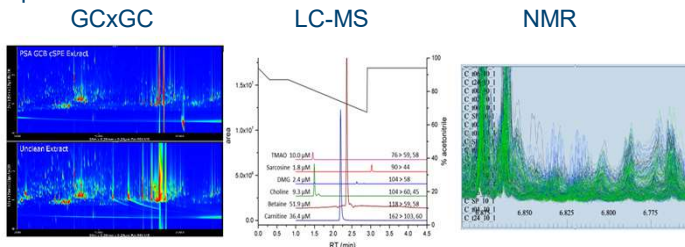


A new workflow combining R packages for statistical analysis of metabolites

Paola G. Ferrario

Motivation (nutrition context)

- In nutrition research: interest in metabolites (amino acids, sugars, fatty acids, hormones, etc...) and their association with some attributes of human beings
- representative research question: Is there any association between age (X) and the metabolites (Y)?
- examples of metabolites spectra measured on the different platforms:



Issues for the biostatisticians

Treatment of p metabolites ($Y_i, i=1\dots p$):

- coming from different area (plasma/urine)
- measured on different platforms, with different precision
- with heterogeneous variances, correlated
- differently distributed
- sometimes with only a few values (tied data)
- sometimes not measurable under a "limit of detection" (left censored)

Moreover:

- not complete measurements (missing values)
- more analytes than probands ($p > N$ design, N sample size)
- claim for association (relation type between X and Y)
- role of other important probands attribute, i.e., sex, BMI (i.e., competing covariates)
- Interest in each metabolite separately or as patterns (univariate vs. multivariate analysis)

Suggested procedure: Model

$$h_i(Y_i) = \beta_i X + \varepsilon_i$$

- β_i : effects (interpretation depending on errors ε_i)
- h_i : different, str. monotone increasing transformation functions, to be estimated by the so called most likely transformation (Hothorn T et al. (2018) Scand J Stat 45(1):110-134)
 - in a likelihood framework (possible even by censored or tied data)
 - facilitated by parametrization of h (basis functions by Bernstein polynomials)
 - advantage: differently distributed (left-skewed, tied,...), complete, and censored metabolites analyzed by unique approach
- For multivariate analysis: joint distribution of test statistics by multiple marginal modeling approach (Pipper CB at al. (2012) J Royal Stat Soc 61(2):315-326)

Implementation

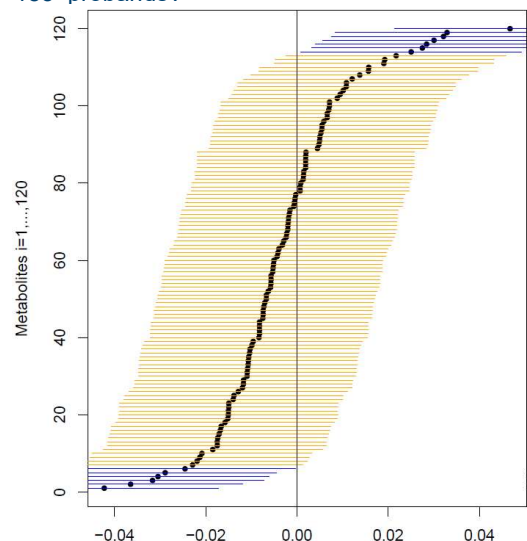
Combining the following CRAN packages in a workflow: *survival*, *tram*, *mlt*, and *multcomp*

Code-availability

R-Code that demonstrates the workflow and illustrative application on available data online as preprint on the BioRxiv (supplementary materials):
<https://www.biorxiv.org/content/10.1101/848812v1>

Illustrative application

- Demonstration of the workflow using public data (Thévenot EA et al. (2015) J Proteome Res, 14(8):3322-3335)
- Is there any association between 120 metabolites and age of 183 probands?



- visualization of the results by confidence intervals, ordered by increasing effects: The blue lines are the metabolites that show association with age

Contact

Paola G. Ferrario, Max Rubner-Institut
Federal Research Institute of Nutrition and Food
Department of Physiology and Biochemistry of Nutrition
Haid-und-Neu-Str. 9, 76131 Karlsruhe, Germany
paola.ferrario@mri.bund.de
ORCID: <https://orcid.org/0000-0003-1769-6799>