# ノンターゲットメタボロミクスデータの利活用に向けた MassBank HumanとMassBank *in silico*の開発

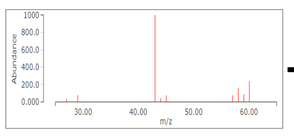
鳥越 大平 (九州大), 早坂 亮祐 (慶應義塾大), 平山 明由 (慶應義塾大), 髙橋 政友 (九州大), 和泉 自泰 (九州大), 松田 史生 (大阪大)

CC BY

## マスバンク

**【ユーザー】**

不明な代謝物質のマススペクトル



検索 →

← 検索

類似のマススペクトルの分子のリスト

https://massbank.jp/

MassBank
High Quality Mass Spectral Database

16278化合物　98525 MS2スペクトル
2024年9月現在
日本質量分析学会スペクトルデータ部会

**【スペクトル提供者】**

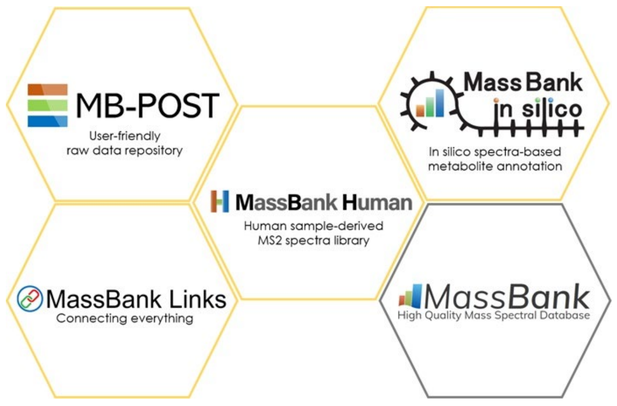
標準化合物を様々なMSの条件で測定したデータ

**問題点**

**実測のサンプルにおいて代謝物質の同定に至るのは10%以下 ⟶ 同定率を向上させたい**

→多くのスペクトルは未知代謝物質のまま

## シン・マスバンクプロジェクト



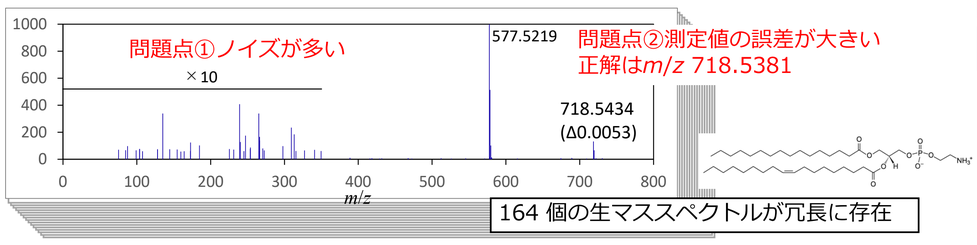

https://shin.massbank.jp/

MassBankレコードの拡充のために、シン・マスバンクプロジェクトを立ち上げ
(JST統合化推進プログラム2023年度採択)

# MassBank Human

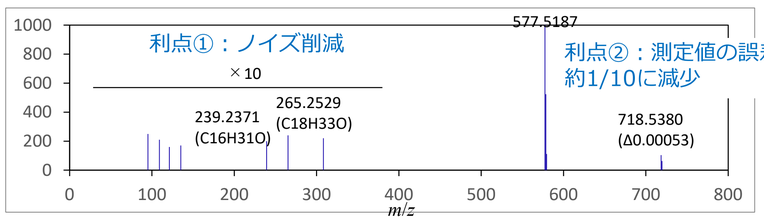
## 公共データベースを利活用したヒト試料由来の低分子化合物マススペクトルライブラリ

酵母リピドミクスで得られたリン脂質の生マススペクトル 164個




平均化後の高品質マススペクトル　PE (16:0/18:1)に構造推定が可能に

平均化 ➡




**平均化することで高品質なマススペクトルを作成可能**

メタボローム データリポジトリ
METABOLOMICS WORKBENCH
MetaboLights
MB-POST

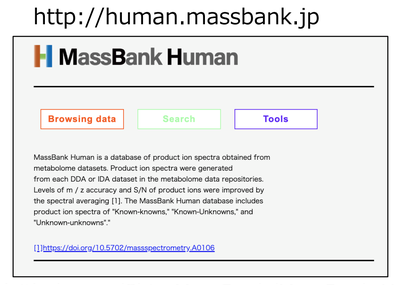
DDA データセット → 平均化作業 → 平均化済スペクトル → 自動アノテーション作業 → 自動アノテーション済スペクトル → キュレーション済スペクトル

Mass Bank in silico

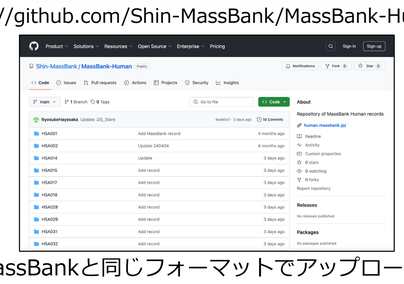
自動アノテーション済スペクトル ↑ Shin-MassBank GitHub ↑ MassBank Human

キュレーション済スペクトル ↑ MassBank GitHub ↑ MassBank High Quality Mass Spectral Database

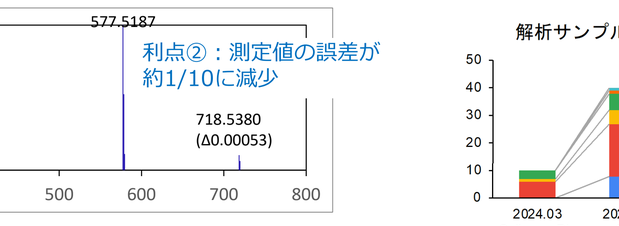
http://human.massbank.jp




https://github.com/Shin-MassBank/MassBank-Human



MassBankと同じフォーマットでアップロード




MS-FINDER

MS-FINDERをはじめ、MassBank フォーマットに対応したソフトウェアでデータ解析可能

# MassBank in silico

## 高精度 *in silico* 予測フラグメントに基づく未知代謝物質の構造アノテーションツール

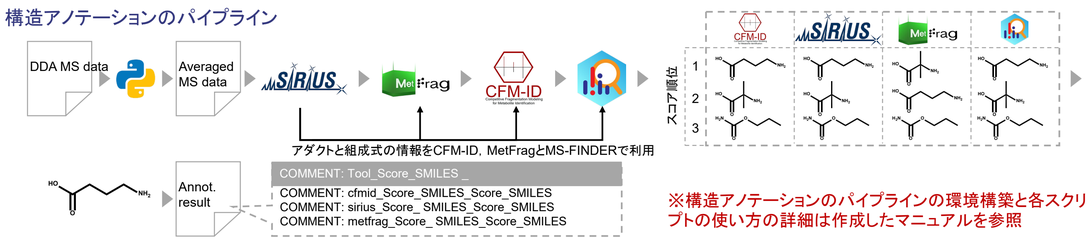
### 構造アノテーションツールの特徴

各構造アノテーションツールの特徴 (O: アリ, X: ナシ)

| | 最低限実装したい機能 | | | | | | 実装したい理想的な機能 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 水素再配置規則 | 全てのin silico fragmentと実測したfrgamentの照合 | ニュートラルロスの情報による構造候補の評価 | 部分構造特異的なプラダクトイオンの情報による構造候補の評価 | 相対強度，結合乖離エネルギー，実測したプロダクトイオンの相対強度，質量許容誤差による構造候補の評価とスコアリング | スペクトルエントロピーによる実測と予測MS/MSスペクトルの照合 | 機械学習によるMS/MSスペクトルの相対強度予測 | 実測したHRMS/MSスペクトルと組成式の照合による予測組成式の評価 | 化合物クラスの推定 | 実測したMS/MSスペクトルから予測した分子FPと構造候補の分子FPデータベースの照合による構造予測 | 天然同位体比による組成式予測 |
| MetFrag | X | O | O | X | O | X | X | X | X | X | X |
| CFM-ID | X | X | O | O | O | X | O | X | X | X | X |
| MS-FINDER | O | O | X | X | O | X | X | X | X | X | O |
| SIRIUS | X | X | O | O | X | X | X | O | O | O | O |

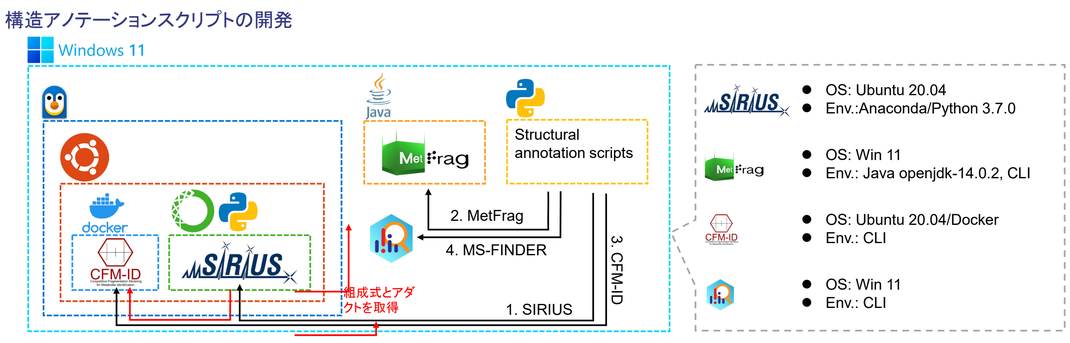
### 構造アノテーションパイプラインの構築

構造アノテーションのパイプライン



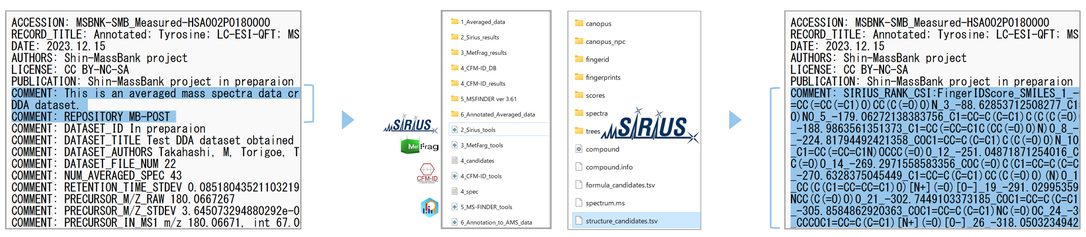

※構造アノテーションのパイプラインの環境構築と各スクリプトの使い方の詳細は作成したマニュアルを参照

構造アノテーションスクリプトの開発



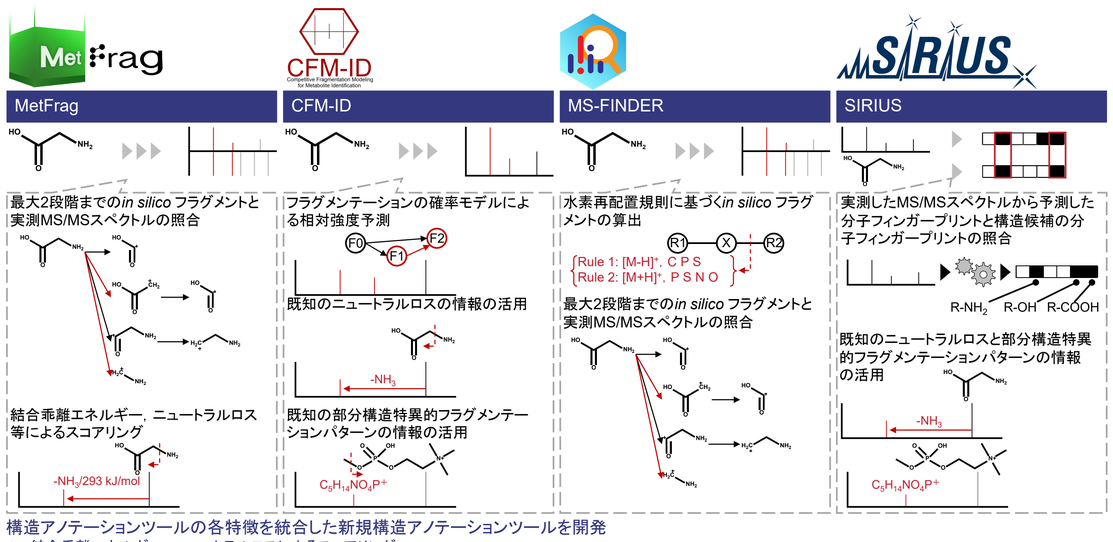

### 構造アノテーションパイプラインを用いた解析の例

解析結果の実例



- 作成したpython scriptを用いて，平均化MSデータからマススペクトルデータを取得
- マススペクトルデータを対象に解析script (例: 2_Sirius_tools) を実行し，構造候補を割り当てる
- 解析結果をまとめるscriptを実行し，平均化MSデータのCOMMENTに割り当てられた構造候補の情報を挿入

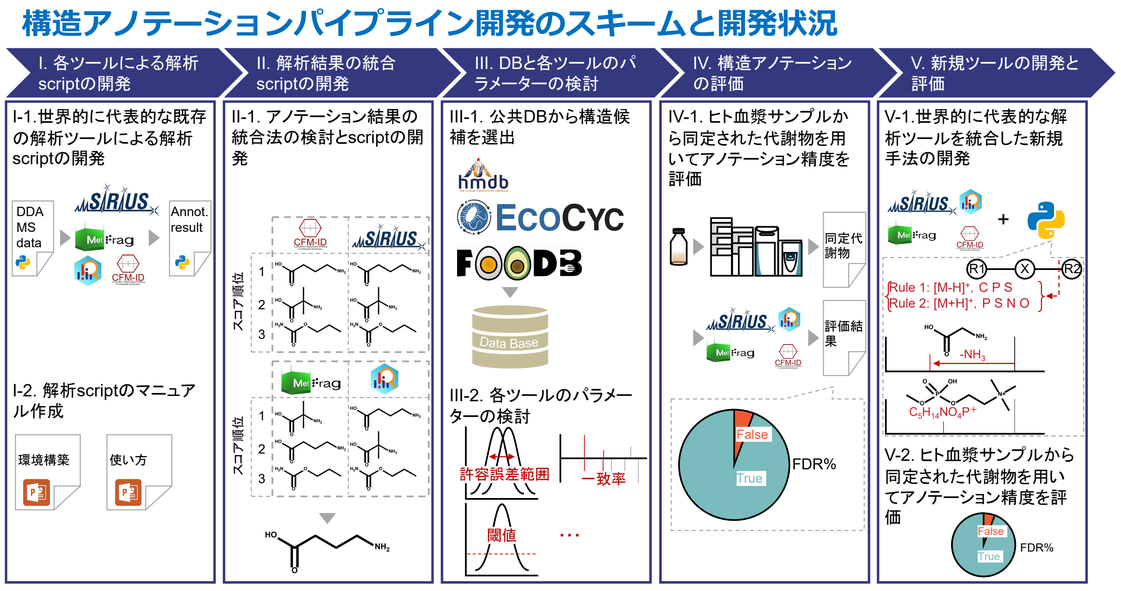
※予備試験としてノンターゲットメタボロームデータ1ファイル（約5300種類の代謝特性が検出）からの構造アノテーションを実施したところ，約8日間必要であった．




構造アノテーションツールの各特徴を統合した新規構造アノテーションツールを開発

- 結合乖離エネルギー，ニュートラルロスによるスコアリング
- 既知の部分構造特異的フラグメンテーションパターンの情報の活用
- 水素再配置規則に基づくin silico フラグメントの算出

### 構造アノテーションパイプライン開発のスキームと開発状況




I-1. 世界的に代表的な既存の解析ツールによる解析scriptの開発
⇒ 完了.

I-2. 解析scriptのマニュアル作成
⇒ 完了.

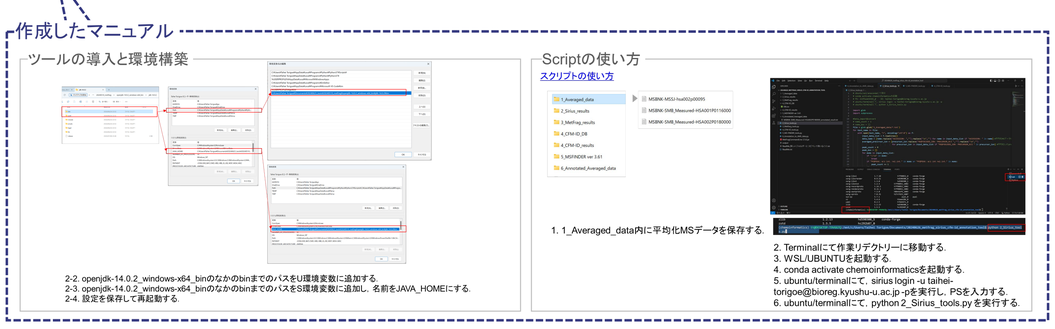
II-1. 各ツールを用いて解析した構造候補のアノテーション方法の検討とscriptの開発
⇒ 各ツールの構造候補を平均化MSデータのテキストに書き込むscriptのみ作成完了. 今後，アノテーション方法を検討.

III-1. 公共DBから構造候補を選出
⇒ 公共DBの情報収集のみ完了. 今後，構造候補の選出を実施.

III-2. 各ツールのパラメーターの検討
⇒ 今後，検討.

IV-1. ヒト血漿サンプルから同定された代謝物を用いてアノテーション精度を評価
⇒ 今後，偽陽性発見率に基づいてアノテーション精度を評価.

V-1. 世界的に代表的な既存の解析ツールの方法を統合したツールの開発
⇒ IVの結果に基づいて検討.

V-2. ヒト血漿サンプルから同定された代謝物を用いてアノテーション精度を評価

作成したマニュアル