Mapping Raw DIA MS Data to OpenSWATH

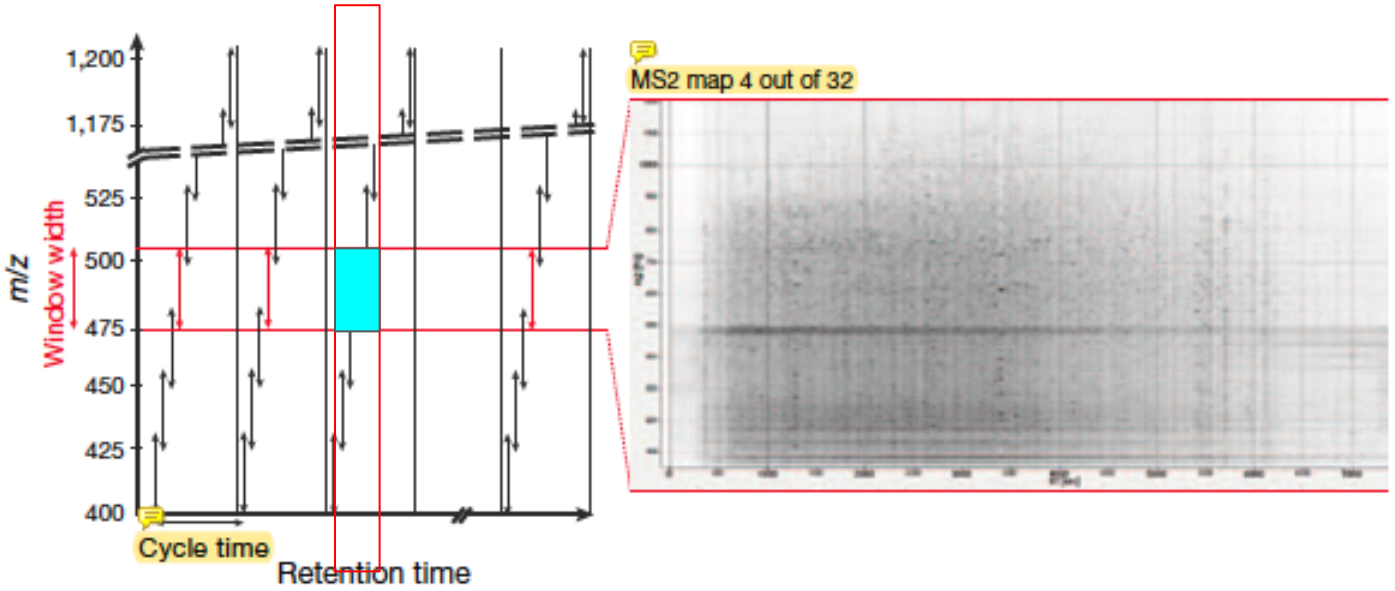
Ira Horecka
Hannes Roest, Rotation 4, Winter 2022
2022-02-04

Conceptual background

- Working with Data Independent Acquisition (DIA)
  - Fragments all precursors within predefined mass to charge (m/z) ratio and retention time (RT) range
  - Creates a complete 2D (m/z to RT) record of the fragment-ion signal from every sample precursor
- Workflow of SWATH-MS (DIA method):
  - Isolate 25 m/z precursor windows
  - Fragment the precursors
  - Analyze the fragment ions on a time-of-flight (TOF) analyzer
- OpenSWATH is an automated software for targeted DIA analysis
  - Accepts SWATH-MS data

Conceptual background

Conceptual background



Conceptual background

Project background

- Map raw DIA MS2 peaks (transitions) that are deemed significant by OpenSWATH
  - Raw transitions → Transitions from **.mzML** file (TimsTOF instrument)
  - Significant transitions → Transitions from **.osw** file with QVALUE < 0.01
- Perform simple data analysis in two categories:
  - Across retention times
    - % raw transitions identified and labeled as significant by OpenSWATH
  - In one retention time window, assess flux in % identified raw transitions across:
    - m/z
    - Intensity
    - Ion Mobility

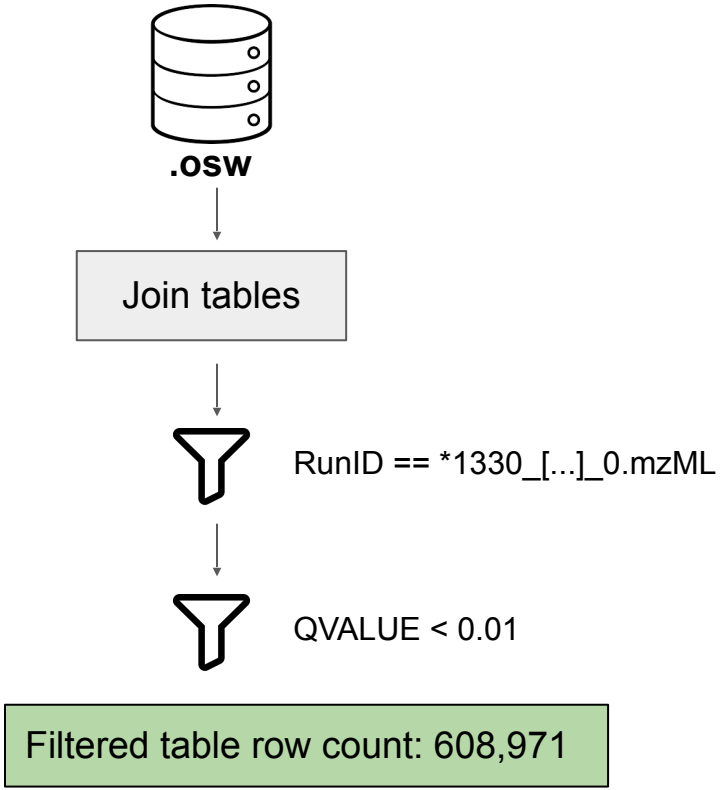
Data gathering and manipulation :: **.mzML**

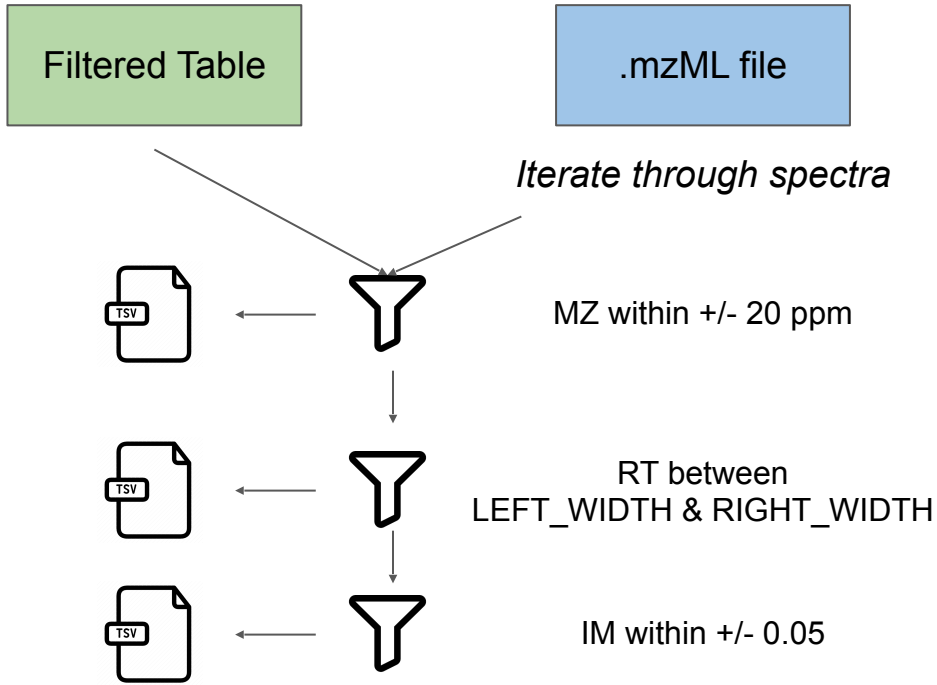
- Raw MS data (**.mzML**) is stored as an XML file
- Josh C. provided an **.mzML** file with the following properties:
  - SWATH precursor window: 700 - 725 m/z
  - RT range: 2400 - 2700 sec
- This window offers a feature-dense region for analysis
- Use `pyopenms` to work with **.mzML** in an object oriented way

Data gathering and manipulation :: **.osw**

- OpenSWATH file (**.osw**) is SQLite-compatible
- Josh C. provided an **.osw** file with six runs, one of which is linked to the **.mzML** file.
- **.osw** tables and columns of interest (bold columns = joining columns):
  - SCORE_MS2 :: [**FEATURE_ID**, QVALUE]
  - FEATURE_TRANSITION :: [**FEATURE_ID**, **TRANSITION_ID**]
  - FEATURE :: [**FEATURE_ID**, RUN_ID, EXP_IM, LEFT_WIDTH, RIGHT_WIDTH]
  - TRANSITION :: [**TRANSITION_ID**, PRODUCT_MZ]
- Joined table row count: 30,066,277

Filtering strategy



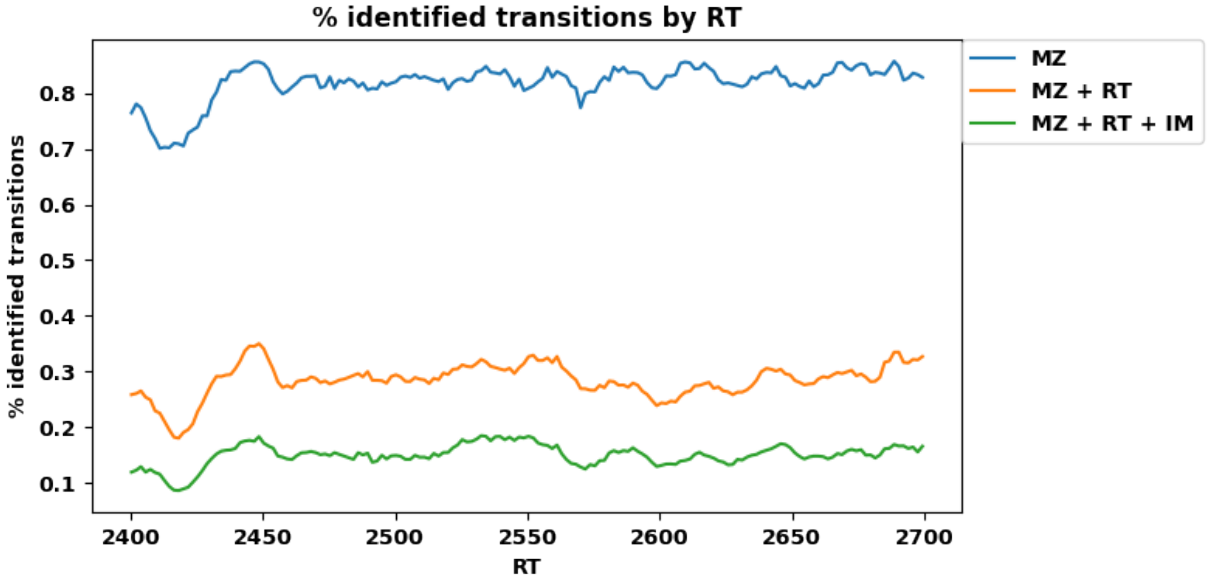


Looking at significant transitions across RT

- The RT window is between 2400 and 2700 sec
- How do % identified transitions compare among levels of filtration?

Looking at significant transitions across RT

- The RT window is between 2400 and 2700 sec
- How do % identified transitions compare among levels of filtration?

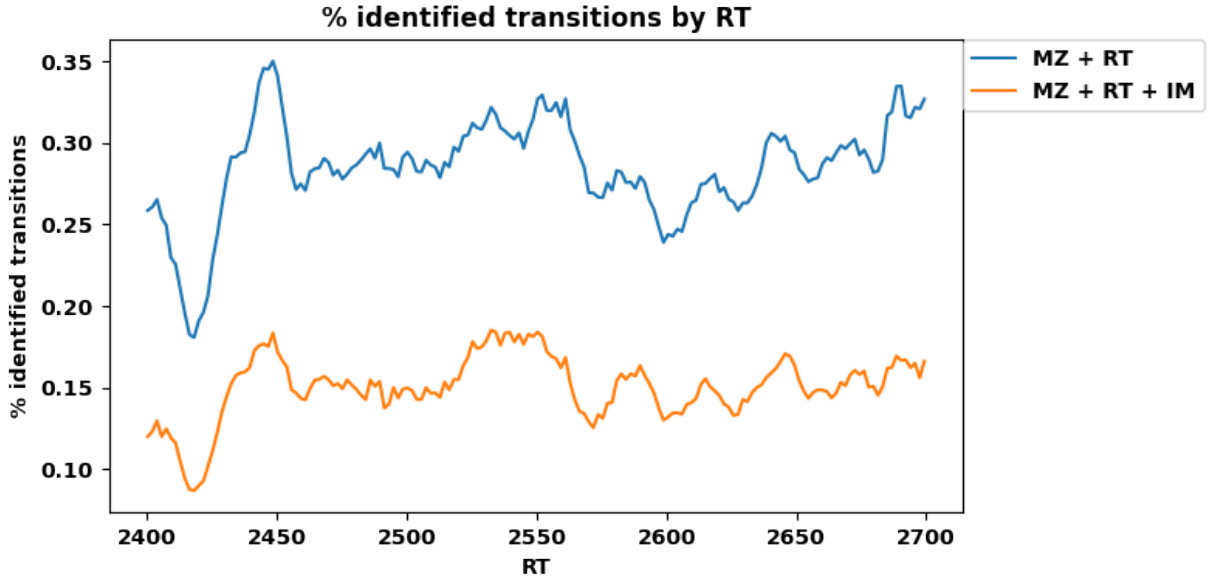


Looking at significant transitions across RT

- The RT window is between 2400 and 2700 sec
- How do % identified transitions compare among levels of filtration?

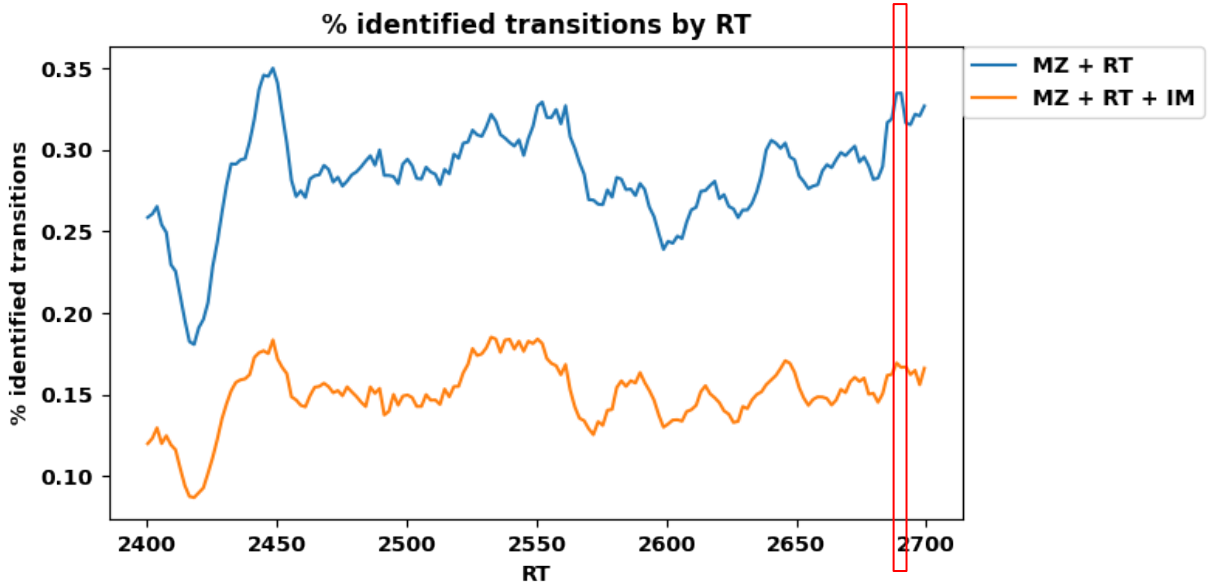
Looking at significant transitions across RT

- The RT window is between 2400 and 2700 sec
- How do % identified transitions compare among levels of filtration?



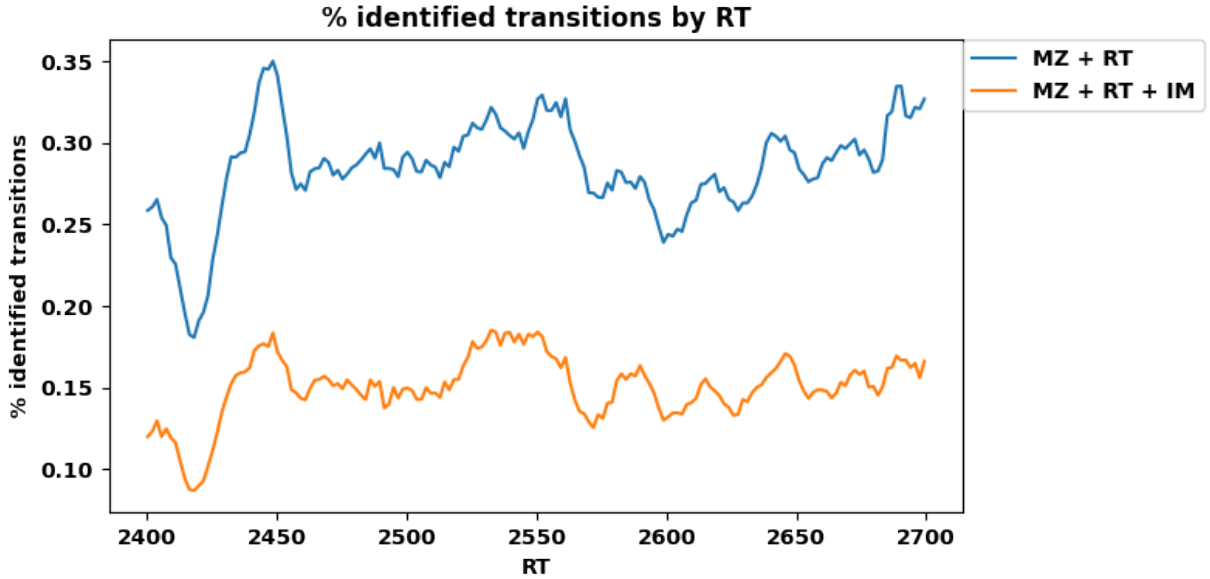
Looking at significant transitions across RT

- The RT window is between 2400 and 2700 sec
- How do % identified transitions compare among levels of filtration?



Looking at significant transitions across RT

- The RT window is between 2400 and 2700 sec
- How do % identified transitions compare among levels of filtration?



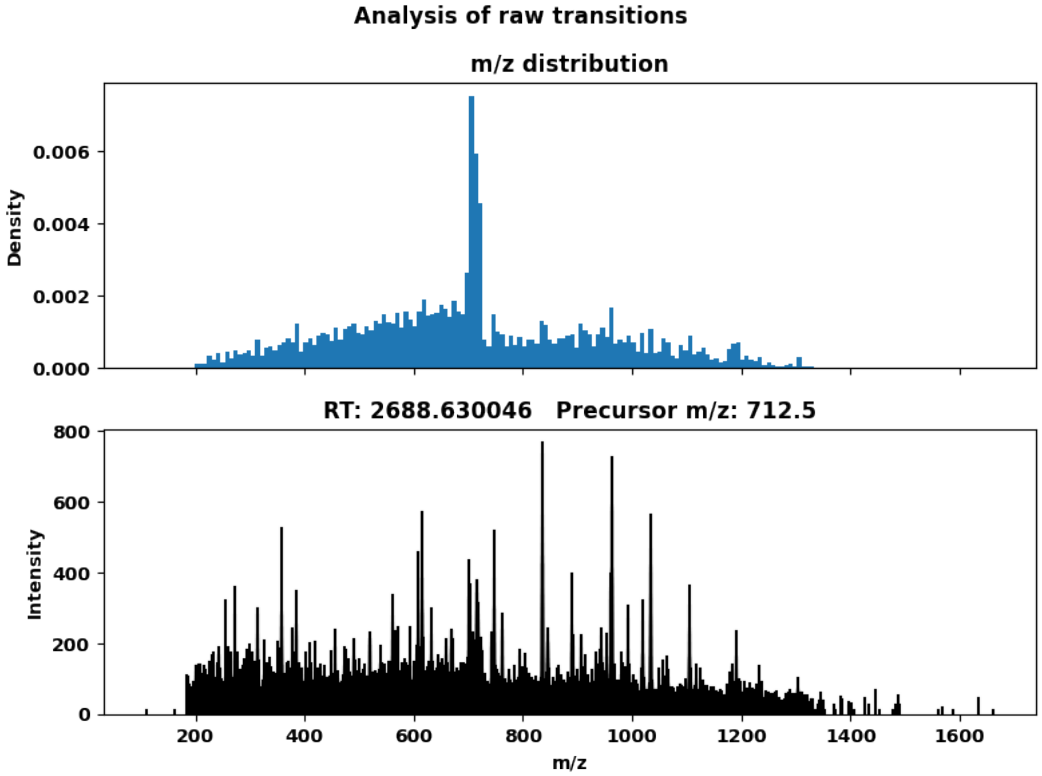
Looking at significant transitions across RT

- The RT window is between 2400 and 2700 sec
- How do % identified transitions compare among levels of filtration?

Looking at significant transitions @ RT: 2688.63

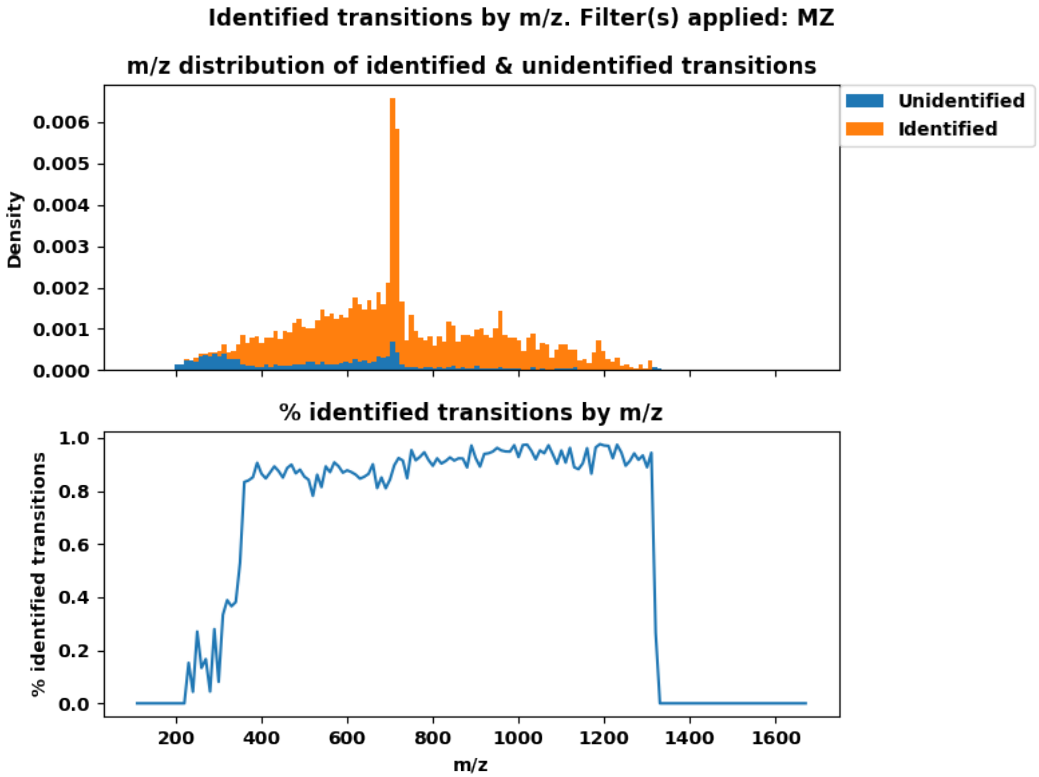
- Isolate spectrum with RT of 2688.63 sec
- Analyze % identified transitions as a factor of:
  - m/z
  - Intensity
  - Ion Mobility

RT: 2688.63 - distribution of transitions as a factor of **m/z**

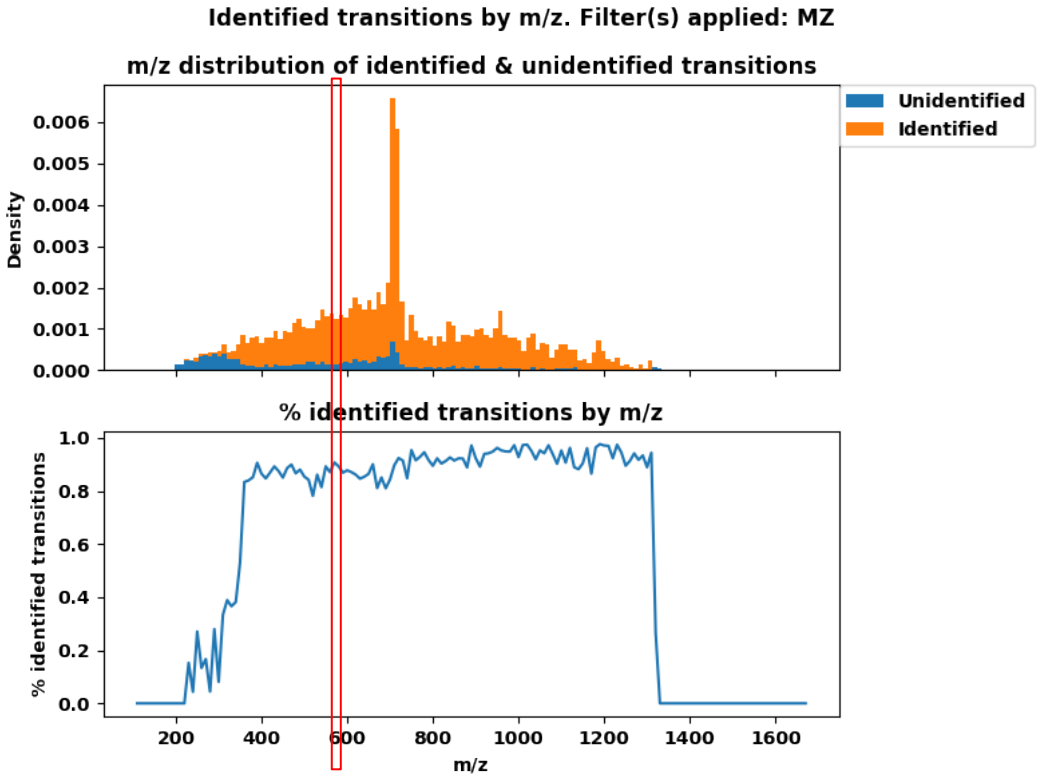


RT: 2688.63 - % identified peaks as a factor of **m/z**

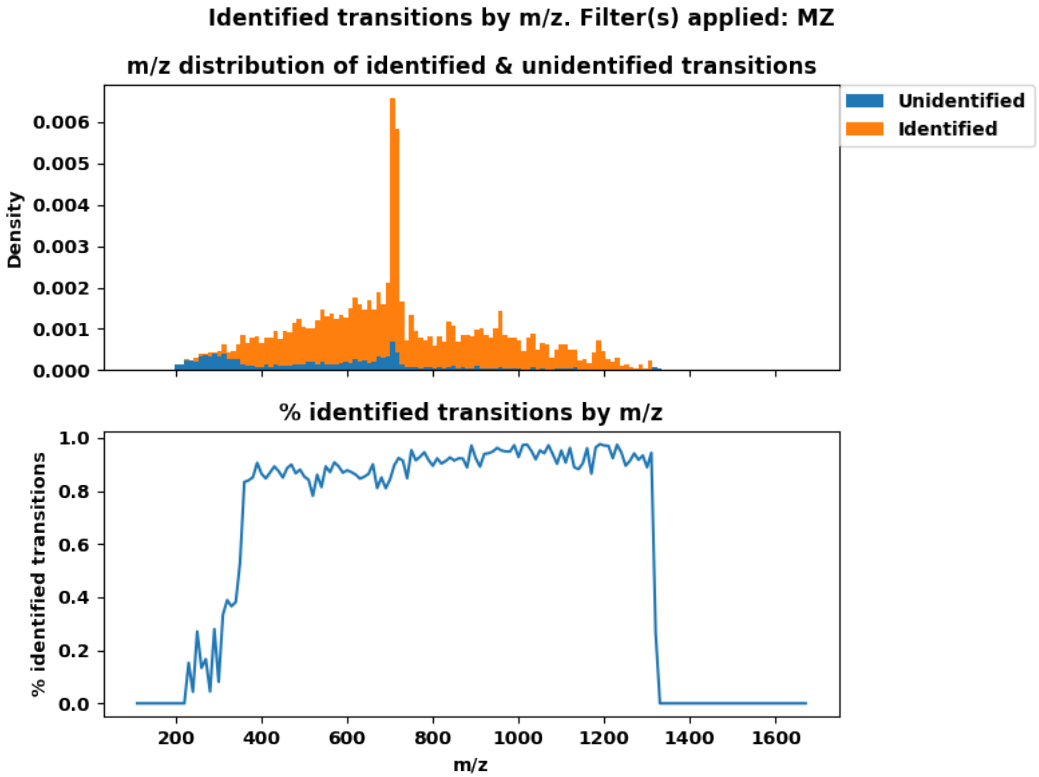
RT: 2688.63 - % identified peaks as a factor of **m/z**



RT: 2688.63 - % identified peaks as a factor of **m/z**

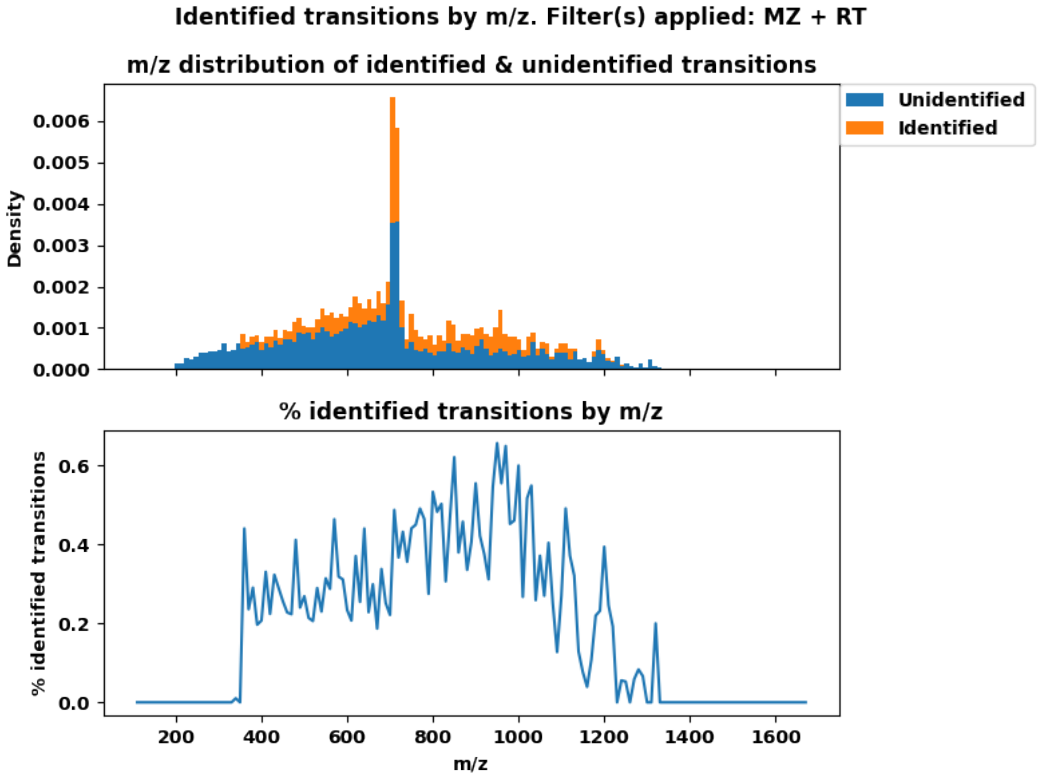


RT: 2688.63 - % identified peaks as a factor of **m/z**



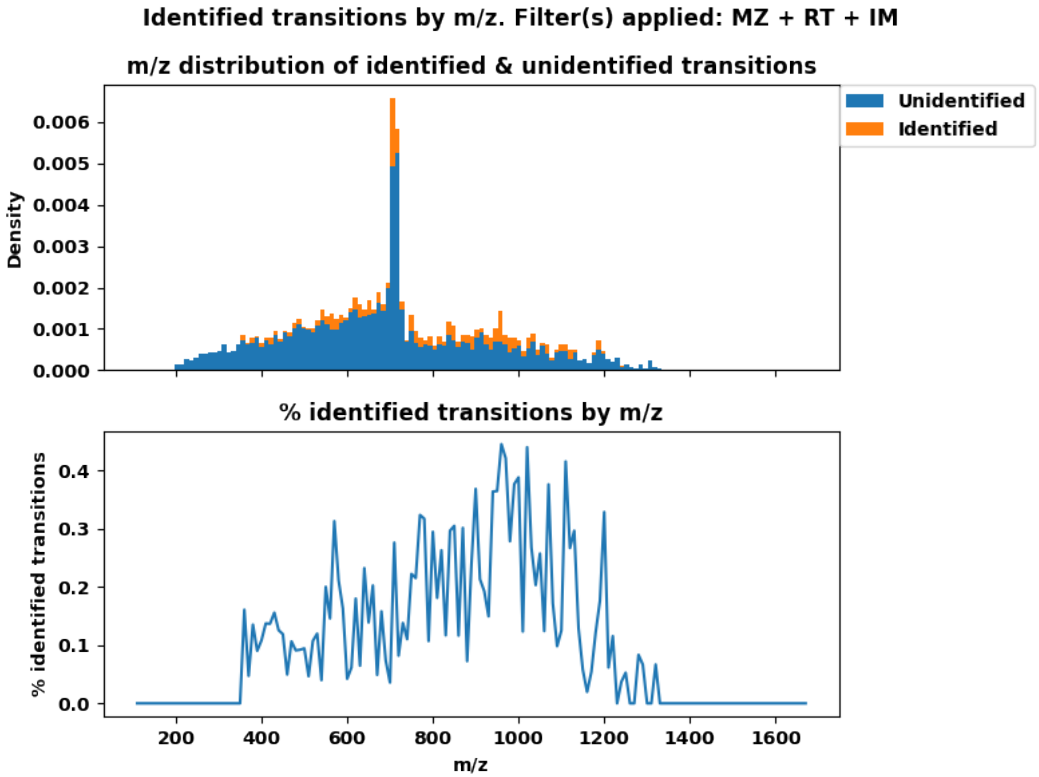
RT: 2688.63 - % identified peaks as a factor of **m/z**

RT: 2688.63 - % identified peaks as a factor of **m/z**



RT: 2688.63 - % identified peaks as a factor of **m/z**

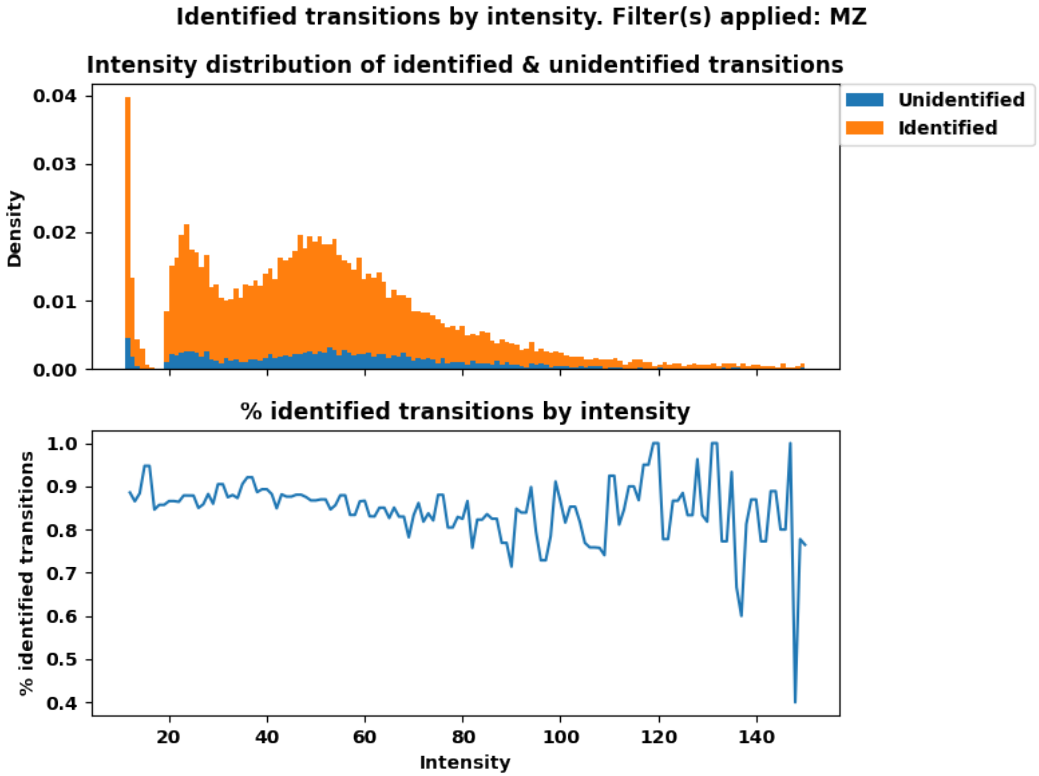
RT: 2688.63 - % identified peaks as a factor of **m/z**



RT: 2688.63 - % identified peaks as a factor of **m/z**

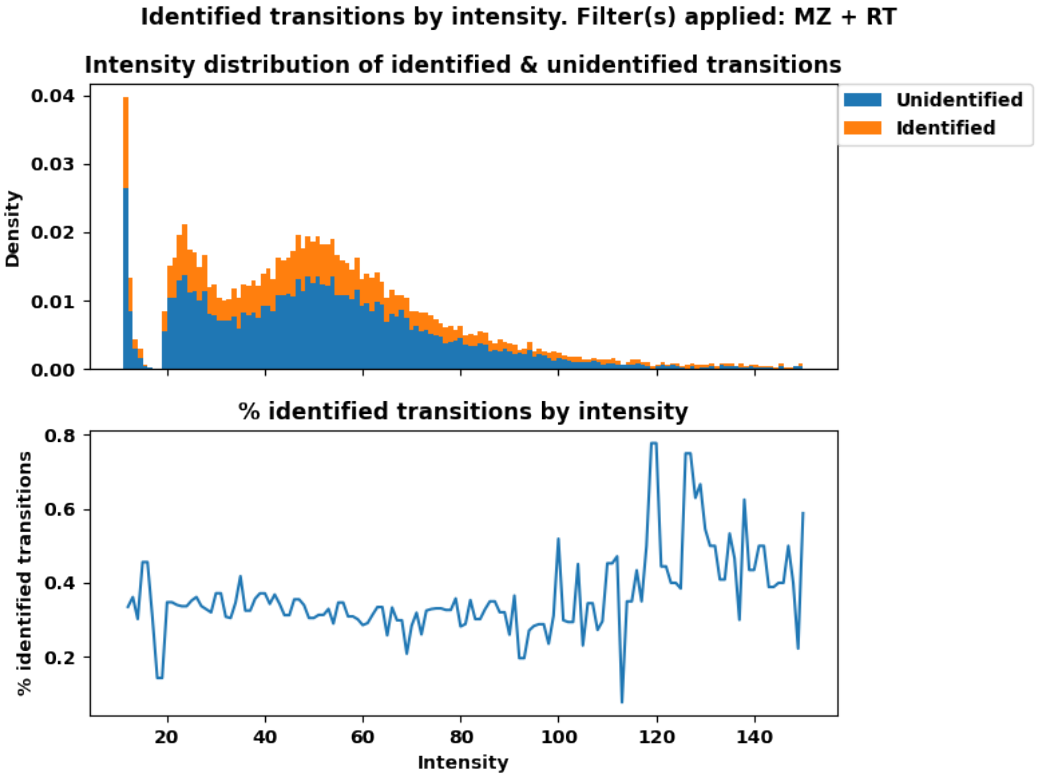
RT: 2688.63 - % identified peaks as a factor of **intensity**

RT: 2688.63 - % identified peaks as a factor of **intensity**



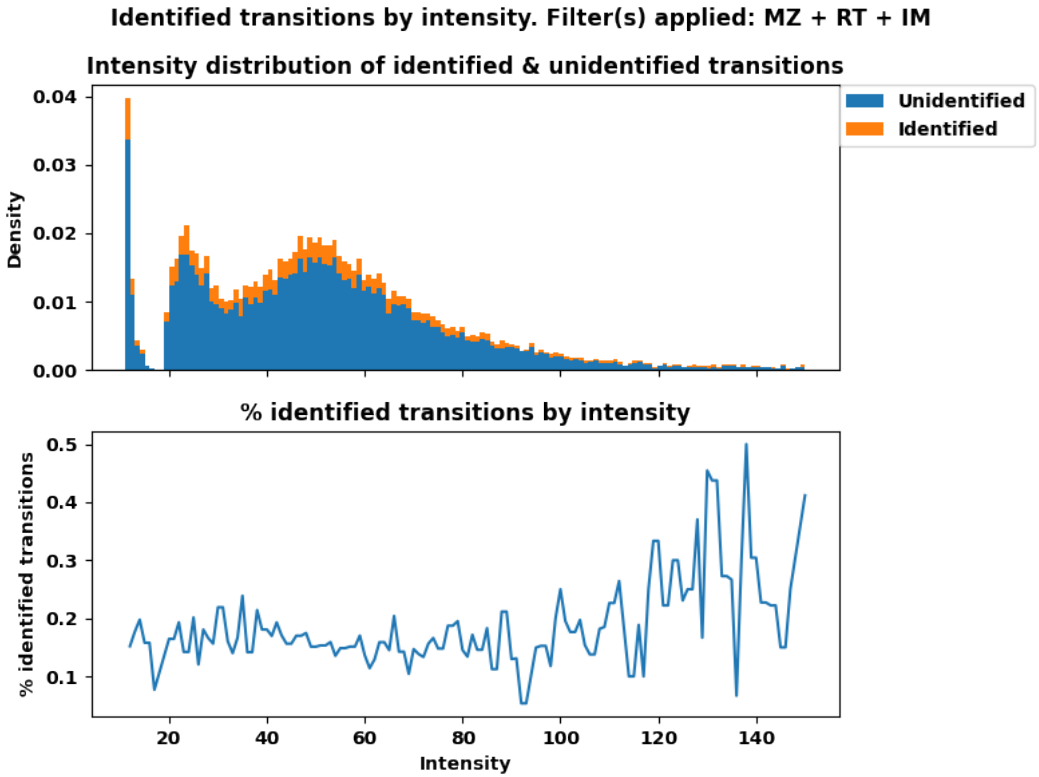
RT: 2688.63 - % identified peaks as a factor of **intensity**

RT: 2688.63 - % identified peaks as a factor of **intensity**



RT: 2688.63 - % identified peaks as a factor of **intensity**

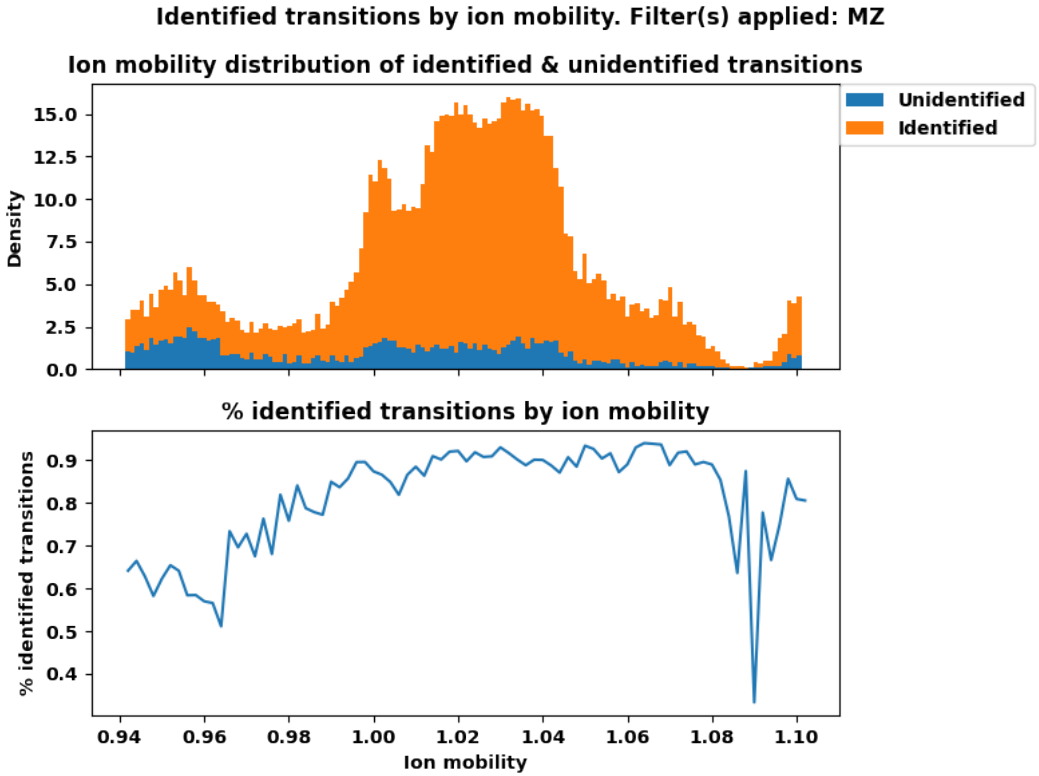
RT: 2688.63 - % identified peaks as a factor of **intensity**



RT: 2688.63 - % identified peaks as a factor of **intensity**

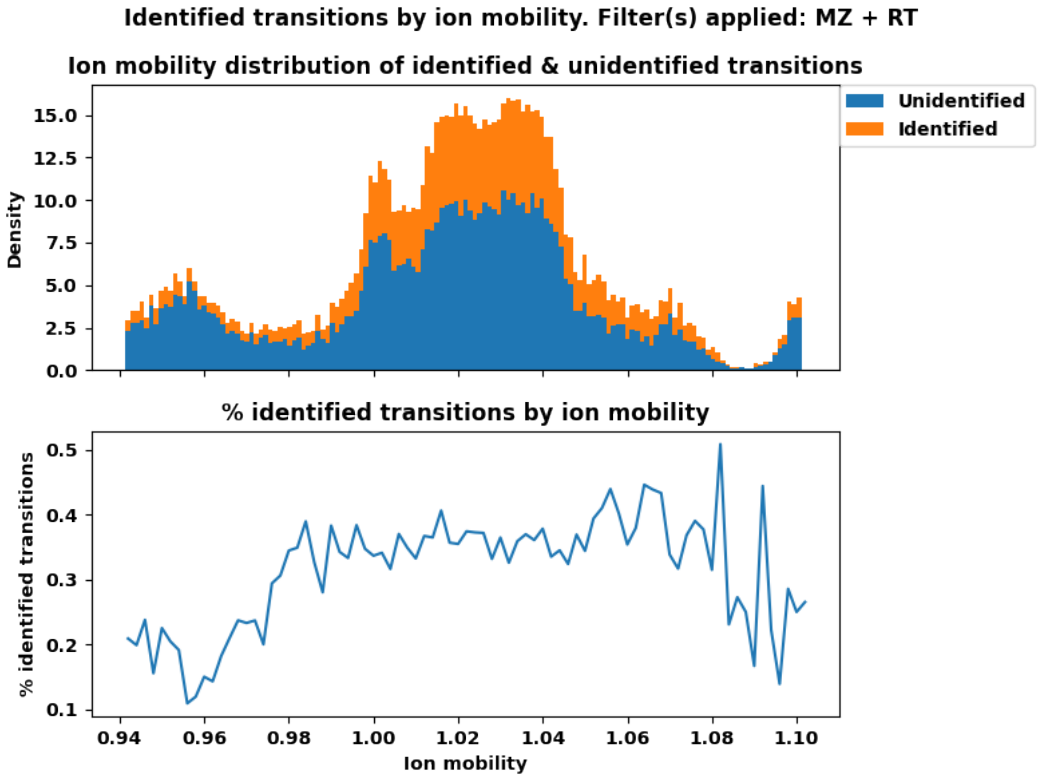
RT: 2688.63 - % identified peaks as a factor of **ion mobility**

RT: 2688.63 - % identified peaks as a factor of **ion mobility**



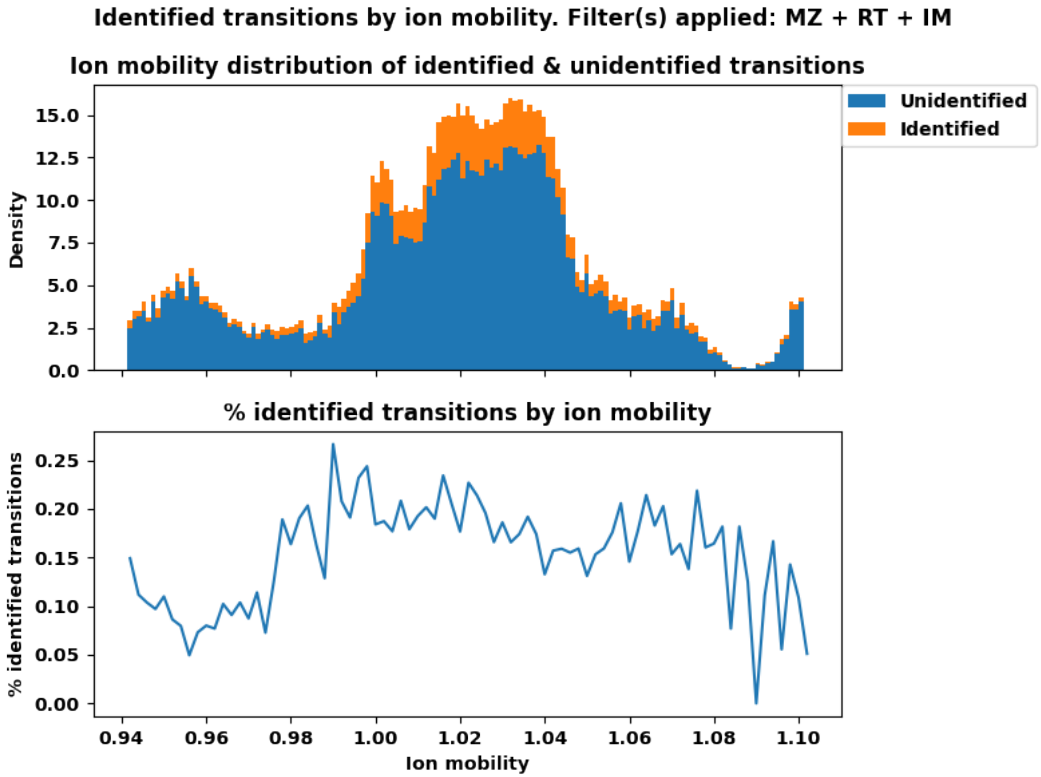
RT: 2688.63 - % identified peaks as a factor of **ion mobility**

RT: 2688.63 - % identified peaks as a factor of **ion mobility**



RT: 2688.63 - % identified peaks as a factor of **ion mobility**

RT: 2688.63 - % identified peaks as a factor of **ion mobility**



RT: 2688.63 - % identified peaks as a factor of **ion mobility**

Summary

- Join tables within **.osw** file
  - Isolate significant transitions (QVALUE < 0.01)
- Map filtered **.osw** table to raw transitions in **.mzML** file
- Mapped raw transitions: mean of significant transitions across RT 2400-2700:
  - m/z (+/- 20ppm) : **81.8%**
  - m/z + RT : **28.5%**
  - m/z + RT + IM (+/- 0.05) : **15.0%**
- Looking at one spectrum (RT: 2688.63)
  - Intensity and ion mobility appear to not influence OpenSWATH's selectivity.
  - m/z appears to influence selectivity
    - Raw transitions appear at 200 m/z, but significant transitions appear after 350 m/z

Fin