Typical LC-MS metabolomics workflow for profiling urine samples of patients with colorectal cancer

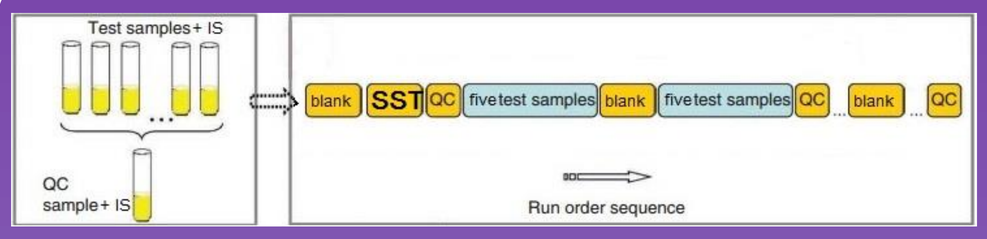
Ivan Plyushchenko [1], Timofey Bolotnik [1], Dmitry Shakhmatov [2], Sergey Achkasov [2], Oleg Sushkov [2], Oleg Shpigun [1], Igor Rodin [1]

[1] Lomonosov Moscow State University, Moscow, Russia

[2] State Scientific Center of Coloproctology, Moscow, Russia

plyushchenko.ivan@gmail.com

Design of experiment & run order



- Dilute & shoot – for sample preparation as most universal and reliable approach
- QC samples - as SST (by controlling reproducibility of retention times, signal intensities and mass accuracy in terms of CV and Pearson`s correlation coefficients) and for optimization of preprocessing procedure.
- Replicates – for checking stability (similarly as for QC)
- Blanks - controlling contaminations, carry-over and flushing column
- Randomization - to ensure that no bias is introduced
- MS tuning and ion source washing – after each batch
- IS (papaverine) – for choosing both time alignment and integration parameters; also for determination of quality control metabolites (with high correlation coefficients for IS)

Short overview of experimental part

- LC-MS/MS QQQ instruments (“Agilent Technologies”, USA)
- Waters Acquity UPLC BEH; 1.7 µm column (“Waters”, USA) with guard column
- 60' long chromatographic gradient for high resolution
- The divert valve was opened to the waste line during first 0.5 and after 45 min
- Creatinine in urine was determined by the kinetic method of Jaffe
- 40 urine samples from 3 groups of patients

Software

- “ProteoWizard” – converting raw data
- “iMet-Q” – integration and peak table construction
- “MetaboAnalyst” – for missing values imputation, univariate, unsupervised statistical analysis and PLS-DA
- “NOREVA” and “NormalizeMets” (R package) – for utilization of different signal drift correction methods
- “Rattle” (GUI for R language) – for supervised statistical analysis
- “Excel” and “MassHunter” – basic operations

Optimization of preprocessing algorithm

- ✓ 2 normalization methods – by creatinine, MSTUS
- ✓ 9 methods of signal drift correction – Contrast, Cubic Spline, Cyclic Loess, EigenMS, Quantile, Total Sum, VSN and 2 types from “NormalizeMets”: RUVrand & RUVrandclast
- ✓ 3 methods of missing values imputation – KNN, Median, Half Minimum

all combinations

Evaluation was carried out by:

1) CV criterion for all peaks in all QC samples from all batches (maximum number of features in 50% cut-off)
2) Maximum values for classification accuracy from different unsupervised methods (Random Forest, Decision Tree, SVM with Linear and Gaussian Radial Basis kernel functions)

The best preprocessing algorithm: HM for MSI, Creatinine normalization, no correction

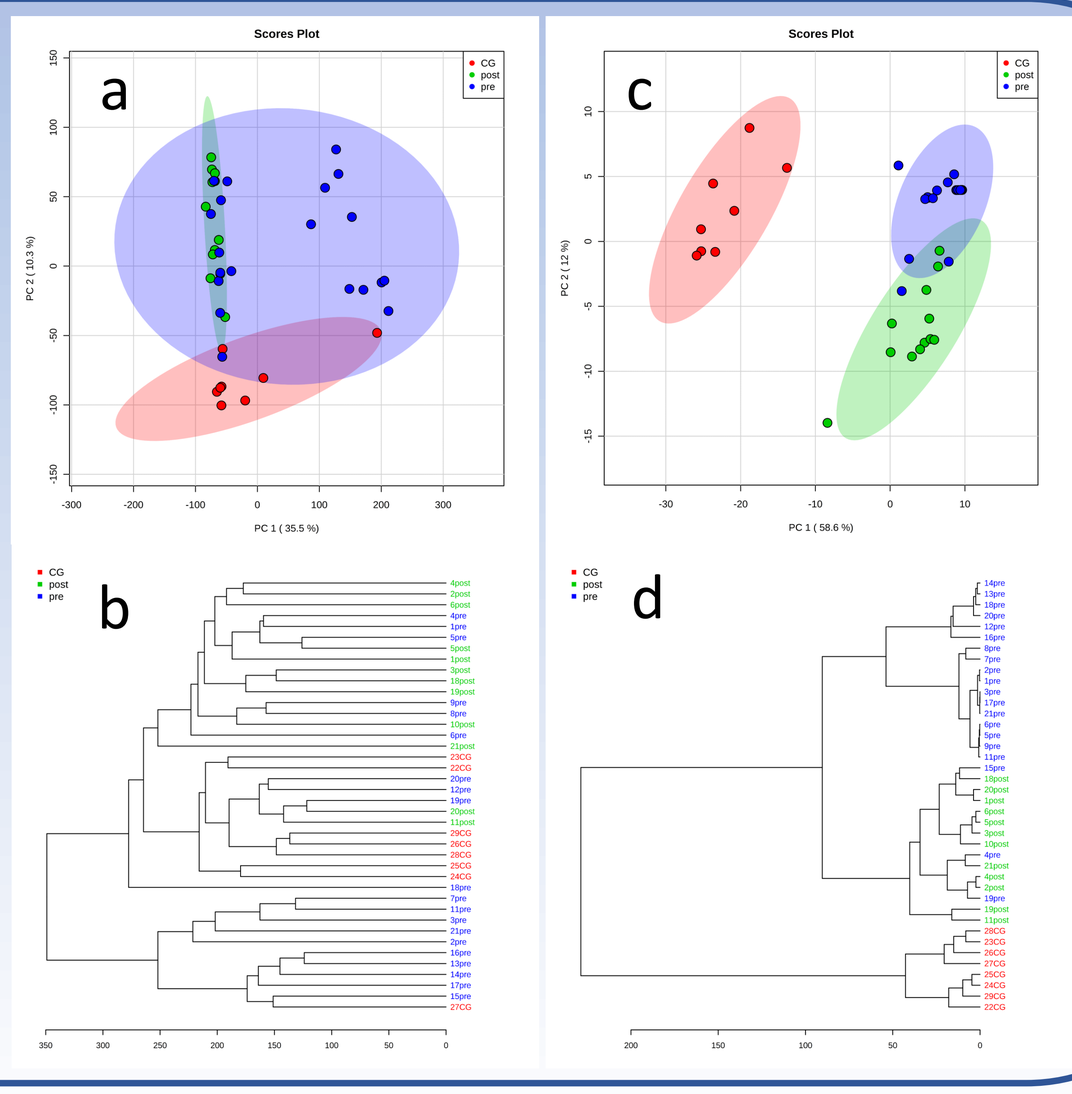
%, of all features	%, CV cut-off	Tree	RF	SVM(rbf)	SVM (lin)	Method
87	50	75.1	91.7	66.6	66.6	Classif. accuracy

Results

<u>for verification of the result, data with chosen combination of preprocessed methods were put to the following procedure:</u>

- Top features with VIP values > 3 from PLS-DA and p-value < 0.001 were extracted from data
- New data with only selected features were subjected to unsupervised statistical analysis (Hierarchical Cluster Analysis; PCA)
- All results were obtained with log transformation

Left side (a, b) – results from raw data in chosen combination, right side (c, d) – data with only selected features



Conclusion

This pragmatic decision procedure for selection and evaluation of preprocessing methods in metabolomics studies can be used in other relatively short studies such as: clinical research, examination of the metabolism of drugs, plant metabolomics, pharmacokinetics, etc.