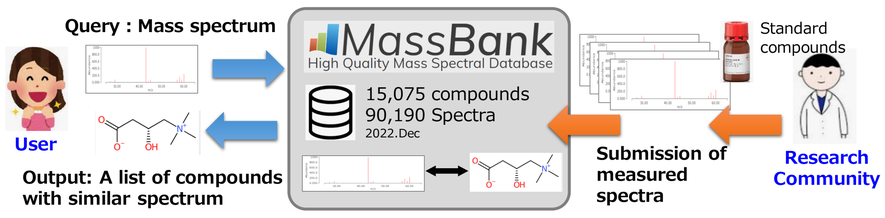
次世代低分子マススペクトルデータベース　シン・マスバンクの構築

○松田史生（大阪大）、奥田修二郎（新潟大）、津川裕司（京都大）、平山明由（慶應大）、和泉自泰（大阪大）、

2026.9.1 正式公開開始！

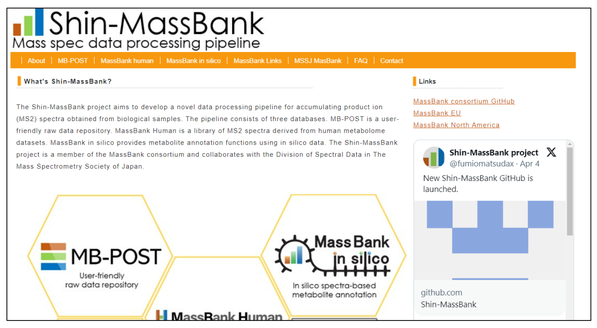
メタボローム解析の代謝物アノテーションは未だ実測スペクトルデータが不足している

- ✓ MassBankは、標準化合物から得られた ＞90,000マススペクトルからなるコミュニティ ベースの質量スペクトルデータベース。



シン・マスバンクプロジェクト
https://shin.massbank.jp/

- ✓ Shin (Neo) -MassBankプロジェクトは、MassBankレコードを拡張するために2023年にスタート。
- ✓ 公開されているメタボロームデータセットから高品質で適切に注釈が付けられたプロダクトイオンスペクトルを生成する再解析パイプラインを開発。
- ✓ MB-POST（生データファイルリポジトリ）、MassBank -Human（ヒトサンプル由来のMS2スペクトルのライブラリ）、MassBank in silico（予測MS2スペクトルのライブラリ）、MassBank Links（ポータルサイト）の4つの新しいWebサイトが現在開発中。
- ✓ シン・マスバンクプロジェクトは、NBDC/JST Japan から支援を受けている。



Shin- MassBankプロジェクトページ。

生メタボロームデータリポジトリ

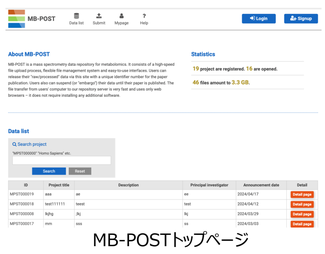
MB-POST

https://repository.massbank.jp/

公開中！

- ユーザーフレンドリーな生メタボロームデータリポジトリ。
- 将来のマルチオミクスサポートのためのプロテオミクスデータリポジトリjPOSTとの連携。

奥田@新潟大学



MB-POSTトップページ

メタボロミクス研究者のシン博士

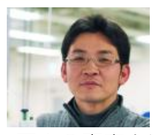
ヒト代謝物のスペクトルライブラリ

MassBank Human

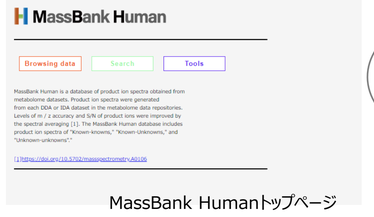
https://human.massbank.jp/

公開中！

- ヒトの DDA データセットから生成された高精度の製品イオンスペクトルのライブラリ。
- ヒトの既知-既知、既知-未知、未知-未知の代謝物のスペクトルは GitHub から入手可能。



平山@慶応大学



MassBank Humanトップページ

彼女はMS2スペクトルが大好きです。

シミュレーションによる予測スペクトルのデータベース

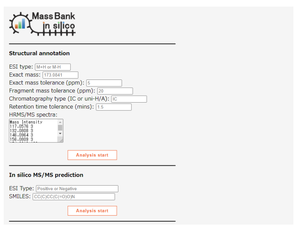
MassBank in silico

公開中！

https://insilico.massbank.jp/

- *in silico*予測 MS/MS スペクトルの生成。
- シミュレーションによる予測スペクトル

和泉@大阪大学



MassBank in silicoトップページ

彼女は断片化に夢中だ。

スペクトルデータを世界と繋ぐ

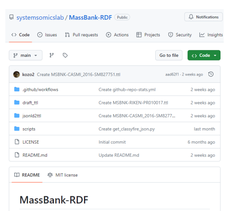
MassBank Links

公開中！

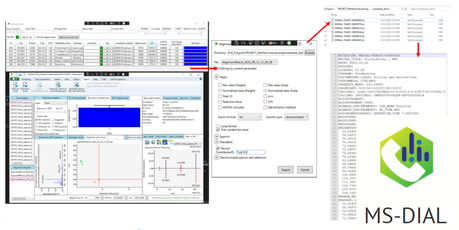
https://github.com/systemsomicslab/MassBank-RDF

- スペクトルデータを化学および生物学情報に接続するためのMassBankレコードの RDF の開発。
- MS-DIAL のMassBankレコード ファイルの読み取りと書き込みサポート。

津川@京都大学

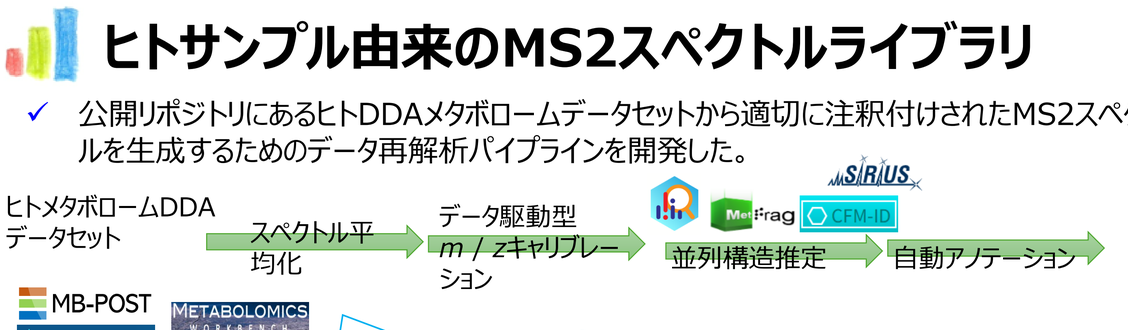


MassBank RDF GitHub ページ



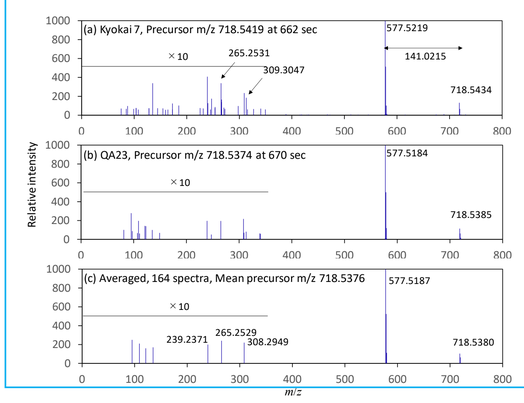
ヒトサンプル由来のMS2スペクトルライブラリ

- ✓ 公開リポジトリにあるヒトDDAメタボロームデータセットから適切に注釈付けされたMS2スペクトルを生成するためのデータ再解析パイプラインを開発した。



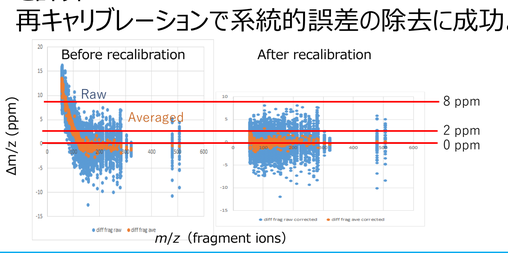
スペクトル平均化

- ✓ 平均化スペクトルは、DDA メタボローム データセット内の類似したプロダクトイオンスペクトルを積算して生成。
- ✓ 平均化で、プロダクトイオンの信頼性が向上。
- ✓ *m / z*値の精度も平均化で改善



データ駆動型*m / z*キャリブレーション

- ✓ Δm /z 値 (m/z (測定値) - m/z (理論値)) を計算
- ✓ 再キャリブレーションで系統的誤差の除去に成功。



自動代謝物アノテーション

- ✓ Schymanski et al. 201のアノテーションレベルを使用。

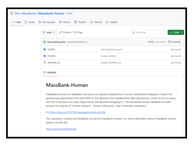
レベル 2a:
2 つ以上のMassBankレコードにヒット (>0.7)
&& FDR<0.1 で SIRIUS5 または 6 がトップヒット。
MetFrag 、CFM-ID、または MS-FINDERでヒットしたトップ10
&&HMDBに存在（既知の固有構造）

レベル3b:
FDR<0.1 で SIRIUS5 または 6 がトップヒット。
MetFrag 、CFM-ID、または MS-FINDERでヒットしたトップ10
&&HMDBに存在（既知の固有構造）

MassBank Human

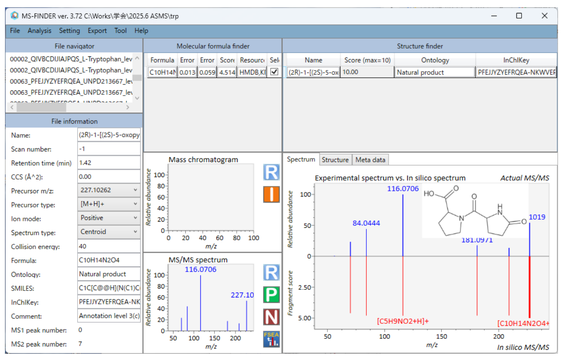
MassBank Human Github

MassBank
High Quality Mass Spectral Database



手動キュレーション

- ✓ スペクトル HSA009R0227047 は、5-オキソプロリルプロリンであると推定。
- ✓ MassBankおよび Human Metabolome Databaseにとっての新規化合物。



MS2スペクトルライブラリの統計

- ✓ レベル2aまたは3の注釈が付いた合計29,268のスペクトルが19のデータセットから生成。
- ✓ すべてのスペクトルデータはGitHubから入手可能

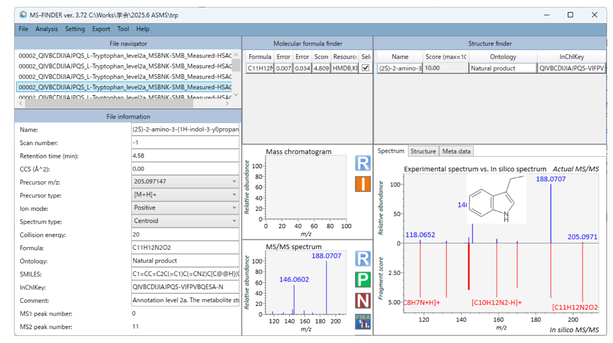
DDAデータセット				注釈付きスペクトルの数		
ID	データリポジトリID	起源	ファイル	レベル2a	レベル3b	レベル3c
HSA001	MPST000004	プラズマ	19	73	293	939
HSA002	MPST000004	プラズマ	19	88	68	263
HSA003	ST002338	ファセル	50	88	365	623
HSA004	ST002338	ファセル	50	26	95	213
HSA005	ST001264	プラズマ	31	3	29	50
HSA006	ST001264	プラズマ	31	26	120	332
HSA007	MTBLS417	プラズマ	60	96	221	436
HSA008	MTBLS417	プラズマ	60	16	66	414
HSA009	ST001171	血清	253	385	1313	4136
HSA010	ST001914	ファセル	101	324	1033	6985
HSA019	ST002044	血清	286	141	221	1052
HSA020	ST002044	血清	286	107	367	914
HSA021	ST002044	血清	286	106	348	1939
HSA022	ST002044	血清	286	205	573	1335
HSA014	MTBLS6402	細胞	27	39	39	122
HSA015	MTBLS6402	細胞	27	56	186	765
HSA031	MTBLS3233	糞便	59	33	67	144
HSA032	MTBLS3233	糞便	60	103	279	612
HSA033	MTBLS3854	プラズマ	50	28	97	271
合計				1943	5780	21545

MS-DIALはMassBankレコード形式をサポート

- ✓ MS-FINDERはMassBankレコードファイルの読み取りと書き込みをサポート。

MassBankレコードファイル

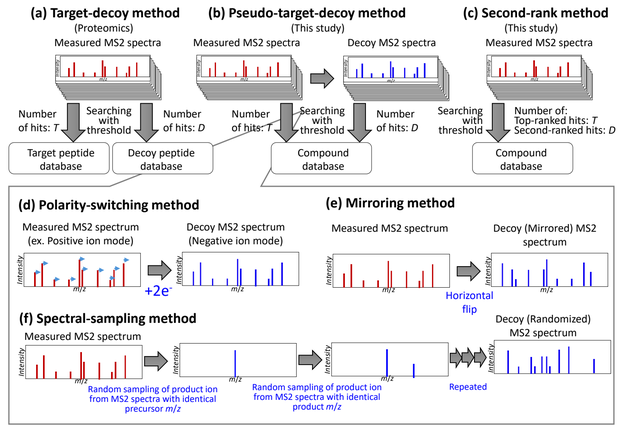
MS-FINDER



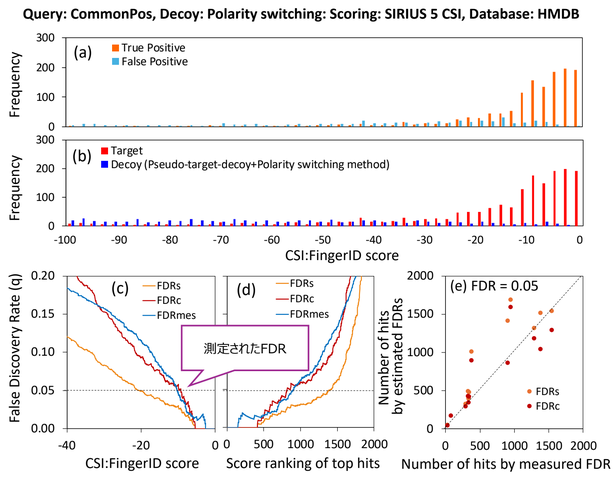
代謝物アノテーション結果のFDR推定法の開発

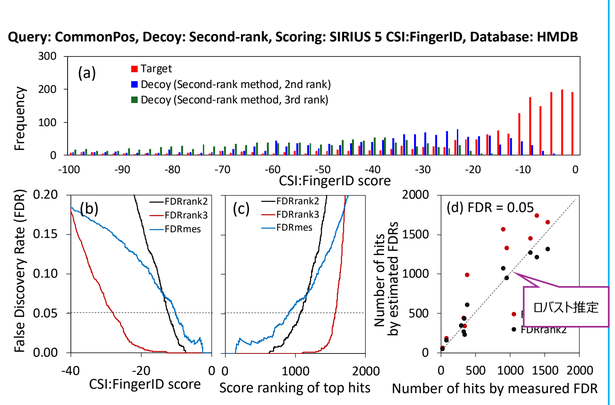
Mass Spectrometry 13（1）、A0155-A0155

- ✓ SIRIUS 5 などの*in silico*ツールを使用すると、プロダクトイオンスペクトルから化合物を直接アノテーション可能。
- ✓ 化合物アノテーション結果中のFDR を推定するための 4 つの簡単な方法を提案し、性能を評価。
- ✓ MassBankデータのアノテーションから、セカンドランク法で推定した FDR が真の FDR＝0.05 に最も近似。
- ✓ しかし、FDR レベルの推定するには、よりロバストな方法の開発が必要です。



FDR推定方法の比較





極性スイッチング法： SIRIUS 5 CSIスコアリング法を用いたMassBankデータのアノテーション結果を用いた評価
(a) 真陽性と偽陽性のスコア分布。(b) ターゲットおよびデコイスペクトルデータのスコア分布 (c、d) FDR、スコア、および順位の関係。
(e) 12組のスペクトルデータセットと化合物データベース間の真のFDRレベルと推定FDRレベルの比較。

セカンドランクの方法: SIRIUS 5 CSIスコアリング法を用いたMassBankデータのアノテーション結果を用いた評価
(a) ターゲットスペクトルデータとデコイスペクトルデータのスコア分布
(b、c) FDR、スコア、ランキングの関係。
(d) 12組のスペクトルデータセットと化合物データベース間の真のFDRレベルと推定FDRレベルの比較。