Fusion of LC–MS data from different measurement sessions using pooled blood plasma as transfer sample

H Draisma[1], F van der Kloet[2], T Reijmers[1], I Bobeldijk-Pastorova[2], J Meulman[3], F Estourgie–van Burk[4,5], M Bartels[5], J van der Greef[1], D Boomsma[5], T Hankemeier[1]

[1] Leiden University, LACDR, Leiden, The Netherlands; [2] TNO Quality of Life, Zeist, The Netherlands; [3] Leiden University, Mathematical Institute, Leiden, The Netherlands;
[4] Department of Paediatric Endocrinology, Institute for Clinical and Experimental Neurosciences, VU University Medical Center, Amsterdam, The Netherlands;
[5] Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands

Contact: **h.draisma@lacdr.leidenuniv.nl**

Introduction

We have implemented a strategy that corrects for batch effects occurring between consecutive measurement sessions with a liquid chromatography–mass spectrometry platform that analyzes lipids. In this case two batches of human blood plasma samples from in total 182 different individuals were analyzed with almost one year in between.

Four different internal standards were used to calculate relative responses for all 59 lipids from five classes that were detected in the samples analyzed in the first (y1) and second (y2) year.

Before batch effect correction, the data from both sessions display separation along the first two principal components upon Principal Components Analysis (PCA) (Figure 1A).

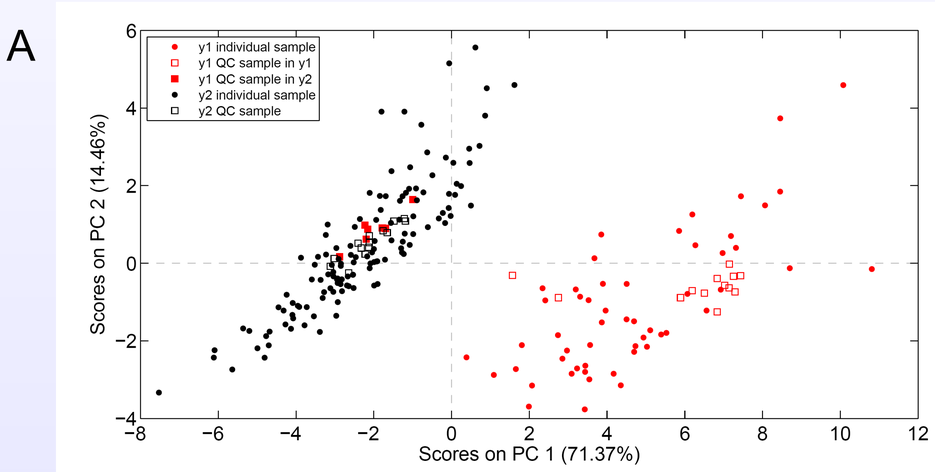

Figure 1. PCA scores (A) and loadings (B) based on relative responses in y1 and y2 before batch effect correction. Data were mean centered prior to PCA. Abbreviations: LPC, lysophosphatidylcholines; PC, phosphatidylcholines; SPM, sphingomyelins; ChE, cholesterol esters; TG, triglycerides.

Batch effect correction

To correct for the offset in the data caused by the time gap between y1 and from y2, we used a sample consisting of pooled study samples from y1 (y1 QC sample) that was measured during both sessions (Figure 2) [1, 2].

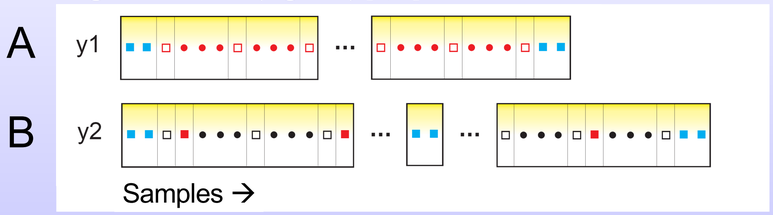

Figure 2. Measurement order in y1 (A) and y2 (B). ■: calibration samples. Remaining symbols in A and B correspond to those in Figures 1A and 3A, respectively: ● y1 individual sample; □ y1 QC sample in y1; ■ y1 QC sample in y2; ● y2 individual sample; □ y2 QC sample

For each of the lipids detected in y1 and y2, a correction factor was defined as the ratio of the median relative responses calculated from the y1 QC sample measurements in both years.

The correction factors of all lipids were in the range (0.67–5.4).

These factors were then applied to correct the relative responses for each lipid in the y2 individual study samples as follows:

$$y2value_{corr} = y2value_{original} \times \frac{median(y1QC\_in\_y1)}{median(y1QC\_in\_y2)}$$

After this standardization, the PCA scores on the first two principal components suggest a marked reduction of the y1–y2 offset (Figure 3A).

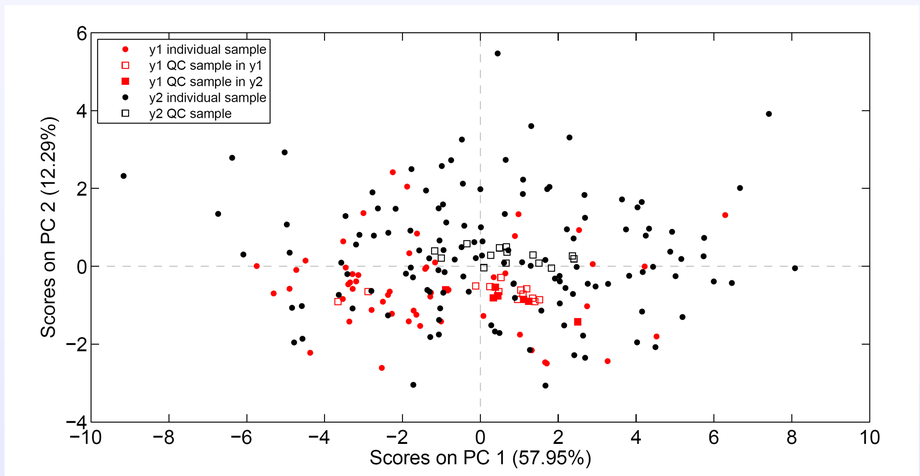

Figure 3. PCA scores (A) and loadings (B) based on relative responses in y1 and y2 after batch effect correction. Data were mean centered prior to PCA. For explanation of abbreviations see the legend to Figure 1.

Conclusion

We have successfully applied a method for batch effect correction that is based on measurement of the same pooled study sample in all batches. This method is particularly useful when there are no reference compounds available for all measured analytes, which is typically the case in e.g. metabolomics studies [3].

References

2. Van der Greef, J. *et al* (2007). J Proteome Res 6, 1540–1559
3. Van der Kloet, F. *et al*. In preparation for Metabolomics
4. Sangster, T. *et al* (2006). Analyst 131, 1075–1078

Universiteit Leiden

Leiden / Amsterdam Center for Drug Research
P.O. Box 9502, 2300 RA Leiden
The Netherlands
Universiteit Leiden
Vrije Universiteit Amsterdam

Netherlands Metabolomics Centre

CENTRE FOR Medical Systems Biology

BioRange activity of NBIC

Vrije Universiteit Nederlands Tweelingen Register