ACHIEVING ACCURATE REPRODUCIBLE RESULTS AND COMPUTATIONAL CHALLENGES IN LARGE SCALE UNTARGETED METABOLOMIC STUDIES

A. Pasamontes, V. Koval, A.C. Dubbelman, S. Bos, M.S. van Vliet, A.C. Harms, T. Hankemeier

Leiden Academic Centre for Drug Research – Leiden University, The Netherlands

INTRODUCTION

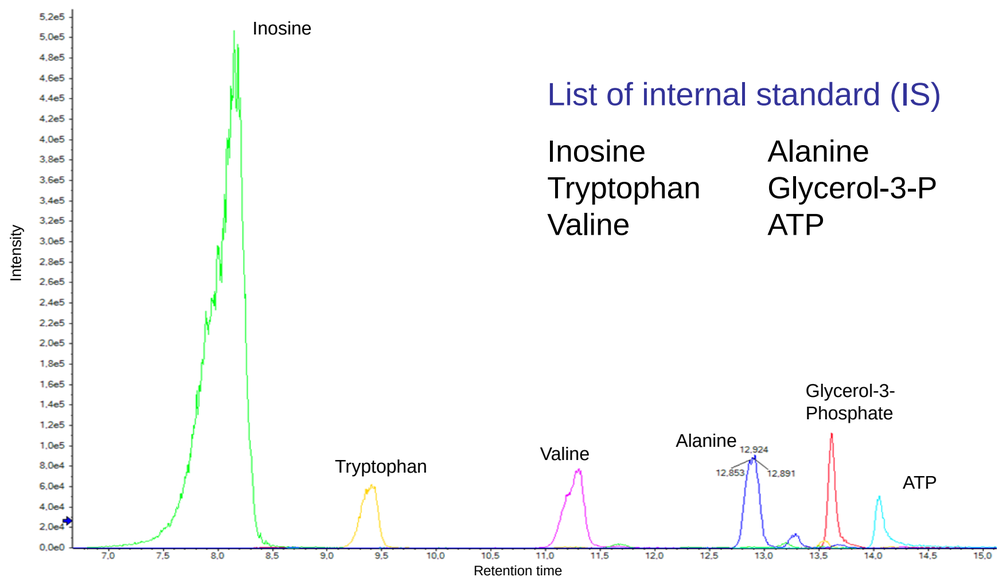
Untargeted feature extraction in metabolomics has been a long standing challenge in HPLC/MS [1]. Another level of complexity is added for mapping features between batches in large scale multi-batch studies (sometimes spread in time). Apart from the complicated task to map features detected in different samples and batches and merging them in a unified output, we also have to deal with analytical drift and possible changes in instrumentation over long running projects.

In this study, we propose a workflow for untargeted metabolomics including several additional techniques to assure consistent results in large studies. This includes library building from pooled samples, inclusion of standards to correct for detector response and retention index correction to facilitate accurate alignment. Previously, the retention time index was applied to GCMS to convert the measured retention time to system independent constants using Kovats retention index [2]. For HPLC, the offset in retention times is less constant over time than in GC, especially when working with less robust separations such as HILIC where the shifts may be unpredictable [3]. We show that with use of appropriate internal standards we can perform retention time alignments, which results in rugged metabolomics profiles and accurate relative quantitation for diverse samples measured in large scale studies.

METHODOLOGY

LC-MS method

A HILIC chromatographic method was coupled with quadrupole-time of flight mass spectrometry (Q-TOF) in positive ionization mode for fingerprint metabolomic studies. The retention time window of interest was between 7 and 15 minutes. The column used was: c-HILIC 100x2.1 mm, 3 µm, 100 Å. The internal standards were selected in order to get a uniform distribution over the retention time window (see Figure 1).



List of internal standard (IS)

Inosine	Alanine
Tryptophan	Glycerol-3-P
Valine	ATP

Fig. 1: Chromatographic coverage of the internal standards

Pipeline

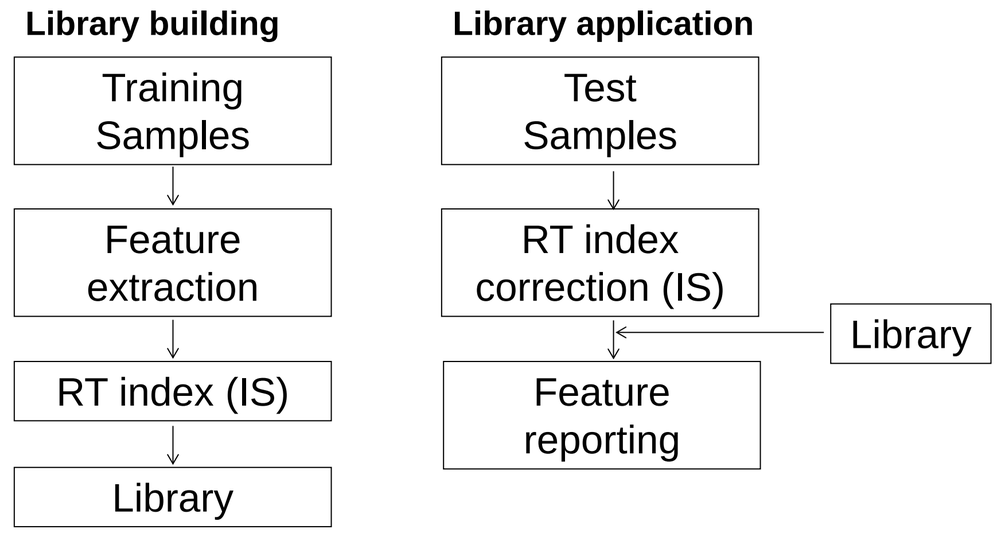


Fig. 2: Schematic overview of the pipeline

Pipeline description

Due to the necessity to combine current batches with future batches, this pipeline (Fig.2) will be applied. During the library building, untargeted analysis will be carried out in order to find all the features/chemicals. Then, retention time (RT) index will be calculated from the internal standards (IS) spiked in the samples. Features and RT index will be added to the library. In the second step, the RT index correction will be applied to the test samples for all the features previously added to the library.

RESULTS AND DISCUSSIONS

RT index (IS) workflow

- *Internal standard selection*
  The internal standards should have a good peak shape (no double peaks due to isomers) to avoid possible confusions during feature peak picking process
- *Spread uniformly the internal standards through run time*
  The internal standards should be spread over the retention time. Ideally one every minute.
- *Building models*
  Linear or non-linear models were tested. See Fig. 5.
  Number of internal standards per model: Change the number of internal standards and see if there is any improvement
- *Validation of the model*
  Validation is carried out using an external data set and evaluating the accuracy and variability of the predictions. See Fig. 3 and 4.

Linear model (6 IS)

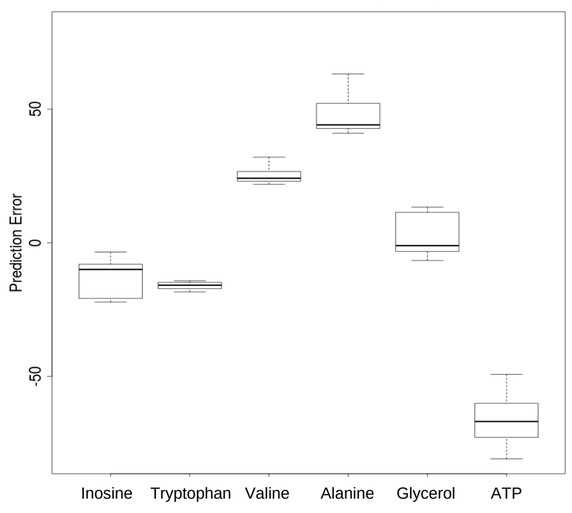


Fig. 3: Prediction error of each IS in a linear model

Non-linear model (6 IS)

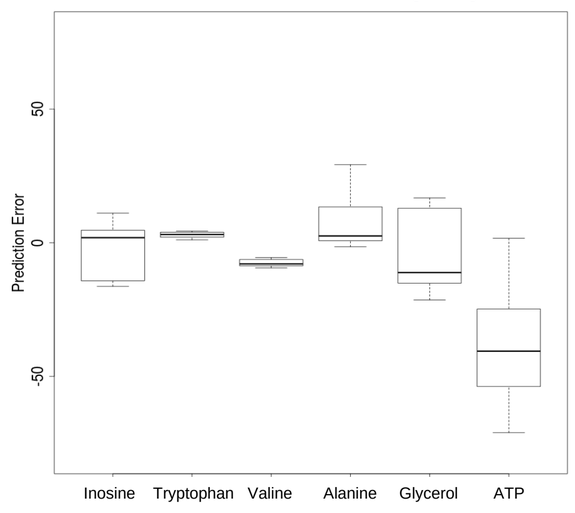


Fig. 4: Prediction error of each IS in a non-linear model

Non-linear versus Linear models

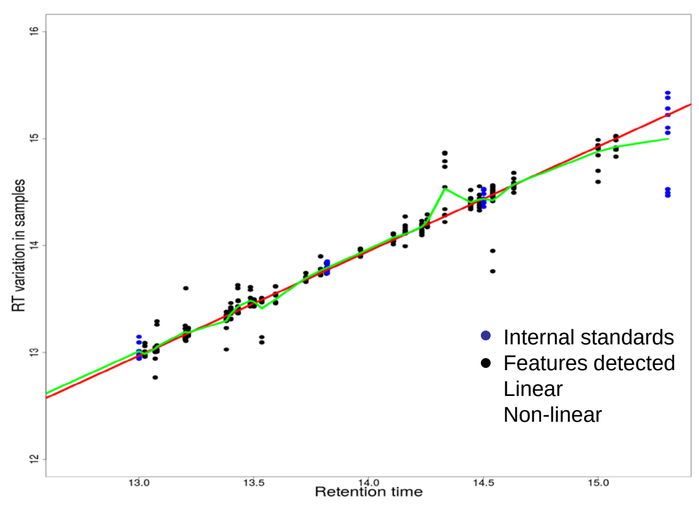


Fig. 5: Linear and Non-linear models for retention time drift

Fig. 5 shows the change in the retention time for different samples and we observed that non-linear regression represents better fit for the retention time. Fig. 3, Fig. 4 show that by using non-linear (quadratic) model for retention index we achieve higher predictive accuracy comparing with linear retention index

CONCLUSIONS

- Retention time alignment is a crucial step in building consistent workflow for metabolomic studies, specially in the large scale multi-batch studies which is prone to analytical changes.
- A non-linear model is superior to linear model alignment and the optimal number of internal standards should be selected.
- For untargeted projects retention time alignment is an important tool to ensure consistency of features found in different batches.
- This workflow will be validated and adjusted for different chemicals classes.

REFERENCES

[1] R. Di Guida, J. Engel, et al "Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling" Metabolomics. (2016) 12, 93.

[2] R. Smith, et al. "LC-MS alignment in theory and practice: a comprehensive algorithmic review" Briefing in bioinformatics, (2013) 16, 10.

[3] K. Podwojski K., et al. (2009). Retention time alignment algorithms for LC/MS data must consider non-linear shifts. Bioinformatics, (2009) 25, 758.

ACKNOWLEDGEMENT

PhenoMeNal is funded by European Commission's Horizon2020 programme, grant agreement number 654241