relative to others in the same dataset. Subsequently, all chosen features are combined during the final modeling phase.

Halfway fusion approaches are often used in multi-omics studies to balance the trade-offs between comprehensive integration and modular analysis. By integrating data at an intermediate stage, halfway fusion allows researchers to leverage the strengths of both early and late fusion strategies while addressing specific challenges associated with multi-omics data analysis.