



# Identification of xenobiotic metabolites using *in silico* tools and non-targeted analysis

Matthew W. Boyce<sup>a,b</sup>, Kristin A. Favela<sup>c</sup>, Alex Chao<sup>b</sup>, Grace Patlewicz<sup>b</sup>,  
Jon R. Sobus<sup>b</sup>, Antony Williams<sup>b</sup>, John F. Wambaugh<sup>b</sup>

<sup>a</sup>Oak Ridge Associated University, Oak Ridge, TN 37830, USA

<sup>b</sup>Center for Computational Toxicology and Exposure, US EPA, RTP, NC 27711, USA

<sup>c</sup>Forensics and Specialty Analysis, Southwest Research Institute, San Antonio, TX 78228, USA

*This presentation does not reflect EPA policy.*

# Guiding NTA with *in silico* predictions

Sample  
Preparation

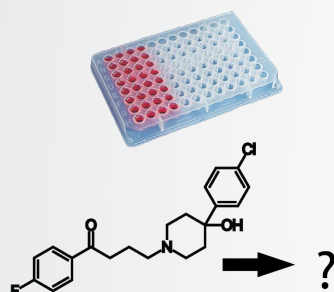
*In Silico* Data  
Generation

Data  
Acquisition

Data  
Processing

Data Analysis

## In Vitro Assay



## *In Silico* data

Aggregate Metabolite  
Predictions

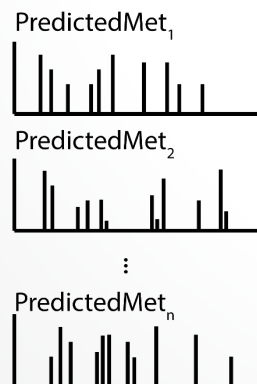
Software

TIMES  
Meteor  
BioTransformer  
QSAR Toolbox

Suspect List

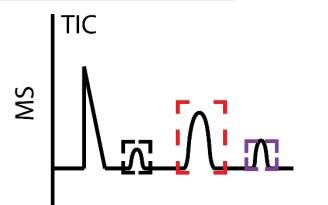
PredictedMet<sub>1</sub>  
PredictedMet<sub>2</sub>  
⋮  
PredictedMet<sub>n</sub>

CFM-ID Predictions

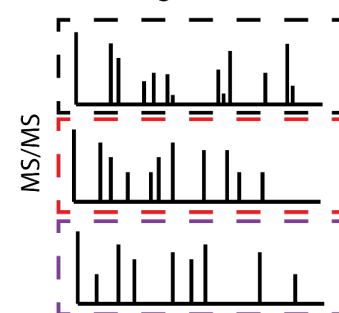


## Non-Targeted Analysis

Parent ion selection



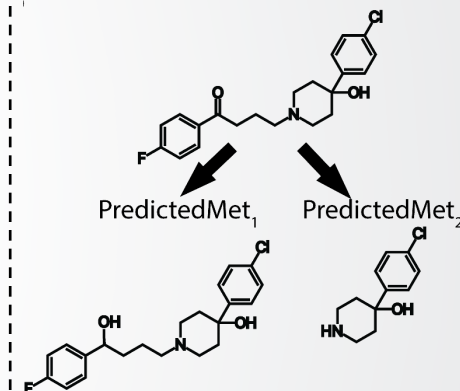
Parent ion fragmentation



## Feature Selection & Data Cleaning

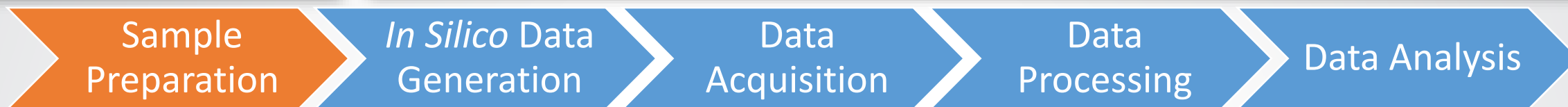
- 1) Peak Selection
- 2) Feature Identification
- 3) Data Cleaning

## Metabolite Assignment





## *In vitro* assay

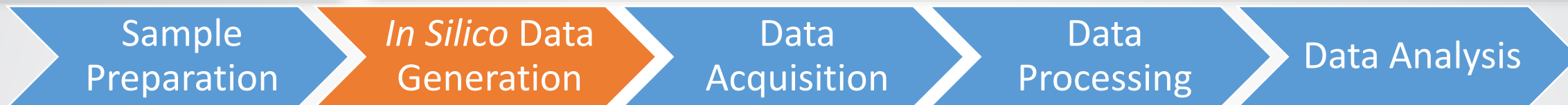


### Metabolite Generation

- Starting compounds metabolized via pooled primary human hepatocytes (10 donors)
  - Three time points: 0, 1, 4h
  - Three sample treatments: Supernatant (post lysis), B-glucuronidase treated, cell pellet
- Standards/Controls
  - Vehicle blank – DMSO
    - Used as blank for MS analysis
  - Standard control – Cell free solution with compound
    - Used to identify retention time window and mass error

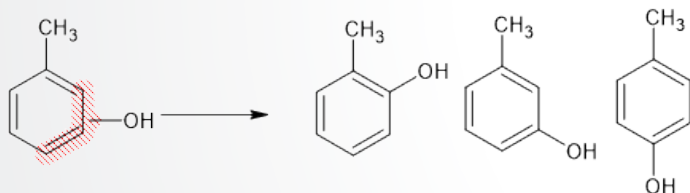


# Compiling a suspect screening list



## Known Metabolites

- Pulled 438 metabolites from 49 papers
- Markush structures were enumerated



## Predicted Metabolites

- Compiled predicted structures from:
  - TIMES
  - BioTransformer
  - QSAR Toolbox
  - Meteor Nexus
- 1,666 predictions in total

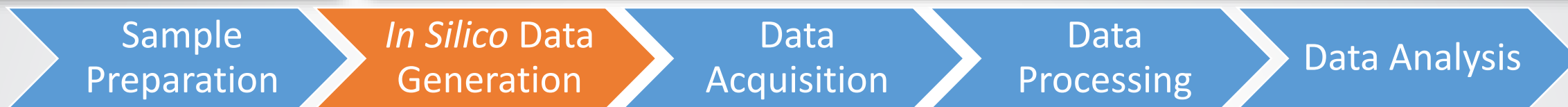


## Suspect Screening List

- 1,808 unique structures used to generate *in silico* MS<sup>2</sup> spectra
- 490 unique molecular formulae for MS<sup>1</sup> formula assignment

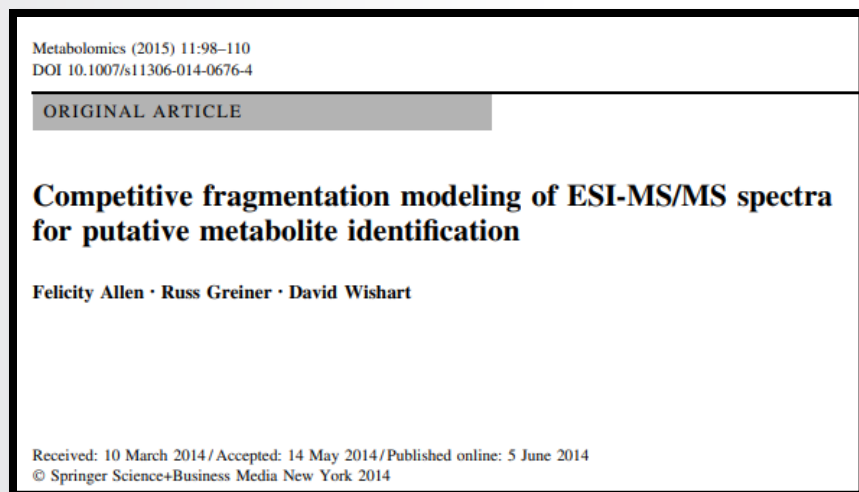


# Generating database of *in silico* MS<sup>2</sup> spectra



**Fragmentation spectra were generated for each predicted metabolite**

**Competitive Fragmentation Modeling-ID (CFM-ID)**



## **Spectra were generated using CFM-ID**

- Reference spectra were generated at three collision energies (CE)
- Data were stored in database to query against for comparisons
- Validated against CASMI datasets for HRMS identification  
*DOI: 10.3390/metabo10060260*
- Applied to ~700,000 chemicals in EPA's CompTox Dashboard  
*DOI: 10.1038/s41597-019-0145-z*



# MS<sup>1</sup> and MS<sup>2</sup> data collection



**LC-qTOF was used to collect high resolution MS<sup>1</sup> and MS<sup>2</sup> data**

## MS<sup>1</sup>

- ESI+ and ESI-
- Range 100 – 1700 m/z
- Used to collect features for identification

## MS<sup>2</sup>

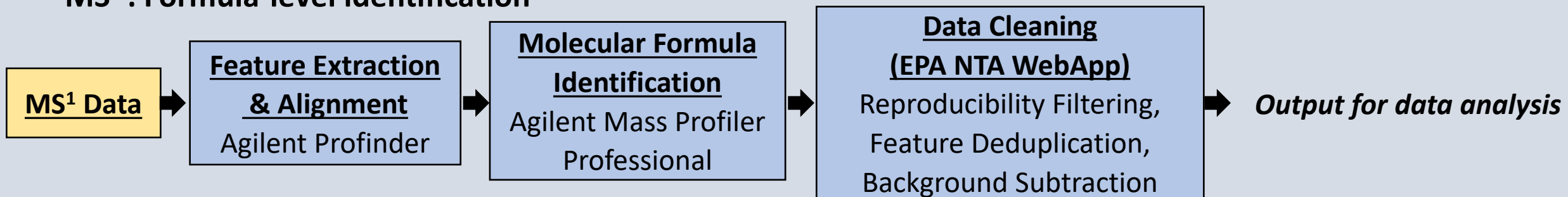
- Data-dependent acquisition (using suspect screening list)
- 1 replicate per treatment per time point
- Used to identify a feature's probable structure



# Data processing steps

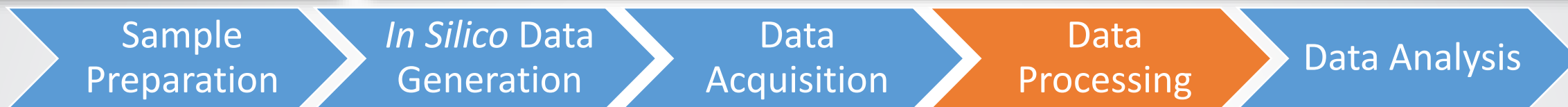


## MS<sup>1</sup> : Formula-level identification

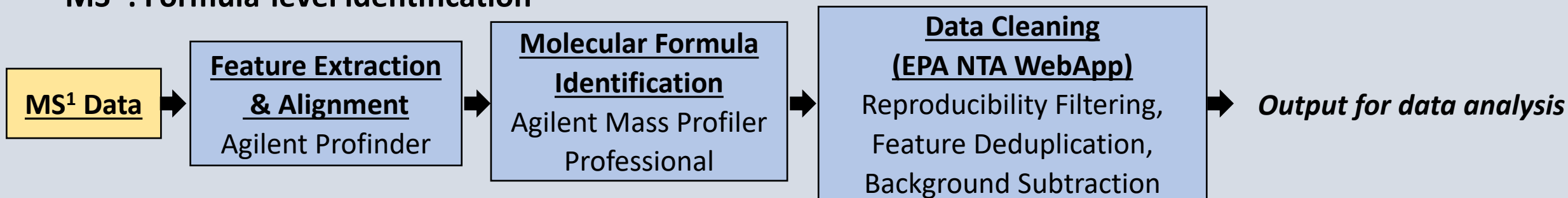




# Data processing steps



## MS<sup>1</sup> : Formula-level identification



## ***Output of MS<sup>1</sup> processing: Annotated features***

### Suspect-Screening matches

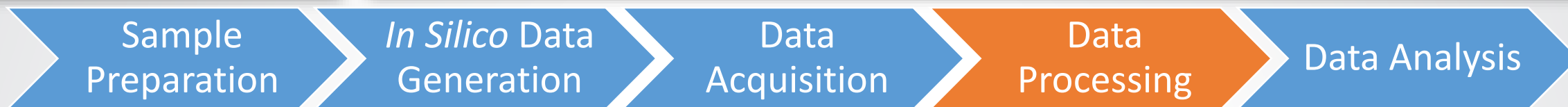
- Identified using suspect list
- Molecular formula with suspected structural assignments

### Features without suspect matches

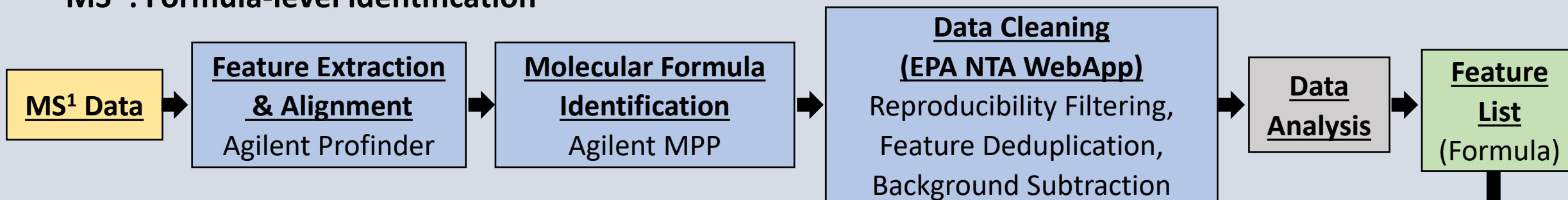
- Formula proposed using Agilent's Molecular-Formula generator
- Formulae with no known structural assignments



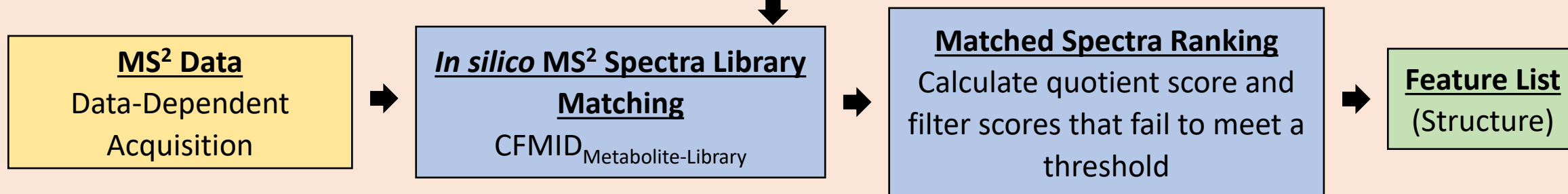
# Data processing steps



## MS<sup>1</sup> : Formula-level identification

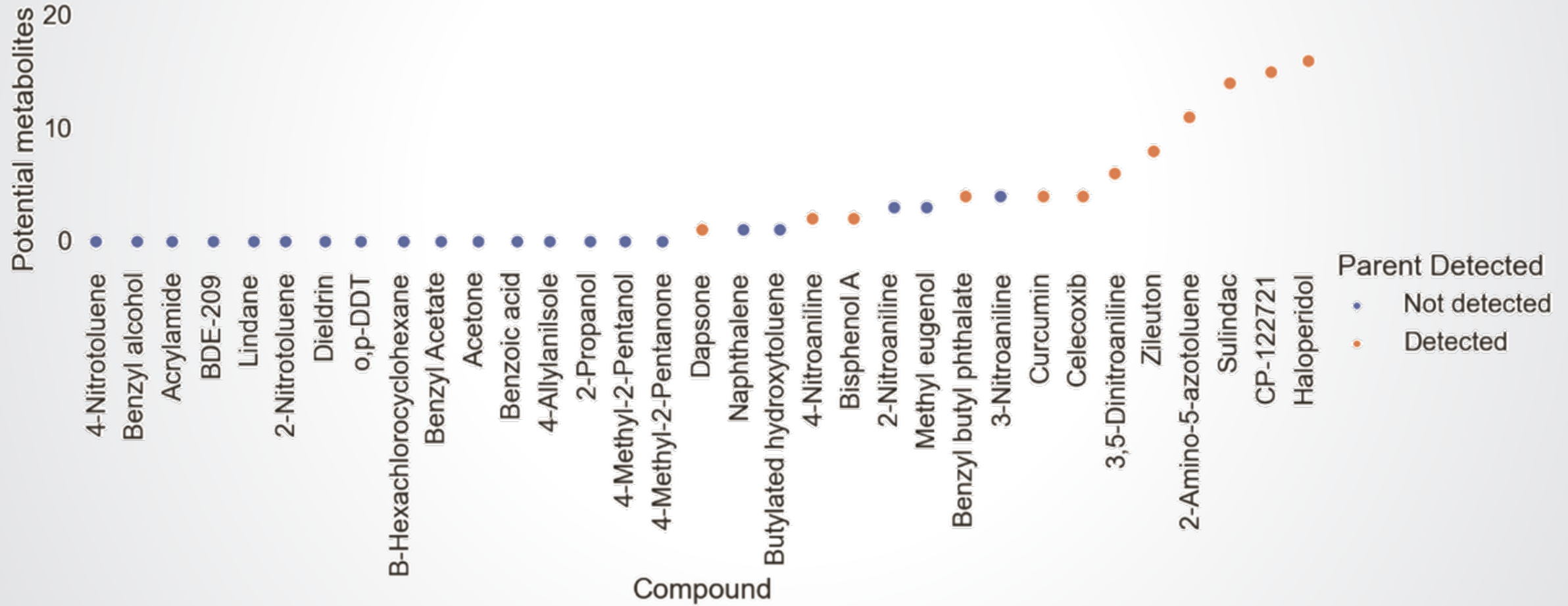


## MS<sup>2</sup> : Structure-level identification



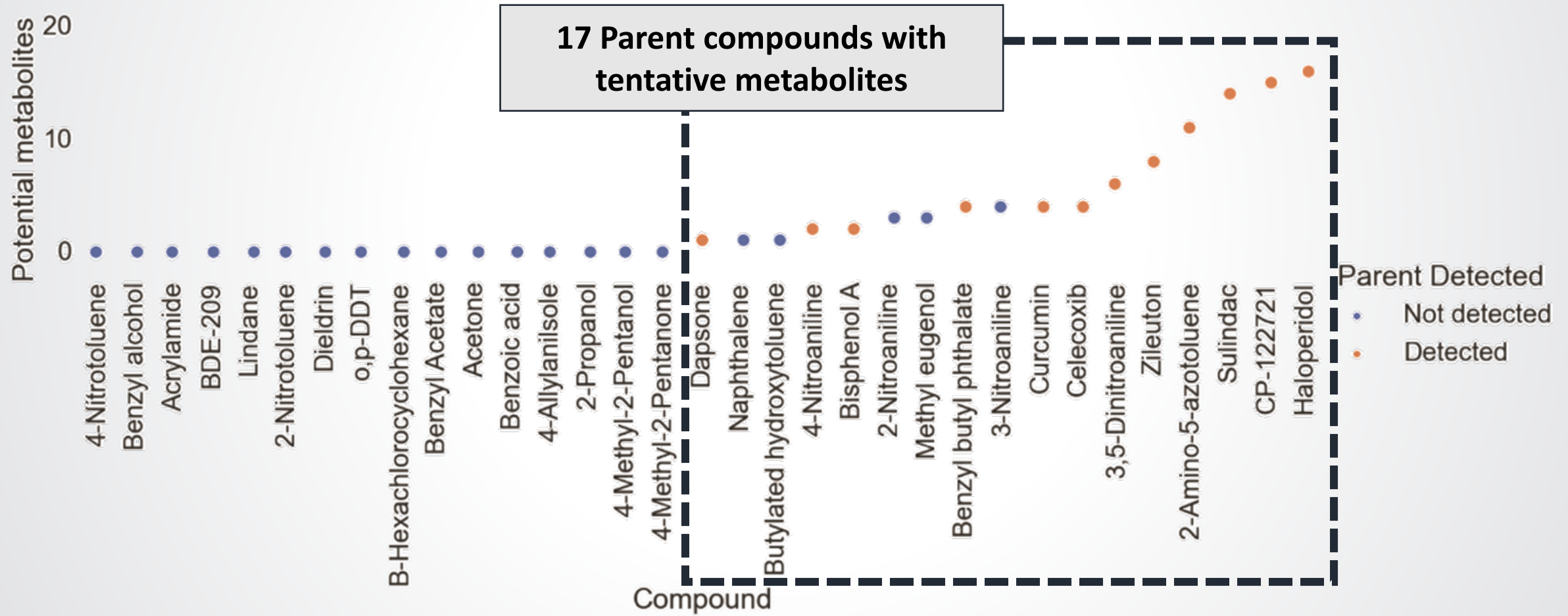


# Which parents have tentative metabolites?



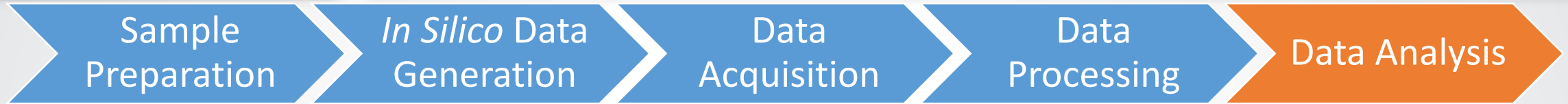


# Which parents have tentative metabolites?



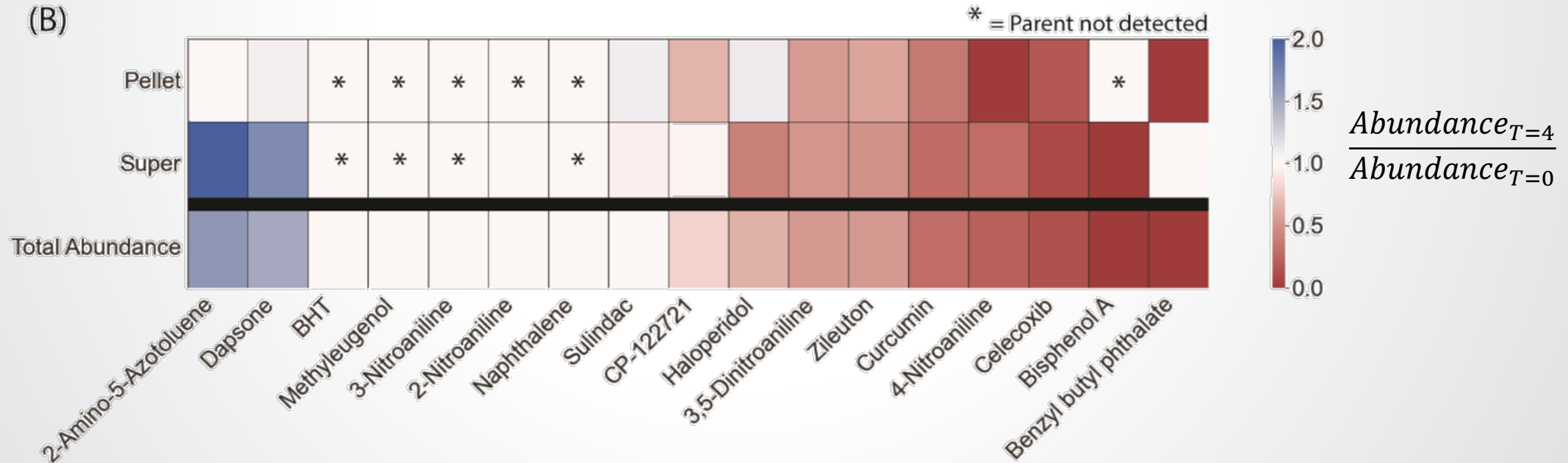


# Which parents are being metabolized?



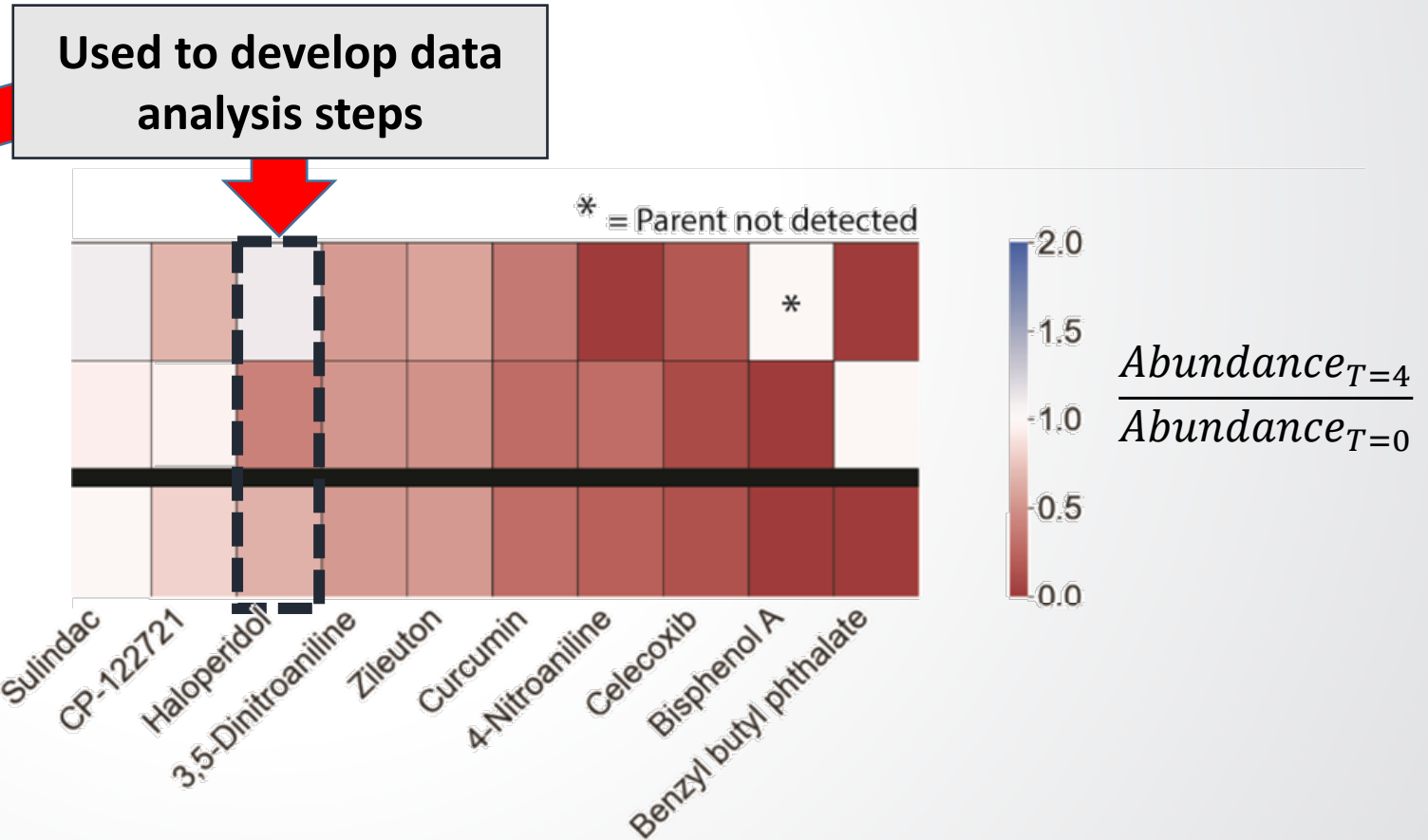
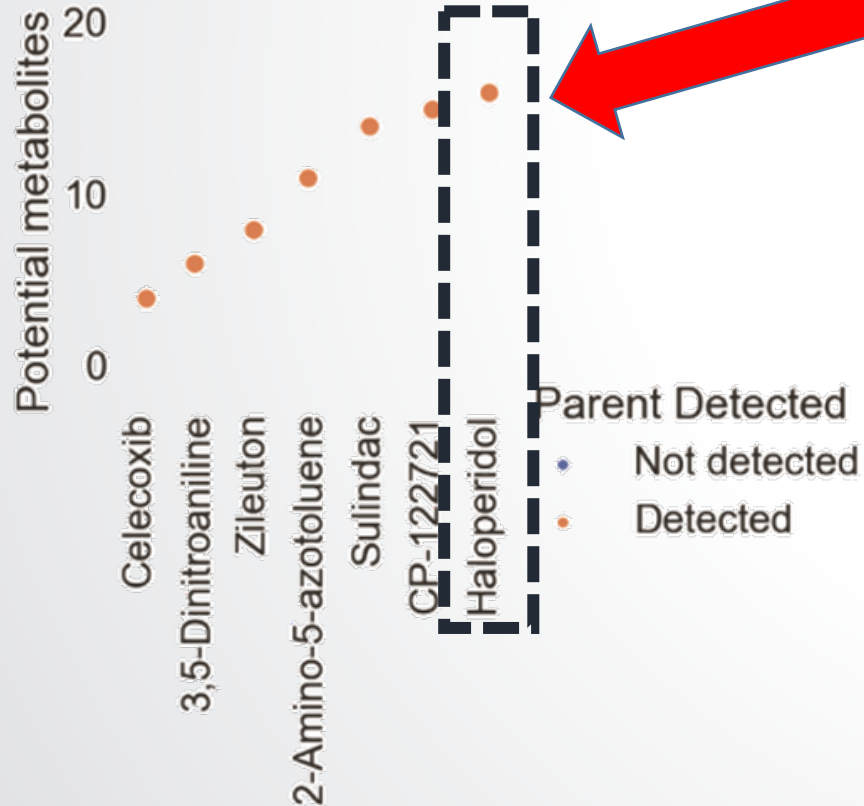
Relative change in parent signal over 4h

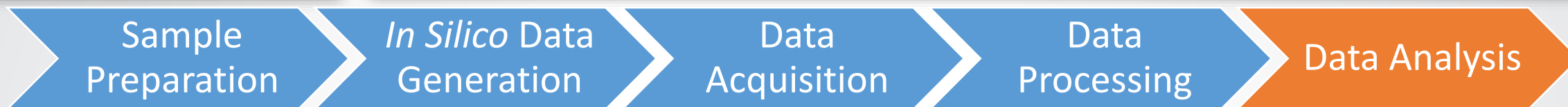
No Change ← → Greatest Decrease





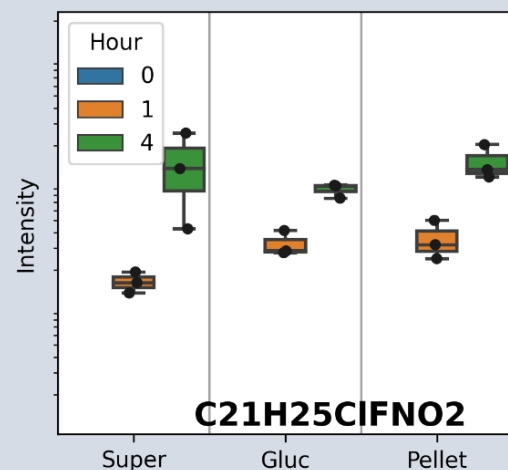
# Which parents are being metabolized?





## MS<sup>1</sup> Analysis Workflow

### 1) Broad feature filtering



*Criteria for selecting features:*

1. Fold-change increase  $\geq 1.5$
2. Appears in a minimum of two time points

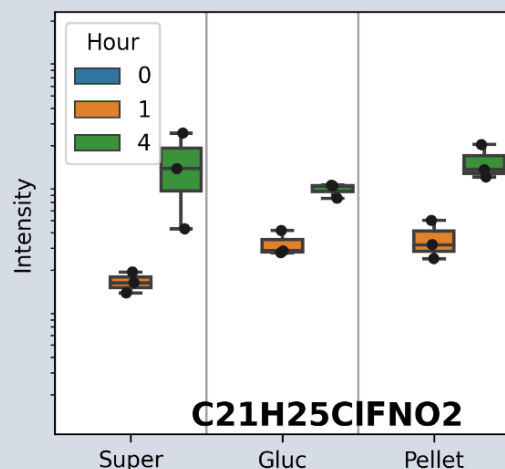


# Identifying relevant features

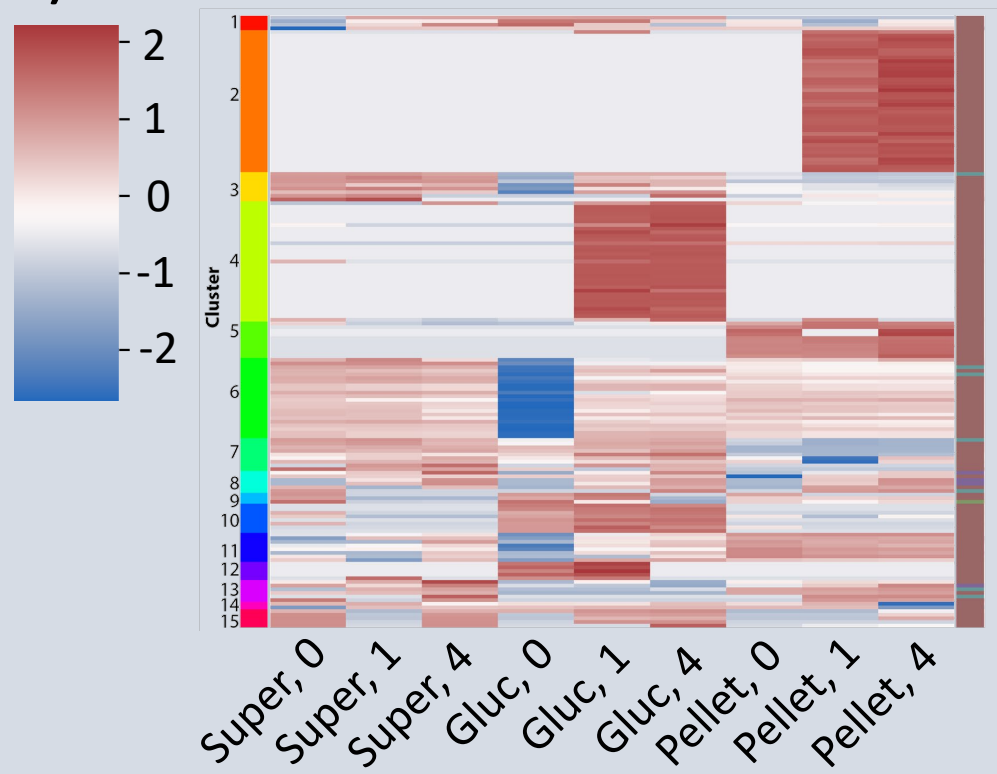


## MS<sup>1</sup> Analysis Workflow

### 1) Broad feature filtering



### 2) Cluster similar features



### Feature Annotation Source

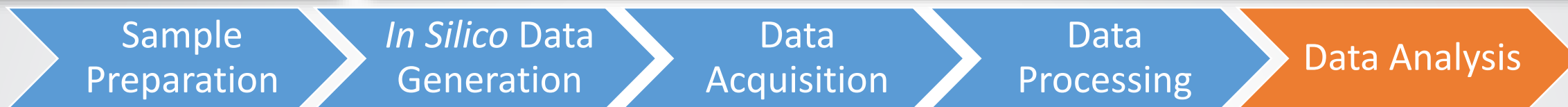
- Molecular Formula Generator
- Haloperidol (parent)
- Haloperidol metabolite
- Endogenous metabolites (human blood plasma)

Criteria for selecting features:

- Fold-change increase  $\geq 1.5$
- Appears in a minimum of two time points

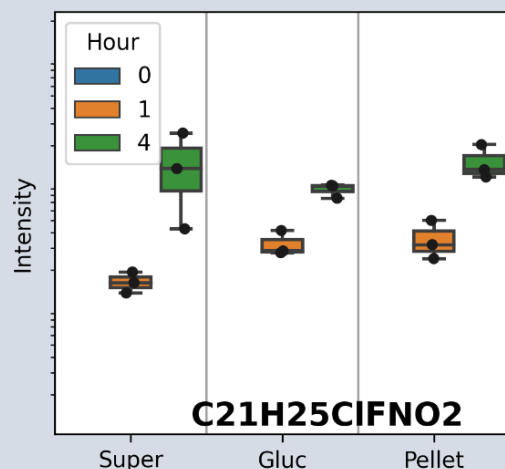


# Identifying relevant features

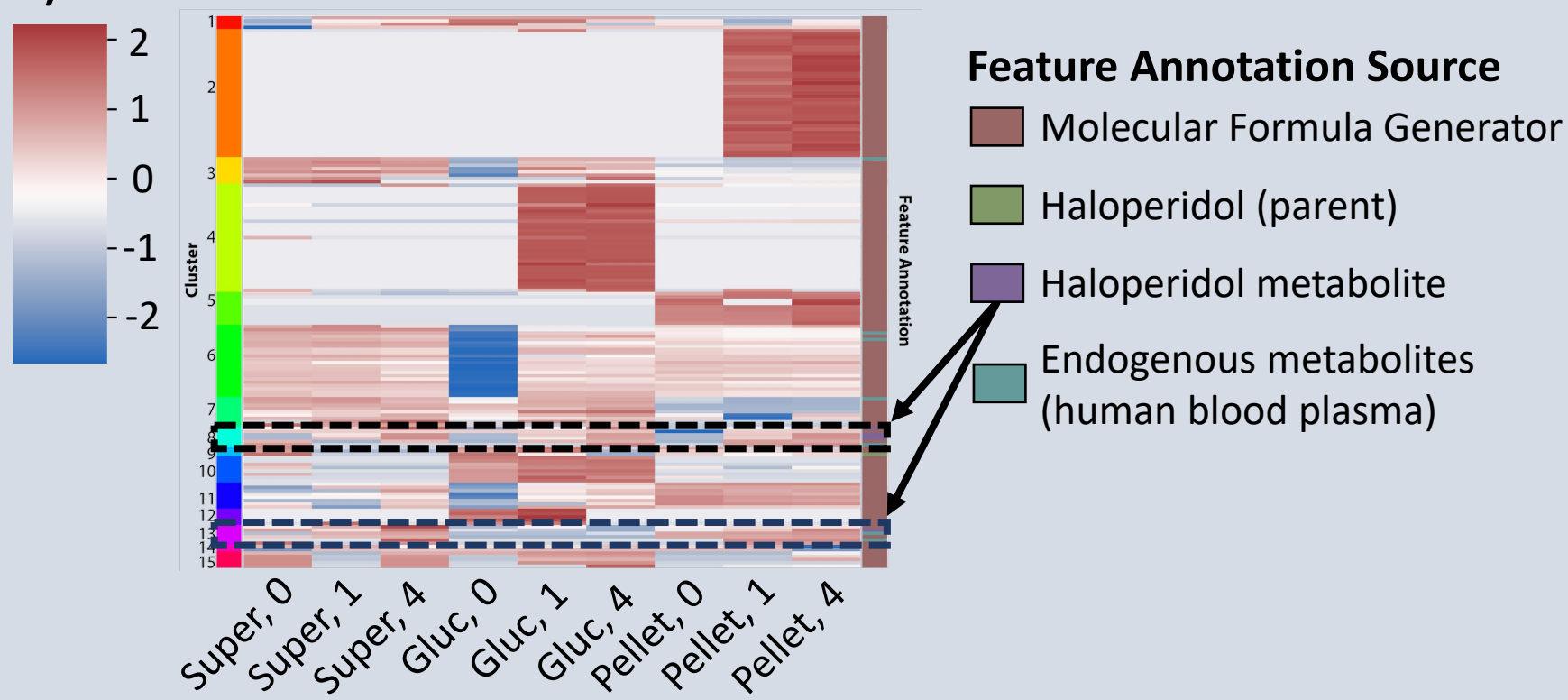


## MS<sup>1</sup> Analysis Workflow

### 1) Broad feature filtering



### 2) Cluster similar features

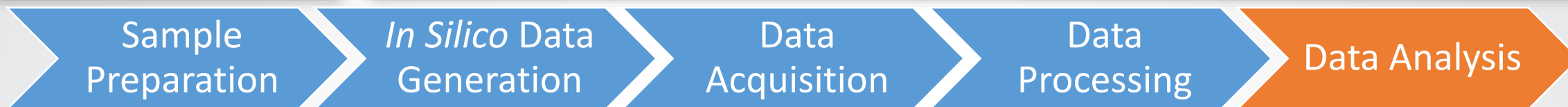


Criteria for selecting features:

1. Fold-change increase  $\geq 1.5$
2. Appears in a minimum of two time points

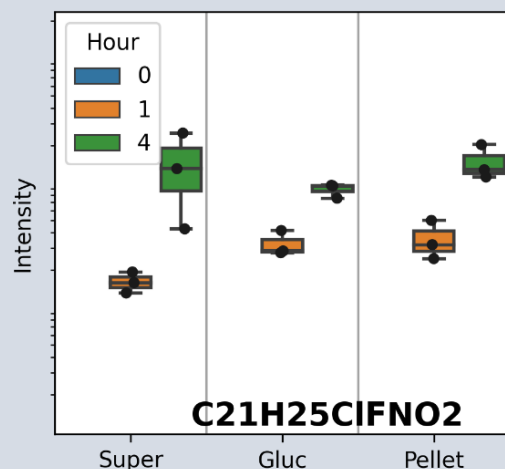


# Identifying relevant features



## MS<sup>1</sup> Analysis Workflow

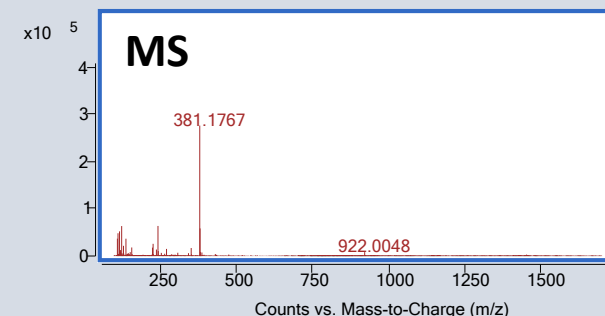
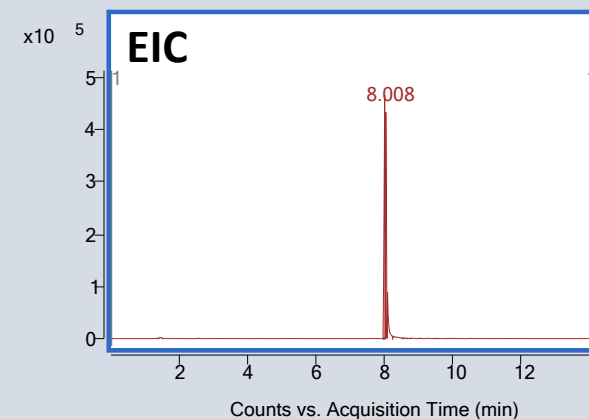
### 1) Broad feature filtering



### 2) Cluster similar features



### 3) Manual Review



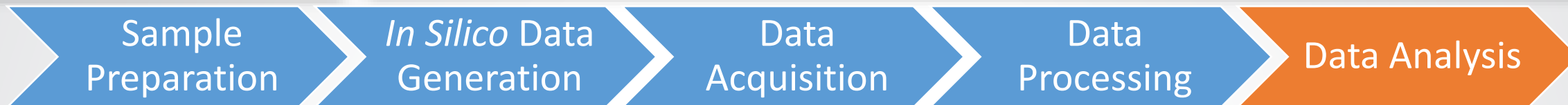
Criteria for selecting features:

1. Fold-change increase  $\geq 1.5$
2. Appears in a minimum of two time points

Super, 0  
Super, 1  
Super, 4  
Gluc, 0  
Gluc, 1  
Gluc, 4  
Pellet, 0  
Pellet, 1  
Pellet, 4



# Assigning structure to features



## MS<sup>2</sup> Analysis Workflow

### CFM-ID Comparisons

1. Precursor ions filtered using features from MS analysis
2. MS<sup>2</sup> data matched against CFM-ID database
3. CFM-ID matches ranked based on similarity values and normalized as a 'Q-Score' (ranging from 0 – 1)



# Assigning structure to features

Sample  
Preparation

*In Silico* Data  
Generation

Data  
Acquisition

Data  
Processing

Data Analysis

## MS<sup>2</sup> Analysis Workflow

### CFM-ID Comparisons

1. Precursor ions filtered using features from MS analysis
2. MS<sup>2</sup> data matched against CFM-ID database
3. CFM-ID matches ranked based on similarity values and normalized as a 'Q-Score' (ranging from 0 – 1)

Q-Score  $\geq 0.75$



### Structure Assignment

	$C_{11}H_{14}ClNO$ 8 matches	$C_{22}H_{15}N_3S$ 1 match	$C_{21}H_{23}ClFNO_2$ 2 matches
Top Match	Q-Score: 1.0 	Q-Score: 1.0 	Q-Score: 1.0 
Suspect List Match	Q-Score = 0.90 	None	Q-Score: 1.0 



# Metabolite identifications

Sample  
Preparation

*In Silico* Data  
Generation

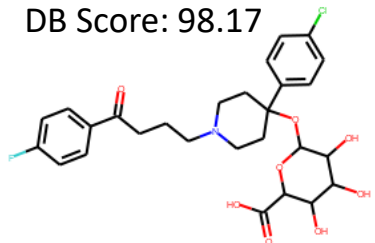
Data  
Acquisition

Data  
Processing

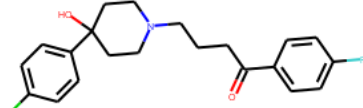
Data Analysis

CFM-ID Match (level 2b):

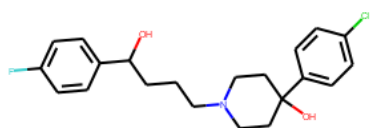
Q-Score: 1.0  
DB Score: 98.17



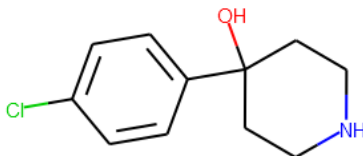
Q-Score: 1.0  
DB Score: 85.04



Q-Score: 1.0  
DB Score: 98.43

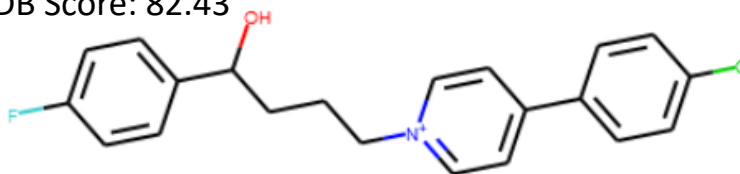


Q-Score: 0.9  
DB Score: 95.83

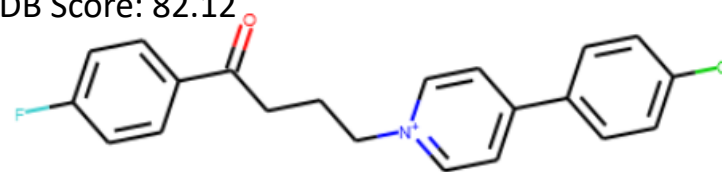


Suspect Match (level 3):

DB Score: 82.43



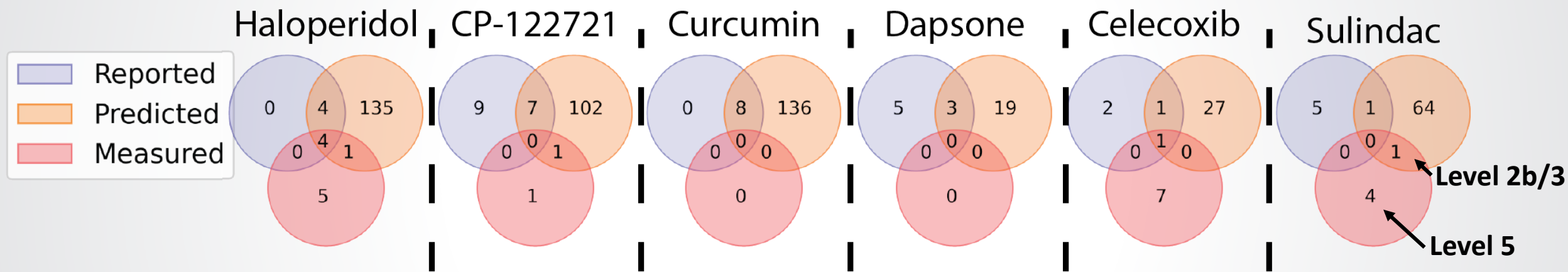
DB Score: 82.12



Predicted Formula (level 5):

- $C_9H_{13}ClN_2O_6$  (278.0816)
- $C_9H_{11}NO_3$  (181.0738 )
- $C_6H_{10}O_3$  (130.0632 )
- $C_{24}H_3ON_2O_{11}$  (522.1851)

# Distribution of tentative identifications



## Summary:

- 3 metabolites not reported in literature, but identified via *in silico* tools
- 5 metabolites identified through all sources
- 17 features without known structures, but related to metabolites



# Acknowledgements

## Environmental Protection Agency

- Grace Patlewicz
- John Wambaugh
- Lucina Lizarraga
- Jon Sobus
- Alex Chao
- Tony Williams
- Chris Grulke
- Ann Richard
- Brian Meyer
- Vicente Samano
- Nancy Baker
- Daniel Chang
- Elin Ulrich

## Southwest Research Institute

- Kristin Favela

## Thermo Scientific

- Jessica A. Bonzo

## Where to reach me

[Boyce.matthew@epa.gov](mailto:Boyce.matthew@epa.gov)



[linkedin.com/in/mwboyce16/](https://www.linkedin.com/in/mwboyce16/)



[orcid.org/0000-0002-3794-1678](https://orcid.org/0000-0002-3794-1678)