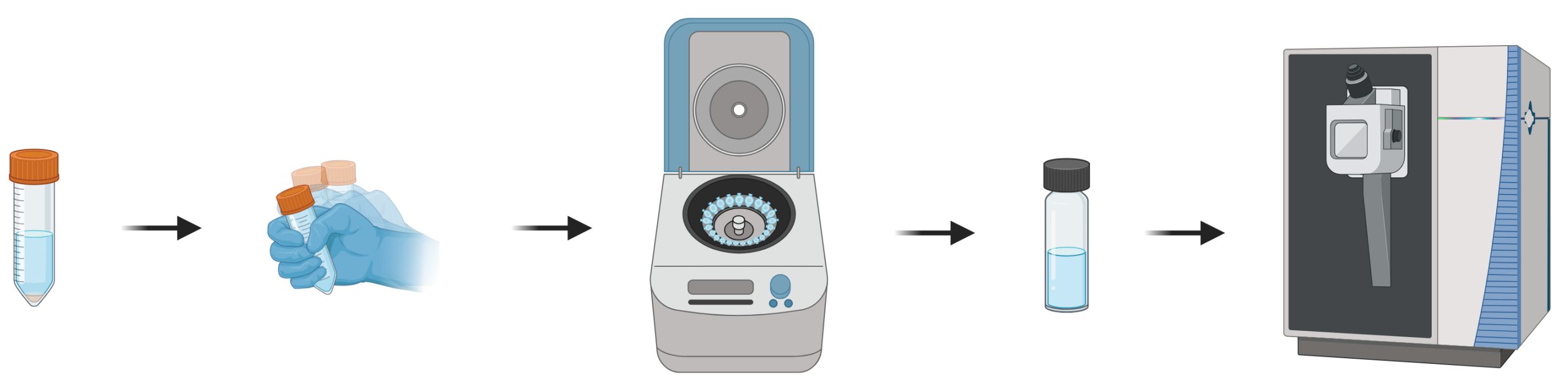
Data processing workflow for untargeted HRMS-based chemical hazard detection in maize

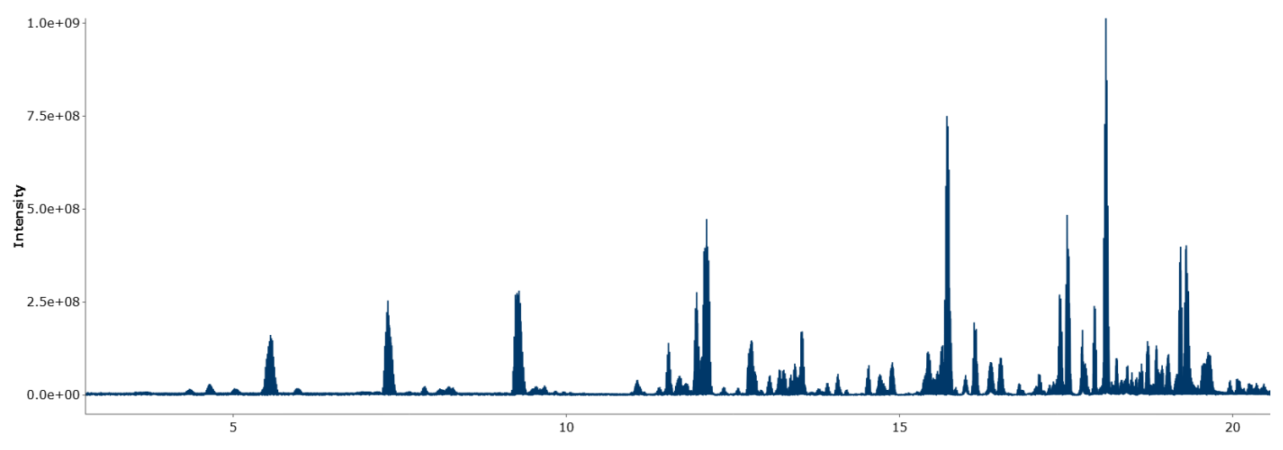
Xenia Pascari, Jakob Kühn, Stefan Weigel
Reference Centre for Food and Feed Analysis, German Federal Institute for Risk Assessment (BfR), Max-Dohrn-Str. 8-10, 10589 Berlin, Germany

High-resolution mass spectrometry (HRMS), when coupled with liquid chromatography has become a powerful tool for characterising small molecules in complex matrices. The data collected on different days, is affected by both systematic and random variability in response, retention time and mass accuracy. This can lead to significant information loss and may interfere with subsequent data analysis in untargeted workflows. We propose a six-step workflow that translates the raw files acquired across five batches into a single feature table, ready for subsequent data analysis.

1 Data acquisition



- **Generic sample preparation** that balances analyte recovery and matrix interferences
- Addition of **native** and **labelled standards** to monitor system performance
- Batch design that includes **QC samples** and **blanks** (system and processing)

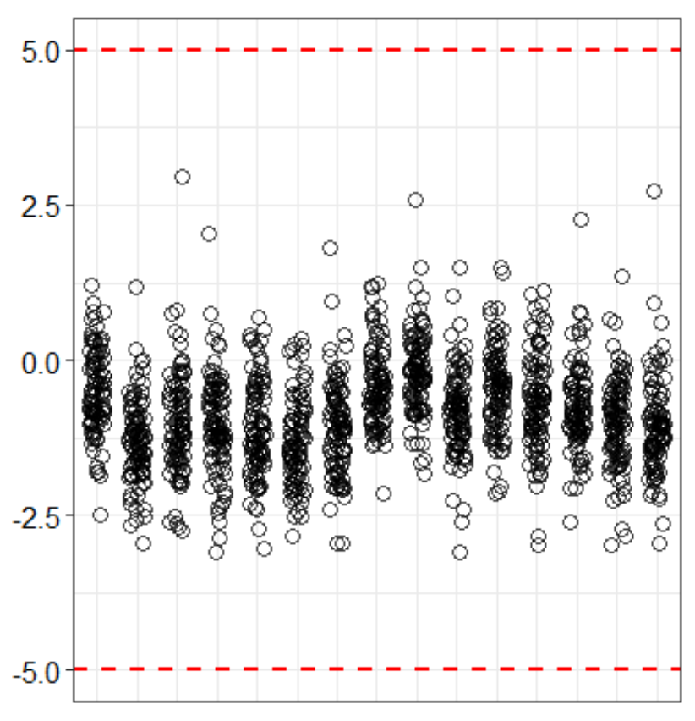


121 maize samples
5 Batches
Vanquish UPLC – Orbitrap Exploris 120
1 run / polarity

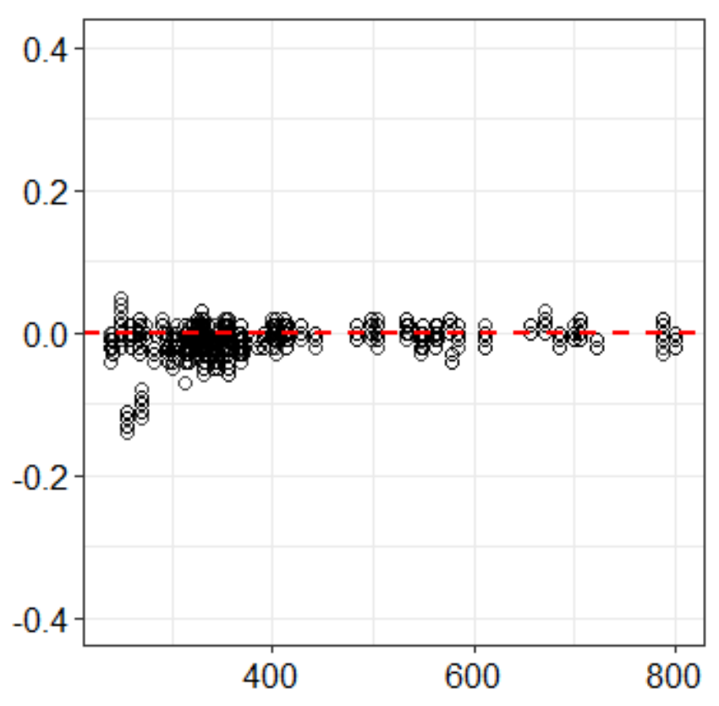
2 Assessment of System Performance

- Monitoring of **instrument performance** using analytical standards
- Assessment of blank injections (instrumental and processing) as well as quality control injections for both intra – and inter – batch stability
- Establishing of acceptance criteria for mass accuracy, retention time, system sensitivity

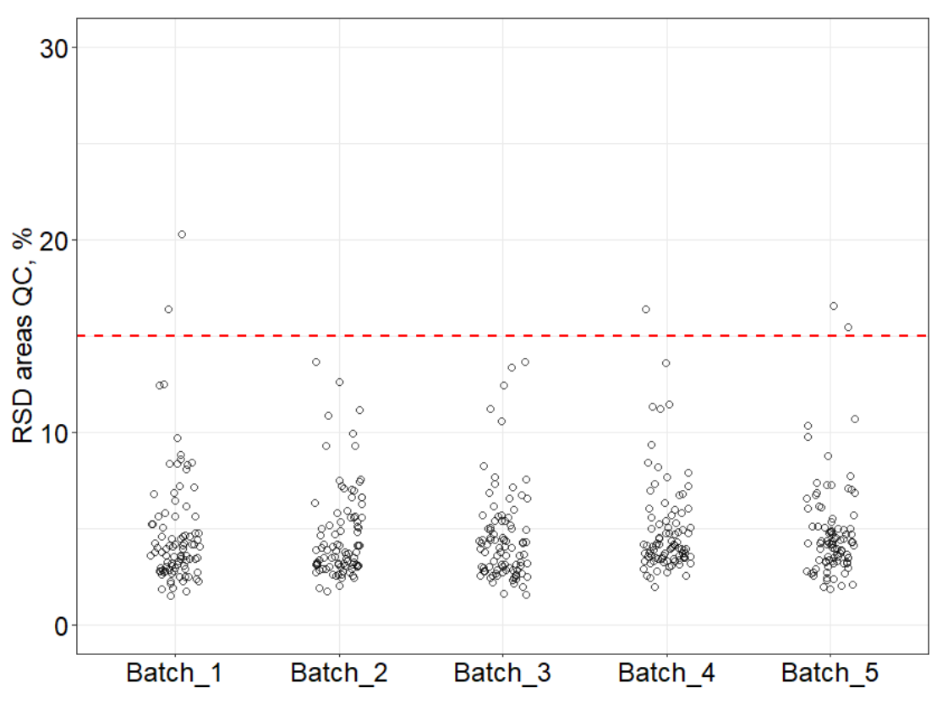
✓ **System and batch performance within acceptance criteria**



Mass accuracy < 5 ppm



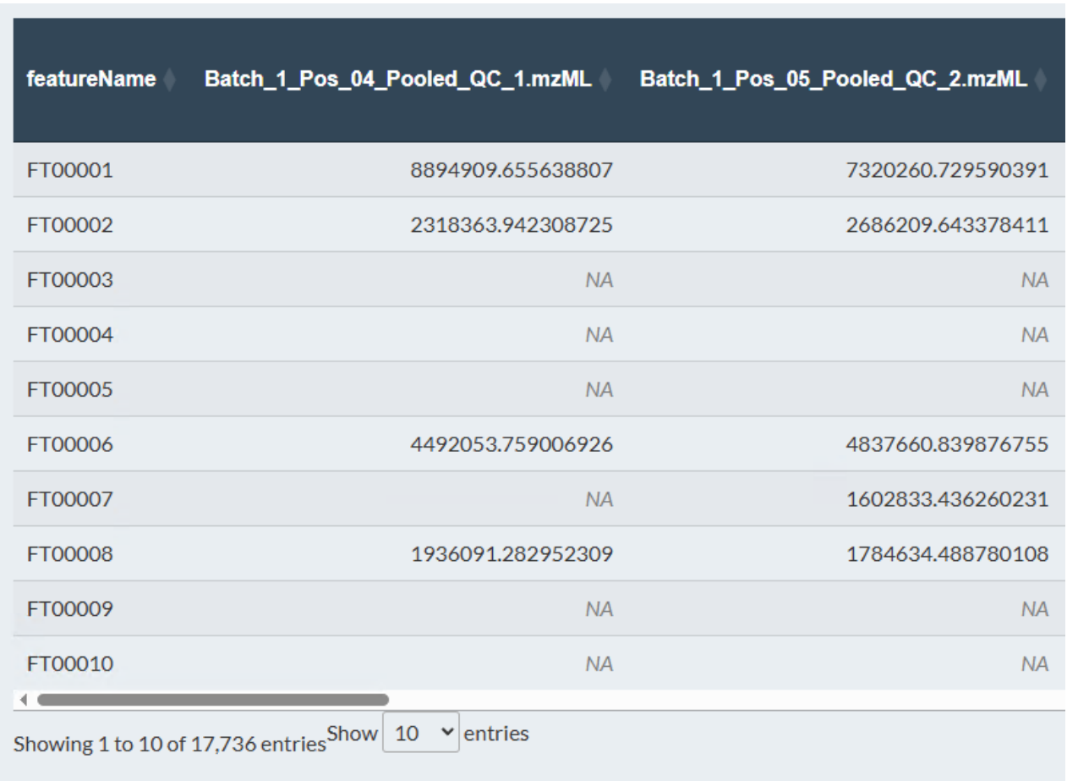
Retention time difference < 0.2 min



Relative Standard Deviation Areas < 15 %

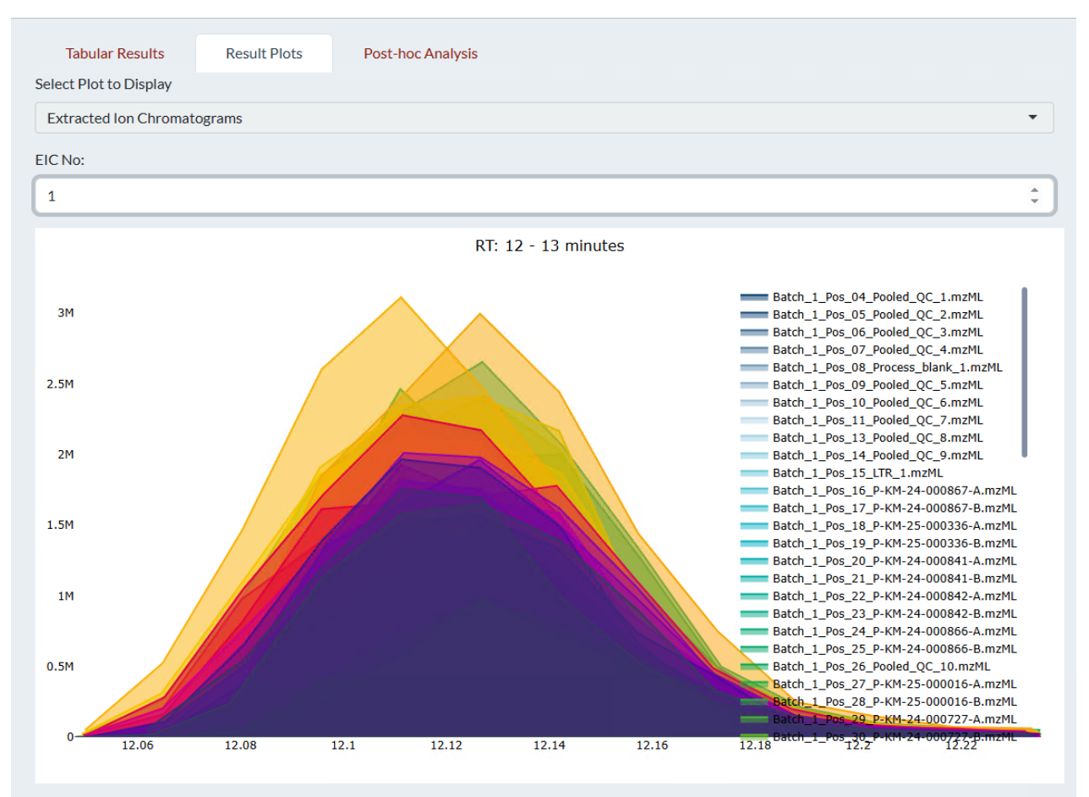
3 Feature Extraction

- In-house developed **user-friendly interface** that integrates open-source tools for feature extraction to support untargeted hazard identification
- Detection and alignment of feature within each batch
- Inspection of extracted ion chromatograms
- Generation of the feature table with m/z, retention time and areas



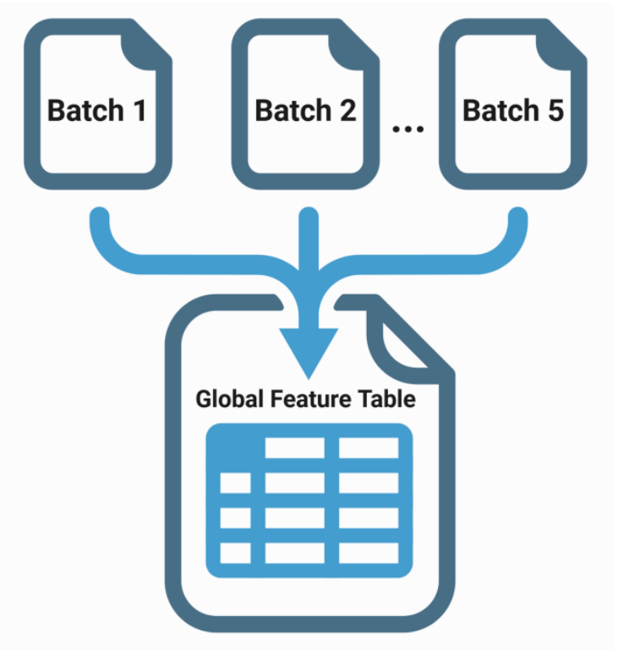
featureName	Batch_1_Pos_04_Pooled_QC_1.mzML	Batch_1_Pos_05_Pooled_QC_2.mzML
FT00001	8894909.655638807	7320260.729590391
FT00002	2318363.942308725	2686209.643378411
FT00003	NA	NA
FT00004	NA	NA
FT00005	NA	NA
FT00006	4492053.759006926	4837660.839876755
FT00007	NA	1602833.436260231
FT00008	1936091.282952309	1784634.488780108
FT00009	NA	NA
FT00010	NA	NA

Showing 1 to 10 of 17,736 entries Show 10 entries



The Application supports the inspection of the extracted features in both tabular form and plots of the extracted ion chromatograms

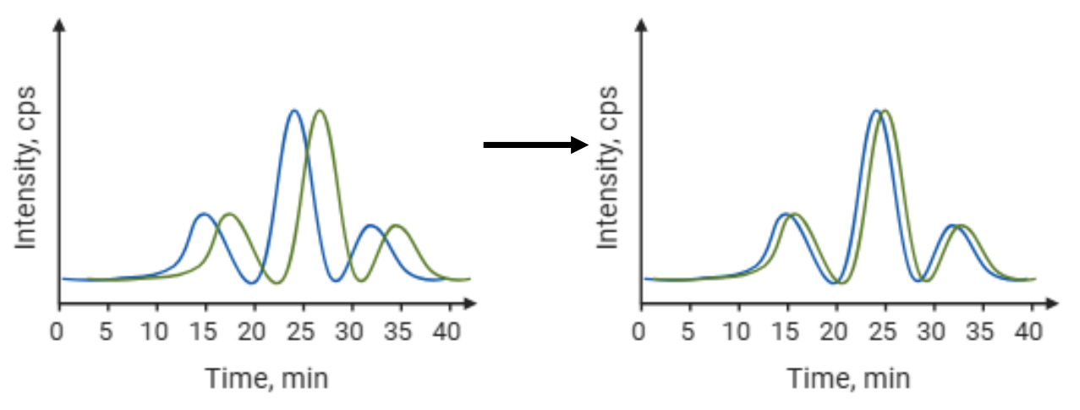
4 Batch combination & Peak Cleaning



- Concatenation of the five batches together using ***xcms***
  Subtraction of features identified in blank injections
- **Comprehensive Peak Characterization (CPC)** – algorithm that identifies and removes low quality peaks

5 Inter-Batch Feature Alignment

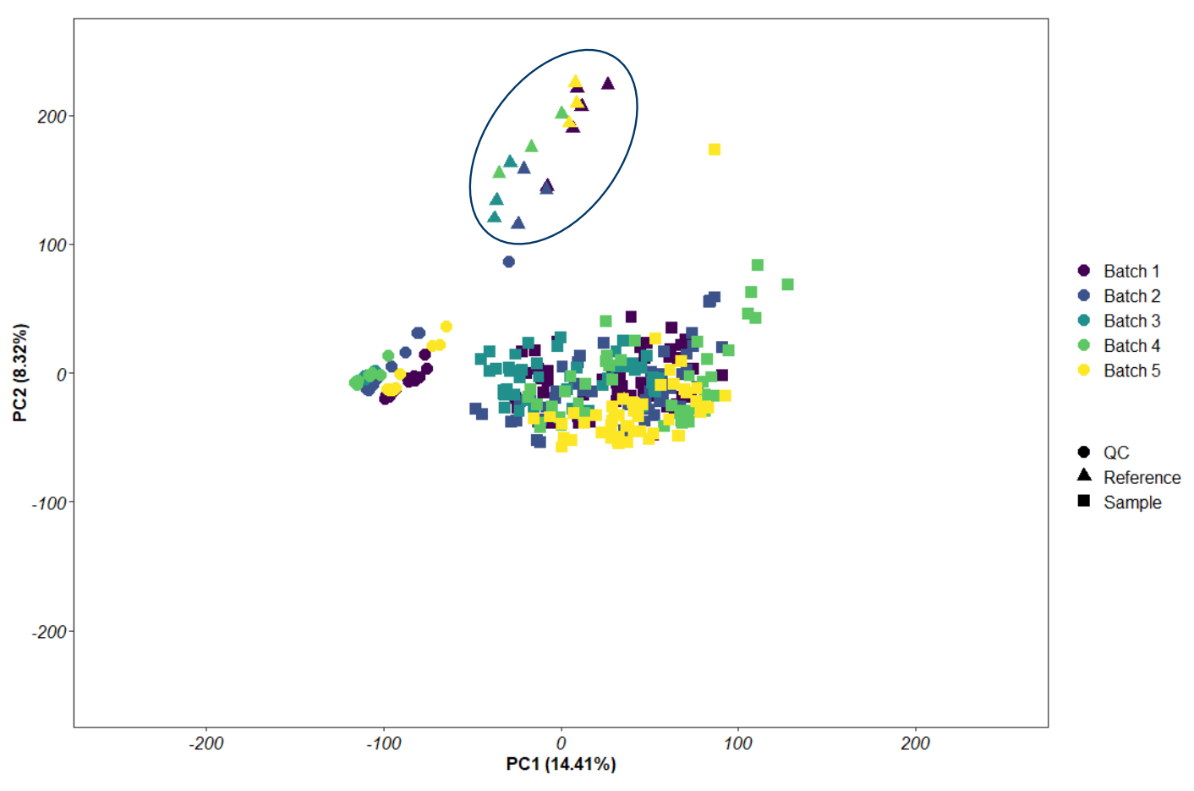
Use of ***BatchCorr*** package to align features based on presence/missingness on batch level and intensity drift correction between all the samples analysed across the five batches.



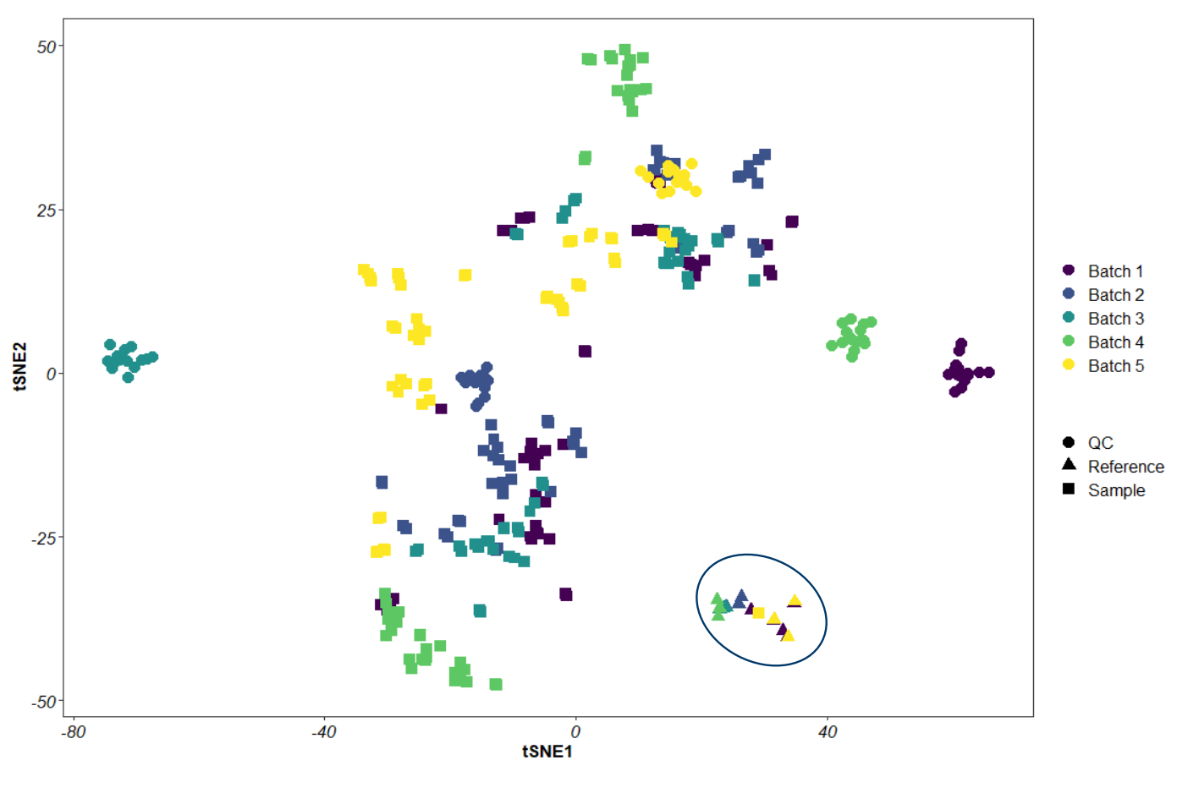
Schematic representation of alignment between samples

6 Data Inspection

- Use of **Principal Component Analysis (PCA)** in the assessment of the quality of QC and Reference samples beyond the spiked analytical standards: groupings of Reference and QC samples are an indication of the system's stability
- Optimization of **t-Distributed Stochastic Neighbourhood Embedding (tSNE)** to illustrate the local structures in the data.
- Grouping of Reference Samples on the **tSNE** projection complements the PCA results in estimating the inter-batch repeatability



Principal Component Analysis (PCA)



t-Distributed Stochastic Neighbourhood Embedding (tSNE)

✓ **High-quality, harmonised and well characterised data → Ready for reliable downstream anomaly search**

References
1. Broadhurst, D., Goodacre, R., Reinke, S.N. et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics 14, 72 (2018). https://doi.org/10.1007/s11306-018-1367-3
2. Brunius, C., Shi, L. & Landberg, R. Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. Metabolomics 12, 173 (2016). https://doi.org/10.1007/s11306-016-1124-4
3. Thrun, M.C., Projection-Based Clustering through Self-Organization and Swarm Intelligence (2017), https://doi.org/10.1007/978-3-658-20540-91