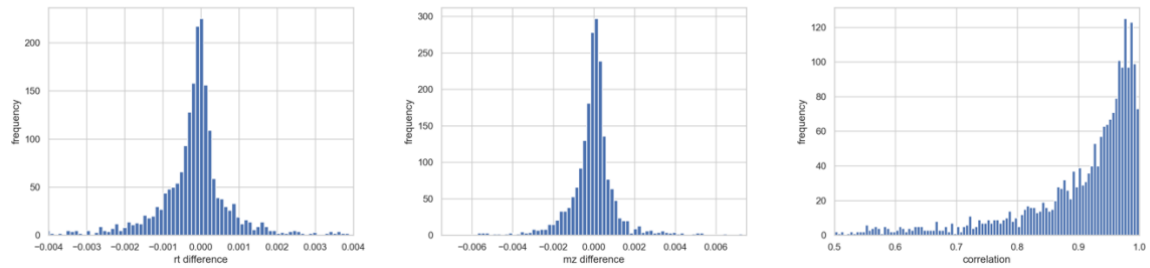
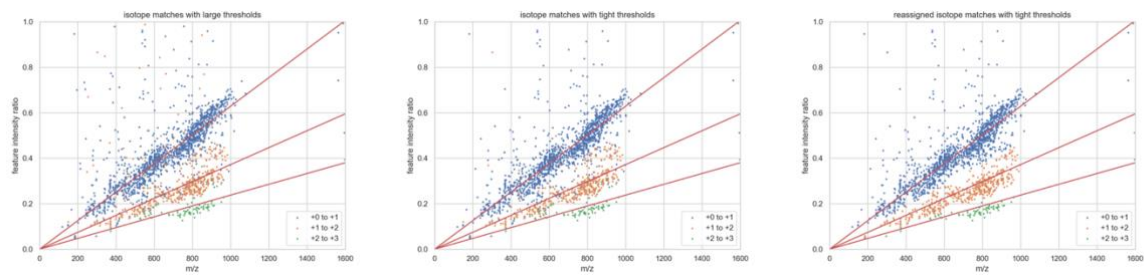


## ISOTOPE



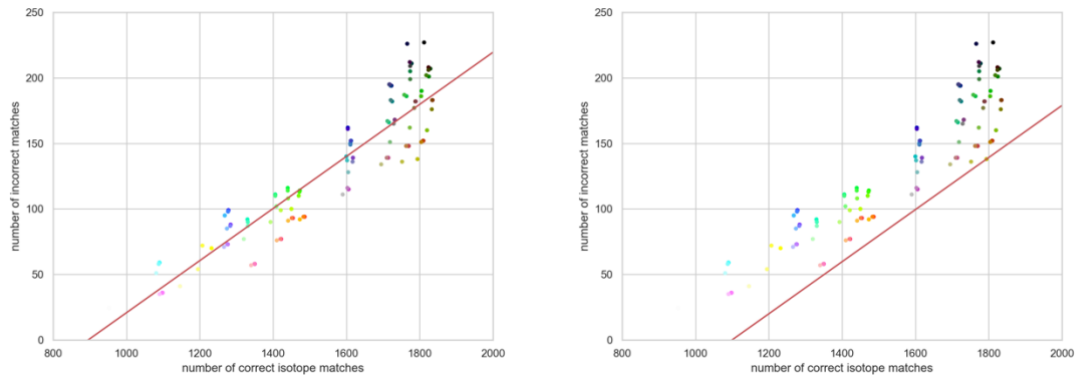
Intra-dataset distribution of rt difference, standardised mz difference and correlation for matched isotopes



Type of isotope matches with large rt, mz and correlation thresholds

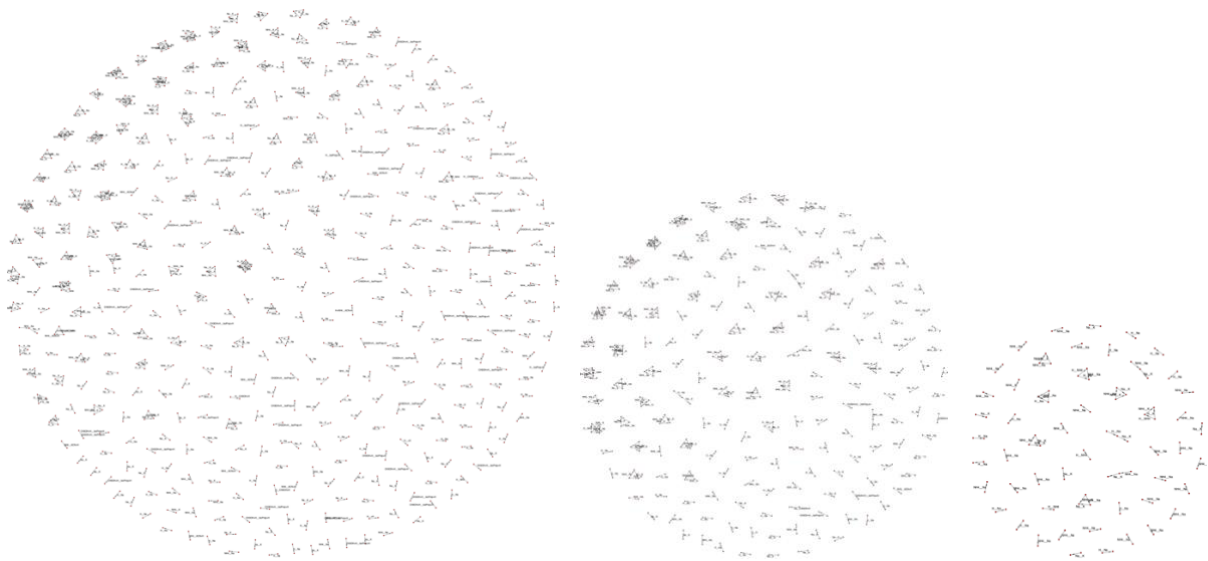
Type of isotope matches with tight rt, mz and correlation thresholds

Reassigned type of isotope matches with tight rt, mz and correlation thresholds

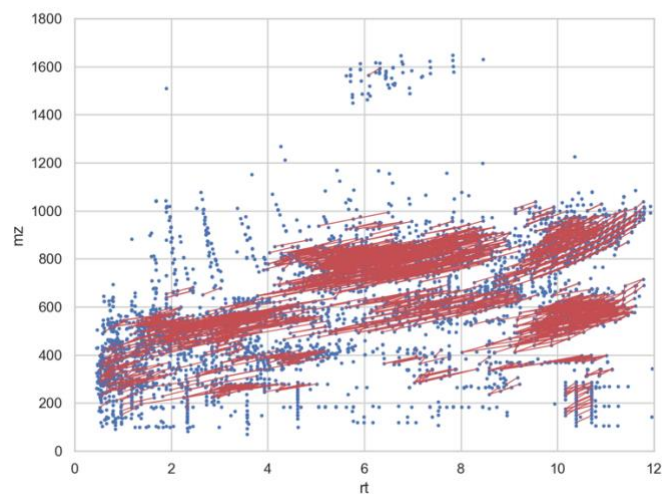


Linear regression to find the best set of thresholds. The colour of each point is RGB encoded thresholds.

## ADDUCT

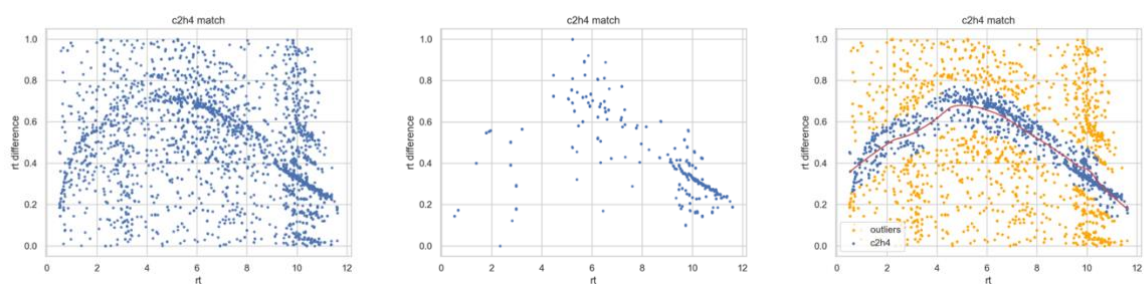


Intra-dataset adduct matches in isotope layer +0, +1, +2 respectively.



Intra-dataset C2H4 match with large thresholds of m/z and rt difference.

## C2H4



C2H4 match with same or unknown adduct type (left)

C2H4 match with same adduct type (right)

LOWESS regression to select high probability C2H4 matches

## WORKFLOW

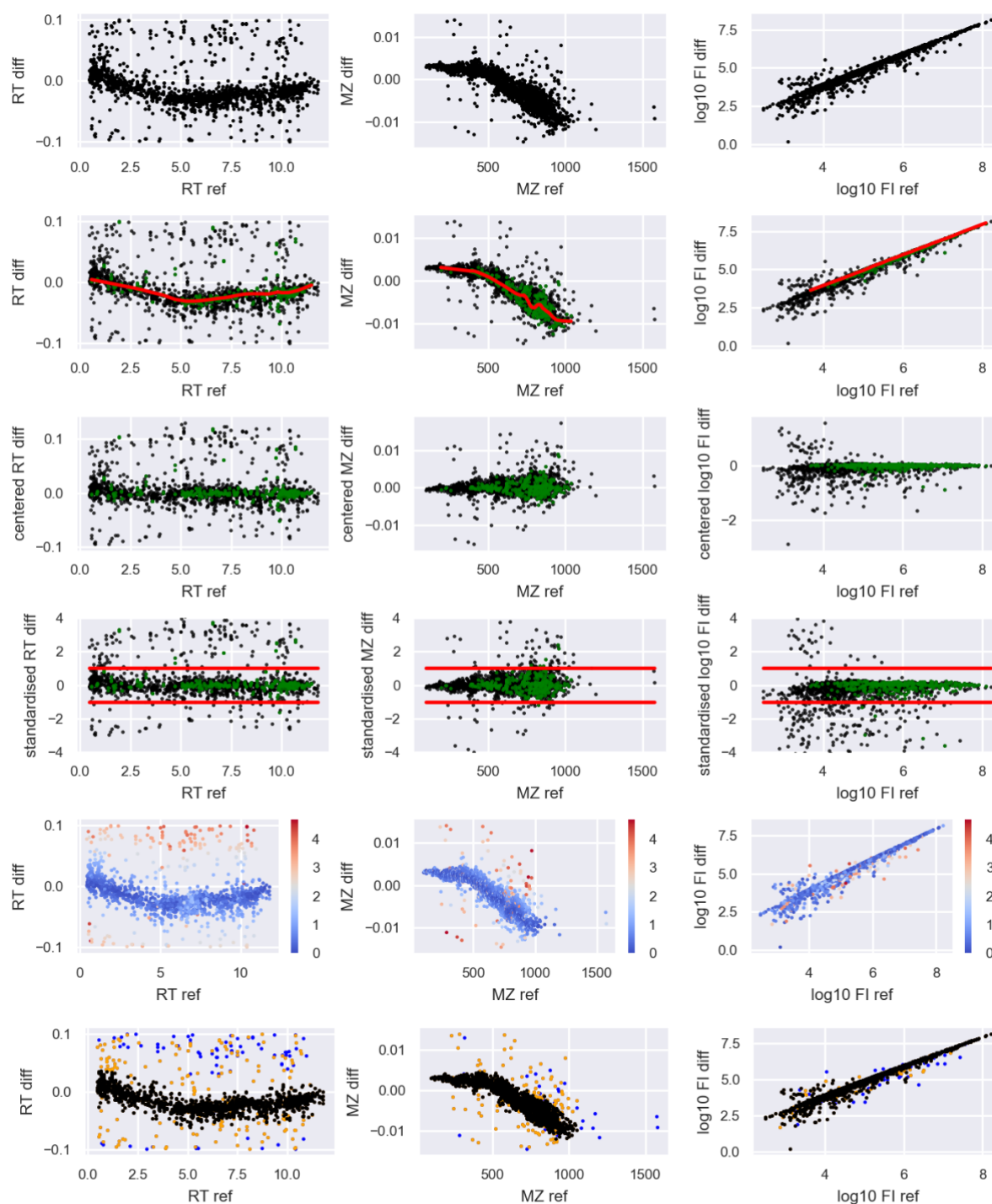


Figure 1: 1<sup>st</sup> row: all possible matches, inter-dataset difference vs reference value in RT, MZ and log10FI domains. 2<sup>nd</sup> row: high probability matches in green and other matches in black, red line is LOWESS regression using only green dots. 3<sup>rd</sup> row: centred difference. 4<sup>th</sup> row: standardised difference by dividing 5MAD. 5<sup>th</sup> row: penalisation score for each match calculated by weighted sum of squared standardised difference of RT, MZ and log10FI. 6<sup>th</sup> row: multiple match in blue, poor match in orange and good match in black.