Introduction to XCMS in R

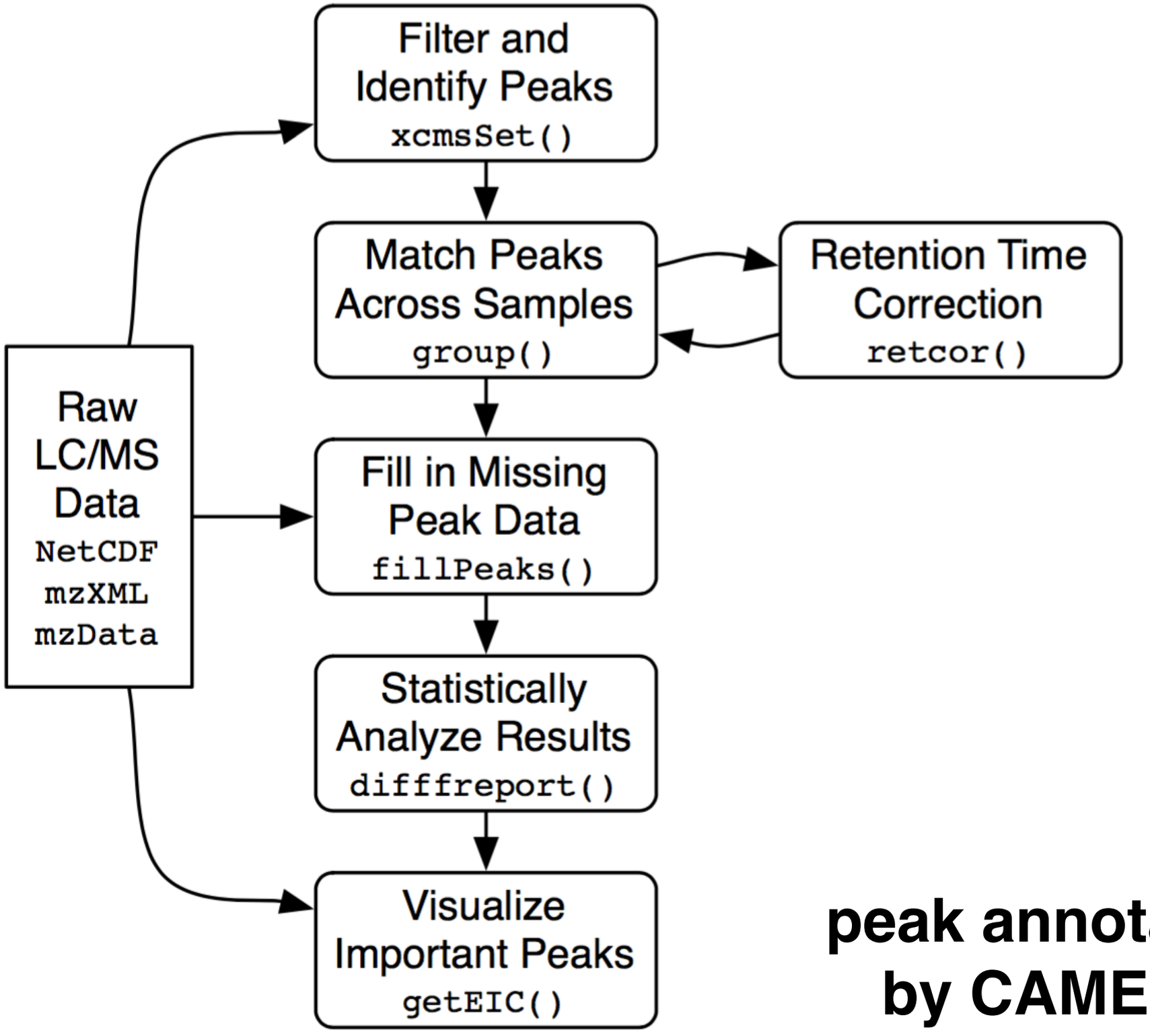
Xiuxia Du, Ph.D.
Department of Bioinformatics and Genomics
University of North Carolina at Charlotte

XCMS in R

- Used for preprocessing untargeted LC-MS and GC-MS metabolomics data.
- Primary related publications:
  - Smith, C.A., Want, E.J., O'Maille, G., Abagyan,R., Siuzdak and G. (2006). "XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification." Analytical Chemistry, 78, pp. 779–787.
  - Tautenhahn R, Boettcher C and Neumann S (2008). "Highly sensitive feature detection for high resolution LC/MS." BMC Bioinformatics, 9, pp. 504.
  - Benton HP, Want EJ, Ebbels TMD (2010). "Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data." BIOINFORMATICS, 26, 2488.
- Information about XCMS R package can be found at

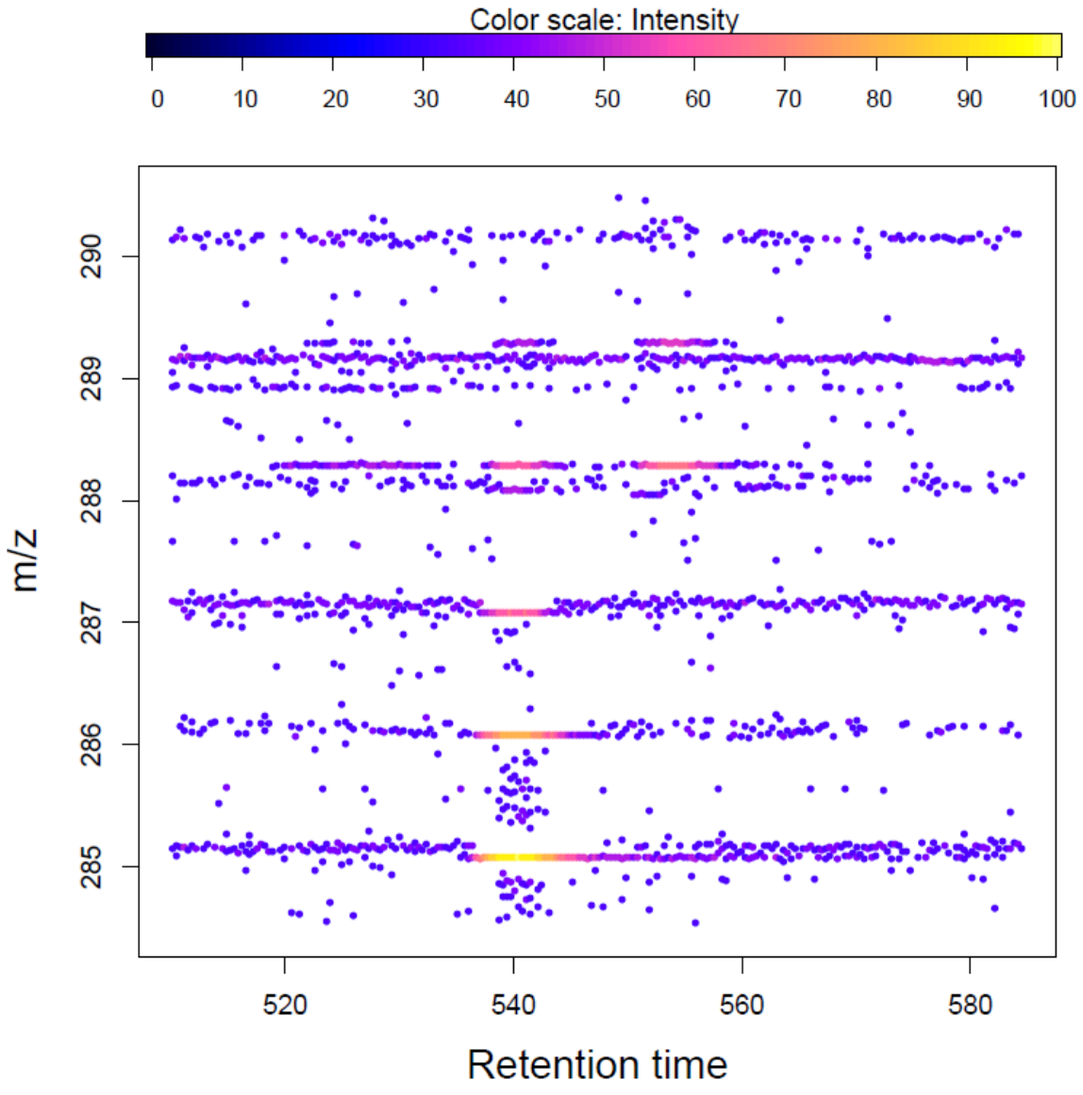
  `https://bioconductor.org/packages/release/bioc/html/xcms.html`

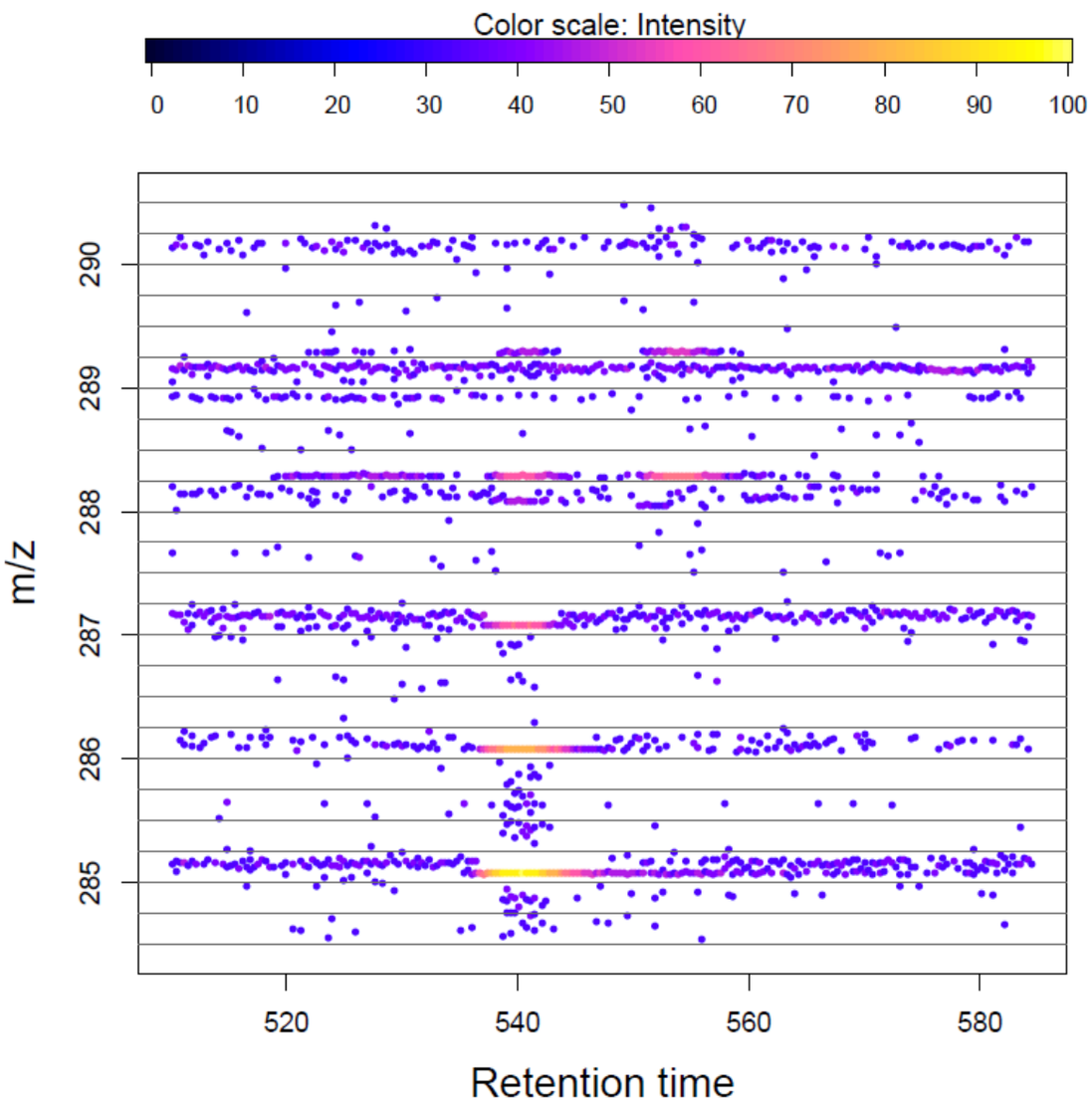
Preprocessing steps



peak annotation by CAMERA

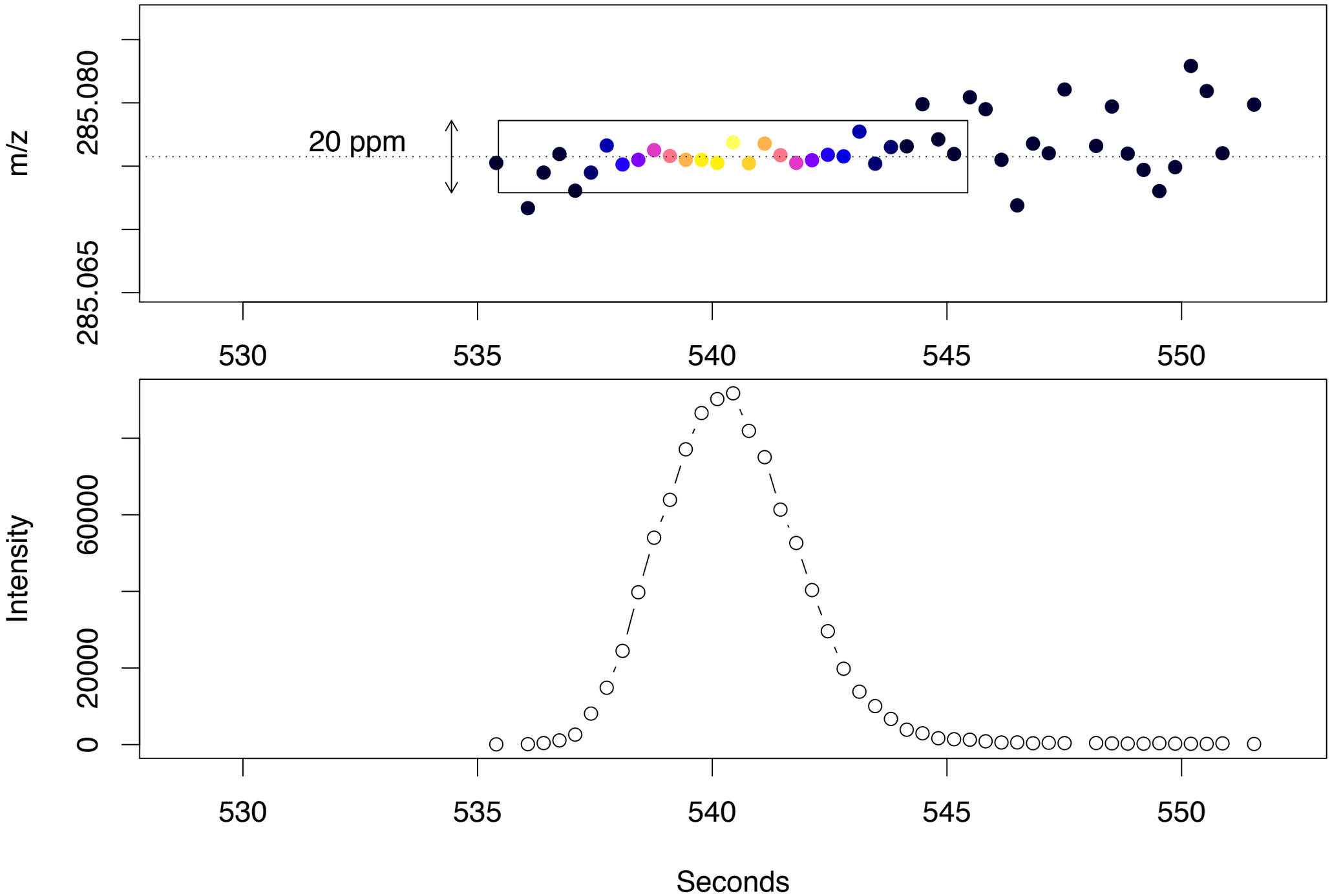
Construct EICs





Tautenhahn, R.; Bottcher, C.; Neumann, S., Highly sensitive feature detection for high resolution LC/MS. *BMC bioinformatics* **2008,** *9*, 504.

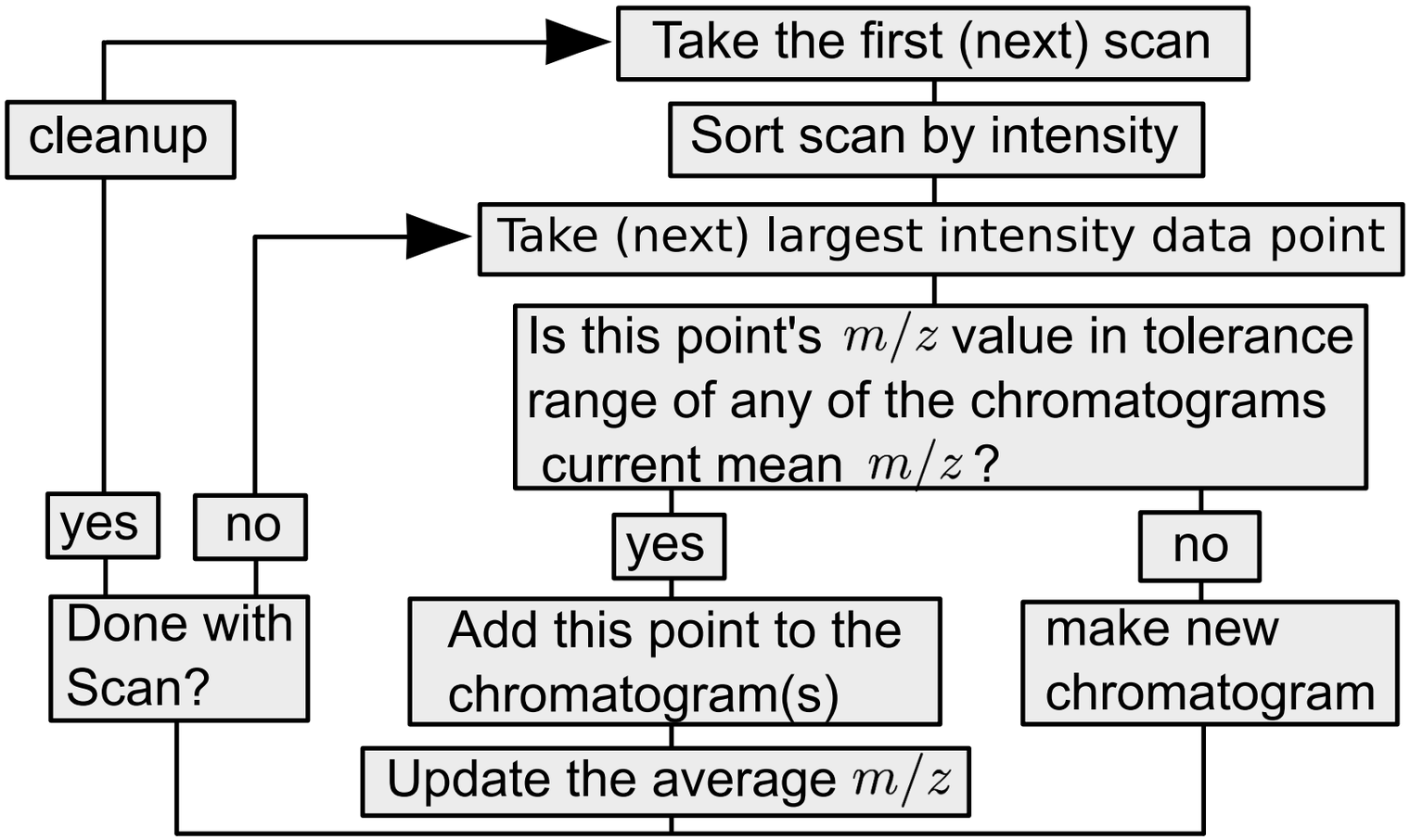
Construct EICs



Tautenhahn, R.; Bottcher, C.; Neumann, S., Highly sensitive feature detection for high resolution LC/MS. *BMC bioinformatics* **2008,** *9*, 504.

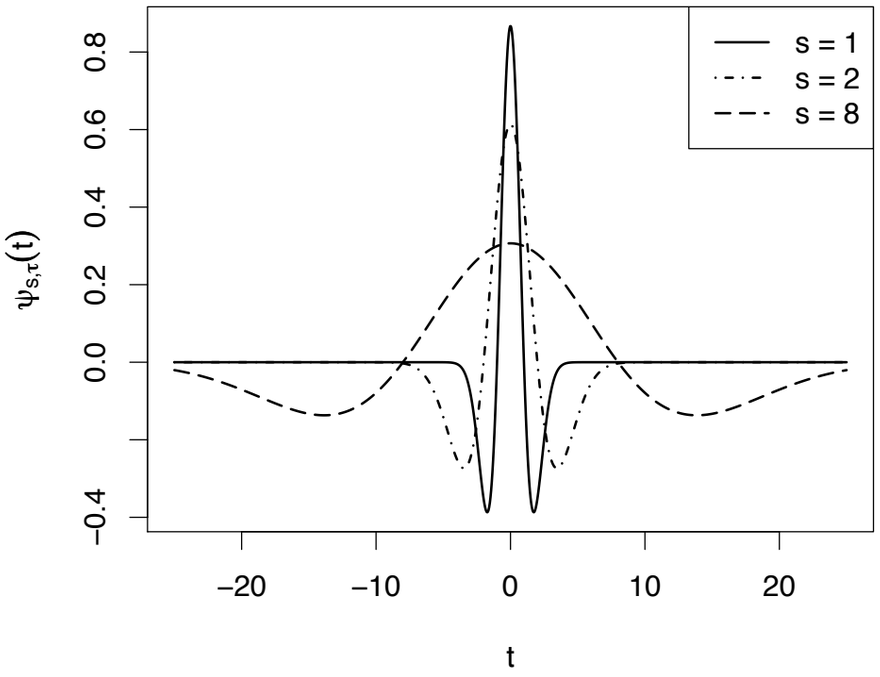
Construct EICs

- Workflow

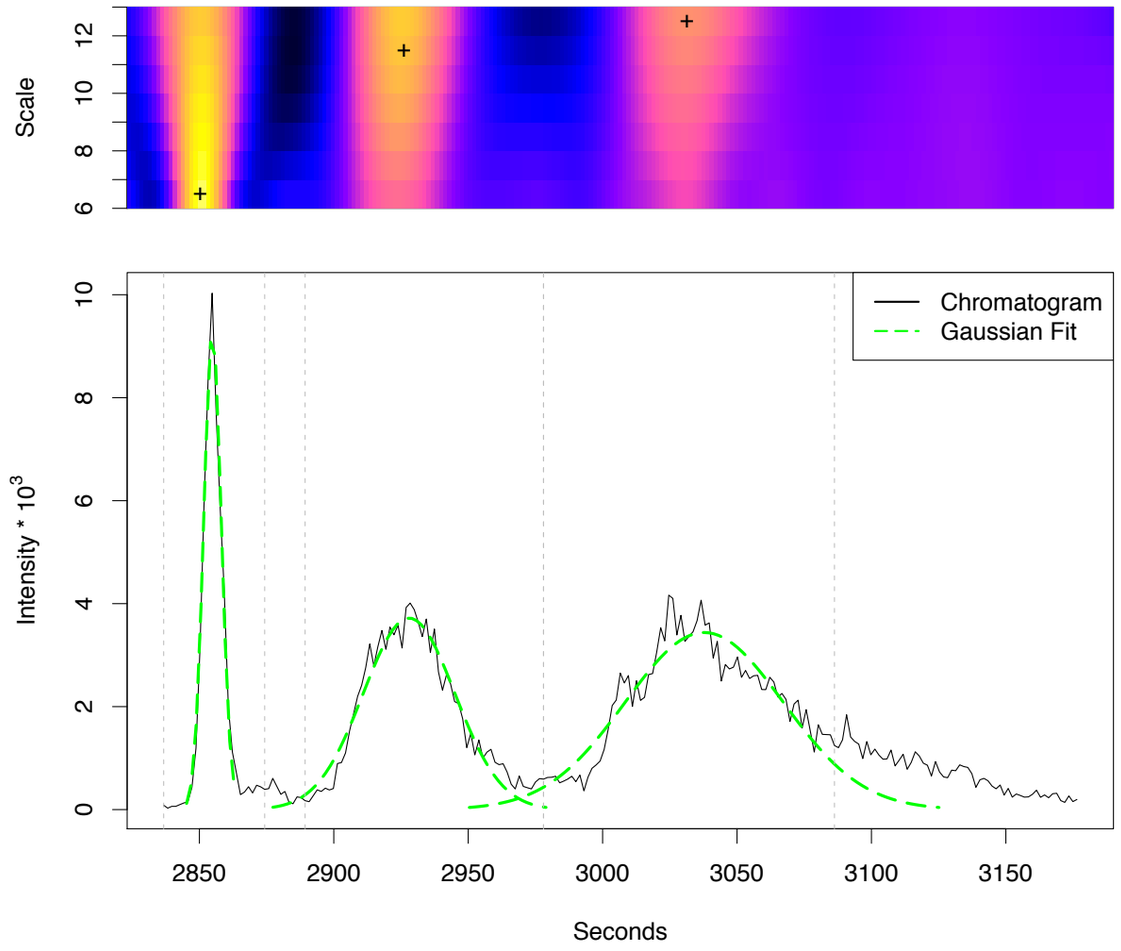


Detect EIC peaks

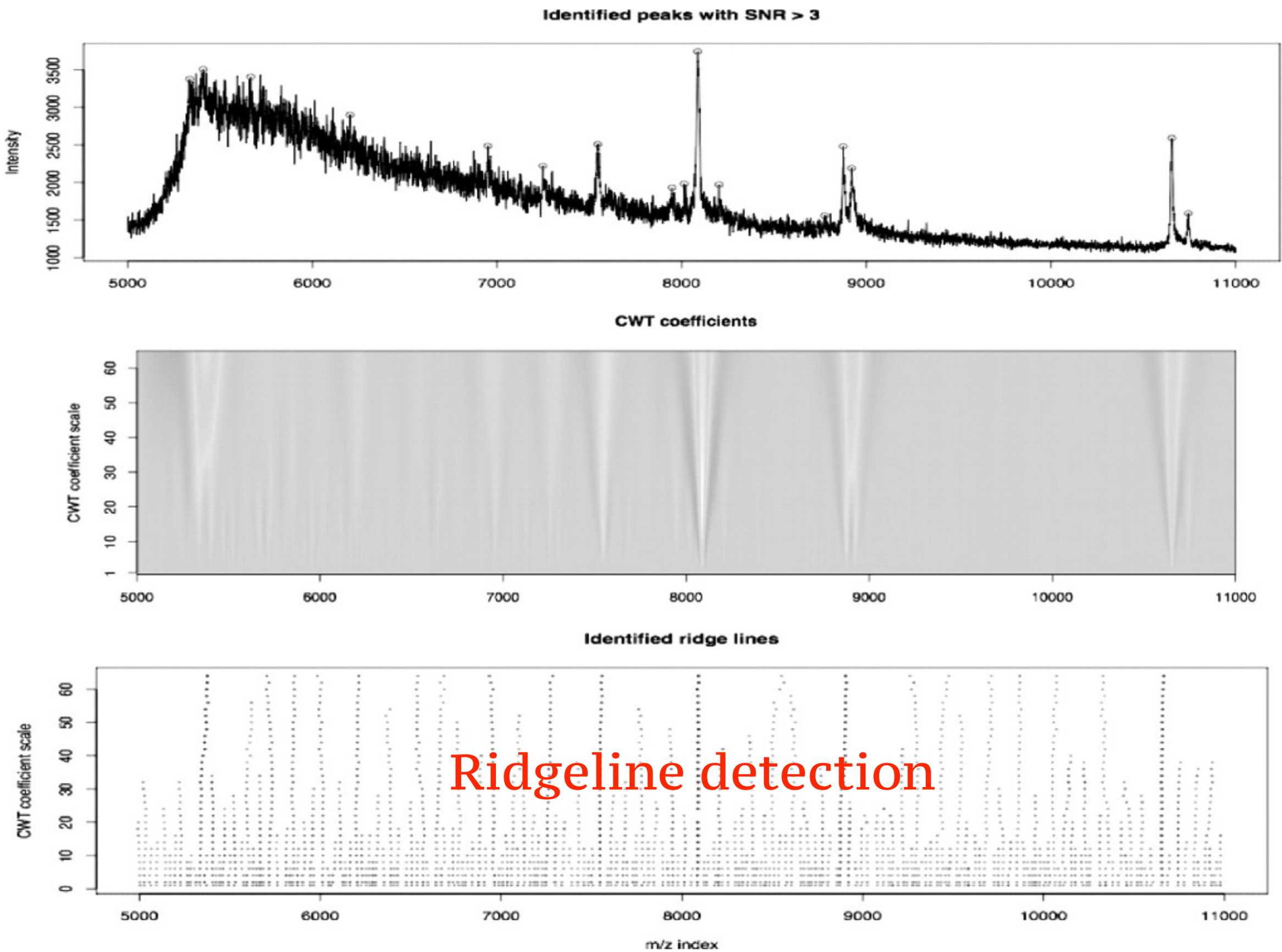
- Use wavelet transform



mexican hat wavelet



Tautenhahn, R.; Bottcher, C.; Neumann, S., Highly sensitive feature detection for high resolution LC/MS. *BMC bioinformatics* **2008,** *9*, 504.



Ridgeline detection

Du P, Kibbe WA, Lin SM: Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics* 2006, 22(17):2059-2065.

Detect EIC peaks

- Results of peak detection

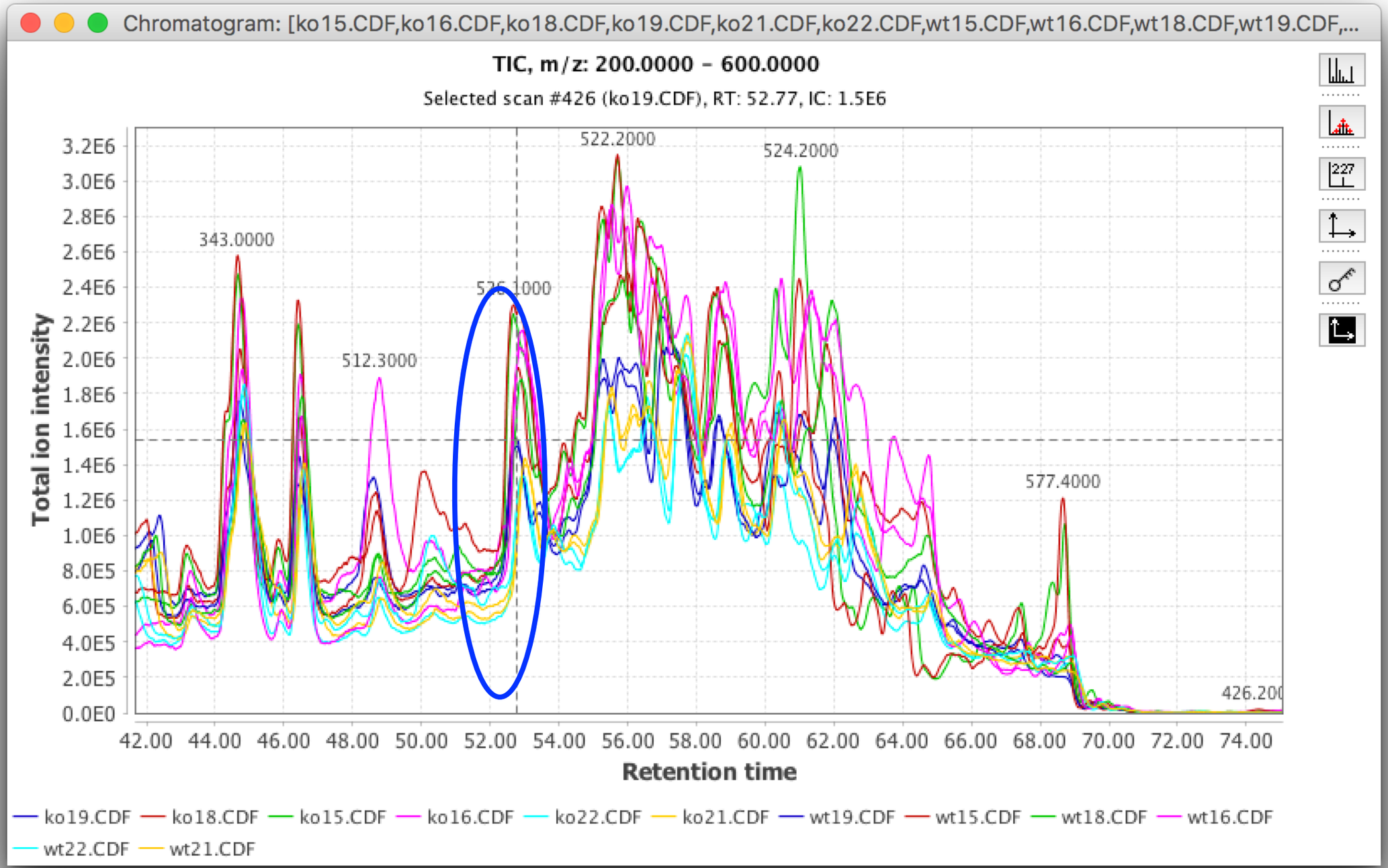
	mz	mzmin	mzmax	rt	rtmin	rtmax	into	intb	maxo	sn	sample
[1,]	412.5000	412.5	412.5	2506.073	2501.378	2509.203	4089.345	4083.085	693	692	1
[2,]	548.4161	548.4	548.5	2506.073	2501.378	2510.768	4482.160	4474.335	646	645	1
[3,]	511.6397	511.6	511.7	2513.898	2509.203	2518.593	3266.155	3258.330	461	460	1
[4,]	214.9000	214.9	214.9	2517.028	2512.333	2520.158	8339.885	8333.625	1467	1466	1
[5,]	564.9000	564.9	564.9	2515.463	2510.768	2520.158	3868.680	3860.855	587	586	1
[6,]	571.5853	571.5	571.6	2518.593	2512.333	2523.287	4065.499	3938.182	526	11	1

Peak information

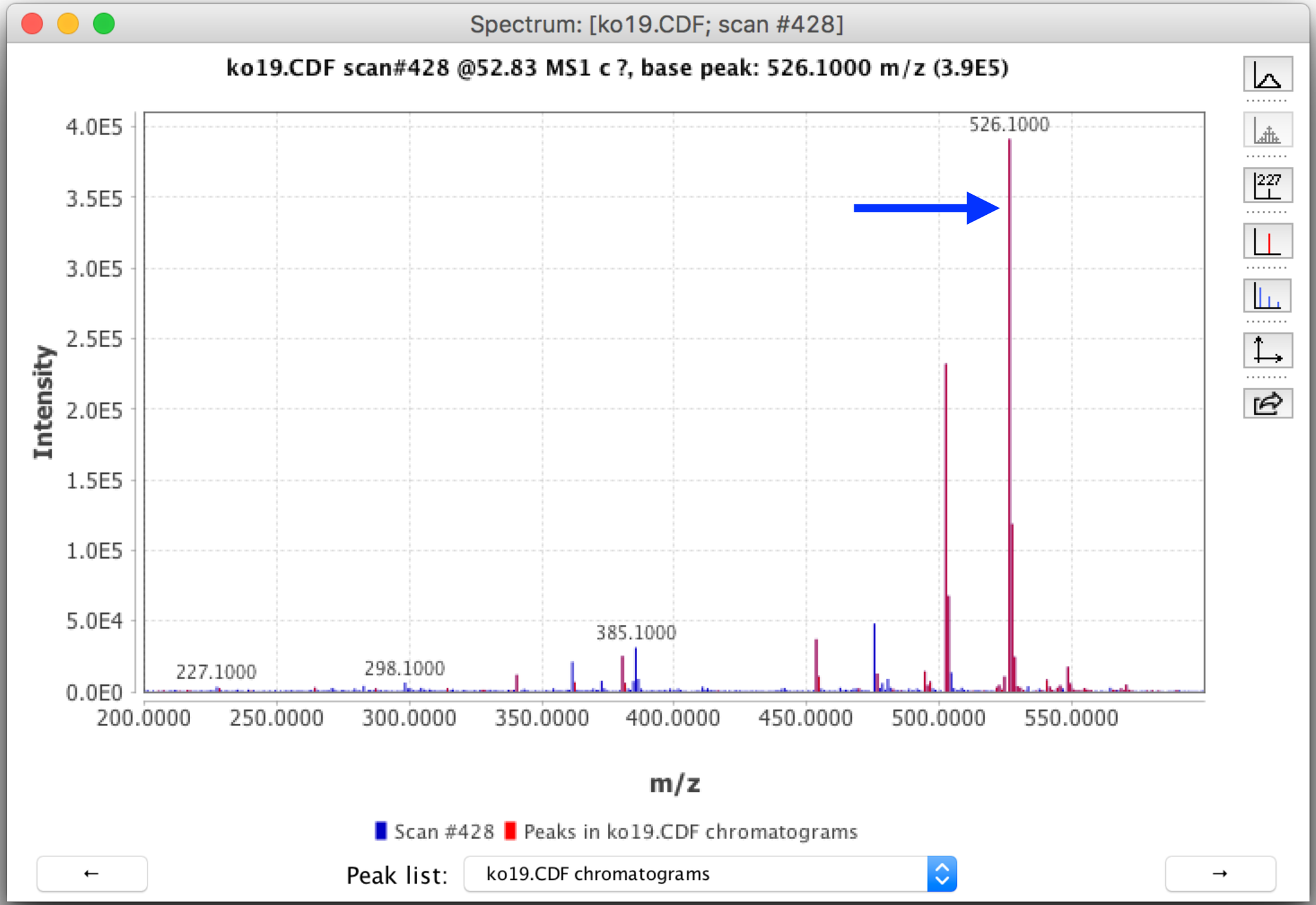
`mz`	weighted (by intensity) mean of peak m/z across scans
`mzmin`	m/z peak minimum
`mzmax`	m/z peak maximum
`rt`	retention time of peak midpoint
`rtmin`	leading edge of peak retention time
`rtmax`	trailing edge of peak retention time
`into`	integrated peak intensity
`intb`	baseline corrected integrated peak intensity
`maxo`	maximum peak intensity
`sn`	Signal/Noise ratio, defined as `(maxo - baseline)/sd`, where `maxo` is the maximum peak intensity, `baseline` the estimated baseline value and `sd` the standard deviation of local chromatographic noise.

Why alignment?

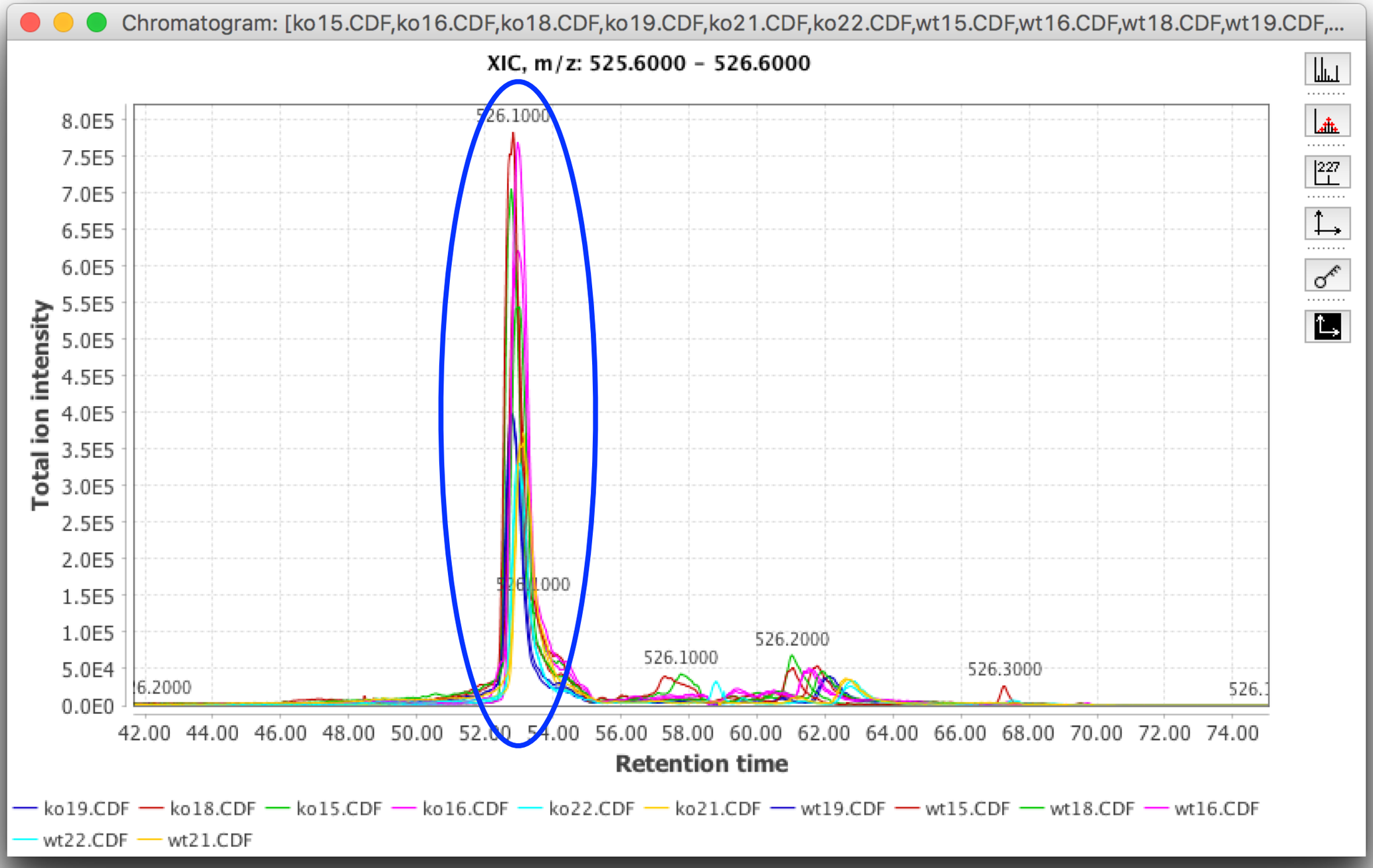
- faahKO



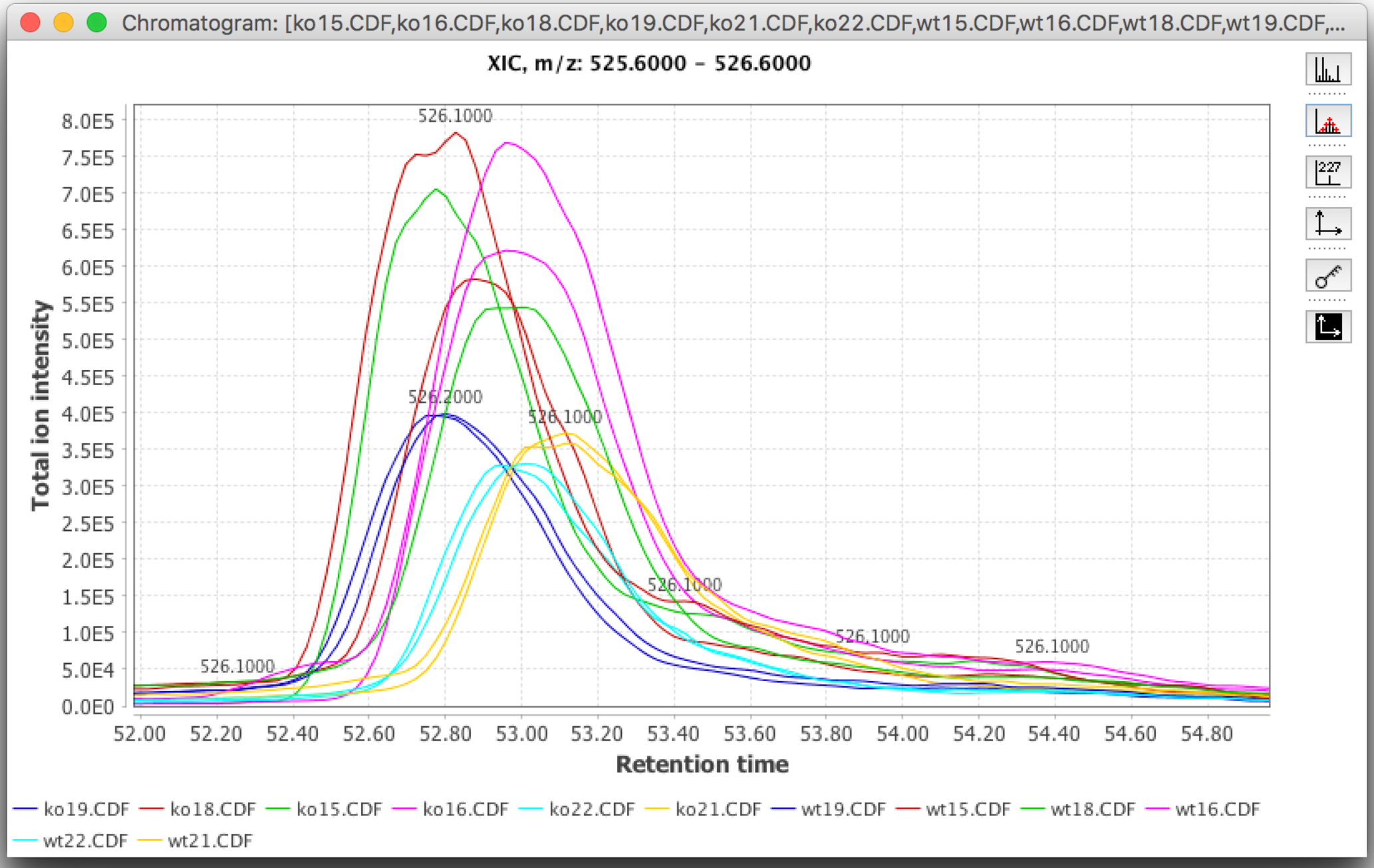
Why alignment?



Why alignment?



Why alignment?

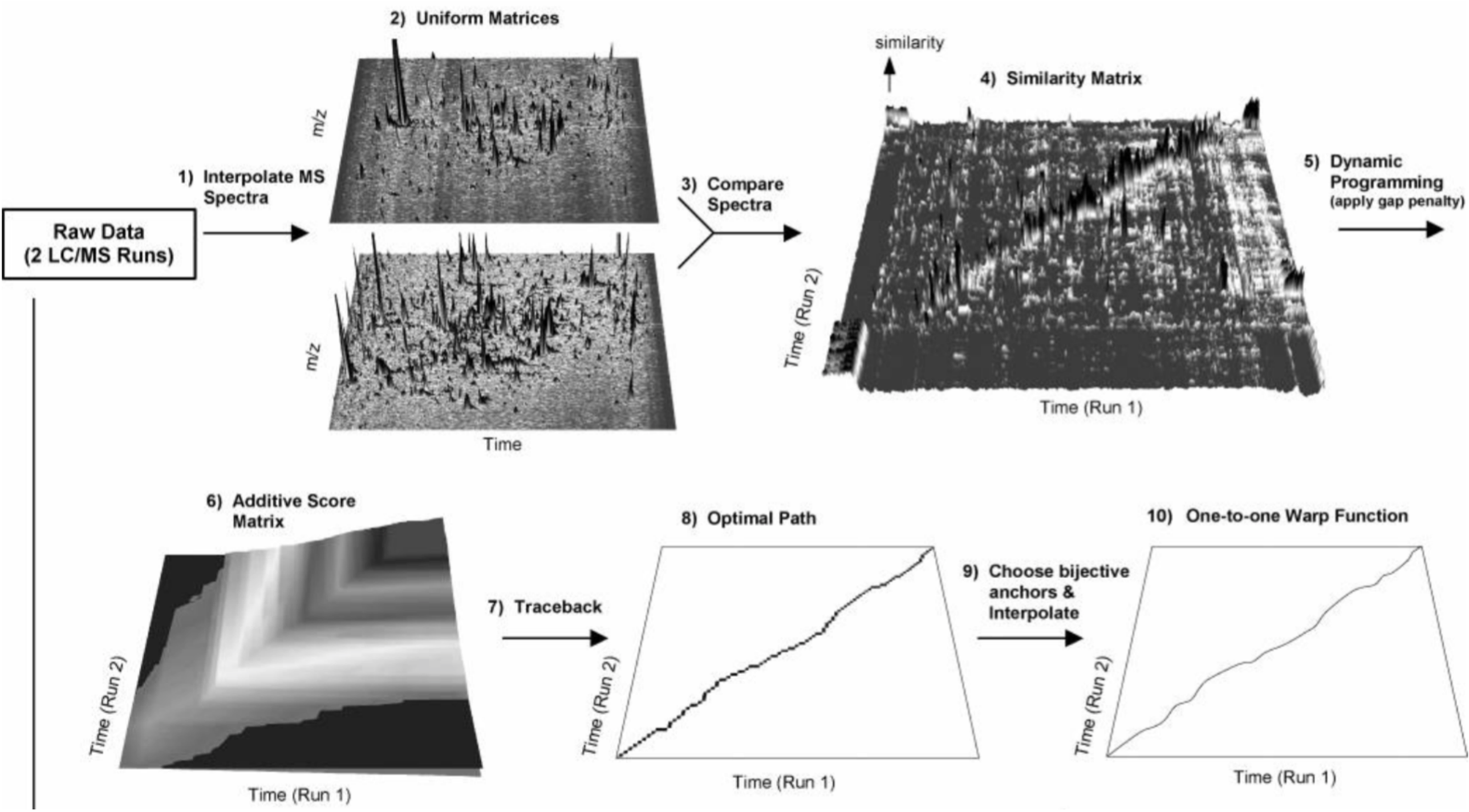


Alignment

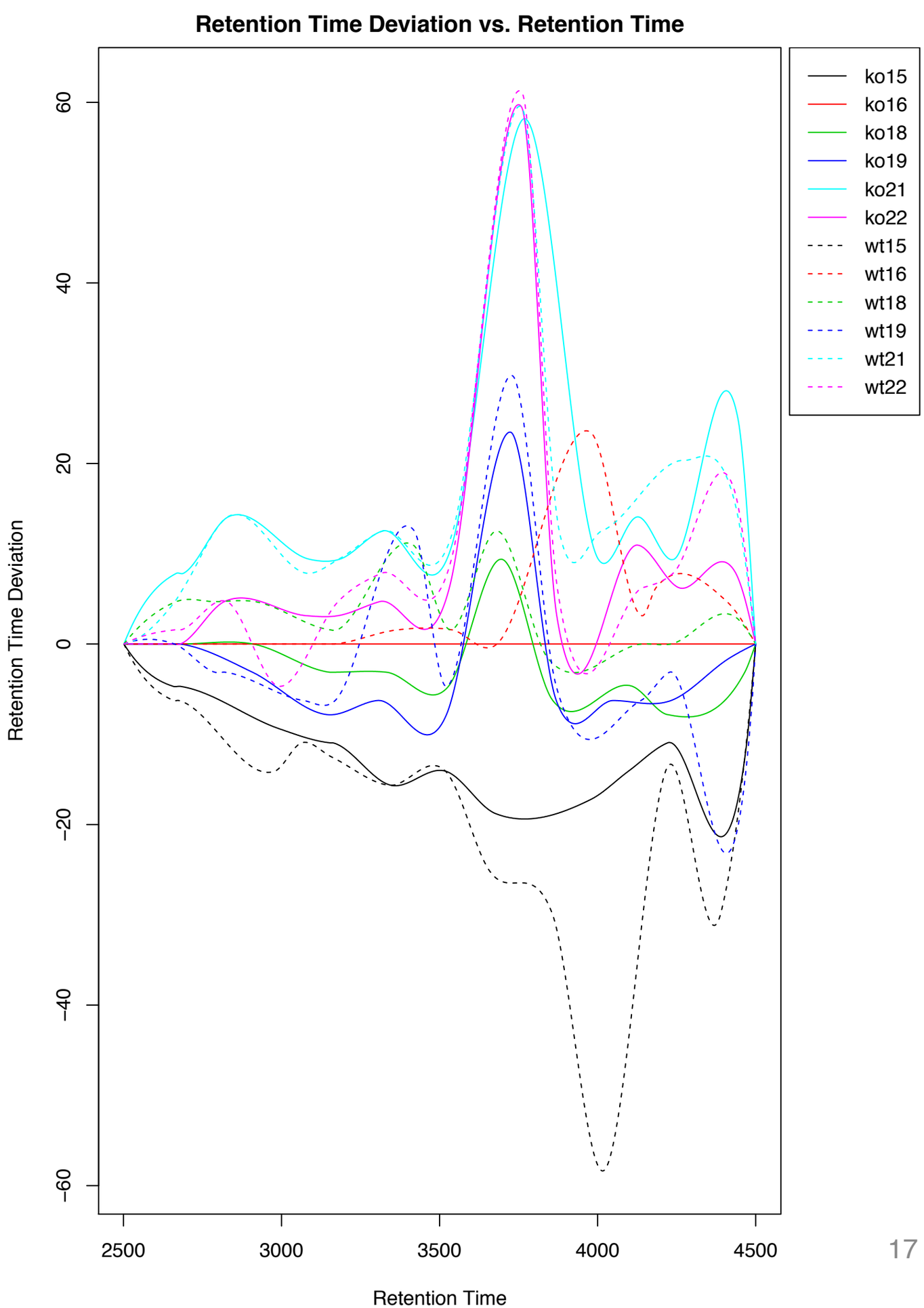
- Algorithm: Obiwarp

- Related publication
  - Chromatographic Alignment of ESI-LC-MS Proteomics Data Sets by Ordered Bijective Interpolated Warping, John T. Prince and, Edward M. Marcotte Analytical Chemistry 2006 78 (17), 6140-6152

Alignment


Raw Data
(2 LC/MS Runs)
1) Interpolate MS Spectra
2) Uniform Matrices
m/z
m/z
Time
3) Compare Spectra
similarity
4) Similarity Matrix
Time (Run 2)
Time (Run 1)
5) Dynamic Programming
(apply gap penalty)
6) Additive Score Matrix
Time (Run 2)
Time (Run 1)
7) Traceback
8) Optimal Path
Time (Run 2)
Time (Run 1)
9) Choose bijective anchors & Interpolate
10) One-to-one Warp Function
Time (Run 2)
Time (Run 1)

Alignment

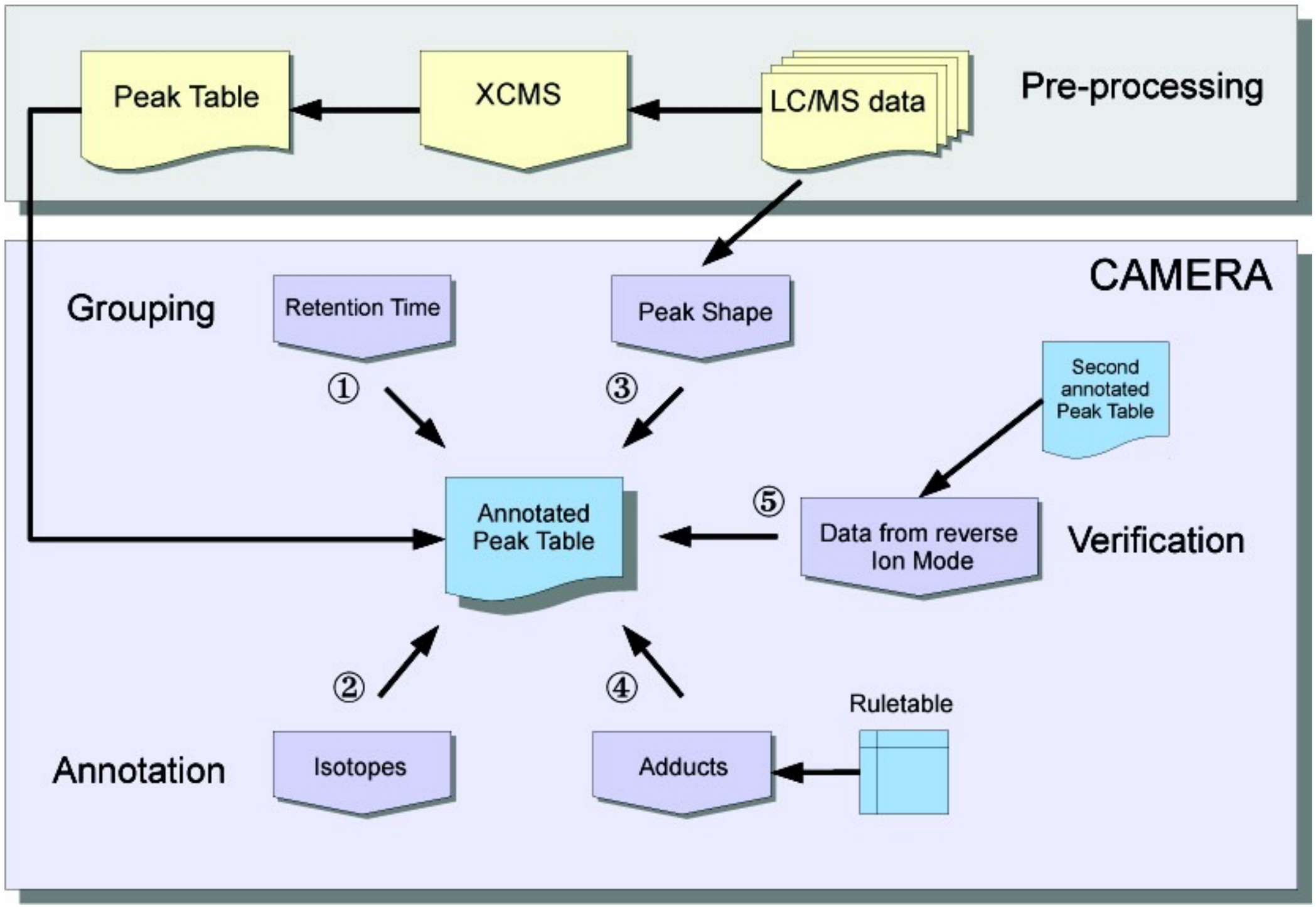

Retention Time Deviation vs. Retention Time
Retention Time Deviation
60
40
20
0
−20
−40
−60
2500
3000
3500
4000
4500
Retention Time
ko15
ko16
ko18
ko19
ko21
ko22
wt15
wt16
wt18
wt19
wt21
wt22

Peak annotation

- Achieved by the *CAMERA* package

- Related publication
  - Kuhl C, Tautenhahn R, Boettcher C, Larson TR and Neumann S (2012). “CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets.” Analytical Chemistry, 84, pp. 283–289. http://pubs.acs.org/doi/abs/10.1021/ac202450g.

Peak annotation



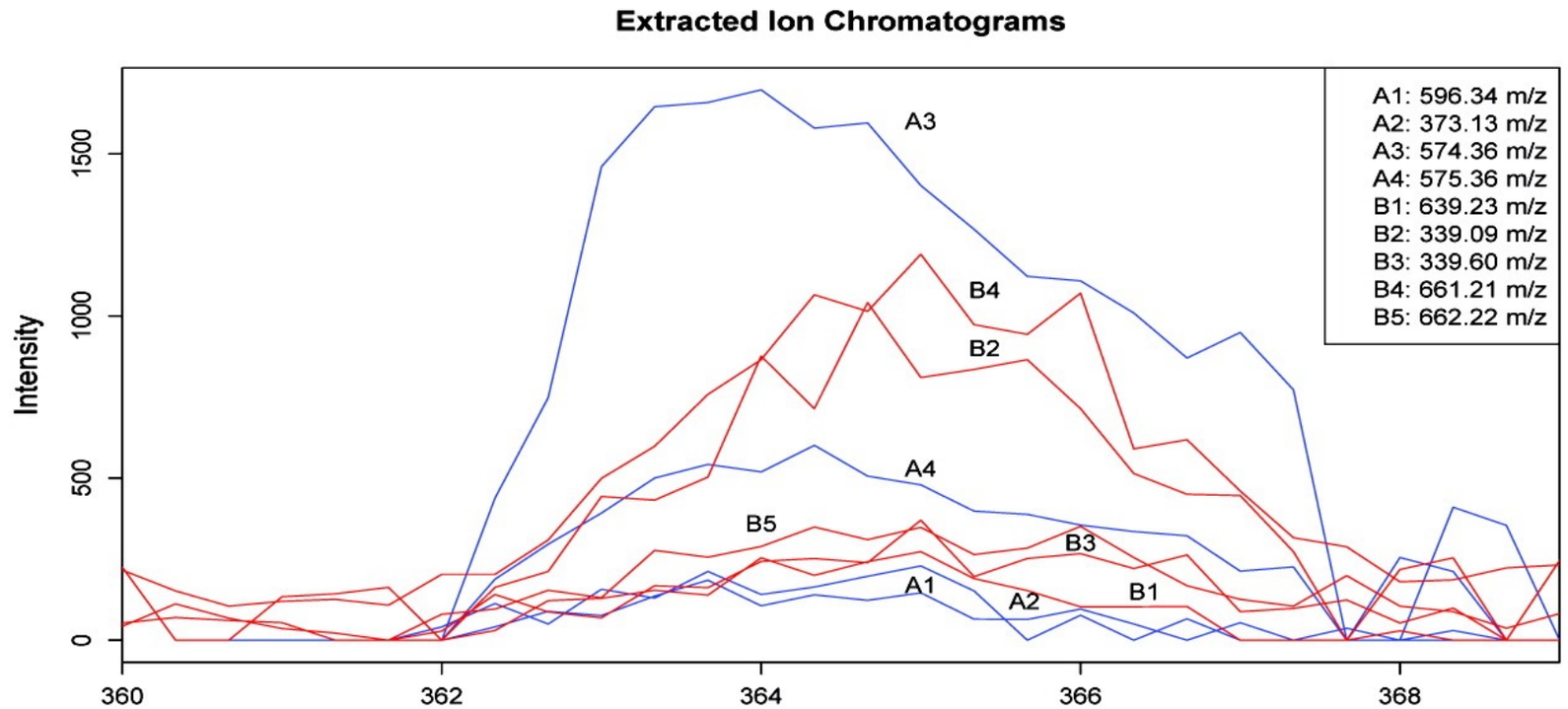
Peak annotation

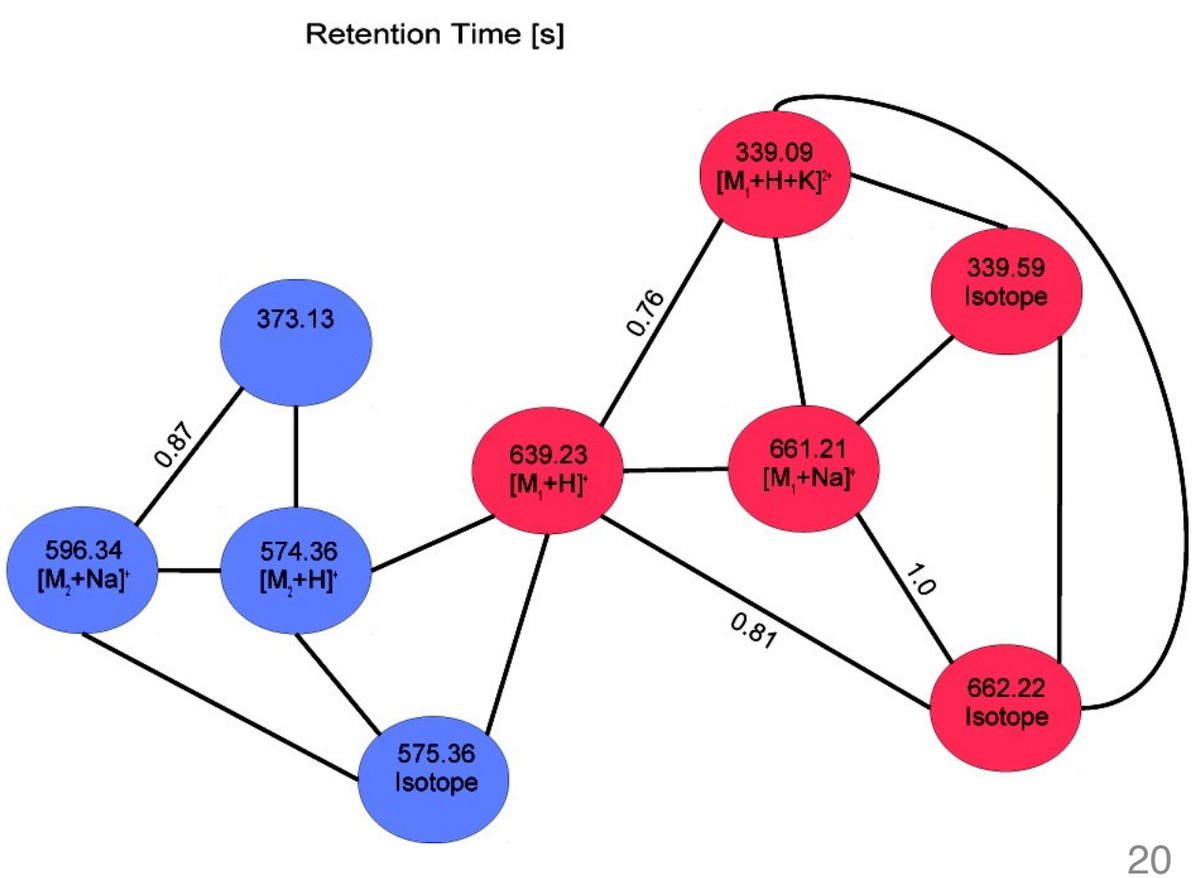
id	m/z	rt [s]
1	339.09	369.1
2	339.59	369.1
3	373.13	368.6
4	574.36	368.7
5	575.36	368.7
6	596.34	368.4
7	639.23	368.7
8	661.21	369.1
9	662.22	369.1

Initial compound spectrum

	1	2	3	4	5	6	7	8
2	1.0							
3	0.34	0.36						
4	0.72	0.67	0.90					
5	0.68	0.47	0.68	1				
6	0.38	0.38	0.87	0.86	0.75			
7	0.76	0.72	0.73	0.82	0.88	0.70		
8	0.94	0.79	0.35	0.70	0.67	0.51	0.79	
9	0.88	0.84	0.46	0.62	0.70	0.62	0.81	1.0

Relationship scoring table





Peak annotation

id	mz	rt	isotopes	adduct	pc
65	176.04	280.09			4
76	136.05	280.43	[14][M+1]1+		5
77	135.05	280.43	[14][M]1+		5
74	153.06	280.43		[M+H]+ 152.05437	5
75	175.04	280.43		[M+Na]+ 152.05437	5
73	197.02	280.76		[M+2Na-H]+ 152.05437	5
78	377.74	286.15			6
79	732.5	286.49			6
83	488.32	286.82		[M+Na]+ 465.33205	7
82	466.34	286.82		[M+H]+ 465.33205	7
...					

Annotation rules

- Primary adduct rules

name	nmol	charge	massdiff	oidscore	quasi	ips
[M+H]+	1	1	1.007276	1	1	1
[M+Na]+	1	1	22.989218	8	1	1
[M+K]+	1	1	38.963158	10	1	1
[M+NH4]+	1	1	18.033823	16	1	1

Annotation rules

Extended adduct rules

name	nmol	charge	massdiff	oidscore	quasi	ips
[M+H]+	1	1	1.007276	1	1	1
[M+2H]2+	1	2	2.014552	2	0	0.75
[M+3H]3+	1	3	3.021828	3	0	0.75
[M+H+Na]2+	1	2	23.996494	4	0	0.5
[M+H+K]2+	1	2	39.970434	6	0	0.5
[M+H+NH4]2+	1	2	19.041099	7	0	0.5
[M+Na]+	1	1	22.989218	8	1	1
[M+2Na]2+	1	2	45.978436	9	0	0.5
[M+K]+	1	1	38.963158	10	1	1
[M+Na+K]2+	1	2	61.952376	11	0	0.5
[M+2K]2+	1	2	77.926316	13	0	0.5
[M+NH4]+	1	1	18.033823	16	1	1
[M+2Na-H]+	1	1	44.97116	34	0	0.5
[M+2K-H]+	1	1	76.91904	60	0	0.5
[2M+H]+	2	1	1.007276	1	0	0.5
[2M+2H]2+	2	2	2.014552	2	0	0.5
[2M+3H]3+	2	3	3.021828	3	0	0.5
[2M+H+Na]2+	2	2	23.996494	4	0	0.5
[2M+H+K]2+	2	2	39.970434	6	0	0.5
[2M+H+NH4]2+	2	2	19.041099	7	0	0.5
[2M+Na]+	2	1	22.989218	8	0	0.5
[2M+2Na]2+	2	2	45.978436	9	0	0.5
[2M+K]+	2	1	38.963158	10	0	0.5
[2M+Na+K]2+	2	2	61.952376	11	0	0.5
[2M+2K]2+	2	2	77.926316	13	0	0.5
[2M+NH4]+	2	1	18.033823	16	0	0.5
[2M+2Na-H]+	2	1	44.97116	34	0	0.25

Thank you!