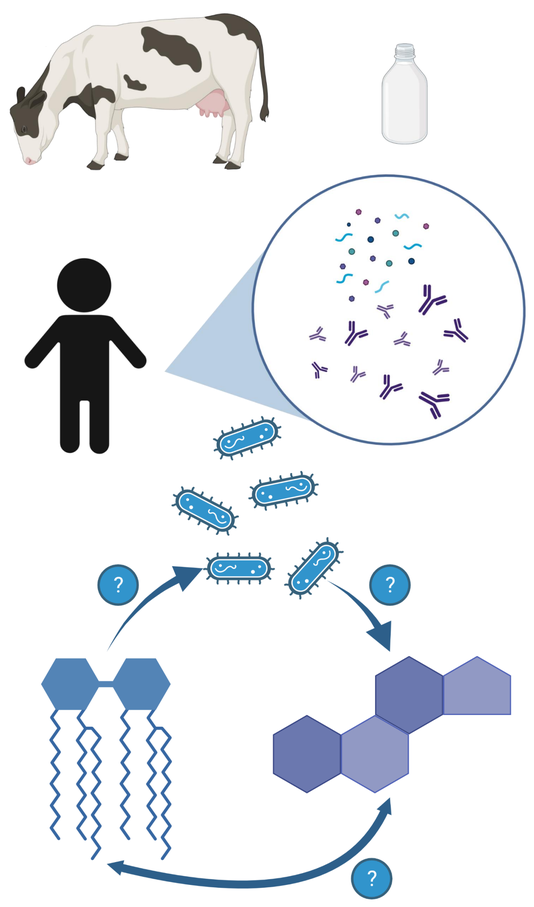
Integrating gut metabolomics and microbiomics towards biomarker discovery in pediatric cow's milk allergy

Pablo Vangeenderhuysen[1], Ellen De Paepe[1], Lieselot Y. Hemeryck[1], Marilyn De Graeve[1], Lynn Vanhaecke[1, 2]

[1] Laboratory of Integrative Metabolomics, Department of Translational Physiology, Infectiology and Public Health, Ghent University, Merelbeke, Belgium
[2] Institute for Global Food Security, School of Biological Sciences, Queen's University, Belfast, Northern Ireland, United Kingdom

Introduction

- **IgE-mediated cow's milk allergy** (IgE-CMA) is the most common and first diagnosed food allergy (FA) in children, affecting 2˜7.5% of infants.
- **Metabolomics** , **lipidomics** and **microbiomics** may uncover novel biomarkers for the diagnosis and prognosis of FAs.
- The generation of **multi-omics data** using different high-throughput methods requires a new approach to analyse and interpret the data.
- **Integration of metabolomics, lipidomics and microbiomics** facilitates discovery of intra- and interrelated variables to aid biological interpretation.



Research Objectives

- To **integrate and analyse** multi-omics data in a (statiscally) meaningful way.
- To **gain insight** in IgE-CMA metabolic mechanisms and uncover novel diagnostic biomarkers.

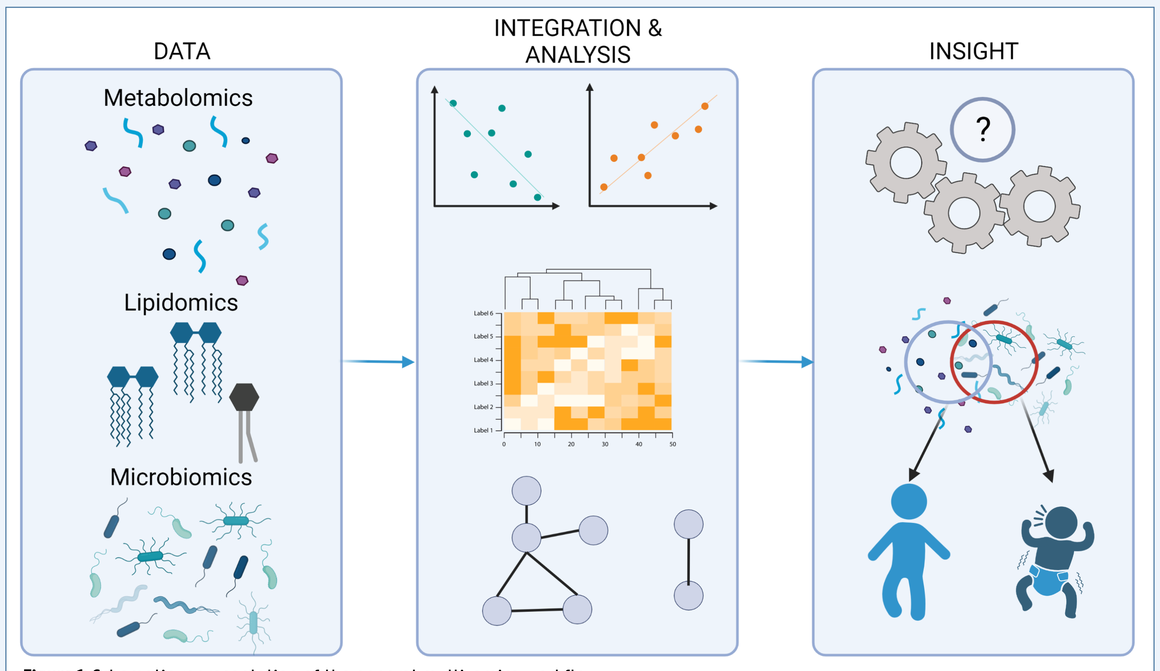


Figure 1: Schematic representation of the general multi-omics workflow.

Material and Methods

- Feces from children:
  - IgE-CMA (n = 21)
  - non IgE-CMA (n = 19)
  - other IgE FAs (n = 11)
  - healthy controls (n = 20)
- Three datasets:
  - LC-MS based metabolomics & lipidomics
  - 16S rDNA microbiome sequencing
- DIABLO[3,4]
  - MixOmics R package
  - Multiblock sPLS-DA
  - Supervised
- MOFA[1]
  - MOFA2 R package
  - Builds upon group Factor Analysis
  - Unsupervised
- Spearman correlation analysis[4]

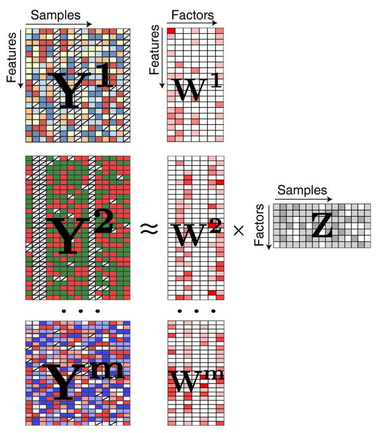


Figure 2: MOFA decomposes M data matrices (Y, ... Y) into a matrix of factors Z, common for all data matrices and M weight matrices (W, ... W). Illustration from Argelaguet et al. (2018).

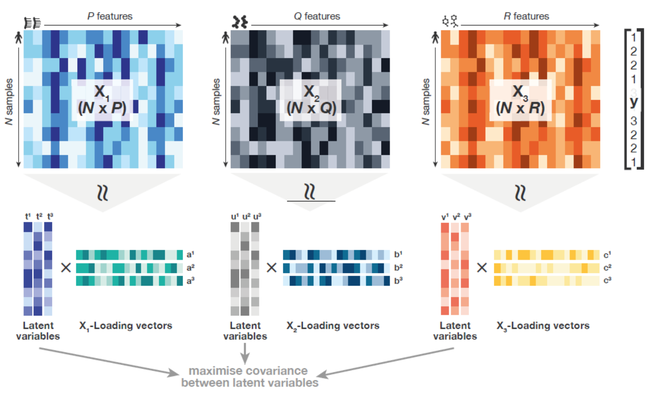


Figure 3: Schematic representation of how DIABLO decomposes the data matrices into a set of latent variables (or latent components) and a set of loading vectors. Illustration from Lê Cao et al. (2021).

Results

- Samples are plotted according to their scores on the first two components of DIABLO (Figure 4) and the first component of MOFA (Figure 5).

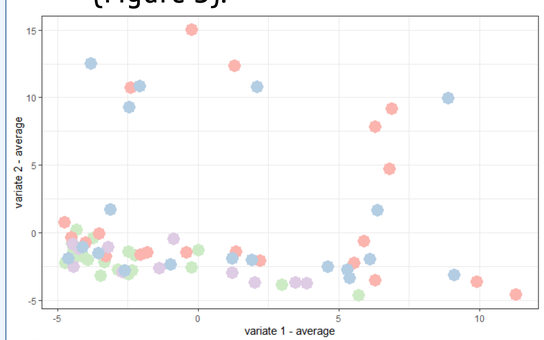


Figure 4: Points represent samples and their values on the two first components of the DIABLO model. Healthy samples are less spread over both components compared to other groups.

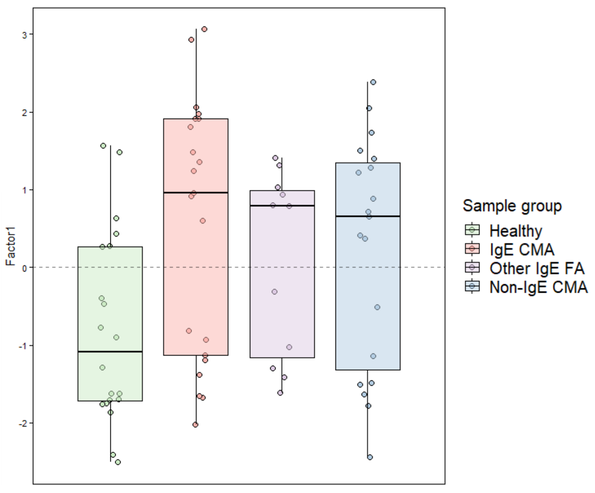


Figure 5: Points represent samples and their values on the first factor (component) of the MOFA model. Healthy samples tend to have lower values for this factor than the other groups.

- Inspecting the weights (or loadings) of the components shows the important variables within the component, allowing for **interpretation of results** (Figure 6 and 7).

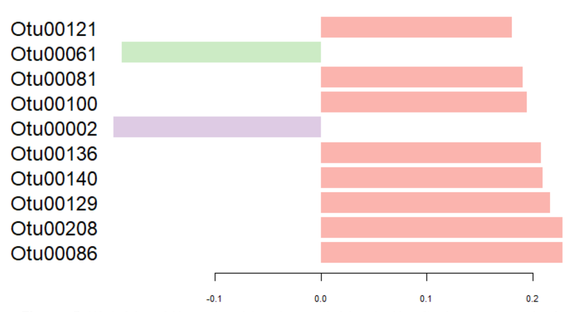


Figure 6: Weights of the microbiomics variables in the first component of the DIABLO model. . The sign of the weight indicates the direction of the effect: a positive weight indicates that the feature has higher levels in the samples with positive factor values, and vice versa. The color indicates the sample group where the variable has the maximum level of expression.

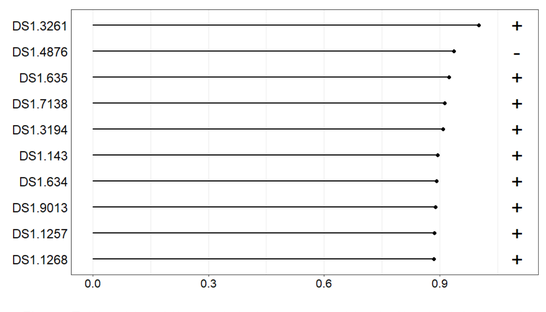
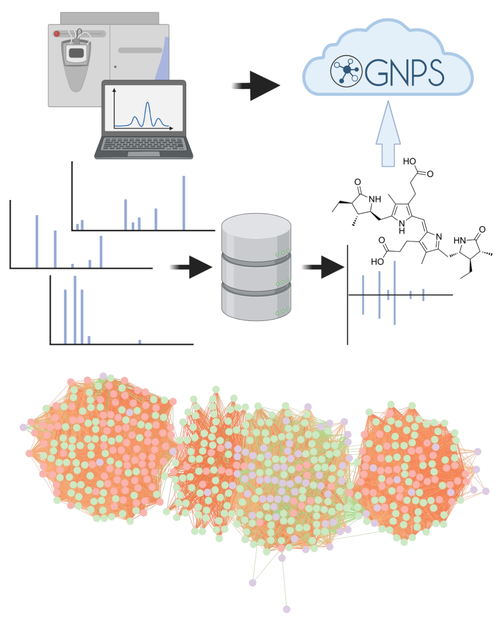


Figure 7: Weights of the metabolomics variables in the first factor of the MOFA model. Interpretation is equivalent to Figure 6.

- Top 10 variables from each dataset in the first component/factor from DIABLO and MOFA were selected: resulting in **51 unique variables**.
- **Spearman correlation analysis** between the metabolome (polar and lipid) and the microbiome resulted in 59140 correlations with an absolute Spearman's correlation coefficient > 0.5.
- In the top 1% correlations, **14 metabolites (mainly bile acids) and 6 OTUs (e.g. *Lactobacillus sp.* and *Enterococcus sp.*)** were found that were present in the 51 variables selected with DIABLO and MOFA.

Conclusion and Future Prospects

- DIABLO and MOFA provide **data-driven and holistic tools** for multi-omics data analysis and biomarker discovery. Dimension reduction is essential to explore high dimensional data.
- Efforts will be made to annotate the metabolites using **MS/MS fragmentation** and to find meaningful microbe-metabolite relationships , ultimately improving our understanding of the molecular mechanisms that drive disease.



CONTACT

Pablo.Vangeenderhuysen@UGent.be
www.linkedin.com/in/pablo-vgd https://github.com/pablovgd https://twitter.com/ljvhaeck
https://www.ugent.be/di/vpi/en/research/limet

REFERENCES

[1]Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., ... & Stegle, O. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Molecular systems biology, 14(6), e8124.
[2]Asnicar, F., Berry, S. E., Valdes, A. M., Nguyen, L. H., Piccinno, G., Drew, D. A., ... & Segata, N. (2021). Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nature medicine, 27(2), 321-332.
[3]Lê Cao, K. A., & Welham, Z. (2021). Multivariate Data Integration Using R: Methods and Applications with the MixOmics Package. Chapman and Hall/CRC.
[4]Singh, A., Shannon, C. P., Gautier, B., Rohart, F., Vacher, M., Tebbutt, S. J., and Lê Cao, K.-A. (2019). Diablo: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics,35(17):3055–3062.