



# pyOPENMS for metabolomics

Eftychia Eva Kontou, Axel Walter

Röst, H.L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., Andreotti, S., Ehrlich, H.-C., Gutenbrunner, P., Kenar, E., Liang, X., Nahnsen, S., Nilse, L., Pfeuffer, J., Rosenberger, G., Rurik, M., Schmitt, U., Veit, J., Walzer, M., Wojnar, D., Wolski, W.E., Schilling, O., Choudhary, J.S., Malmström, L., Aebersold, R., Reinert, K., Kohlbacher, O. *OpenMS: A flexible open-source software platform for mass spectrometry data analysis.* Nature Methods, vol. 13, 2016. [doi:10.1038/nmeth.3959](https://doi.org/10.1038/nmeth.3959)

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



novo  
nordisk  
fonden

# Computational Metabolomics/Proteomics

Hundreds of ***experimental methods*** and protocols

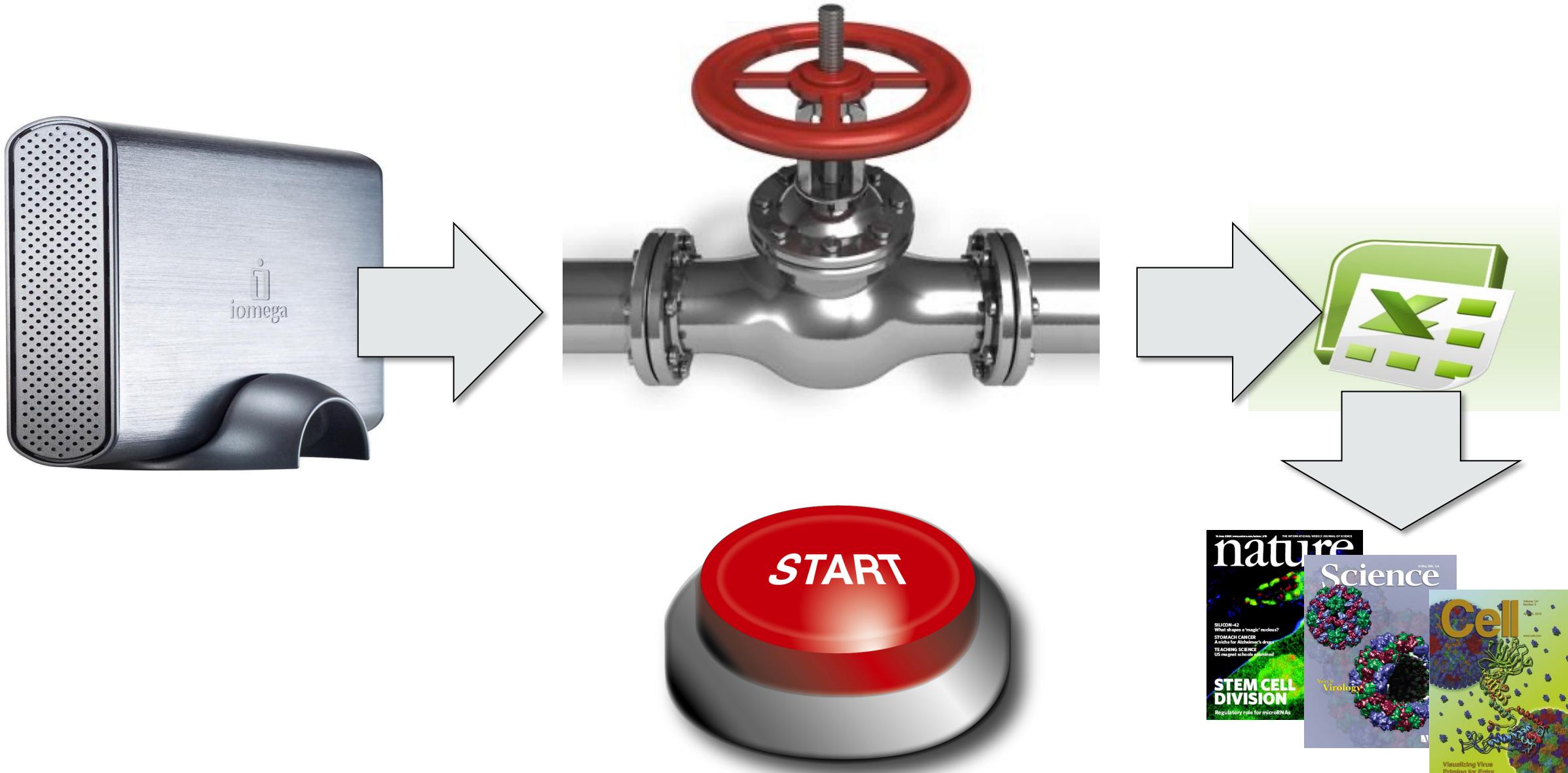
Many ***computational methods*** for

- Identification
- Quantification
- Statistical analysis

often tailored to experimental method

**Goal: Extraction of maximum biological knowledge**

# Bioinformatics – The Holy Grail



# Mass Spectrometry Workflows

Are diverse:

- Proteomics/metabolomics/interactomics, ...
- Separation methods/Acquisition
- with/without label
- Experimental designs

→ **A single workflow or tool is not enough**

**Solution:** A set of tools that can be combined to highly flexible workflows



# **What is OpenMS?**

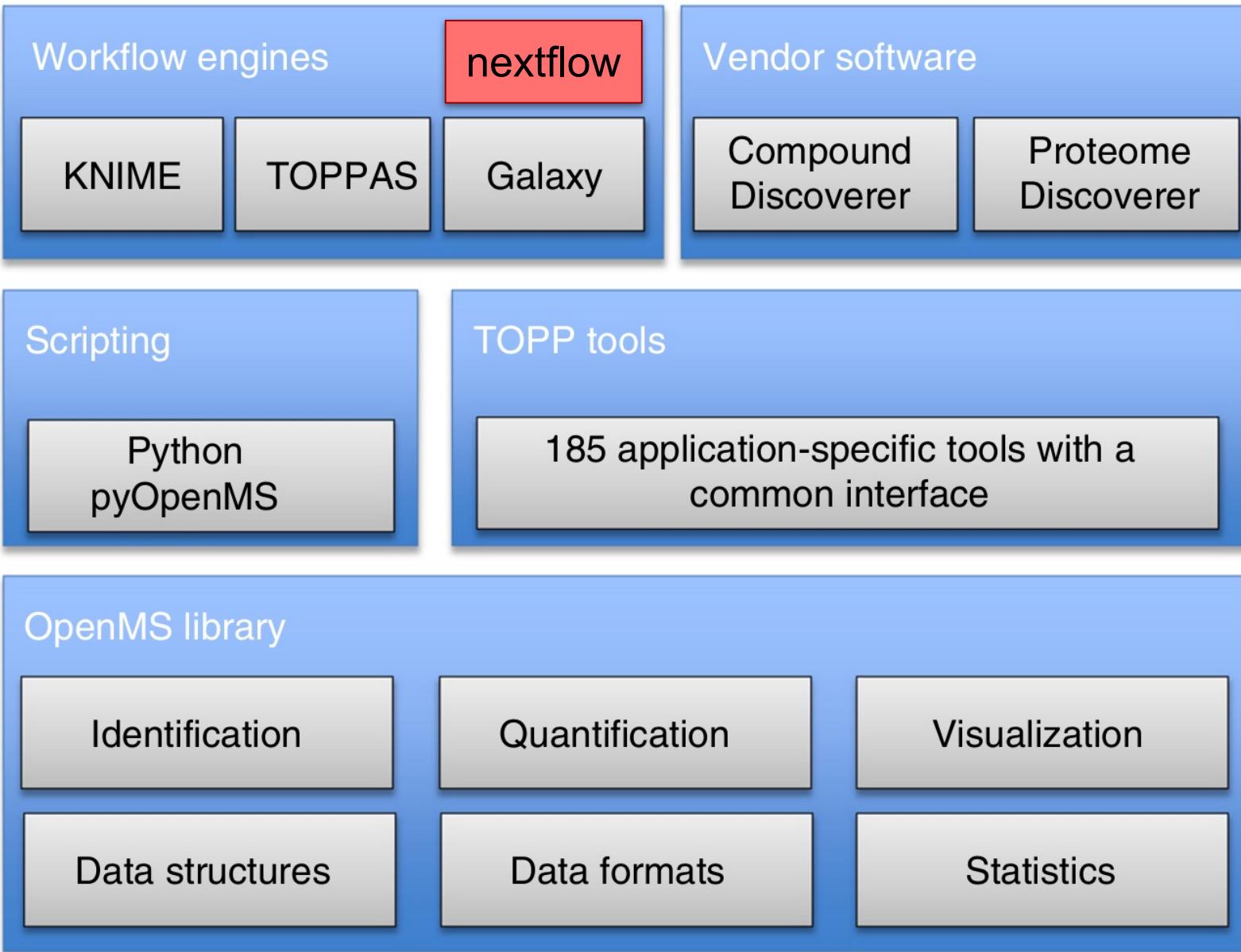
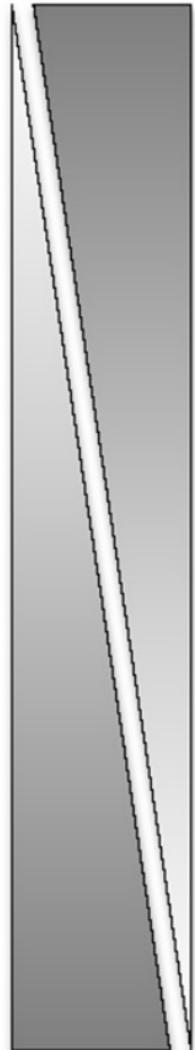
# What is OpenMS?

OpenMS is:

- Computational **library**, **tools**, and **workflows** for the analysis of mass spectrometry data
- **Open-source** and free of charge
- **Vendor-independent**: Supports vendor-formats through conversion with proteowizard library
- **Standards**: Supports all major file formats (PSI)

# The OpenMS framework

Usability and abstraction

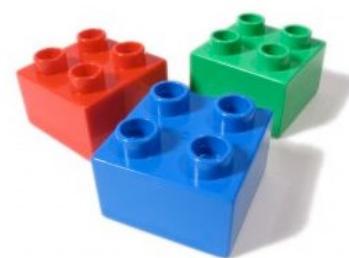


Extensibility

# OpenMS Tools

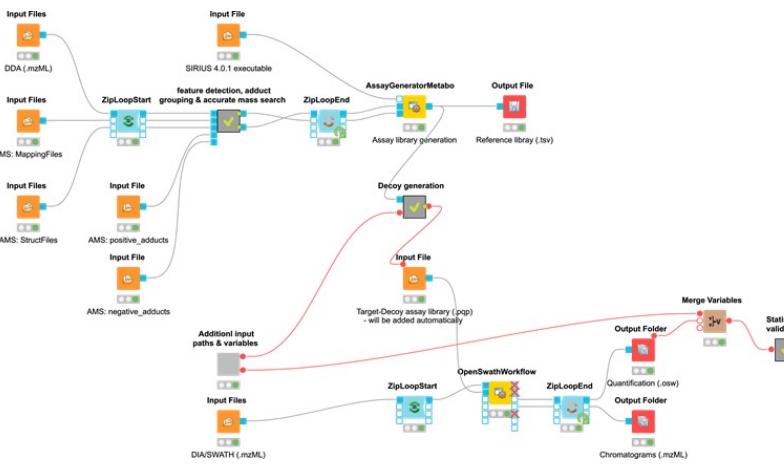
## TOPP tools – The OpenMS PiPeline tools

- > 180 Building blocks: One application for each analysis step
  - All applications share **identical user interfaces**
  - Exchange data in **PSI standard formats**
- Easily integrated in various **workflow systems**
- **Support for other tools (e.g., SIRIUS, GNPS...)**



# OpenMS workflows

Chaining of tools to build workflows  
Facilitate *reproducible* and  *reusable* research



Different workflow systems:

**nextflow**

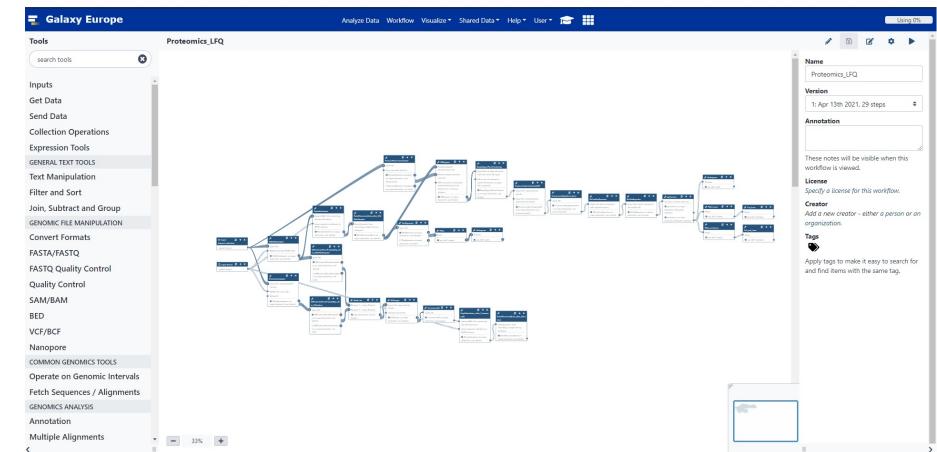
```
/*
 * STEP 1 - generate reversed decoy database
 */
process generate_decoy_database {
    input:
        file fastafile from input_fasta.mix(appended.fasta)

    output:
        file "${fastafile.baseName}_decoy.fasta" into (fastafile_decoy_1, fastafile_decoy_2)

    when:
        !params.skip_decoy_generation

    script:
        ===
        DecoyDatabase -in ${fastafile} \\
            -out ${fastafile.baseName}_decoy.fasta \\
            -decoy_string DECOY_ \\
            -decoy_string_position prefix
        ===
}
```

**Galaxy PROJECT**



# Scripting with OpenMS: pyOpenMS



- Compiled language (fast!)
- Memory management
- Nontrivial to learn
- Verbose syntax
- Dependencies  
(challenging for developers, easy for users)

- Wraps OpenMS
- Interactive scripting
- Easy to learn
- Dependencies easy to install



launch binder

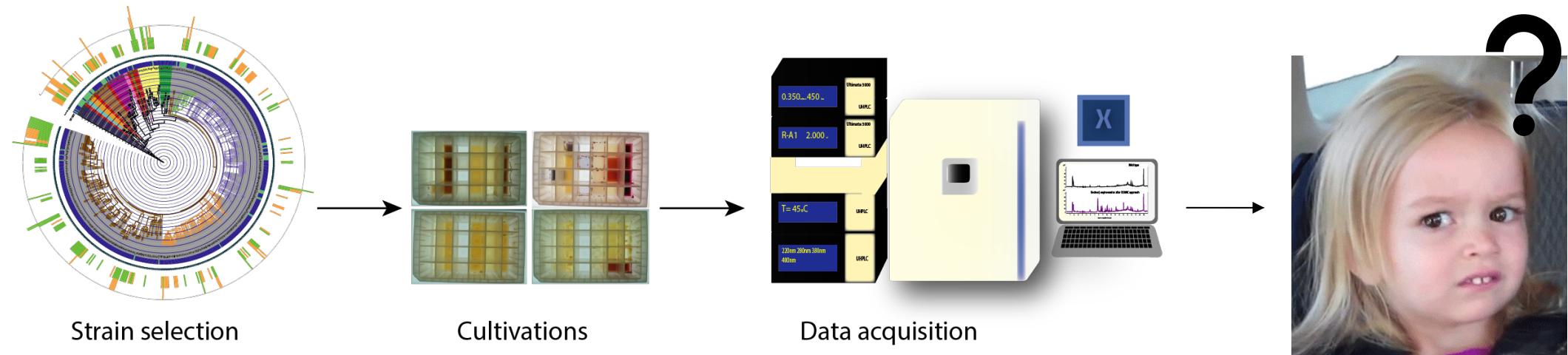
# Brief intermission..

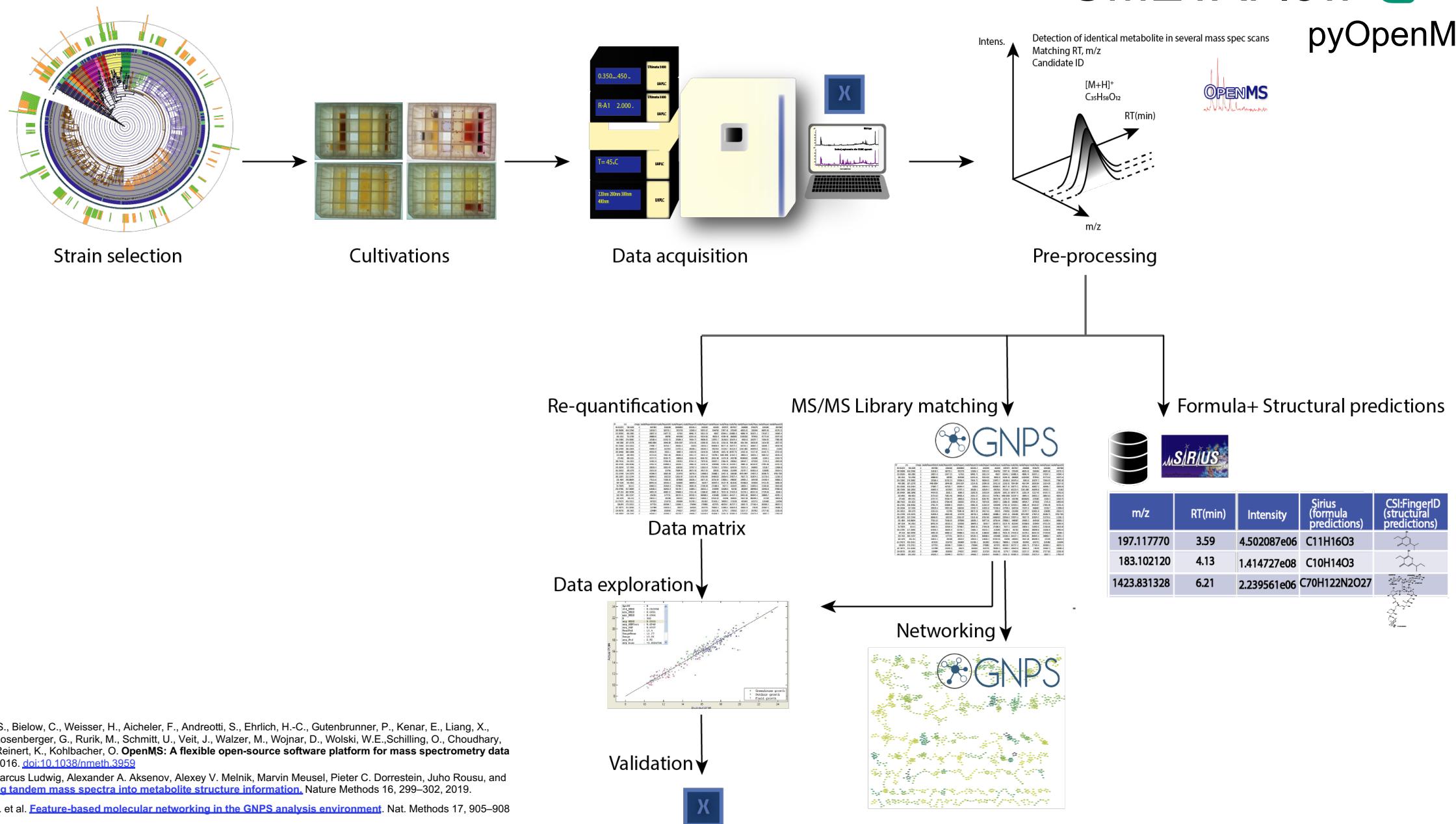
- <https://github.com/axelwalter/pyopenms-workshop>



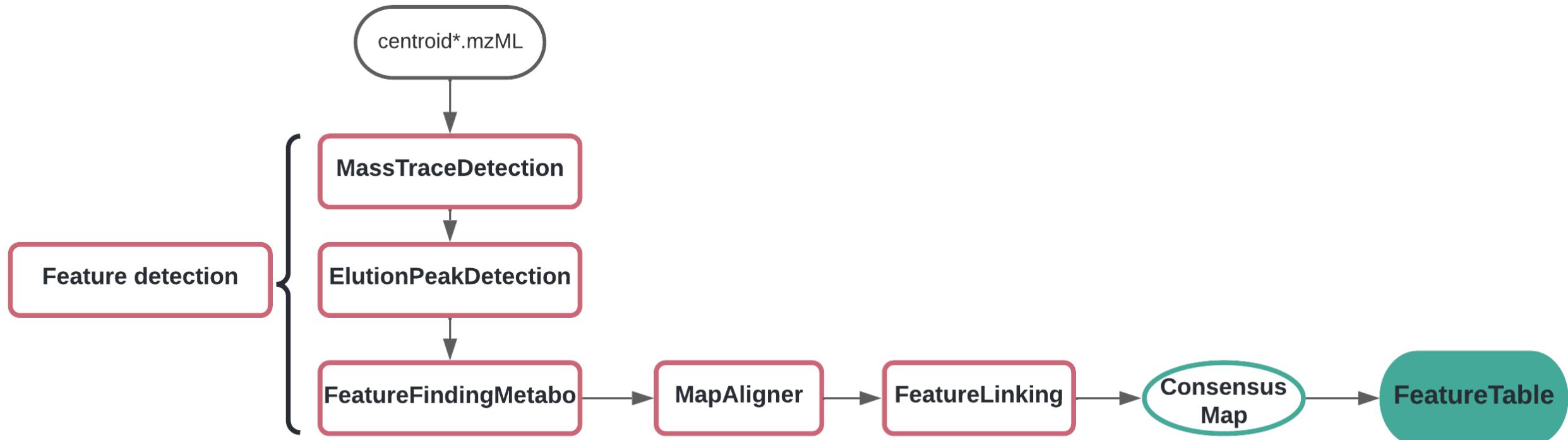
- Click on the

# Metabolomics for Specialized Metabolites, Ph.D. (Efi)



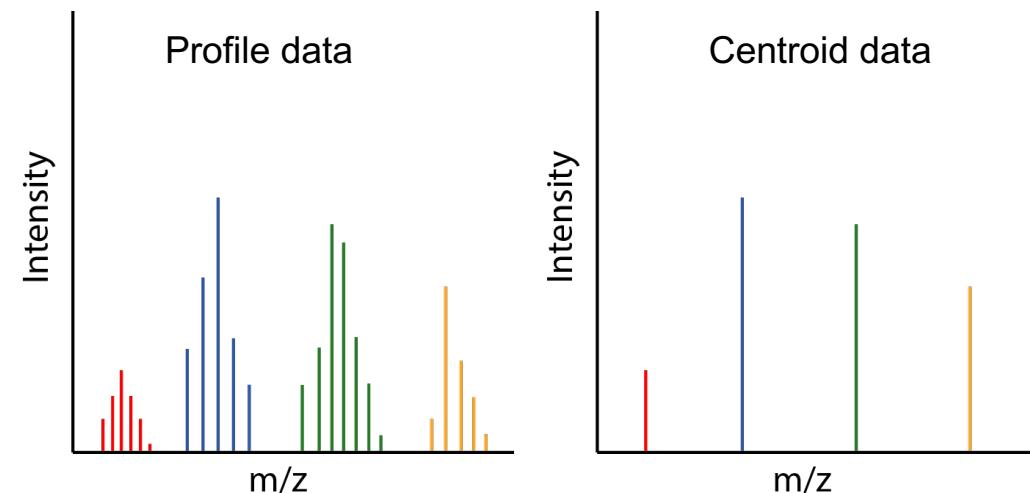
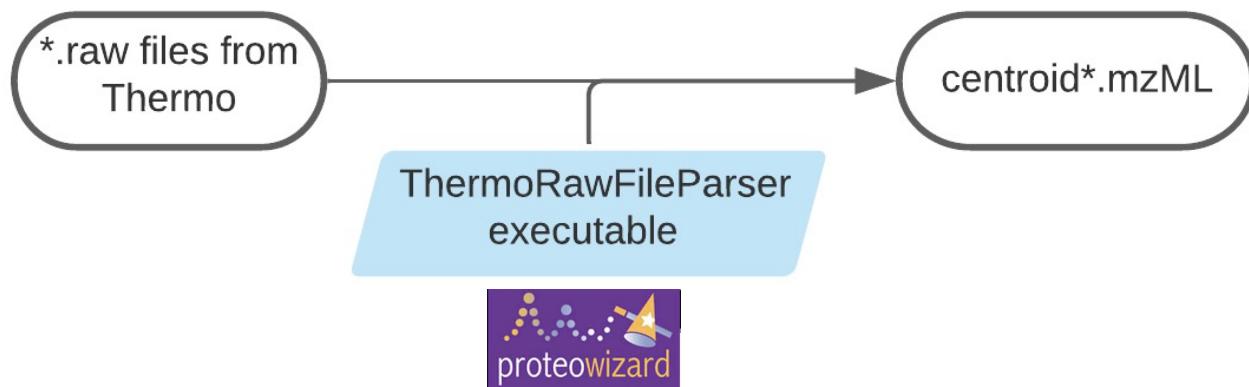


# Pre-processing: From raw data to a table of metabolic features

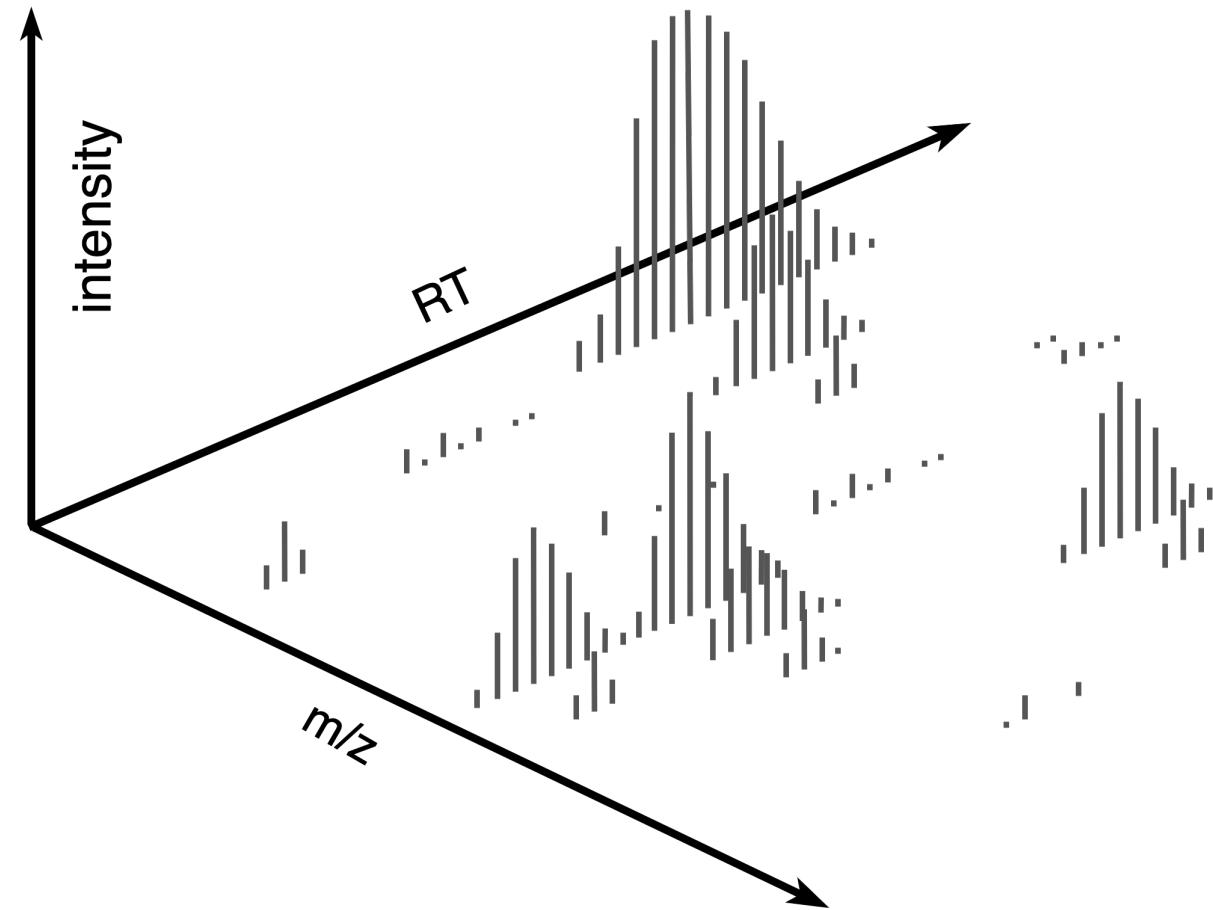


# Data conversion

Vendor-format files to an **open community-driven format**

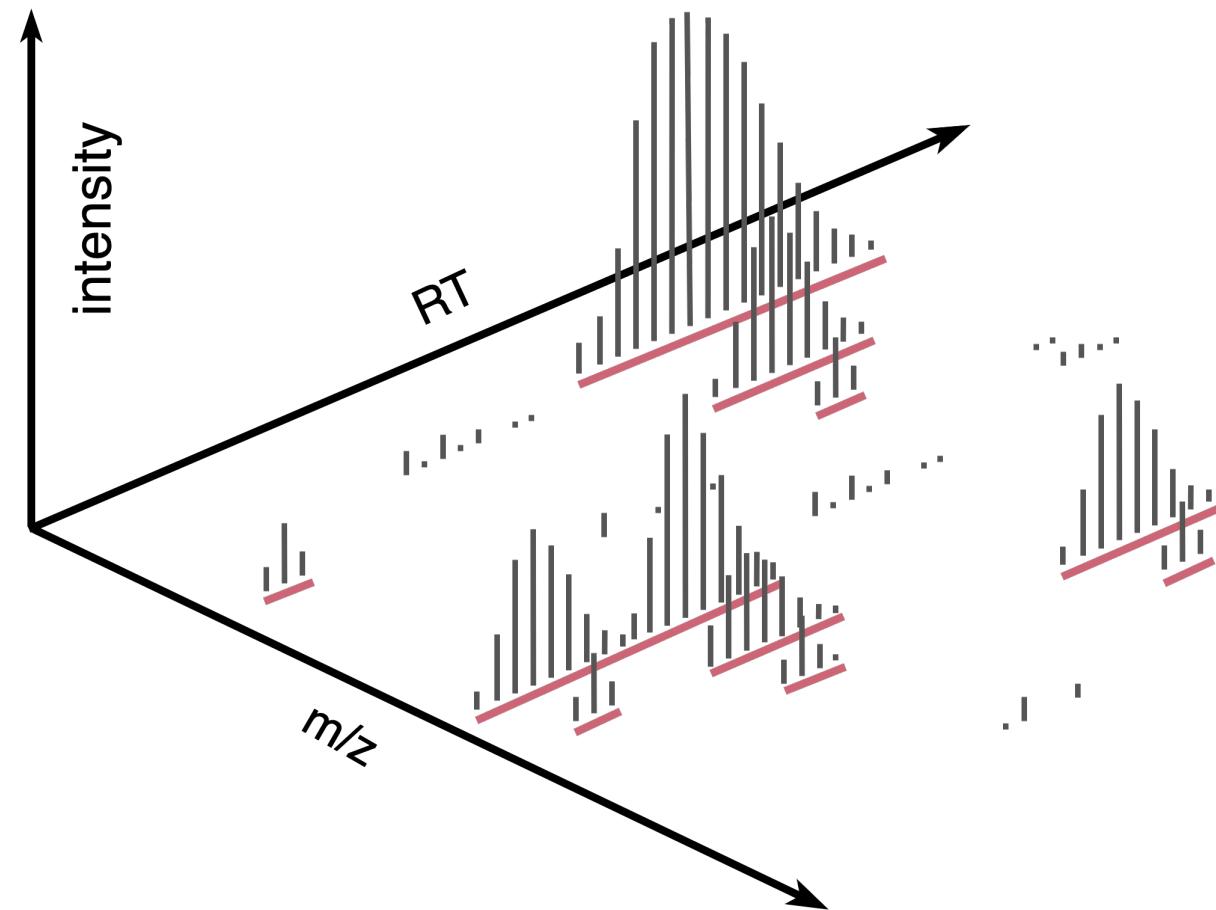


# Feature detection (raw mzML)



# 1. Mass trace detection

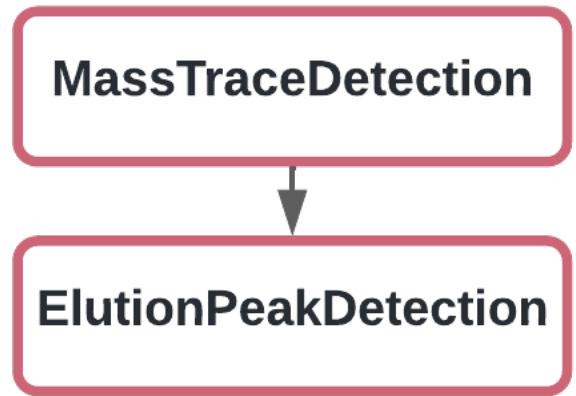
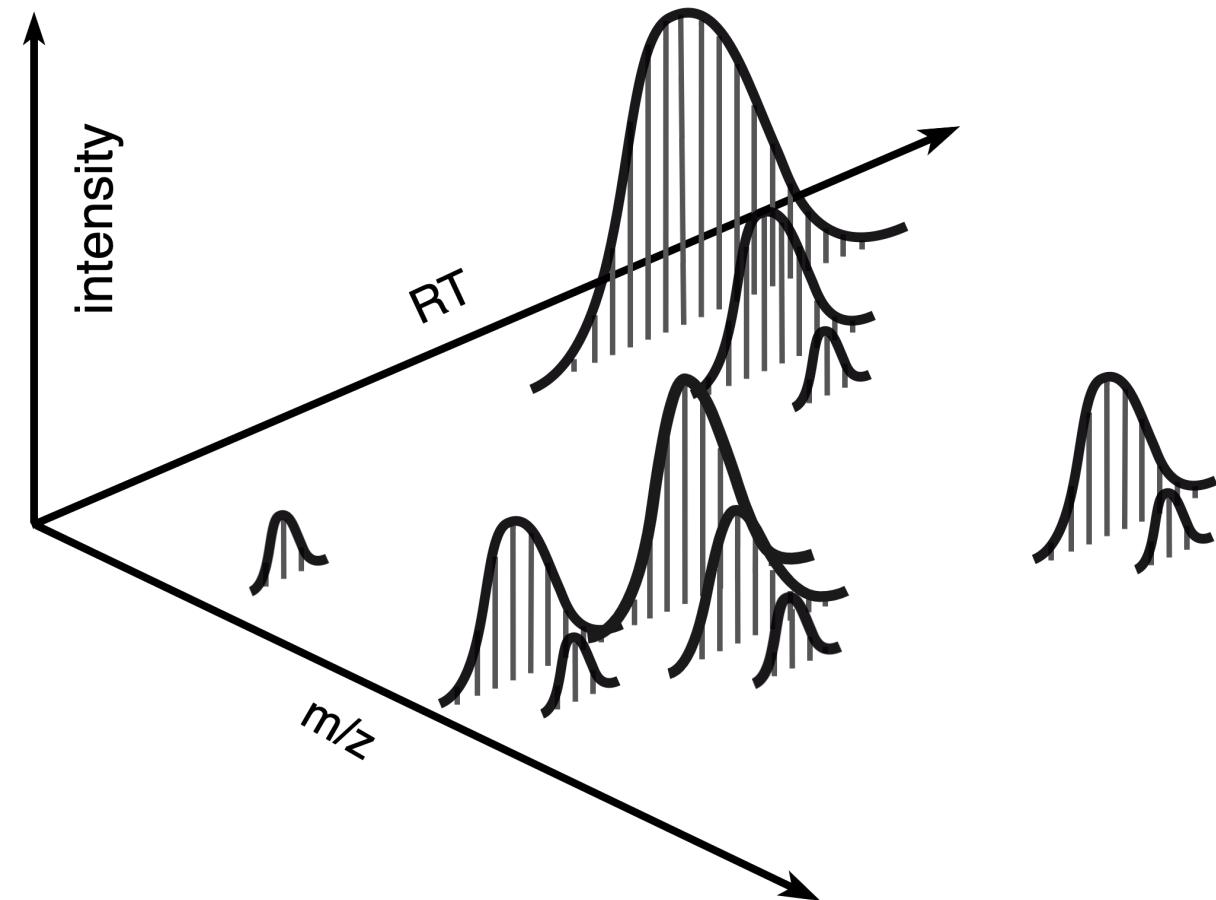
A mass trace extraction method that gathers peaks similar in  $m/z$  and moving along retention time.



MassTraceDetection

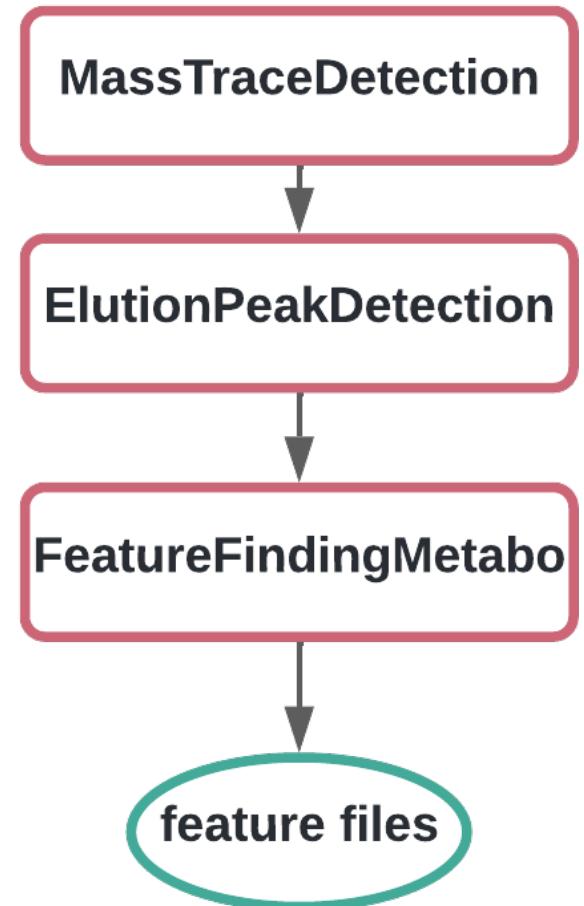
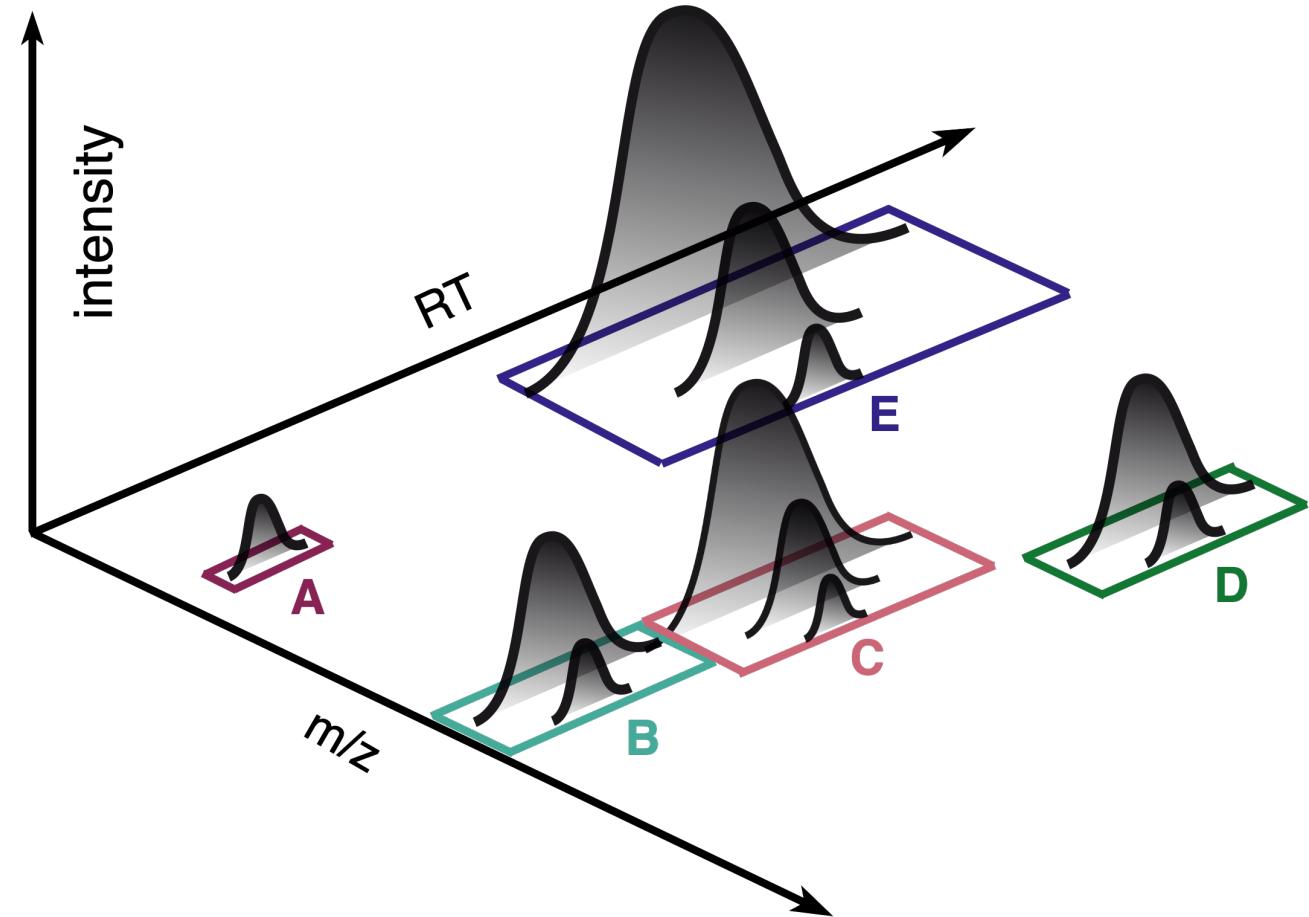
## 2. Elution peak detection

Extracts chromatographic peaks from a mass trace.

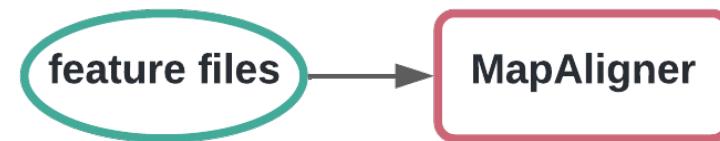


### 3. FeatureFinder

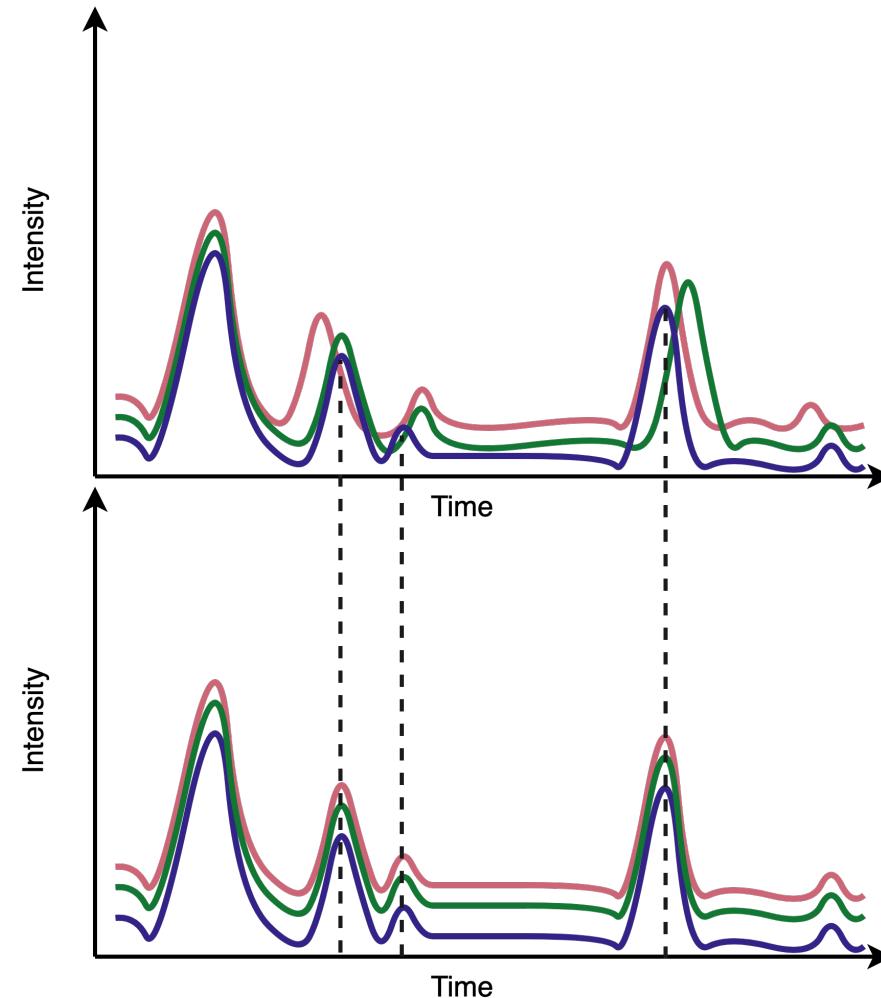
Assembles metabolite features from singleton mass traces.



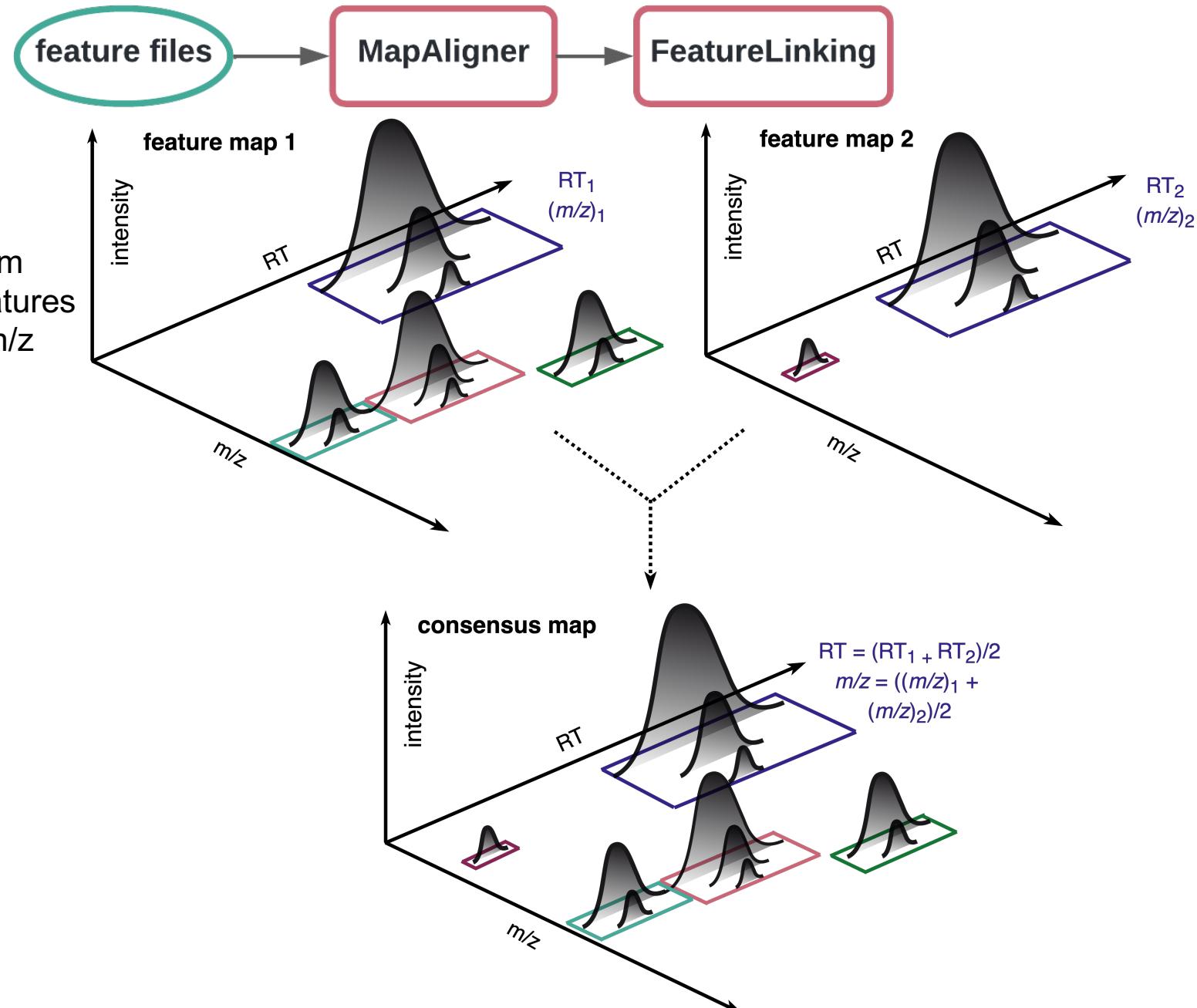
## 4. Map alignment



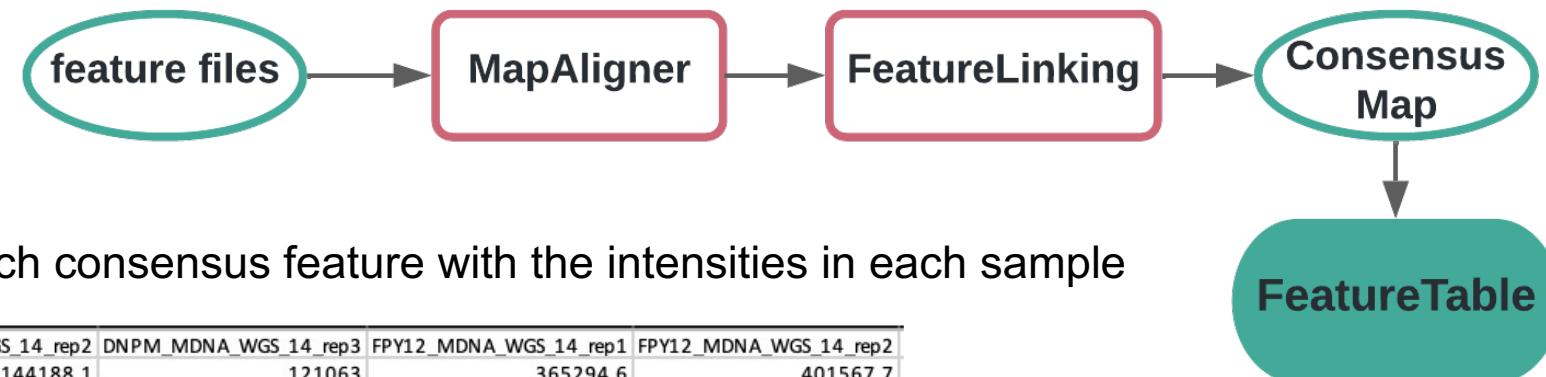
Performs a linear retention time alignment, to correct for linear shifts in retention time.



## 5. Feature linking



## 6. Table of features

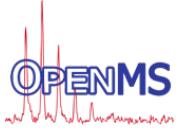


Dataframe with m/z and RT information from each consensus feature with the intensities in each sample

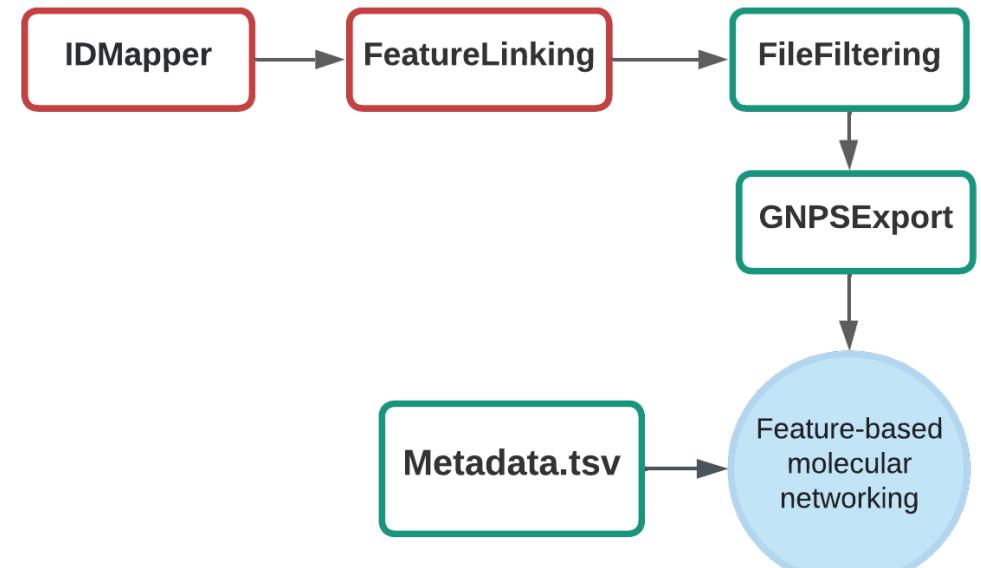
| mz       | RT      | DNPM_MDNA_WGS_14_rep1 | DNPM_MDNA_WGS_14_rep2 | DNPM_MDNA_WGS_14_rep3 | FPY12_MDNA_WGS_14_rep1 | FPY12_MDNA_WGS_14_rep2 |
|----------|---------|-----------------------|-----------------------|-----------------------|------------------------|------------------------|
| 416.2637 | 347.19  | 136219.6              | 144188.1              | 121063                | 365294.6               | 401567.7               |
| 1441.839 | 389.335 | 540440.3              | 58655.34              | 408005.8              | 1033793                | 5176443                |
| 966.6142 | 394.971 | 284705.5              | 149096.8              | 140170                | 29289.28               | 247274.7               |
| 684.2365 | 67.7326 | 124401.3              | 32565.46              | 28127.74              | 932534.5               | 1040187                |
| 1423.831 | 405.735 | 2249281               | 1614707               | 1636310               | 2324051                | 37097830               |
| 705.9128 | 388.712 | 1686535               | 1109222               | 1095907               | 51761.03               | 838351.3               |
| 787.9419 | 394.949 | 115049.2              | 131840.9              | 76413.18              | 349033.8               | 7193028                |
| 1020.645 | 392.771 | 413206.9              | 266186.6              | 284185.1              | 195431.9               | 2686807                |
| 1047.42  | 398.641 | 272859.3              | 432108.1              | 533610.7              | 274664.6               | 7865175                |
| 1162.709 | 411.64  | 183676.2              | 131980.2              | 166135                | 148731.6               | 3300304                |
| 719.9283 | 410.999 | 695653.4              | 534247.3              | 484984.5              | 487374.9               | 4233446                |
| 1000.358 | 403.499 | 196254.7              | 153696.9              | 186016.1              | 29293.06               | 58768.09               |
| 206.1534 | 475.921 | 61302.61              | 22137.95              | 21526.22              | 15900.39               | 14530.26               |
| 819.9009 | 401.672 | 114797.2              | 106675.7              | 107820.8              | 399065.8               | 3710466                |
| 801.4605 | 391.291 | 61343.05              | 88985.27              | 61679.61              | 192527.3               | 5325955                |
| 1180.719 | 395.529 | 34702.9               | 61262.11              | 26275.12              | 104274.9               | 3261669                |
| 657.2782 | 61.5897 | 618277.2              | 380894.1              | 367958.9              | 8468848                | 28313430               |
| 988.6195 | 388.865 | 148853                | 137188.4              | 340571.3              | 6243.7017              | 1518407                |
| 807.2531 | 259.516 | 22518.6               | 15335.66              | 23392.15              | 2158250                | 1640954                |
| 1442.844 | 390.495 | 441195.7              | 370601.3              | 353536.6              | 877375.1               | 40558810               |

# Suggestions after pre-processing

- **SiriusAdapter:** Generate formula and structural predictions with **SIRIUS** and **CSI:FingerID** (MS1& MS2) 

- **Adduct annotation:** Assign adducts & reconstruct neutral masses 

- **GNPSExport:** Generate all the files necessary for Feature-Based Molecular Networking (FBMN) and  Ion Identity Molecular Networking (IIMN)



<https://pyopenms.readthedocs.io/en/latest/index.html>

Kai Dührkop, Markus Fleischauer, Marcus Ludwig, Alexander A. Aksenov, Alexey V. Melnik, Marvin Meusel, Pieter C. Dorrestein, Juho Rousu, and Sebastian Böcker. [SIRIUS 4: Turning tandem mass spectra into metabolite structure information](#). Nature Methods 16, 299–302, 2019.

Nothias, L.-F., Petras, D., Schmid, R. et al. [Feature-based molecular networking in the GNPS analysis environment](#). Nat. Methods 17, 905–908 (2020).

# Hands-on time!

- <https://github.com/axelwalter/pyopenms-workshop>



- Click on the

# Visit our documentation:

<https://pyopenms.readthedocs.io/en/latest/>

The screenshot shows a dark-themed documentation page for pyopenms. On the left sidebar, under 'GETTING STARTED', there are links to Installation, Import pyopenms, Getting help, and First look at data. Under 'MASS SPECTROMETRY CONCEPTS', there are links to MS Data, Chemistry, Peptides and Proteins, Oligonucleotides: RNA, Fragment spectrum generation, and a link to GitHub.

The main content area has a light background. At the top, there are navigation links: a home icon followed by '» Feature Detection', a 'TRY ONLINE WITH' button, a 'BINDER' button, and an 'Edit on GitHub' button.

## Metabolomics - untargeted

For the untargeted detection of small molecule features we can use the `FeatureFindingMetabo` with prior `MassTraceDetection` and `ElutionPeakDetection`.

```
from pyopenms import *
from urllib.request import urlretrieve

gh = "https://raw.githubusercontent.com/OpenMS/pyopenms-docs/master"
mzML_path = gh + "/src/data/FeatureFinderMetaboIdent_1_input.mzML"
urlretrieve(mzML_path, "ms_data.mzML")

exp = MSEExperiment()
MzMLFile().load("ms_data.mzML", exp)
```