

The use of untargeted metabolomics for analysing metabolite biomarkers in colorectal cancer: methodological and statistical challenges

Name	Anne J.M.R. Geijsen, MSc anne.geijsen@wur.nl
Function	PhD candidate Nutrition and Disease, Human Nutrition (Wageningen University)
Institute	Metabolomics and Molecular Informatics, European Bioinformatics Institute (Hinxton, Cambridge, UK)
Supervisor institute	Dr. Reza Salek reza.salek@ebi.ac.uk
Period	January – March 2017

Background

Colorectal cancer (CRC) is the second most common cancer in women and the third most commonly diagnosed cancer in men worldwide. Yet, not a lot is known about the role of lifestyle factors during and after treatment of CRC. Lifestyle recommendations are only present for cancer prevention in the Netherlands. Lifestyle modifications after diagnosis could potentially benefit disease-free survival and quality of life of CRC survivors. A higher risk for recurrence and mortality was found after consuming higher amounts of a Western dietary pattern compared to a prudent diet therefore dietary patterns might play a role in CRC survival. However, no research was conducted investigating the association between plasma metabolites and dietary patterns in CRC survivors. To investigate whether plasma metabolites or metabolite signatures, i.e. groups of metabolites, are associated with general dietary patterns in CRC patients.

Visit

Evaluating the diet with all its food parameters already results in a lot of data.

Analysing hundreds to thousands of individual metabolites logically leads to an even larger dataset containing massive amounts of information. To use the full potential of both

these giant datasets, knowledge from multiple fields is necessary. The Metabolomics and Molecular Informatics group at the European Bioinformatics Institute (EMBL-EBI) in Cambridge has key expertise in informatics and metabolomics.

Methods

Untargeted metabolomics analysis on pre-surgery blood samples was performed prior to the visit at the Biomarkers Group at IARC, Lyon (France). Feature picking was performed at IARC using Agilent MassHunter and Mass Profiler Professional software. Features with intensity values below 10 000 in all samples were removed from subsequent analysis (*noise filter*). Afterwards, features with missing values in 80% of samples were excluded for further analysis.

Demographic and clinical data were abstracted from medical records. In addition, several questionnaires, i.e. Food Frequency Questionnaire (FFQ), were filled out by the participants prior to surgery. The total intake of 236 food items (g/d) and total energy intake (kcal/d). 236 food items were grouped into 48 food groups and relative intakes were

recalculated to g/d per 1000kcal. Principal Component Analysis (PCA) was used to investigate data-driven dietary patterns among CRC patients. Dietary patterns were divided into tertiles to investigate the relationship with metabolites. The R packages *ropls* and *biosigner* were used to investigate the relationship between gender, early/late stage and features, respectively. In addition ANOVA Simultaneous Component Analysis (ASCA) was used to investigate the role of gender, early/late stage and the interaction of gender and early/late stage with features.

Workflow4Metabolomics (W4M) was used to investigate the role of each dietary pattern with single features, using ANOVA. All analyses were corrected for multiple testing, using False Rate Discovery (FDR).

Preliminary results

ASCA showed a significant relationship between gender and some features. Unfortunately ASCA analysis needs a balanced design meaning that, for example, the amount of women has to be equal to men resulting in the loss of statistical power.

PCA resulted in four dietary patterns, namely a Prudent Veggie, Carnivore, Fast Food and Mediterranean pattern, see Figure 1. *Biosigner* can only be used for two-level factors while dietary patterns were divided into tertiles, e.g. three-level factor. Therefore only W4M was used to investigate possible associations between dietary patterns and features. Preliminary analyses resulted in 22 significant features (1% of the investigated features) for the Carnivore dietary pattern. Two and six features were related with the Mediterranean and Prudent Veggie pattern, respectively.

Gained knowledge

Unfortunately the visit was too short to do a lot of statistical analyses. The visit was focussed on exploring (statistical) methods for untargeted metabolomics.

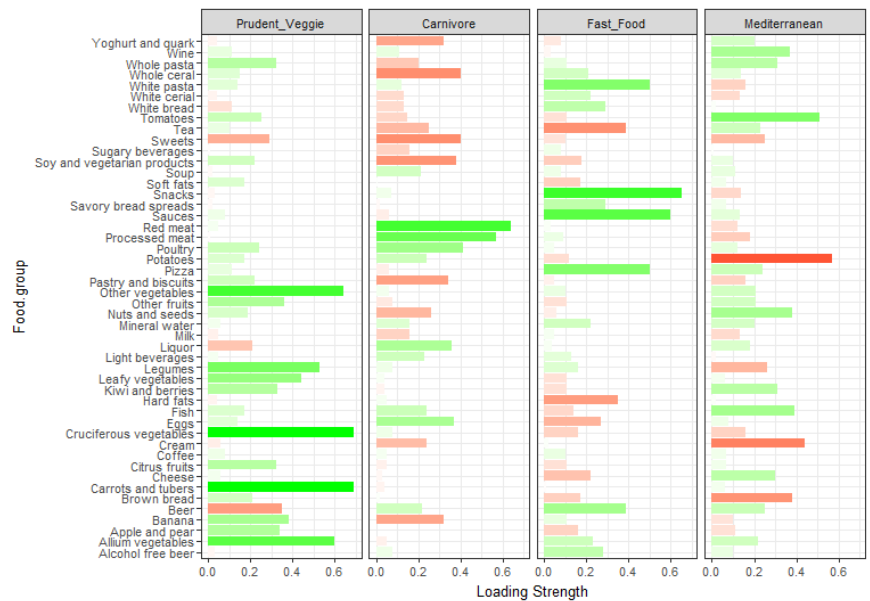


Figure 1. Loadings of food groups of the dietary patterns

There are a lot of choices in the pre-processing steps of untargeted analyses. These will be investigated further after the visit and statistical analyses will continue.

Overall the visit gave a lot of new insights and was considered to be very helpful. Gained insights, knowledge and techniques will help the division of Human Nutrition at Wageningen to get a broader understanding of the usage of (untargeted) metabolomics techniques in nutritional epidemiology. Knowledge and techniques will be passed to other scientists in the field of nutritional epidemiology to build an expertise in nutritional metabolomics.

Acknowledgements

I would like to thank the whole Metabolomics and Molecular Bioinformatics group for the nice atmosphere at EMBL-EBI. Specifically, I would like to thank Reza Salek and Rachel Spicer for the discussions about my project and their -very helpful!- thoughts on metabolomics analysis.

I would like to thank the René Vogels Stichting for the financial support, giving me the opportunity to grow on both work and personal level. In addition, I would like to acknowledge all participants of the COLON study, the hospitals involved in recruiting and the financial support of KWF and ZonMW that made the COLON study possible.