Supporting information for: Cluster-wise peak detection and filtering based on spatial distribution to efficiently mine mass spectrometry imaging data

Jonatan O. Eriksson,[†] Melinda Rezeli,[†] Max Hefner,[†] Gyorgy Marko-Varga,[†] and Peter Horvatovich[*,‡,†]

†*Lund University, Department of Biomedical Engineering, Lund*

‡*University of Groningen, Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, Antonius Deusinglaan 1, 9713 AV, Groningen, the Netherlands*

E-mail: p.l.horvatovich@rug.nl

Peak picking algorithm

1. For each spectrum in the data set append the m/z-values of the centroids of its peaks to a global m/z list applying only a very gentle threshold on minimum peak intensity. This will result in a master m/z list mz_{all} containing many million mz-values. Sort mz_{all}.

2. Identify m/z-clusters in mz_{all} using a a one-directional graph approach where a m/z-value is connected to the next consecutive value if there is a non-zero value within a distance, d_c, proportional to the theoretical peak width of the instrument (this threshold varies across the m/z-range according to the theoretical peak width of the mass analyzer).

3. Investigate the distribution of the peak m/z values within each cluster with Kernel Density Estimation (KDE). The kernel bandwidth is optimized (using MATLAB's *ksdensity* function) for each cluster.

4. Identify local maxima on the KDE curve and add the corresponding m/z values to the data set peak list, mz_{ref}: given the list of local maxima, p_{kde}, with all local maxima on the KDE curve: (i) take the highest maximum (the most dense signal) mz_{max} and add it to mz_{ref} (ii) remove p_{max} and all other maxima within distance $\Delta m/z$ from mz_{max} from p_{kde}, repeat (i) - (ii) until p_{kde} is empty or mz_{max} is below a threshold proportional to the initial mz_{max}, .

5. Align each spectrum to the reference spectrum, m_{ref} to generate the ion images, here we use a simple nearest neighbor alignment method but other methods can be used as well. Given the i-th centroided spectrum, s_i, and aligned spectrum, sa_i, for each mz_j in mz_{ref} find the closest m/z value in s_i, if this value is within a maximum distance set $sa_{i(j)}$ to its corresponding intensity value, if not, set $sa_{i(j)}$ to zero.

6. Remove ion images with too few non-zero pixels.

7. Compute scores based on spatial characteristics for the remaining ion images, i.e. SC from Palmer et al. or VE from Fonville et al. Remove ion image whose scores are below user defined thresholds.

Mass drift

Table S1: Mass drift

Compound	Mass	Mass Error NAME [ppm]	Mass Error Cardinal [ppm]
Ipratropium	332.223	4.29	28.87
Vatalanib	347.107	2.16	194.88
Erlotinib	394.177	3.39	164.73
Sunitinib	399.220	2.40	NA
Pazopanib	438.171	2.68	26.06
Gefitinib	447.160	3.73	NA
Sorafenib	465.094	3.36	4.23
Dasatinib	488.164	2.23	14.06
Imatinib	494.267	1.61	NA
Dabrafinib	520.109	2.30	198.85
Lapatinib	581.143	1.00	40.18
Trametinib	616.086	2.03	NA

Spiked rat liver section

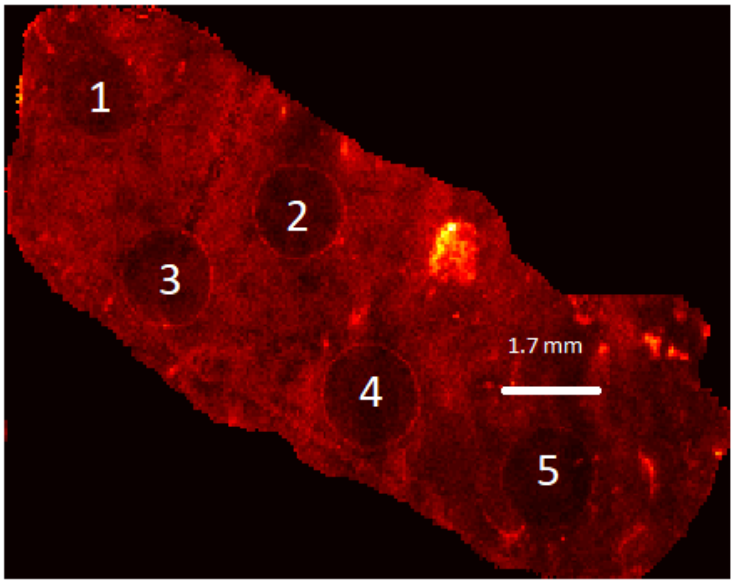


Figure S1: A 181 x 247 pixel ion image the rat liver section at $m/z = 186.0750$. The five spiked-in spots are visible against the background signal from the tissue.

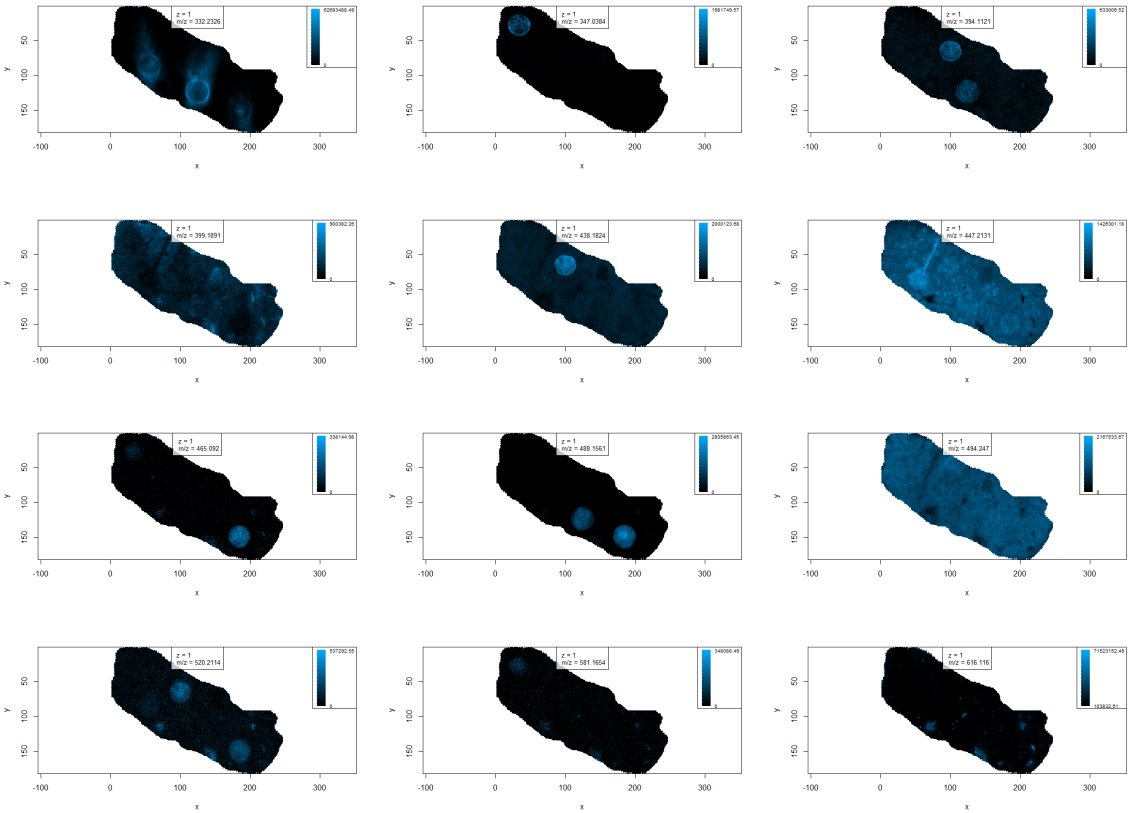
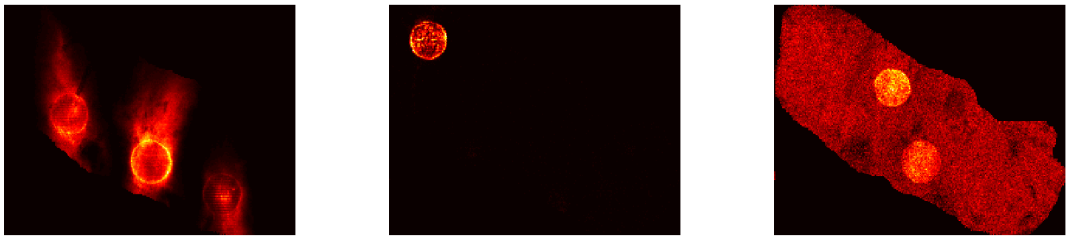
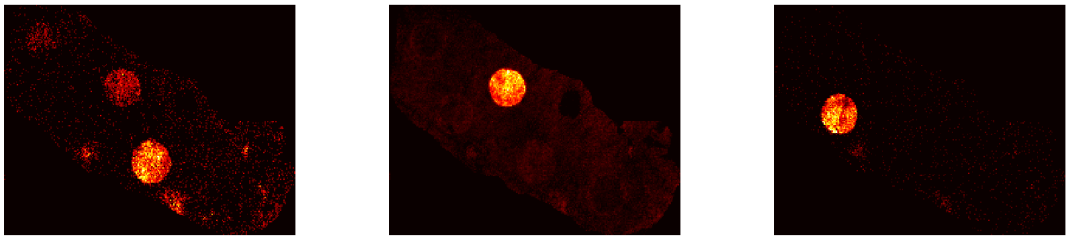


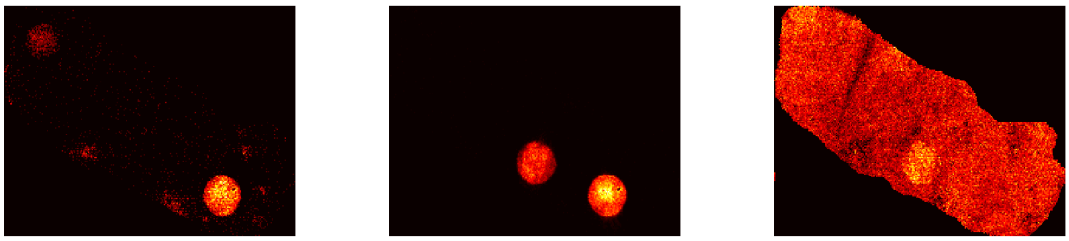
Figure S2: Images of spiked-in drug compounds generated with Cardinal.



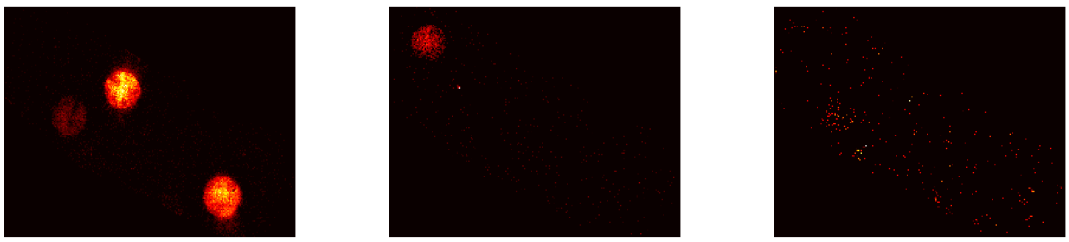
(a) m/z 332.1750 (b) m/z 347.0750 (c) m/z 394.1250



(d) m/z 399.1750 (e) m/z 438.1250 (f) m/z 447.1250

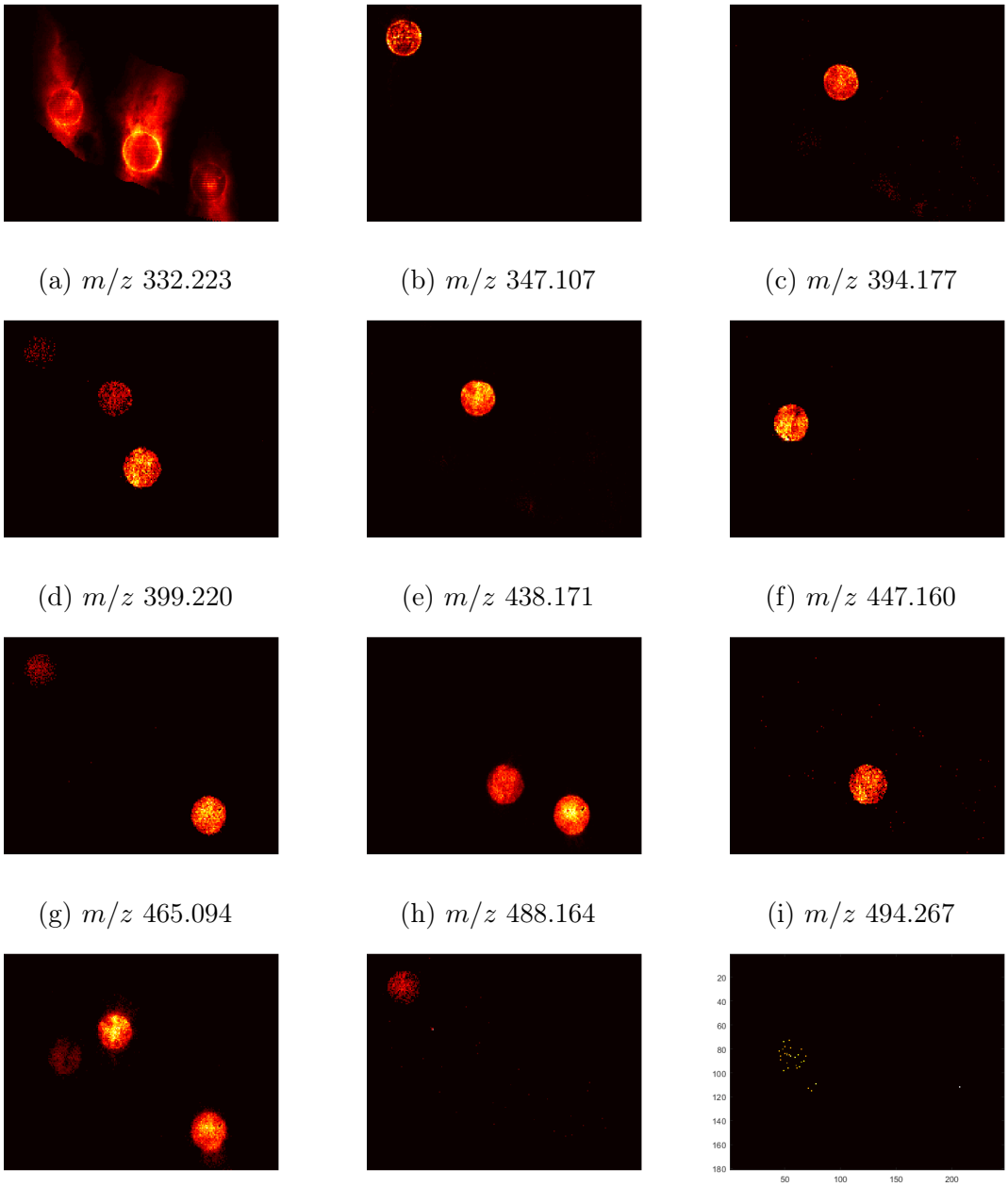


(g) m/z 465.0250 (h) m/z 488.1250 (i) m/z 494.2250



(j) m/z 520.0750 (k) m/z 581.0750 (l) m/z 616.0250

Figure S3: Images of spiked-in drug compounds generated by slicing.



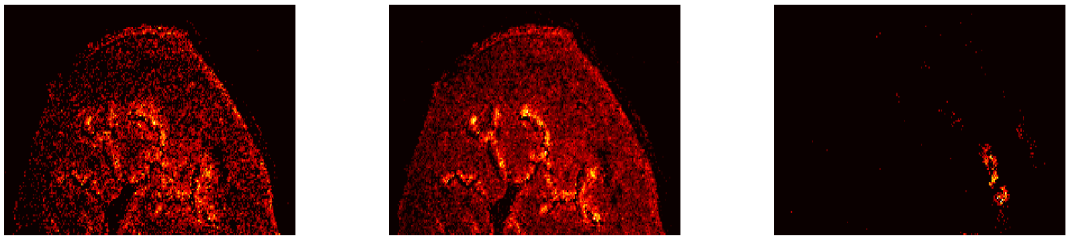
(a) m/z 332.223 (b) m/z 347.107 (c) m/z 394.177

(d) m/z 399.220 (e) m/z 438.171 (f) m/z 447.160

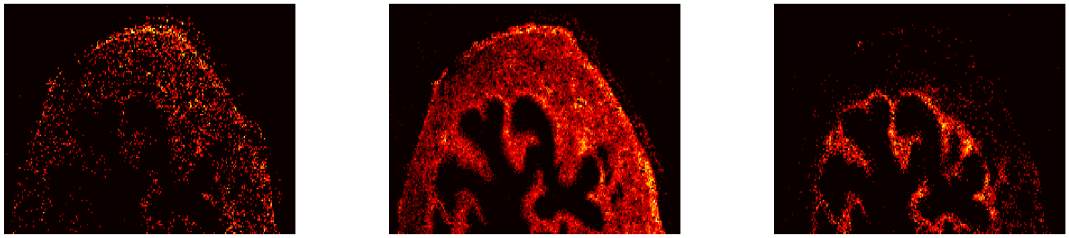
(g) m/z 465.094 (h) m/z 488.164 (i) m/z 494.267

(j) m/z 520.109 (k) m/z 581.143 (l) m/z 616.086

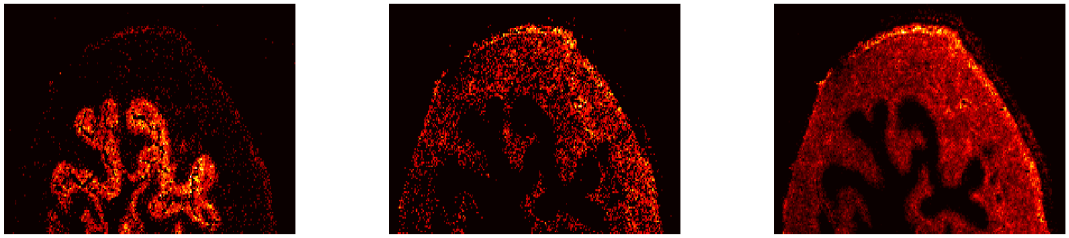
Figure S4: Images of the spiked-in drug compounds generated using KDE cluster-wise peak picking.



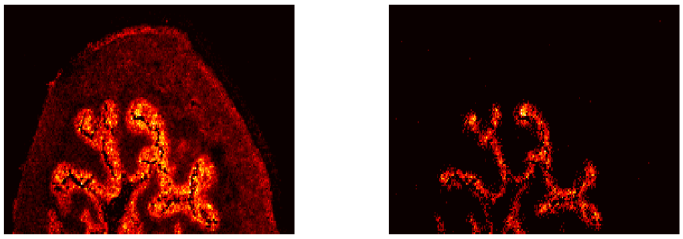
(a) m/z 535.2957 (b) m/z 562.3270 (c) m/z 616.1767



(d) m/z 713.4518 (e) m/z 742.5307 (f) m/z 743.5482



(g) m/z 770.5068 (h) m/z 770.5653 (i) m/z 772.5253



(j) m/z 798.5410 (k) m/z 812.5566

Figure S5: Ion-images of known compounds in the mouse bladder data set KDE cluster-wise peak picking.

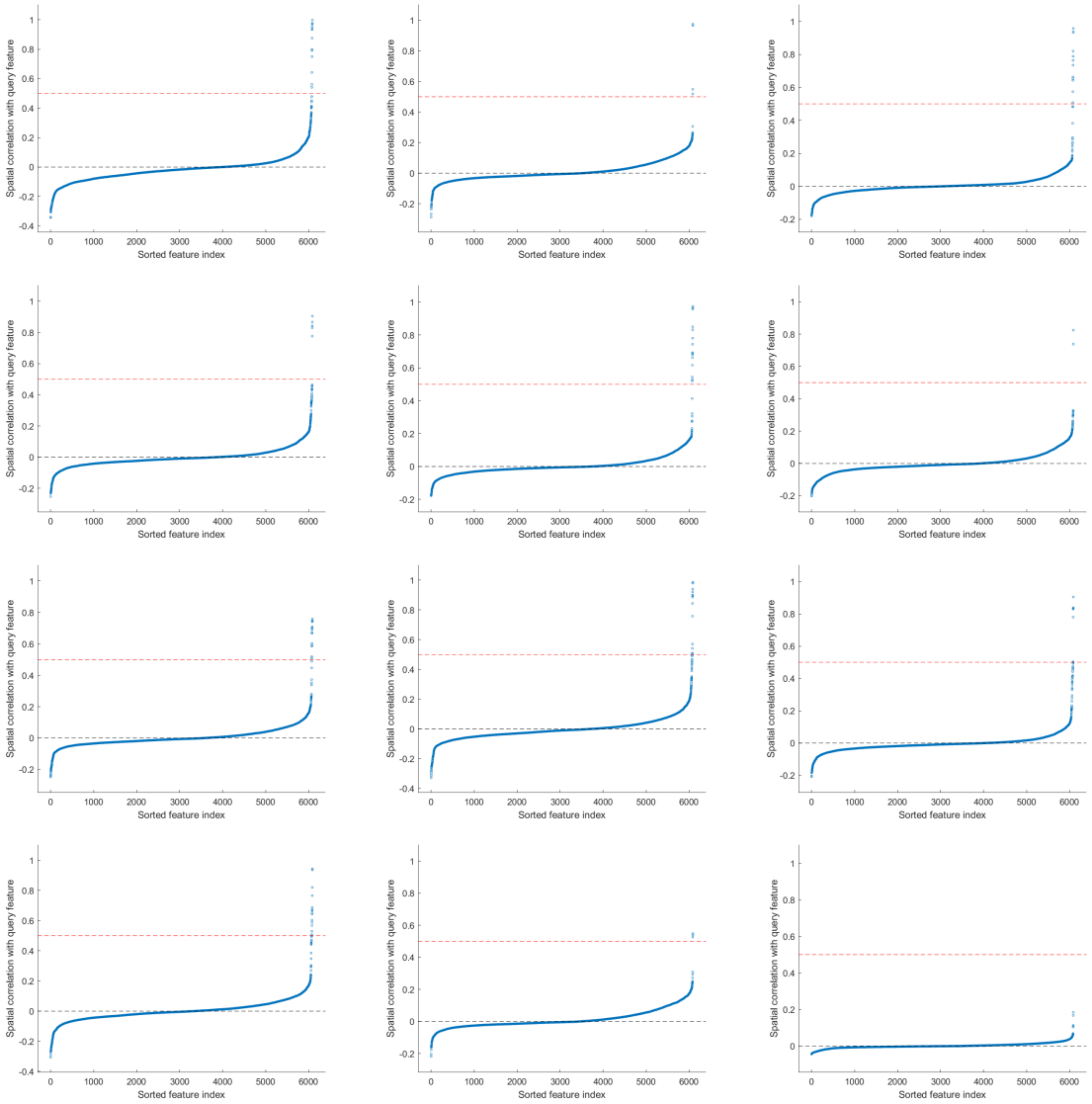


Figure S6: Correlation queries for spiked-in drug compounds using the images generated by our cluster-wise peak detection.

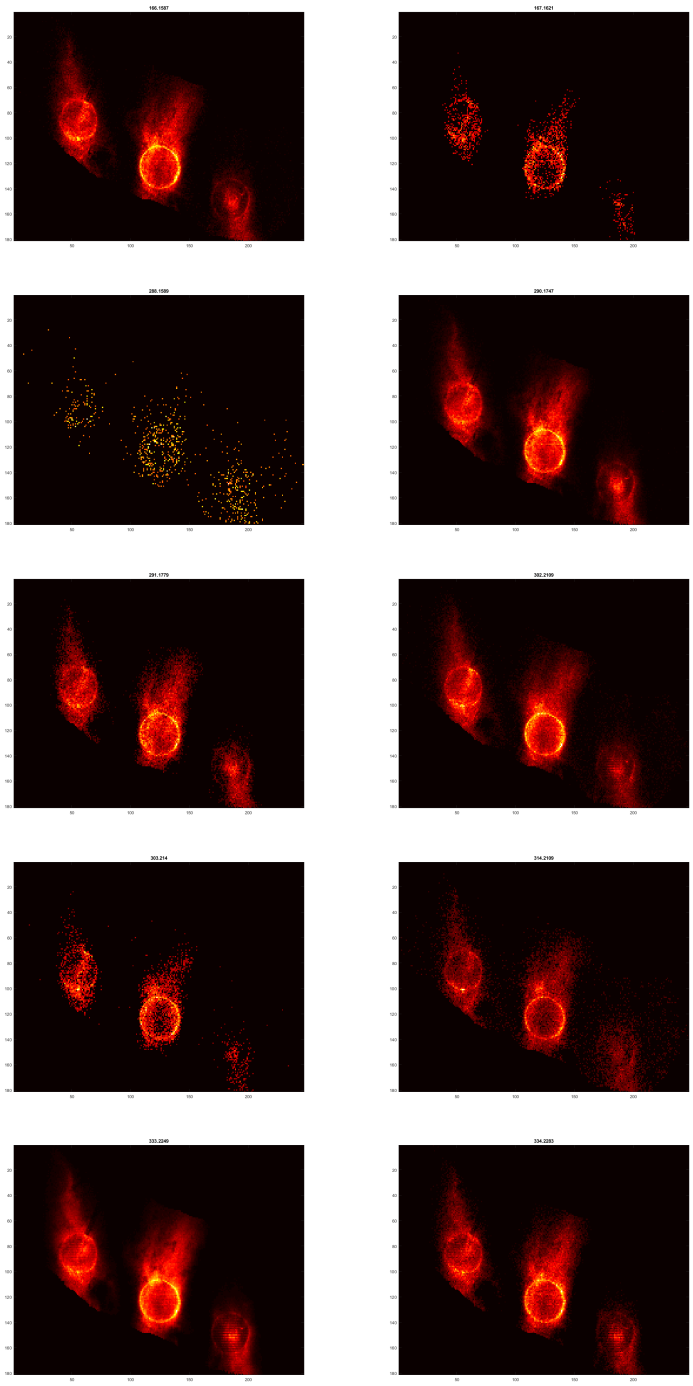


Figure S7: Fragments and isotopes for $m/z = 332.223$.

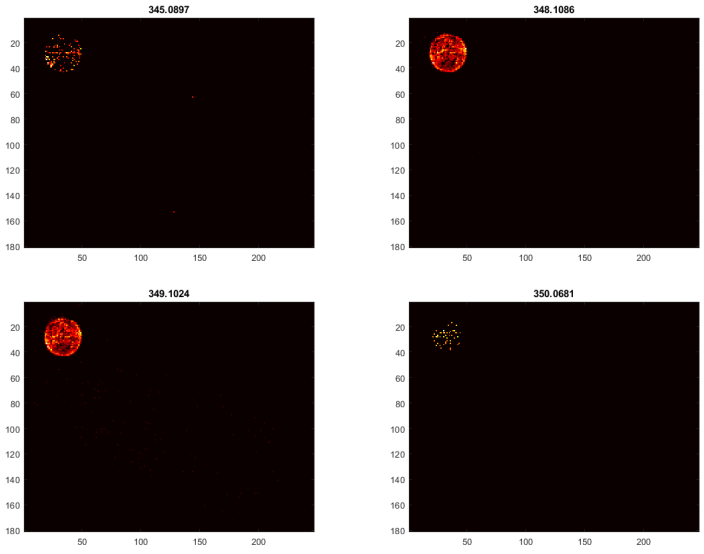


Figure S8: Fragments and isotopes for $m/z = 347.107$.

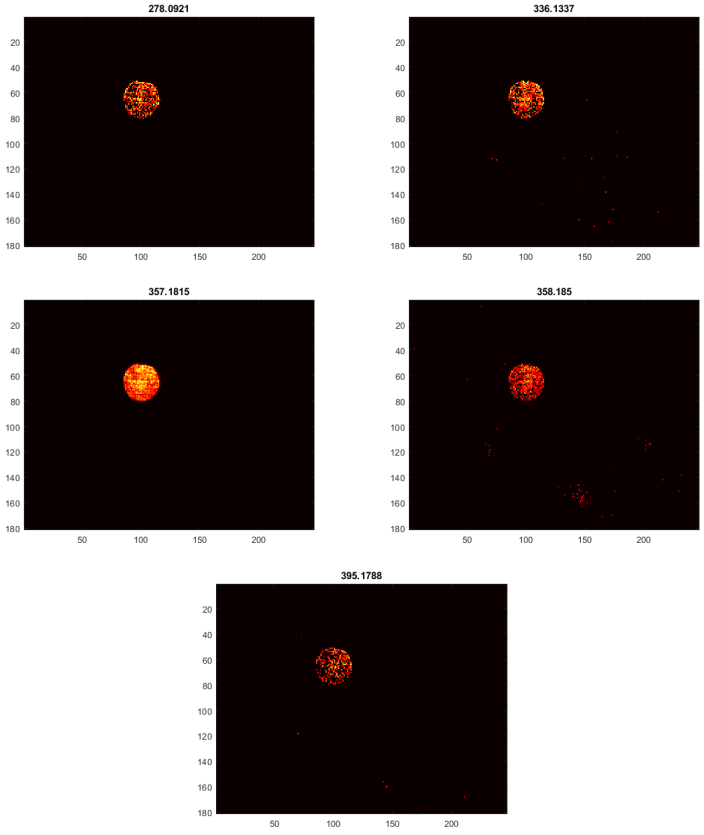


Figure S9: Fragments and isotopes for $m/z = 394.107$.

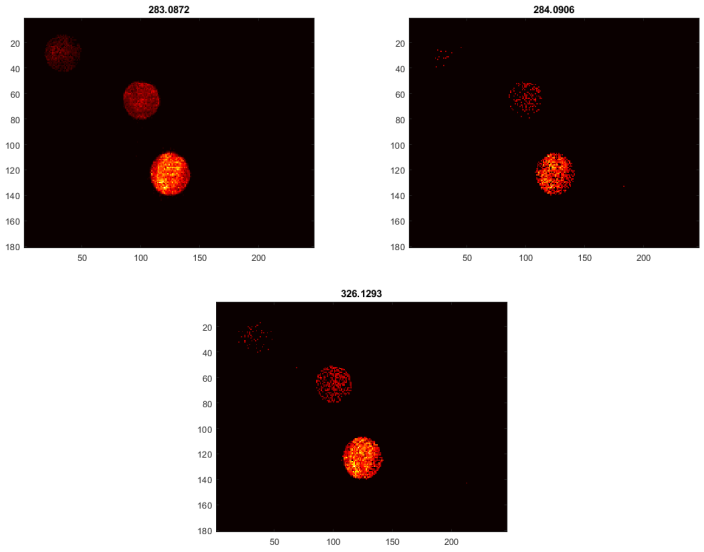


Figure S10: Fragments and isotopes for $m/z = 399.107$.

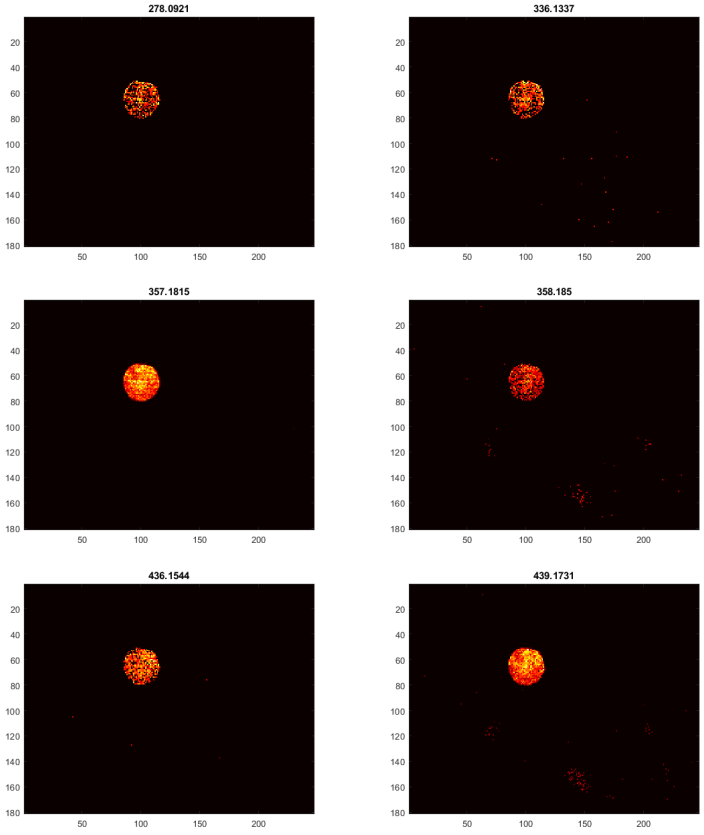


Figure S11: Fragments and isotopes for $m/z = 438.107$.

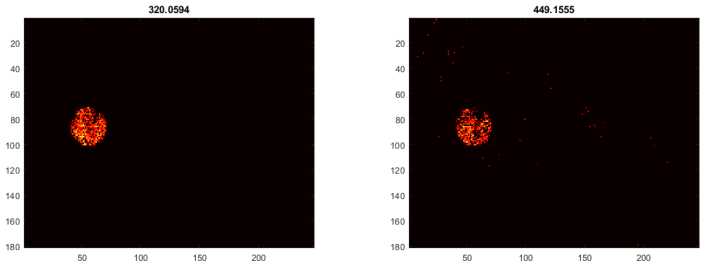


Figure S12: Fragments and isotopes for $m/z = 447.107$.

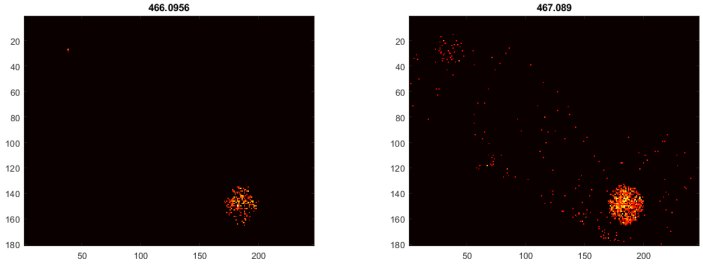


Figure S13: Fragments and isotopes for $m/z = 465.107$.

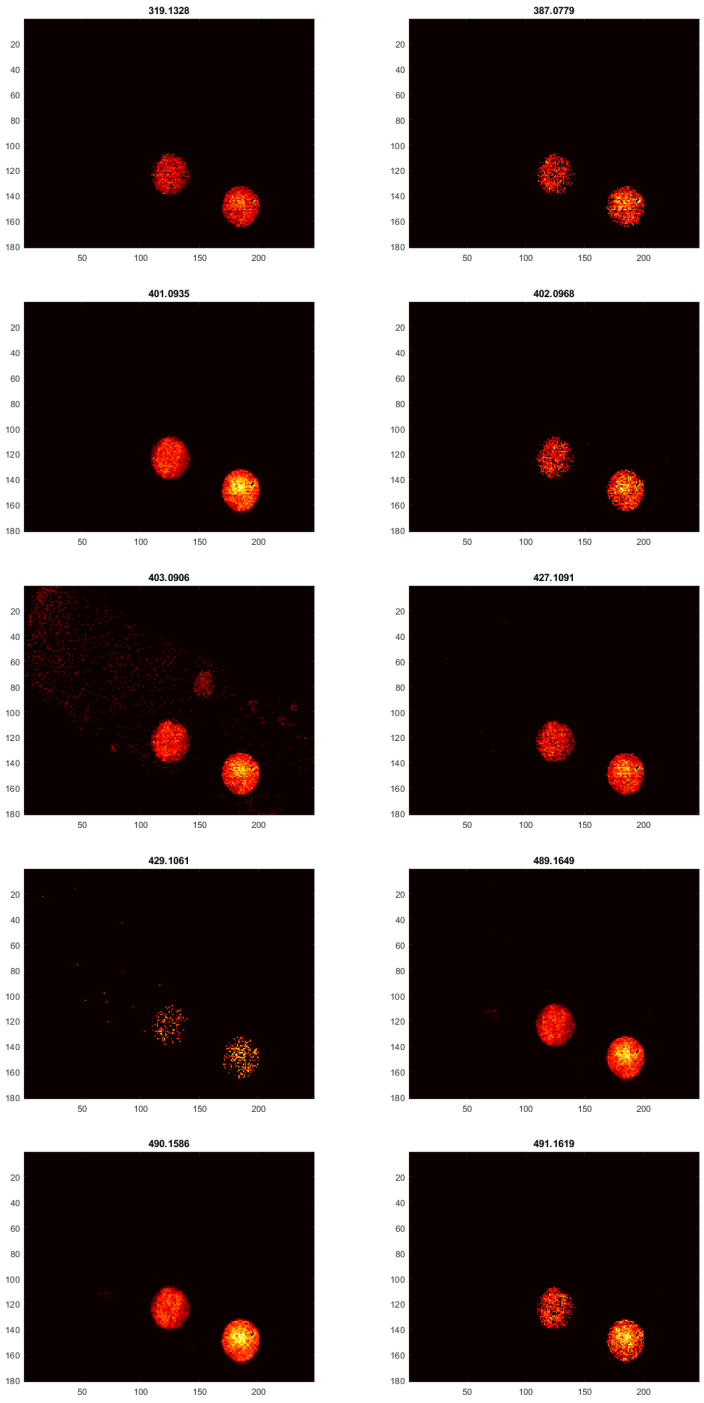


Figure S14: Fragments and isotopes for $m/z = 488.223$.

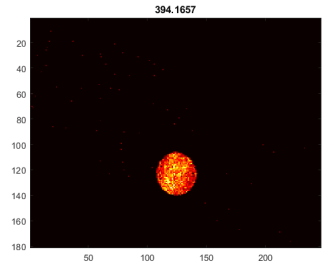


Figure S15: Fragments and isotopes for $m/z = 494.223$.

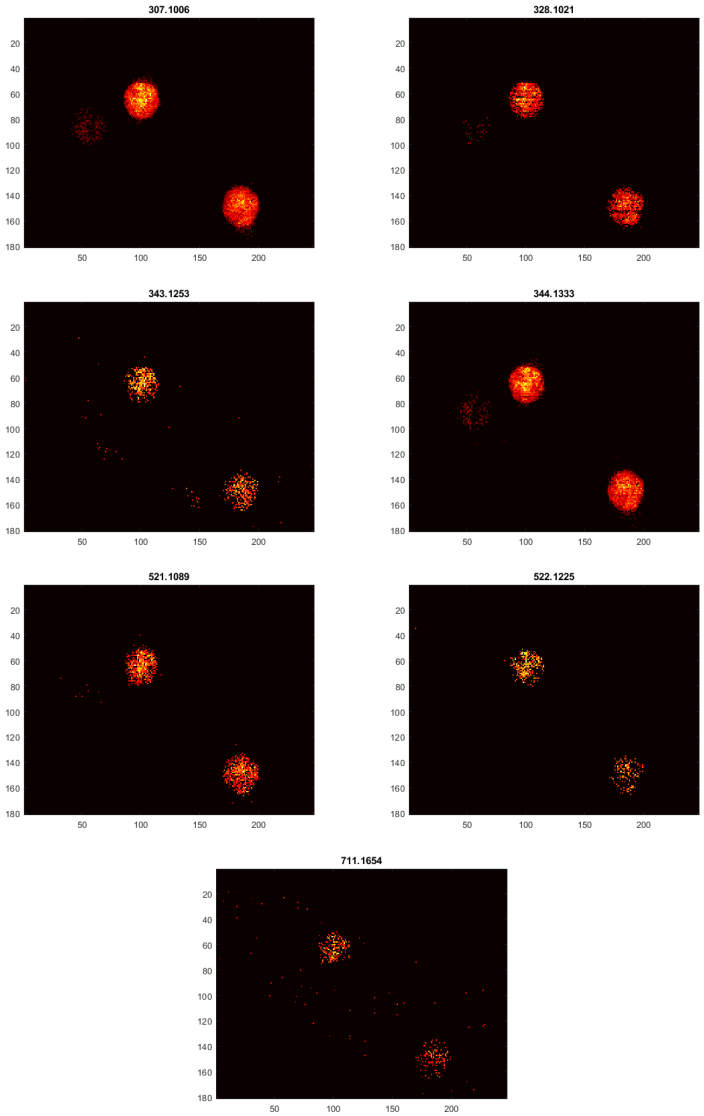


Figure S16: Fragments and isotopes for $m/z = 520.223$.