Integrated Compound Annotation

Sadjad Fakouri-Baygi

March 5, 2023

1 Introduction

Integrated Compound Annotation (ICA) developed by the Integrated Data Science Laboratory for Metabolomics and Exposomics (IDSL.ME) is a novel metabolomics data integration tool for gaining a comprehensive understanding of the chemical space in a study. ICA is capable of integrating data from multiple sources including chemical structure information from MS2 analyses, chemical structure information from MS1 analysis using composite spectra analysis, and MS1 molecular formula annotation using isotopic profiles on the aligned table level.

MS2 analyses such as Data-Dependent Acquisition (DDA) and Data-Independent Acquisition (DIA) are commonly used to obtain structural information about compounds, such as the fragmentation patterns of specific ions. This information can be used to identify known compounds or to predict the structures of unknown compounds, especially when precursor ions are known. On the other hand, even without precursor values, composite spectra analysis on MS1 data can provide in-source fragmentation patterns to complement MS2 analysis.

By integrating chemical structural data from both MS2 and MS1 analyses, along with molecular formula annotation on the aligned table level, ICA provides a more thorough understanding of the chemical space. The chemical space is then aggregated on the aligned table level to identify novel metabolites, to characterize the functions of known metabolites, and to explore the relationships between different metabolites and metabolic pathways.

2 Sample Code

```
ICA_IntegrationAlignedTables(
        IPA_peak_alignment = "...",
        MS2_aligned_spectra_table = "...",
        MS2_MSP_folder = "...",
        MS2_FSDB = "...",
        CSA_aligned_spectra_table = "...",
        CSA_MSP_folder = "...",
        CSA_FSDB = "...",
```

```
UFA_aligned_molecular_formula = "...",
output_directory,
plotVenn = TRUE)
```

This function, `ICA-IntegrationAlignedTables` takes in the following parameters including `IPA-peak-alignment`, which is the peak alignment table from Intrinsic Peak Analysis (IPA), `MS2-aligned-spectra-table` and `MS2-MSP-folder`, which contain the aligned MS2 spectra table and the folder containing MS2 MSP files, respectively. Likewise, `CSA-aligned-spectra-table` and `CSA-MSP-folder`, which contain the aligned composite spectra analysis (CSA) spectra table and the folder containing CSA MSP files, respectively. The aligned molecular formula table from IDSL.UFA may be also provided for `UFA-aligned-molecular-formula`. `output-directory` is the directory where the output files will be stored. `plotVenn` to visualize overlaps among various annotation methods.