Adaptation of Skyline to Analyze Untargeted Metabolomics Data Collected on GCMS Instrument

Pawel Sadowski, PhD

Laboratory Manager, CARF Proteomics Laboratory

Skyline User Meeting

June 02, 2019

CARF Proteomics Laboratory

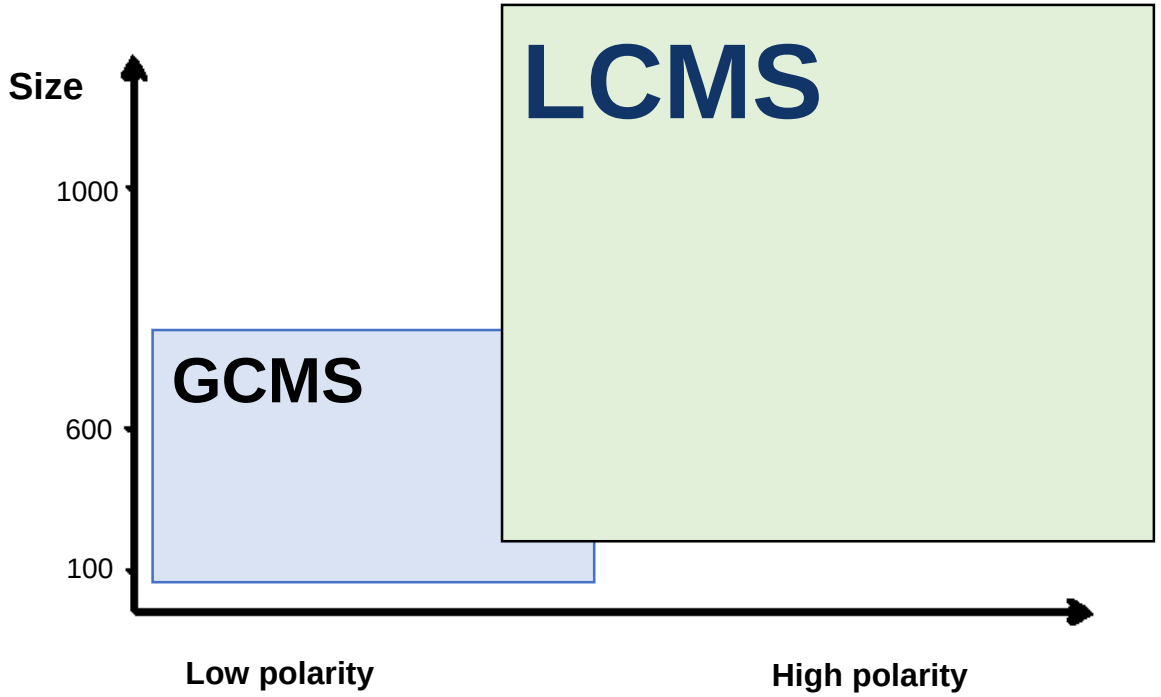
- Central Analytical Research Facility (CARF) is a multi-user multi-laboratory facility with instrumentation ranging from next generation sequencing, microscopy, vibrational spectroscopy, X-ray diffraction through mass spectrometry (elements and isotopes, proteomics, lipidomics)
- Clients: *internal (QUT) and external researchers, industry*
- CARF Proteomics Laboratory:
  - SWATH-MS based quantitative proteomics (use Skyline)
  - Targeted LCMS (use Skyline)
  - Targeted and untargeted GCMS (pesticides, metabolites, PAHs, TPHs, PCBs)

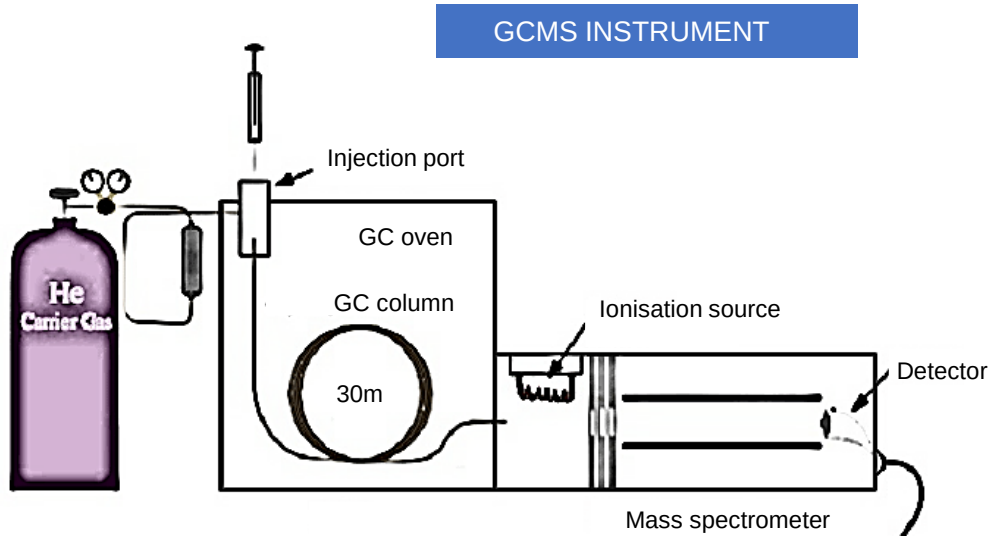
Veterinary and animal research	+	SWATH-MS	=	Advancing clinical veterinary medicine and diagnostics

Metabolomics

	+	-
LCMS	High sensitivity High resolution	Unstable metabolites Limited databases
GCMS	Well-established Reproducible Sufficient sensitivity Good resolution Comprehensive databases	Complex sample processing
NMR	Noninvasive Saving samples High specificity and resolution	Complex data analysis Limited dynamic range and sensitivity Expensive hardware
Other		

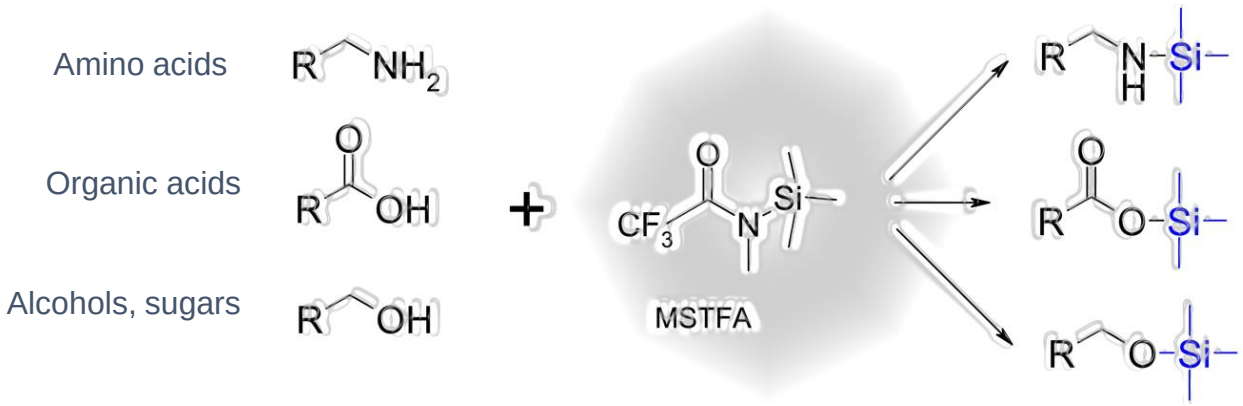
GCMS-based metabolomics

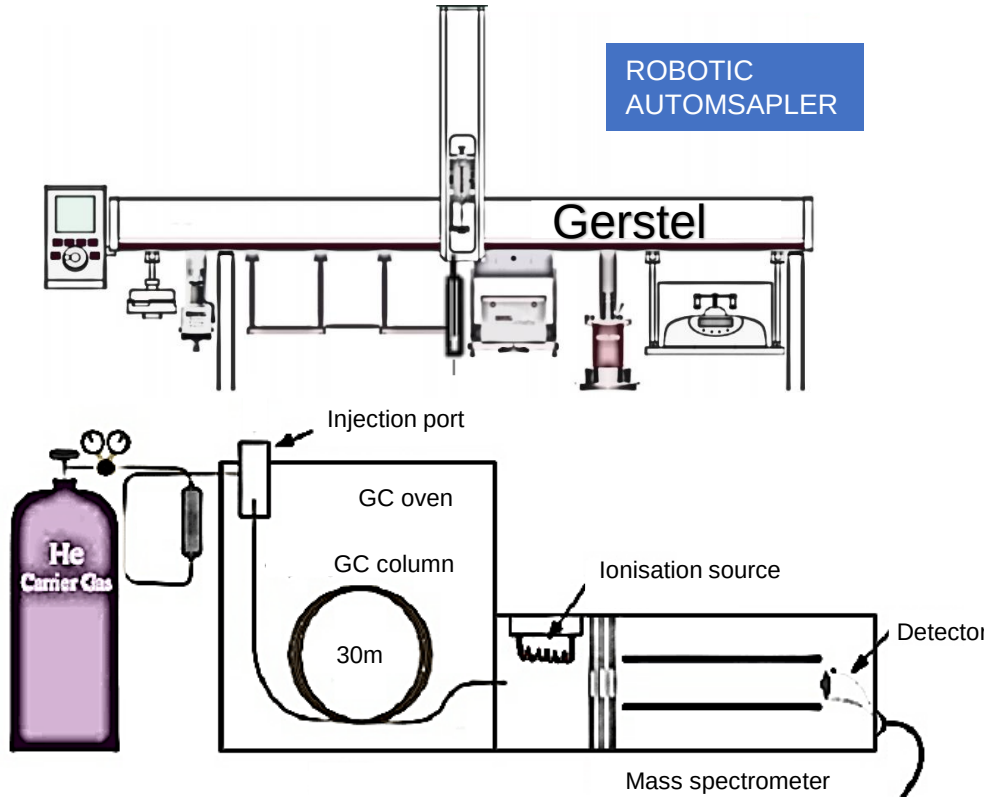




Derivatisation

Lowers boiling point extending the range of compounds amenable to GCMS analysis



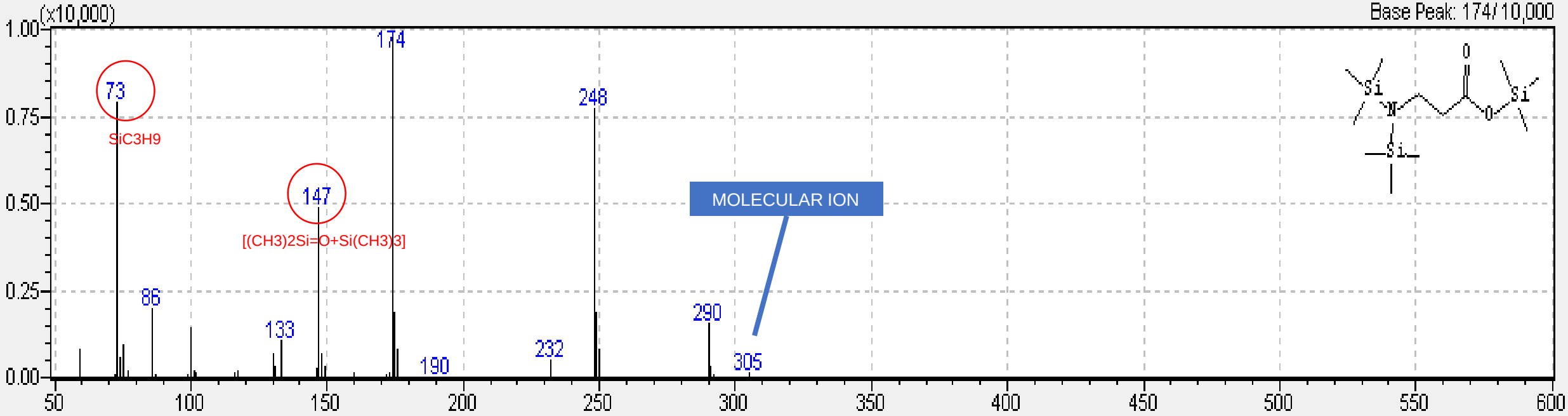


GCMS data

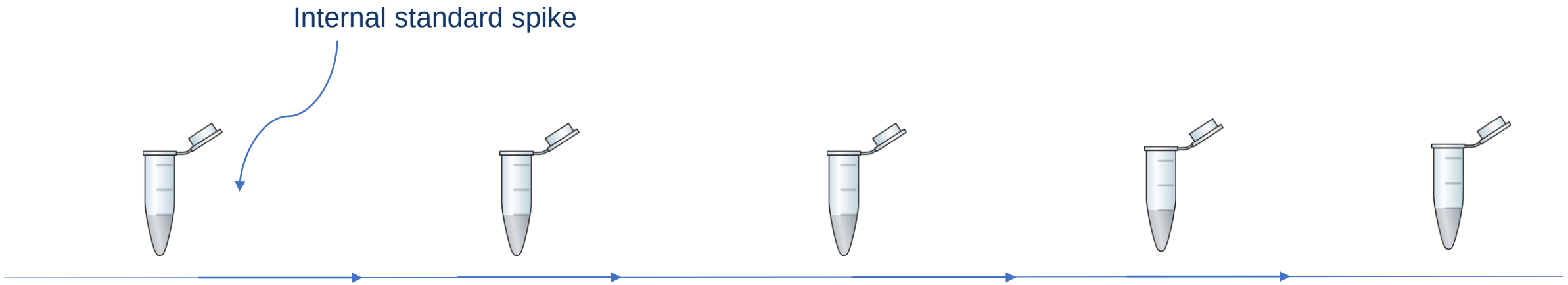
Databases:

i.e. NIST,
Wiley,
Golm

Beta-Alanine



Workflow



- Metabolite extraction and protein removal
- Lipid removal
- Locking stereochemical configuration and derivatisation
- Data acquisition
- Data processing and analysis

Randomization!

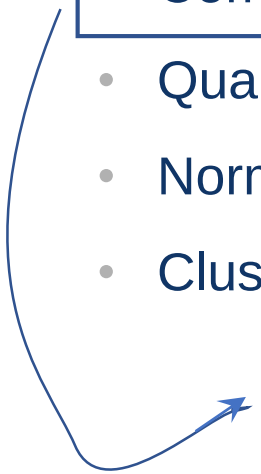
Data acquisition

Batch set up includes:

- PBQC
- Procedural blank
- QC sample
- RT standard
- Samples (possibly injected different concentrations)
- Sweep runs

Data processing and analysis

- Peak detection/integration (extraction of chromatographic features)
- Peak deconvolution
- RT alignment
- Extraction of most abundant ions per chromatographic peak
- Compound identification
- Quality assessment
- Normalization
- Clustering and differential analysis



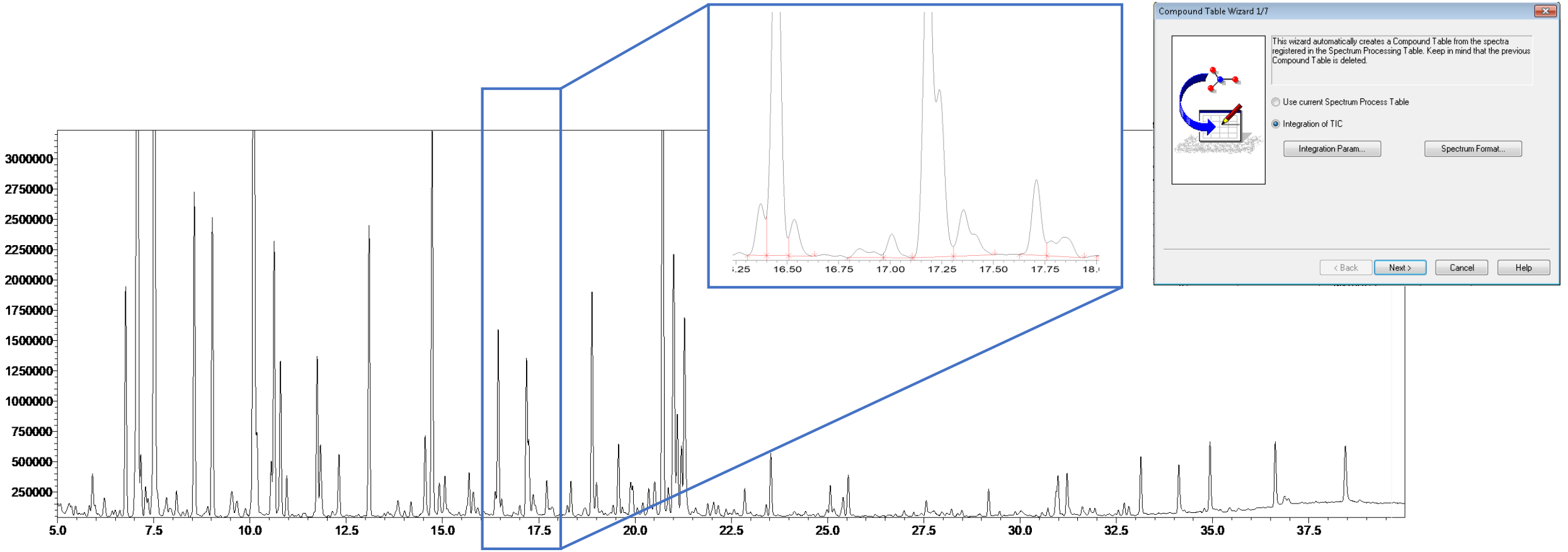
Our intention was to have one universal tool and one set of rules applied across different ‘omics’

Building transition list in Skyline

Can be combined with compound identification.

Automated peak integration

Automated ion extraction



TIC of PBQC sample

Building transition list in Skyline

Re-formatting .csv file into Skyline readable format and removing unwanted ions.
Precursor m/z is made up and set higher than m/z range used during data acquisition.

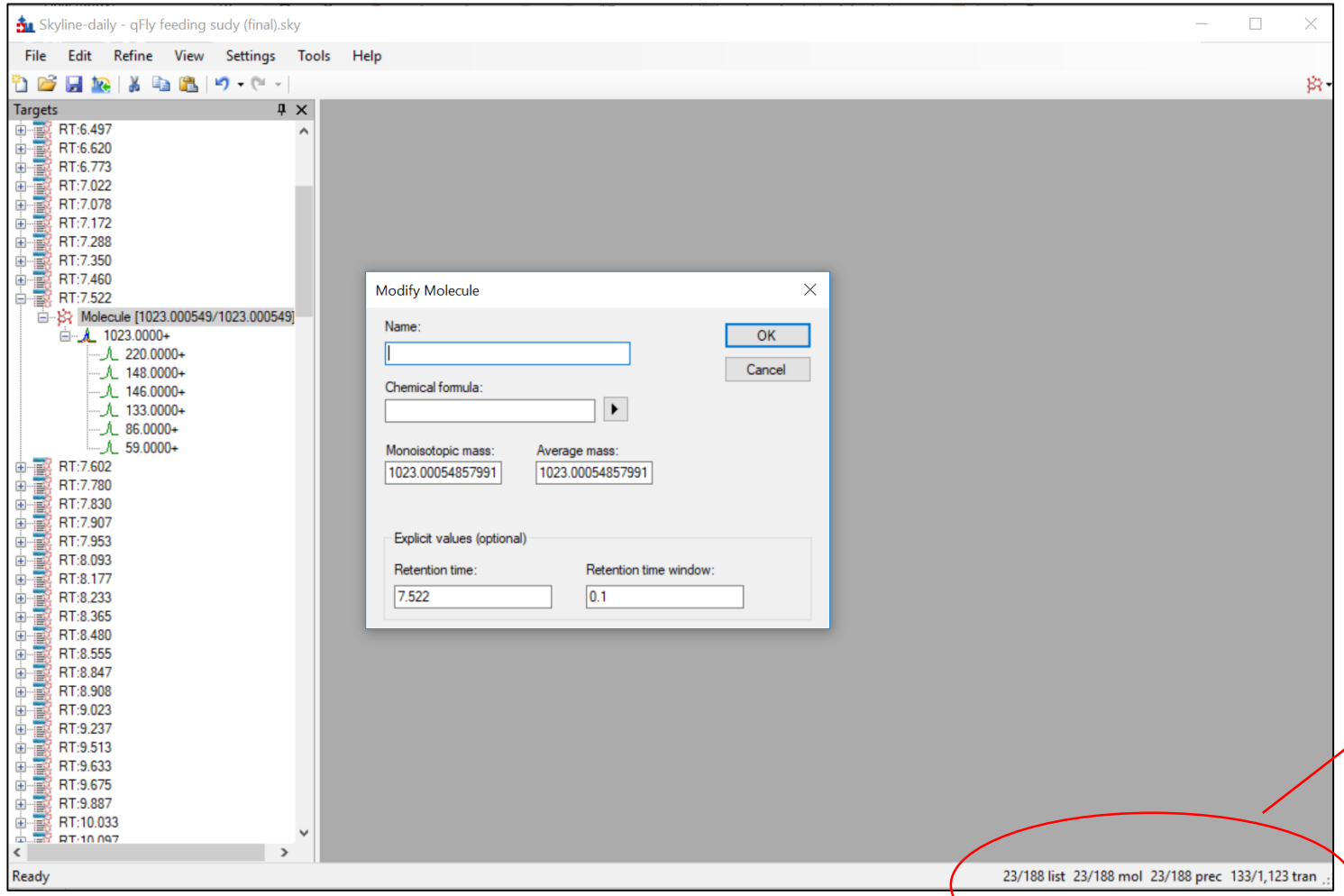
Molecule name ..... Product ions ..... Retention time

	A	B	C	D	E	F	G	H
1	RT:5.078	73	158	59	89	130	203	5.078
2	RT:5.305	73	147	57	77	174	151	5.305
3	RT:5.370	73	117	147	131	75	102	5.37
4	RT:5.470	207	73	152	208	209	295	5.47
5	RT:5.828	207	73	57	72	55	69	5.828
6	RT:5.912	73	117	147	191	148	75	5.912
7	RT:5.983	147	75	73	100	190	148	5.983
8	RT:6.218	73	75	173	147	117	131	6.218
9	RT:6.425	73	117	89	75	118	74	6.425
10	RT:6.500	72	76	75	55	59	74	6.5
11	RT:6.618	73	157	75	103	69	207	6.618
12	RT:6.773	116	73	147	117	74	75	6.773
13	RT:7.075	73	116	258	75	147	89	7.075
14	RT:7.165	102	73	147	55	97	83	7.165
15	RT:7.288	55	83	97	154	82	124	7.288
16	RT:7.350	73	147	130	174	188	75	7.35
17	RT:7.520	147	73	133	59	148	86	7.52
18	RT:7.832	174	73	86	142	59	175	7.832
19	RT:7.920	147	73	148	57	117	75	7.92
20	RT:8.092	147	281	73	282	148	283	8.092

	A	B	C	D	E	F	G
		Precursor m/z	Precursor charge	Product m/z	Product charge	Retention time	Retention time window
1	RT:5.078	1000	1	X 73	1	5.078	0.1
2	RT:5.078	1000	1	158	1	5.078	0.1
3	RT:5.078	1000	1	59	1	5.078	0.1
4	RT:5.078	1000	1	89	1	5.078	0.1
5	RT:5.078	1000	1	130	1	5.078	0.1
6	RT:5.078	1000	1	203	1	5.078	0.1
7							
8	RT:5.305	1001	1	X 73	1	5.305	0.1
9	RT:5.305	1001	1	X 147	1	5.305	0.1
10	RT:5.305	1001	1	57	1	5.305	0.1
11	RT:5.305	1001	1	77	1	5.305	0.1
12	RT:5.305	1001	1	174	1	5.305	0.1
13	RT:5.305	1001	1	151	1	5.305	0.1
14							
15	RT:5.370	1002	1	X 73	1	5.37	0.1
16	RT:5.370	1002	1	117	1	5.37	0.1
17	RT:5.370	1002	1	X 147	1	5.37	0.1
18	RT:5.370	1002	1	131	1	5.37	0.1

Building transition list in Skyline

Edit > Insert > Transition list > Import .csv file



Ions extracted for 188 chromatographic features

Importing GCMS data

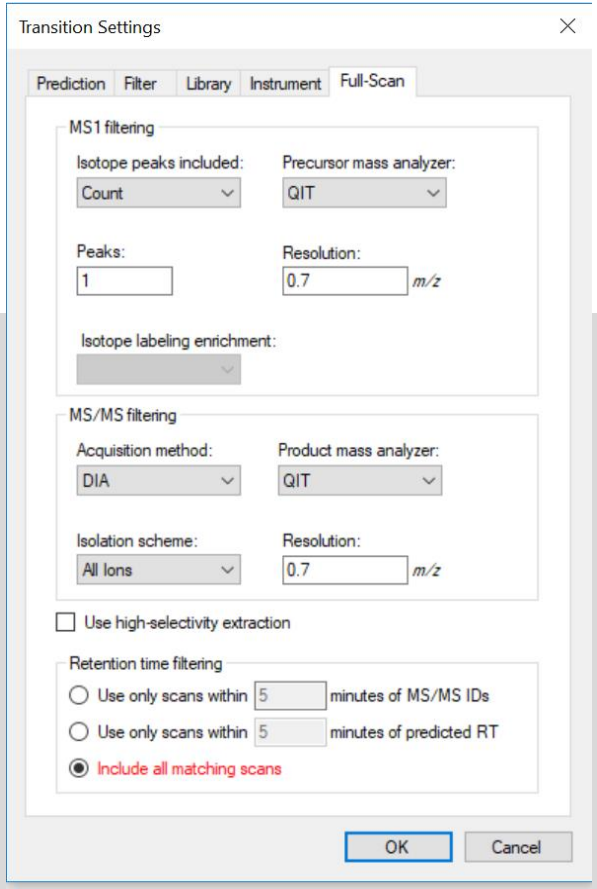
. qgd (Shimadzu)

.xml

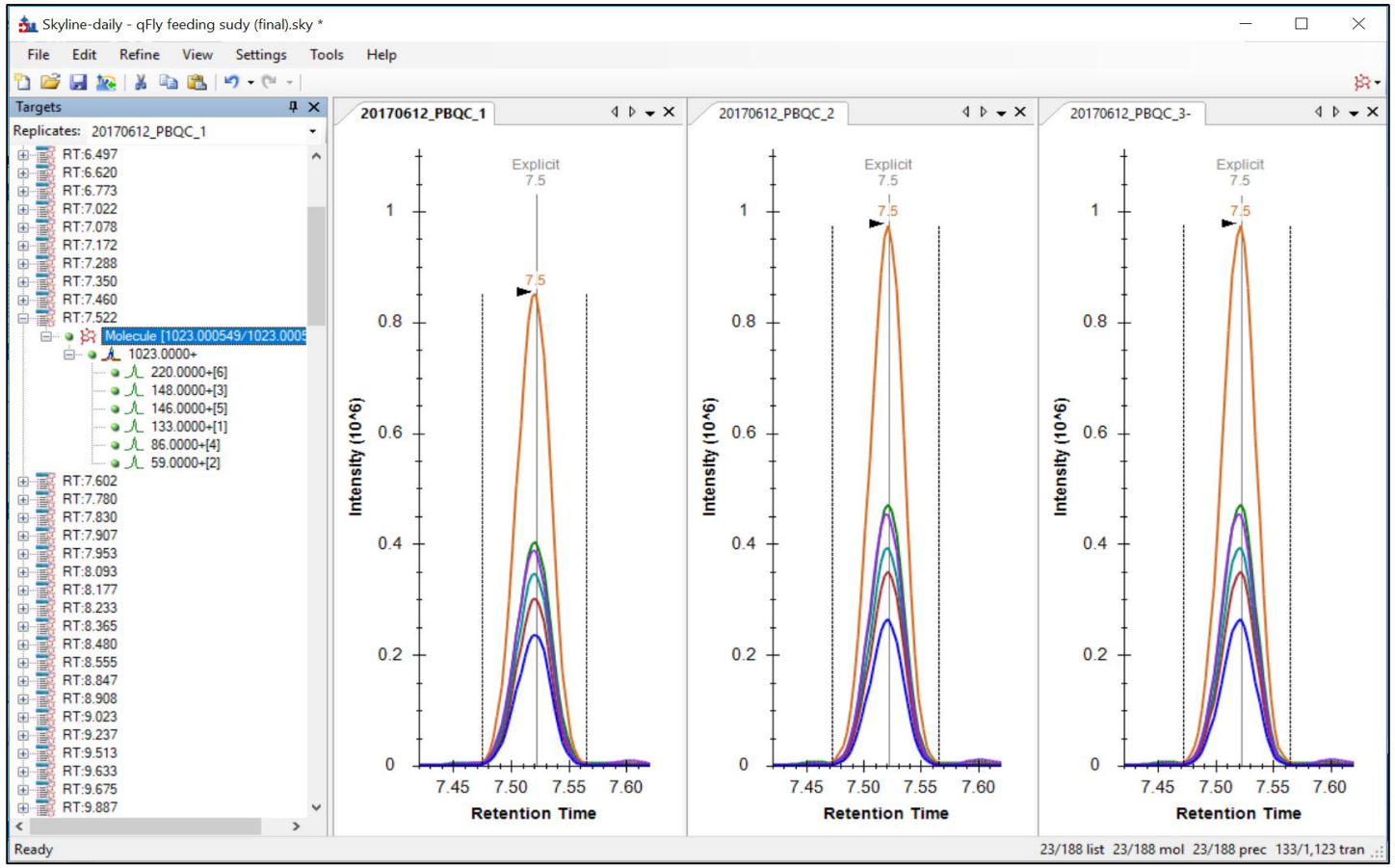
.raw (Thermo)

Skyline

Import settings



Importing GCMS data



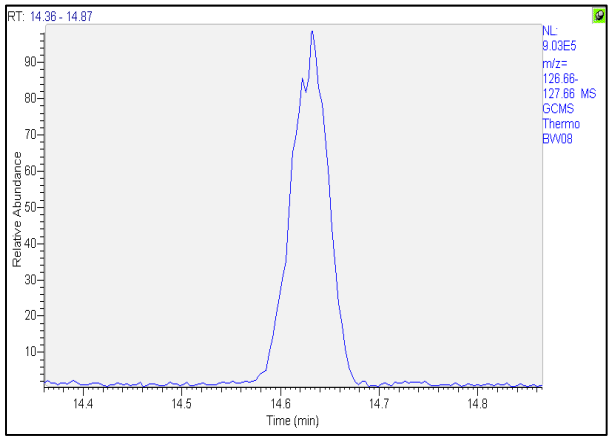
Targeted data extraction
MRM-like quantitation

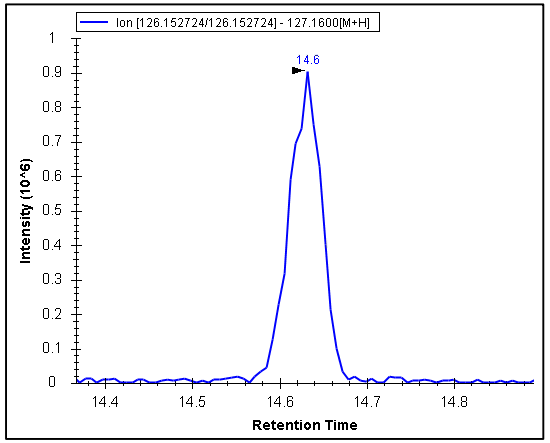
Importing GCMS data

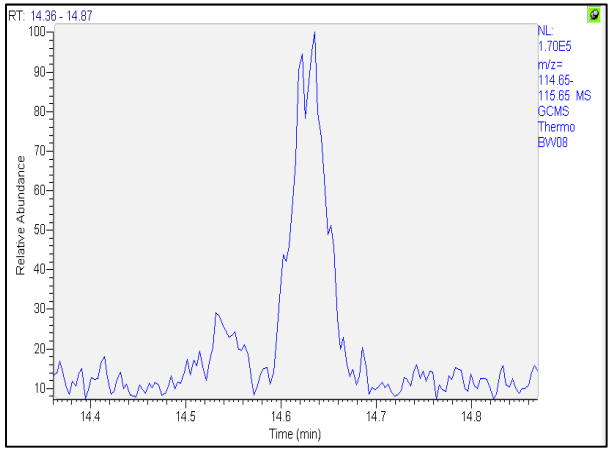
Missing data points (.raw)

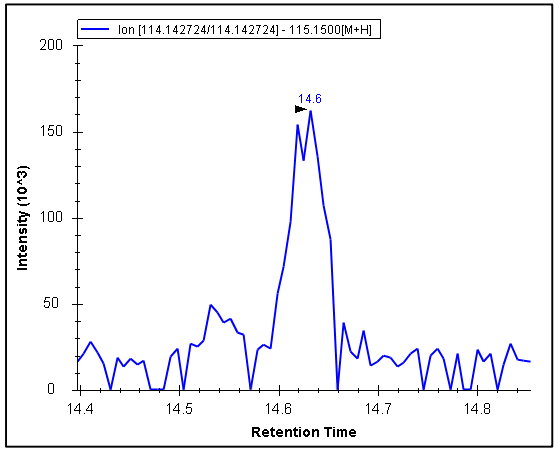
Xcalibur

Skyline

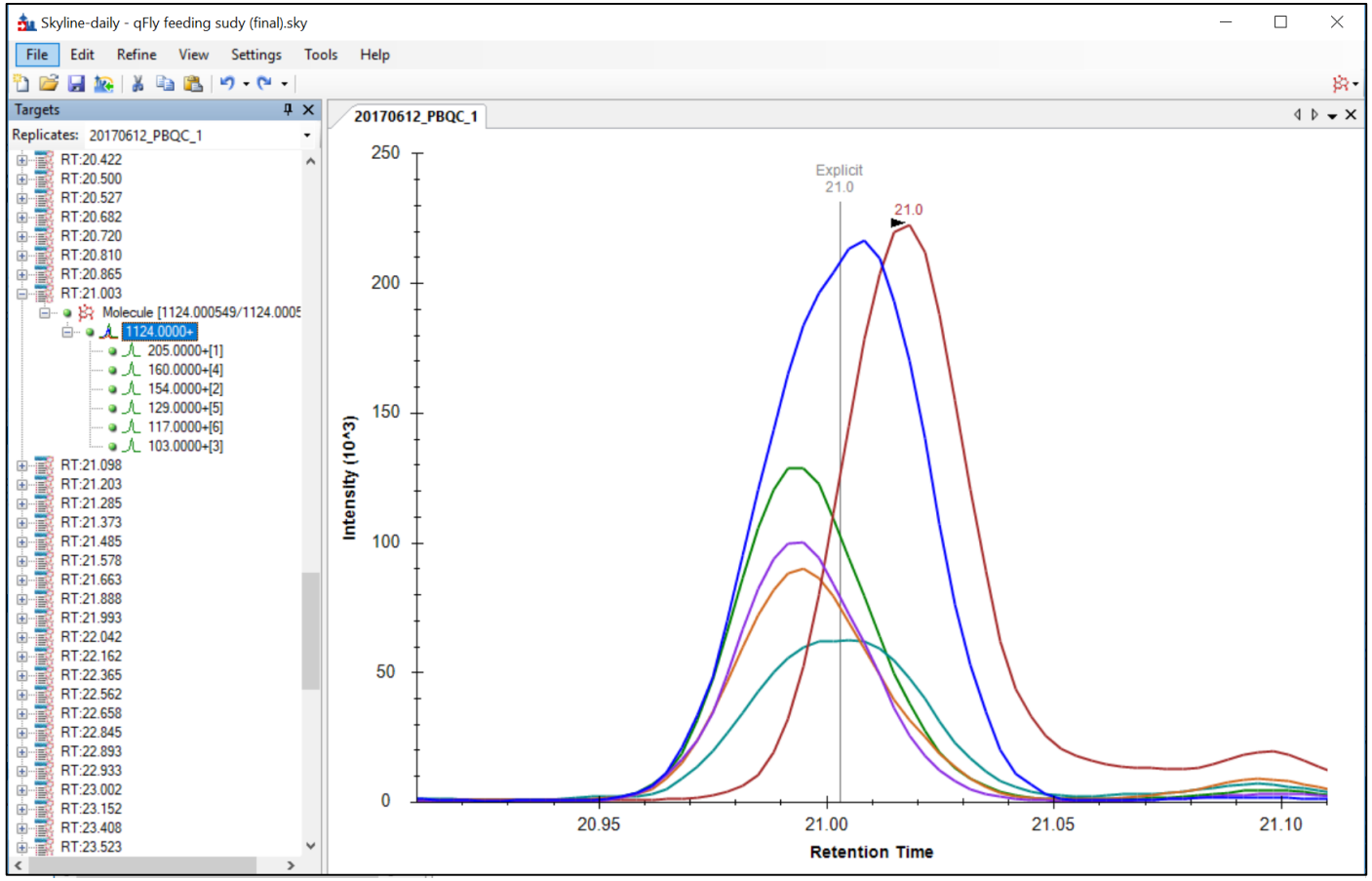




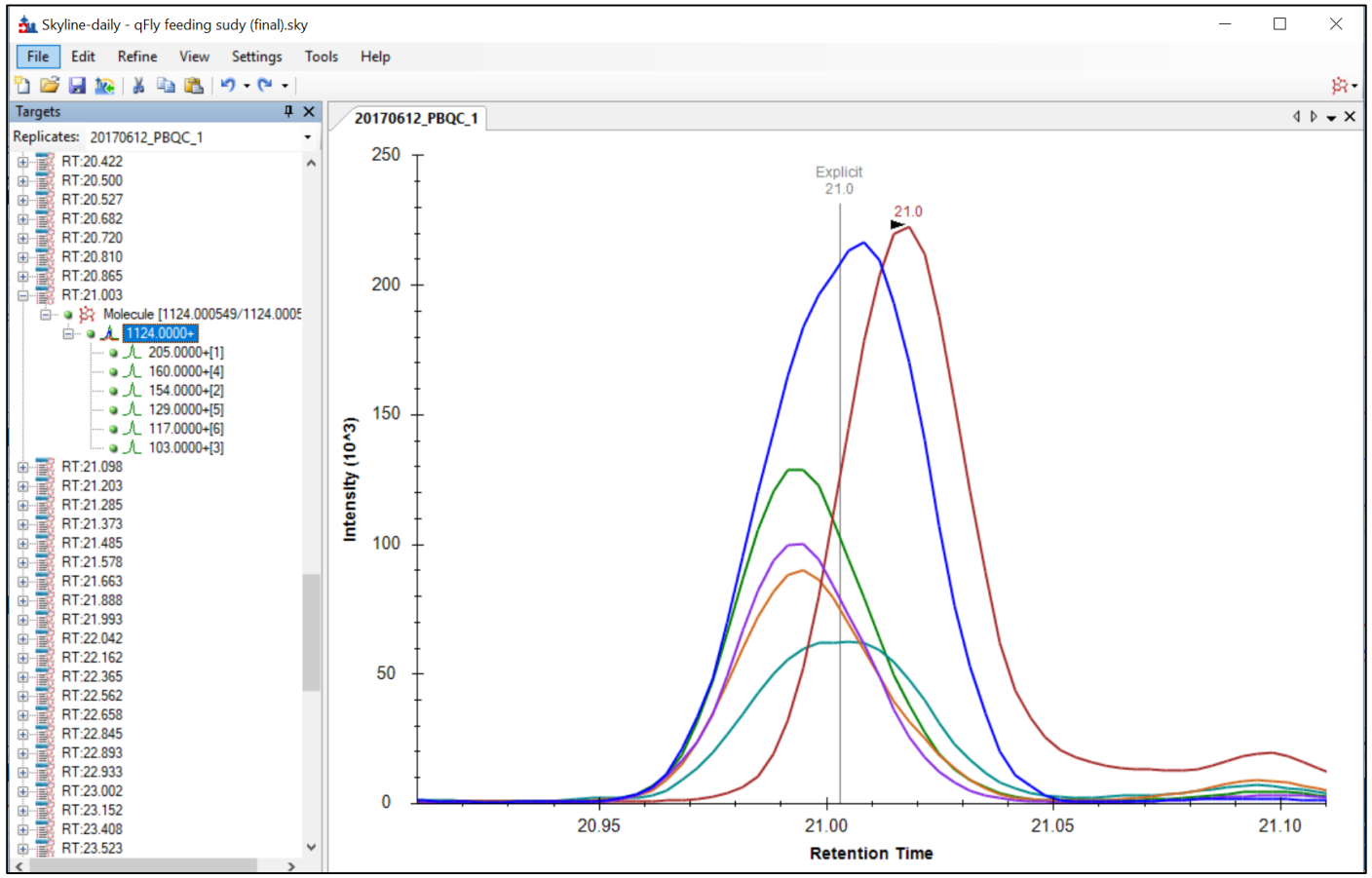


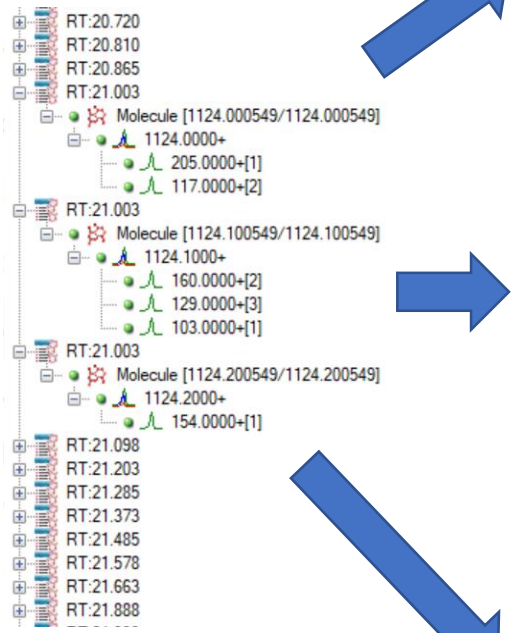


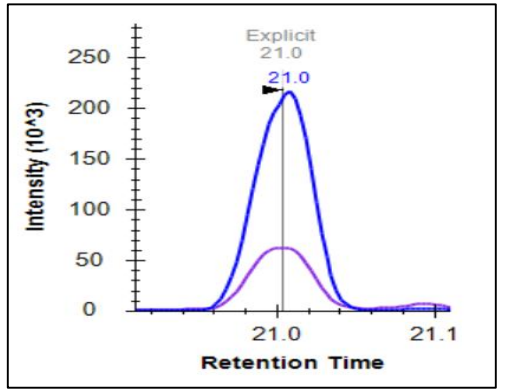
Deconvolution

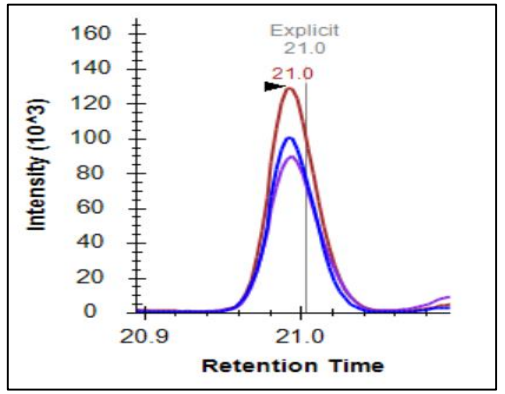


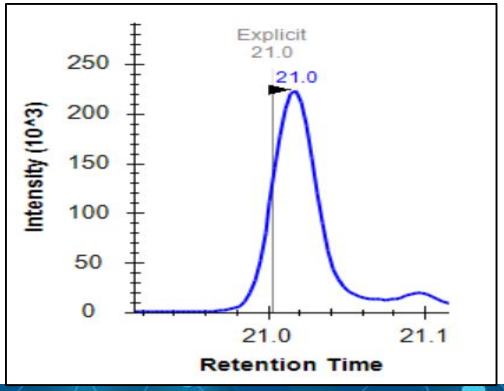
Deconvolution



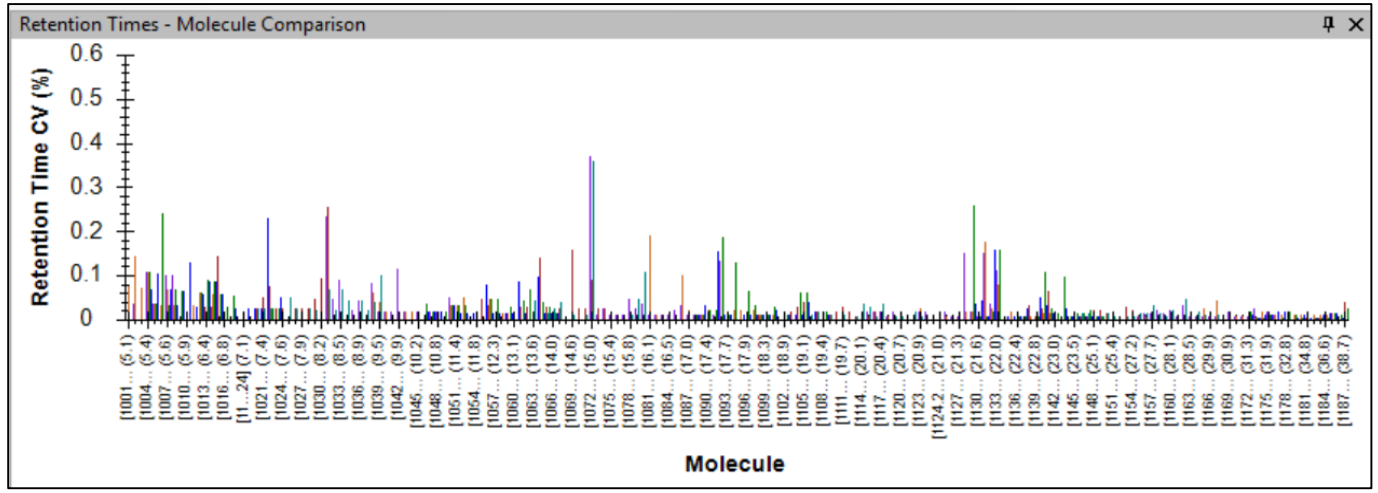


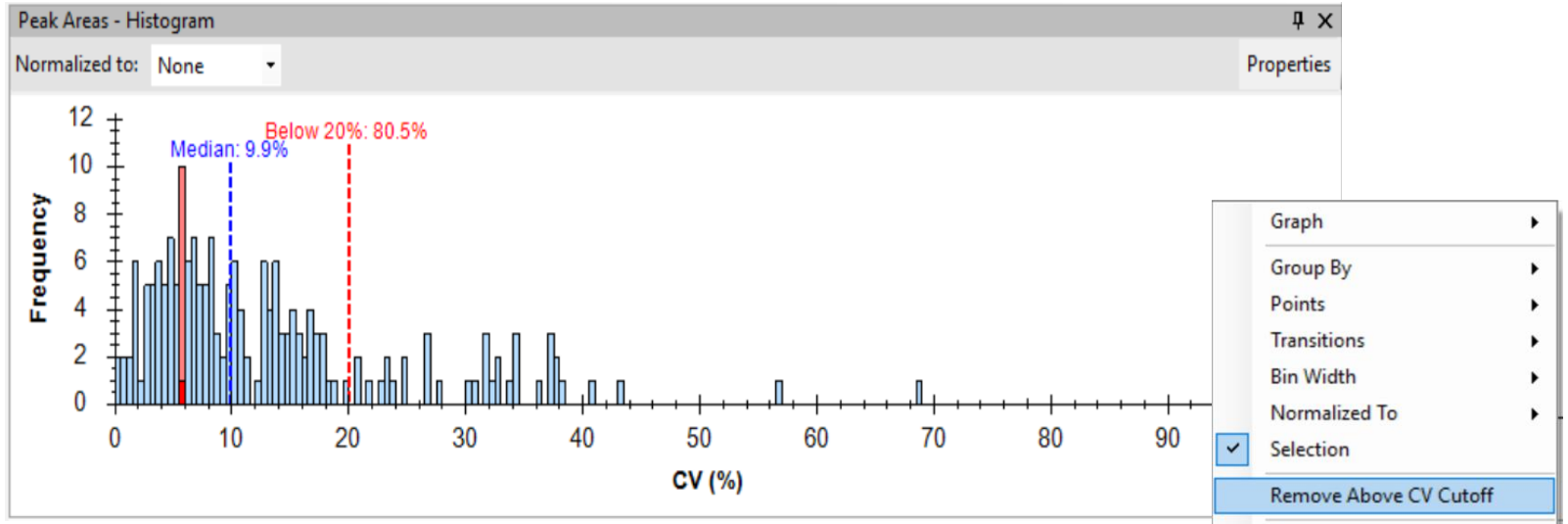






QC check and additional refinement

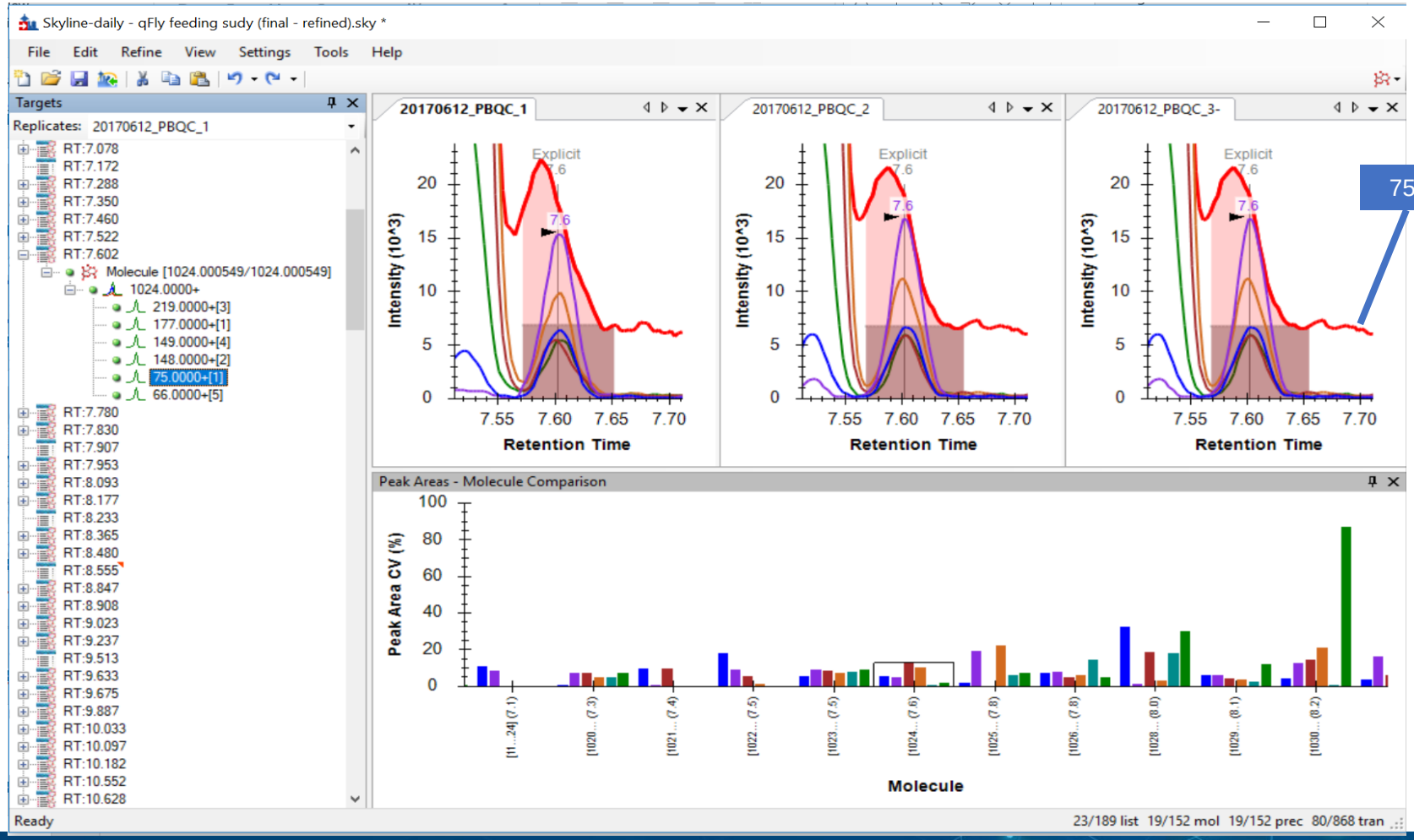




CV < 20%
S/N > 3
Min 2 ions per peak

QC check and additional refinement

Remove non-specific high background ions

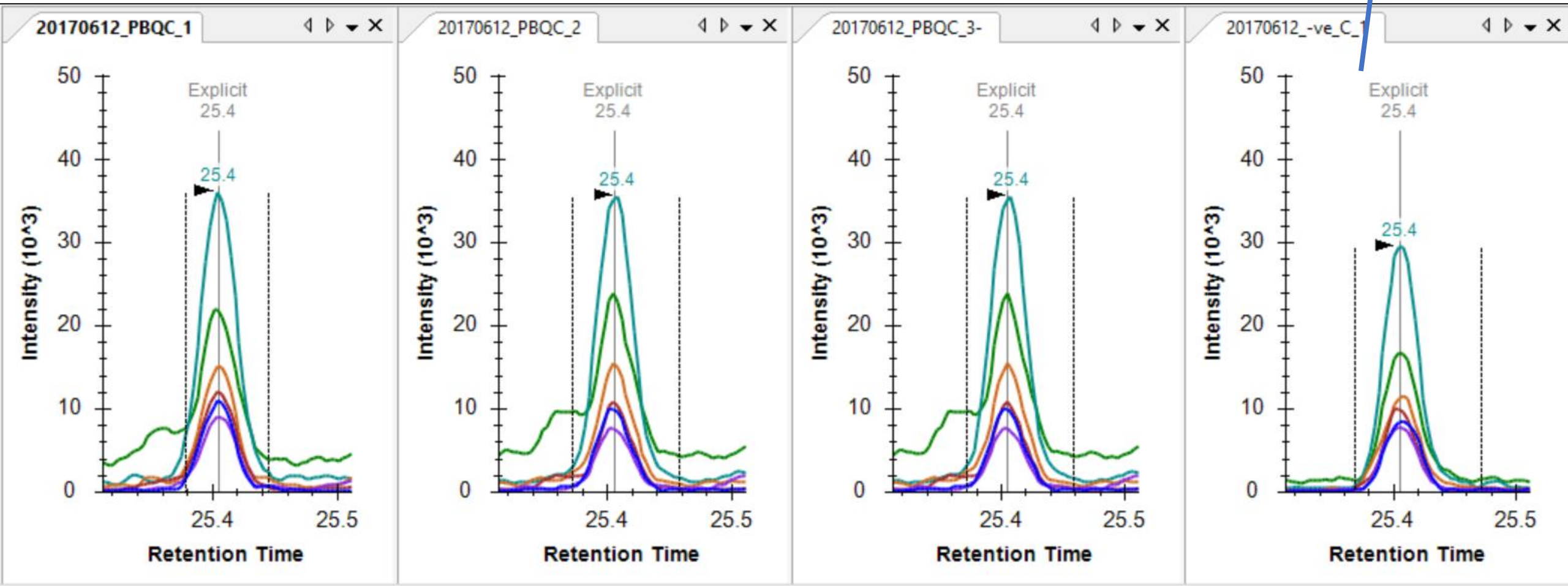


73 (TMS and septa bleed)
75
147 (TMS)
207 (column bleed)

QC check and additional refinement

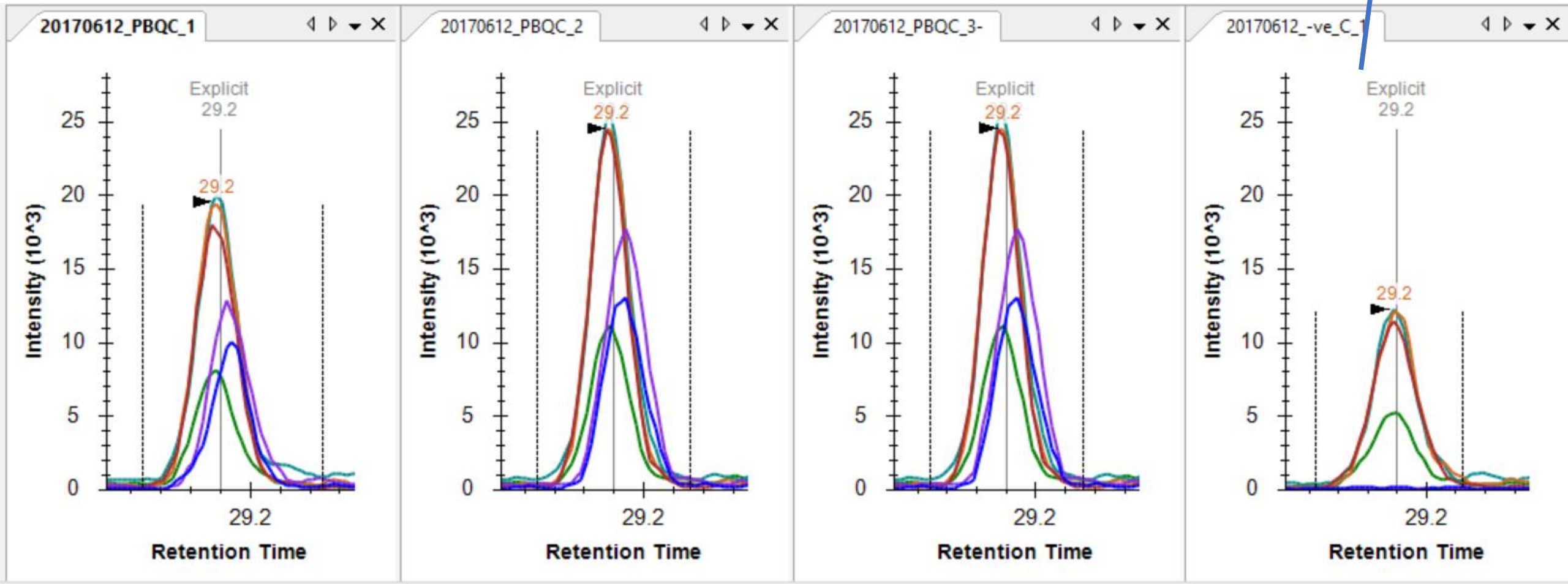
Remove compounds present in procedural blank

PROCEDURAL BLANK



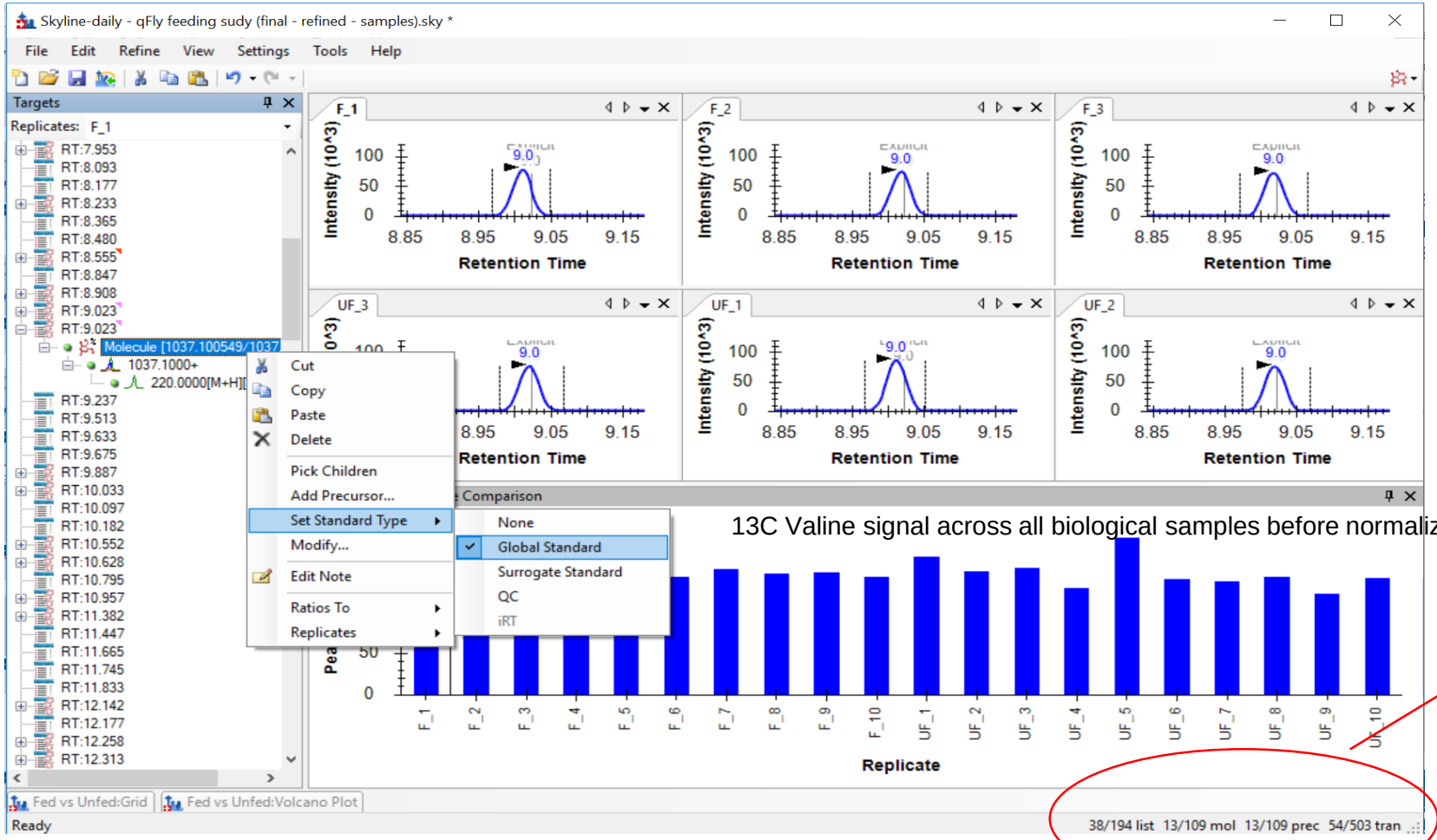
QC check and additional refinement

Remove compounds present in procedural blank



Normalisation

13C Valine and 13C sorbitol

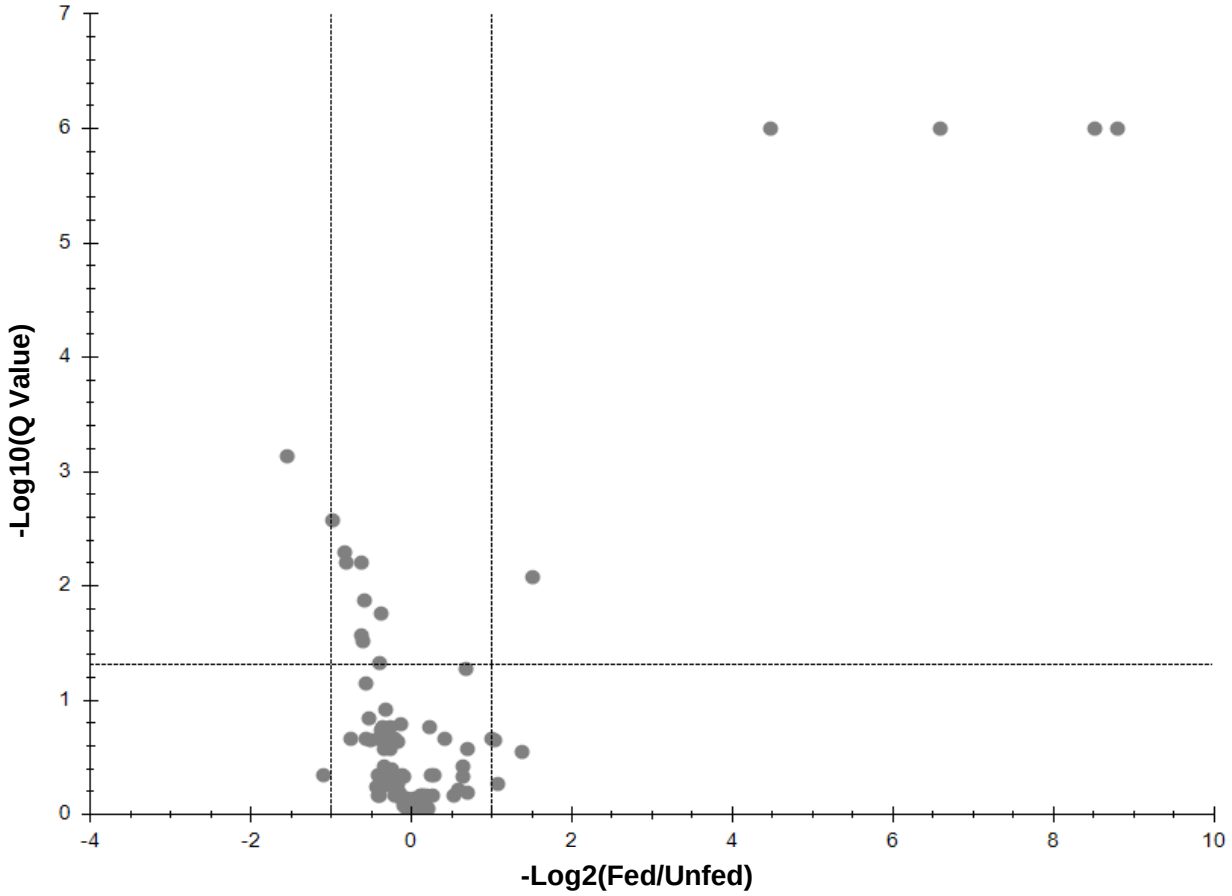


13C Valine signal across all biological samples before normalization

Final refined document has ions extracted for 108 chromatographic features.

Volcano plot

Each point represents a compound (chromatographic feature)

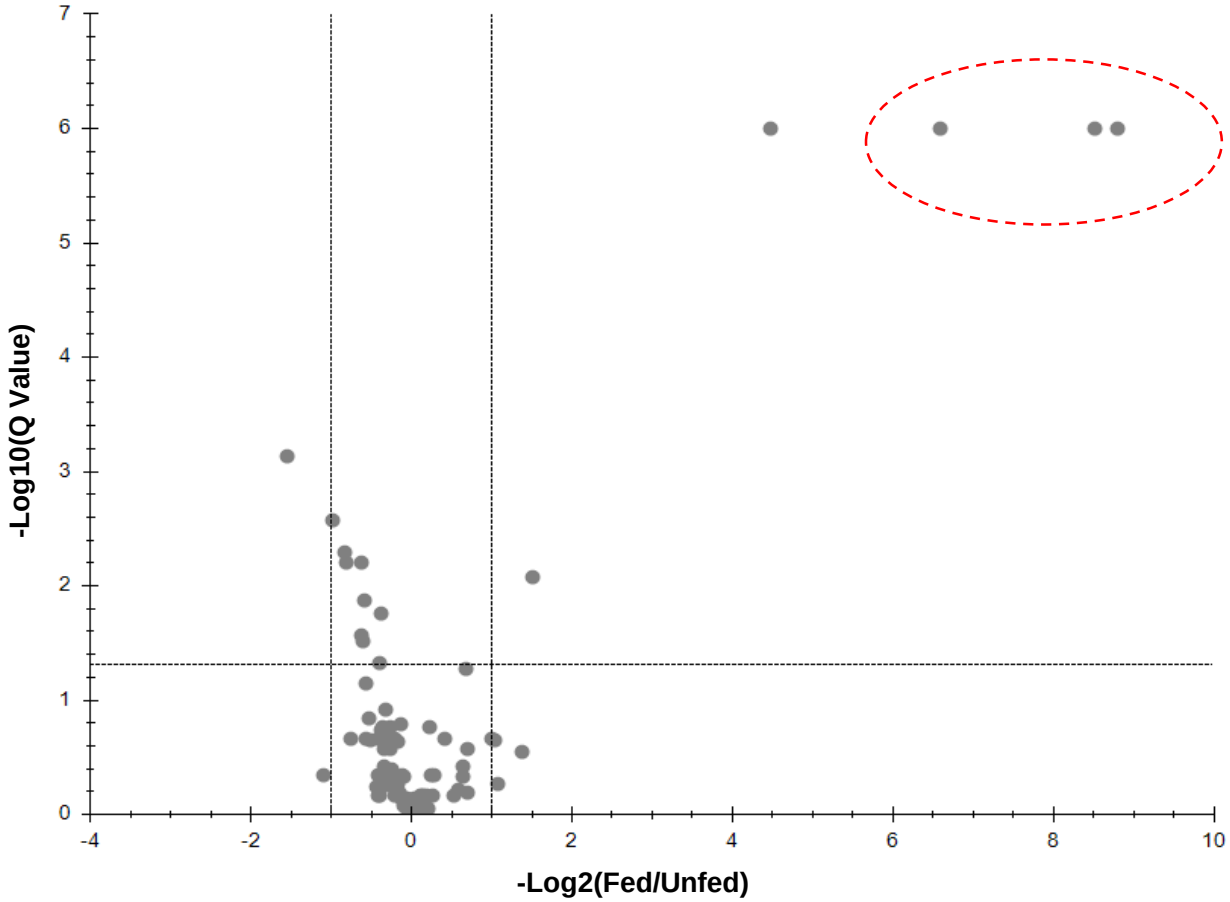


Q value = BH-adjusted p-value

Joel Herring

Volcano plot

Each point represents a compound (chromatographic feature)

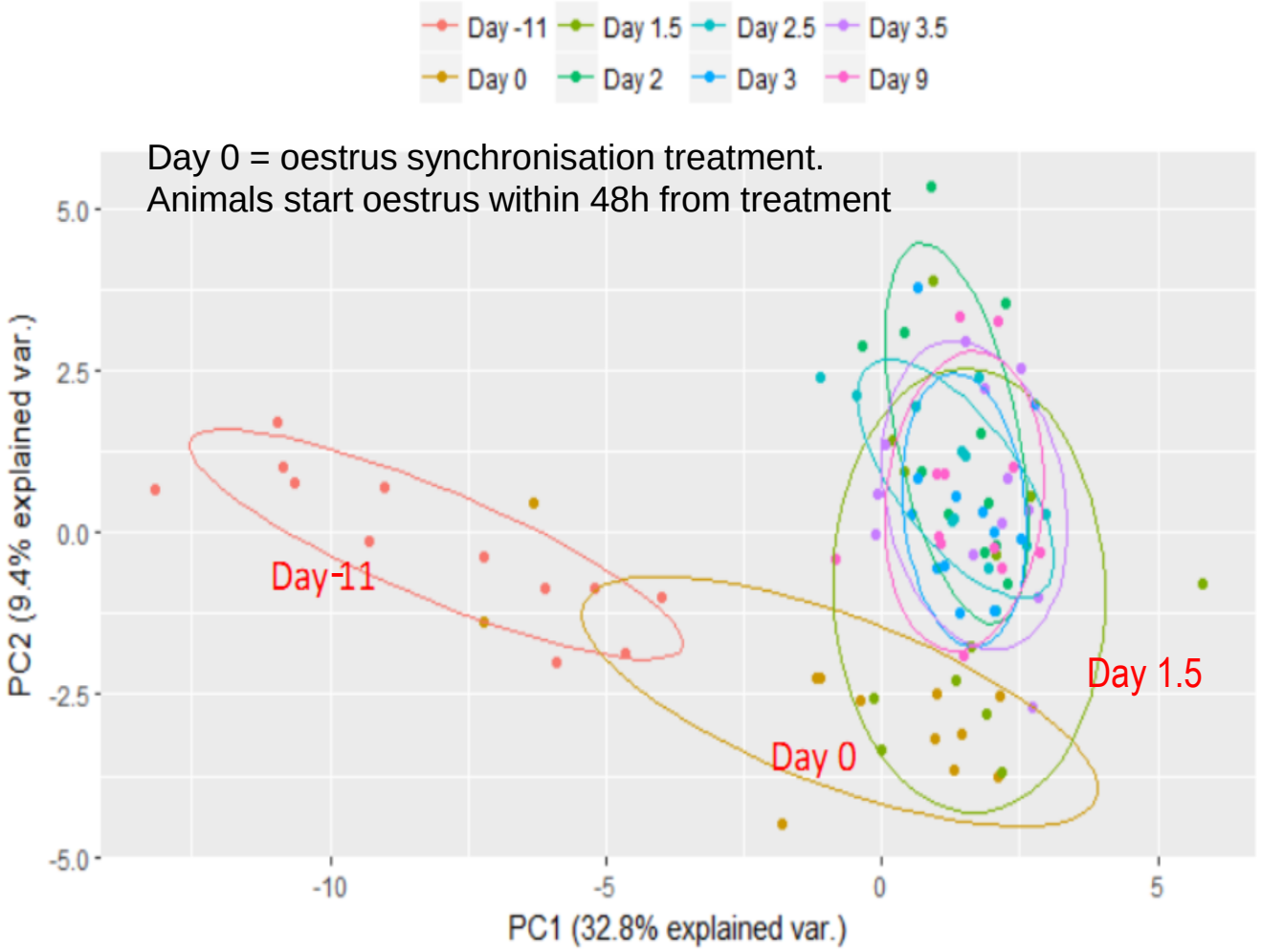


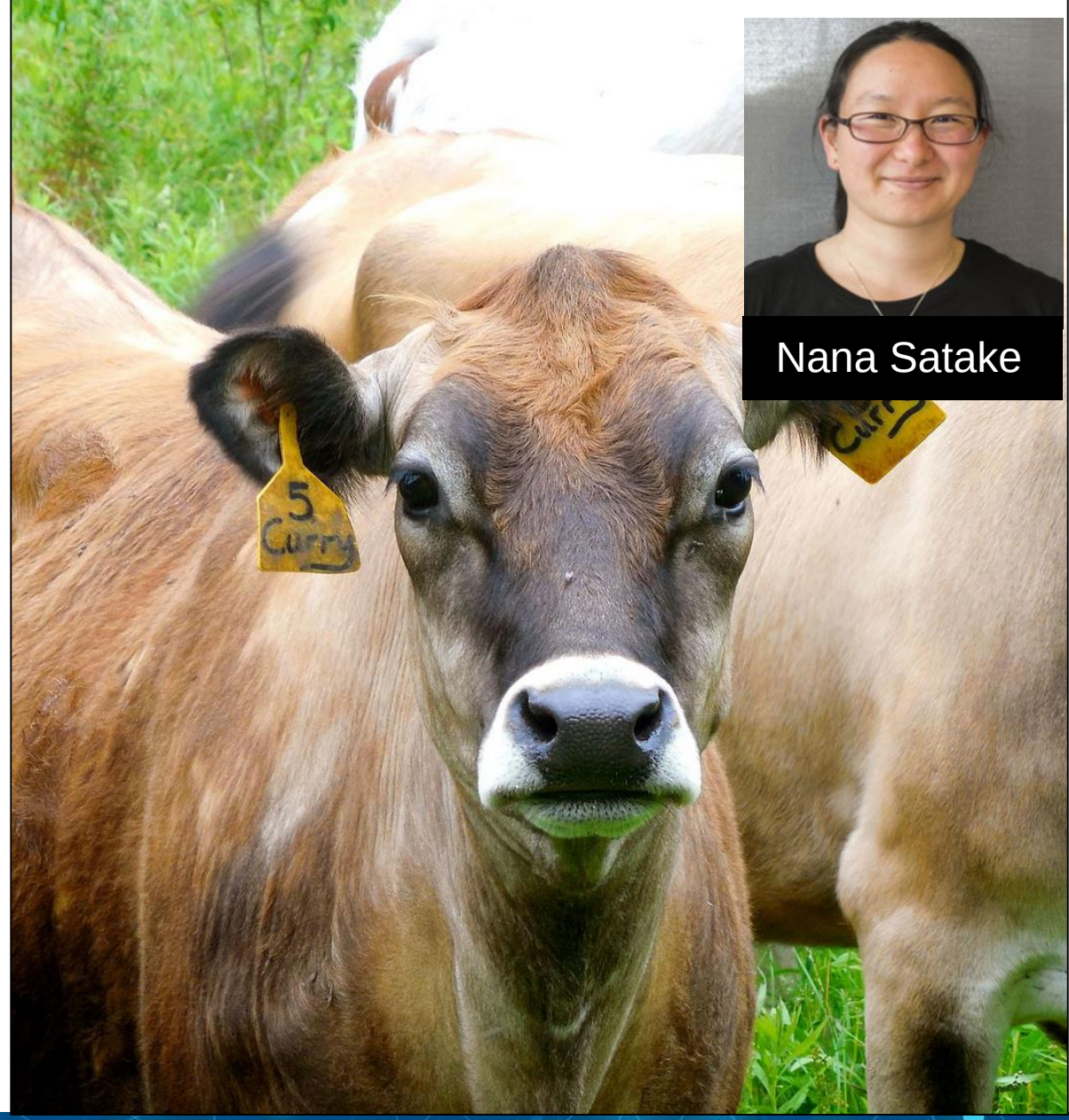
Q value = BH-adjusted p-value

Joel Herring

PCA plot

Each point represents a sample (an animal)





Nana Satake

Summary

- Established pipeline for processing metabolomics data in Skyline that is applicable to any GCMS data
- GCMS-based metabolomics data is among the most challenging to analyze
- The pipeline includes:
  - evaluation of instrument performance prior and during sample acquisition (PBQC)
  - targeted extraction of GCMS data including manual deconvolution and interference removal
  - quantitative assessment of 150-250 chromatographic features and their statistical analysis

- The pipeline has not yet been optimized for analysis of large datasets over the period of multiple days where significant RT shift or intensity differences may occur because it does not include RT alignment and correction based on PBQC.
- Work in progress (reduce manual steps, test different ways of processing i.e. using individual ions rather than sum of the ions).
- Skyline allows to standardize workflows across different ‘omics’

Acknowledgments

- Mr Joel Herring
- Dr Nana Satake
- Dr Tommaso Villa
- Mrs Silvia Gemme
- Prof. Tony Clarke

- Melbourne node of Metabolomics Australia

- Skyline Team

ADVERTISEMENT

Skyline workshop at CARF
9-13 September, 2019

Location: Brisbane, Australia
Trainers: B MacLean, E Borràs, C Chiva, B Searle, M MacCoss

Program: Mostly focused on SWATH-MS data analysis but will include small molecule component with GCMS data analysis.

QUT a university for the real world