Identifying the dietary signatures of arthritis through metabolomics

Carlos Cobas[2], Marianne Defernez[1], E. Kate Kemsley[1,2*], Marcel Lachenmann[2], Vera Martos-Riaño[3], Gwen Le Gall[1*], Alex MacGregor[1*], Jeremy Pirt[1], Dorna Varshavi[1], Ethan Wachowiak[1]

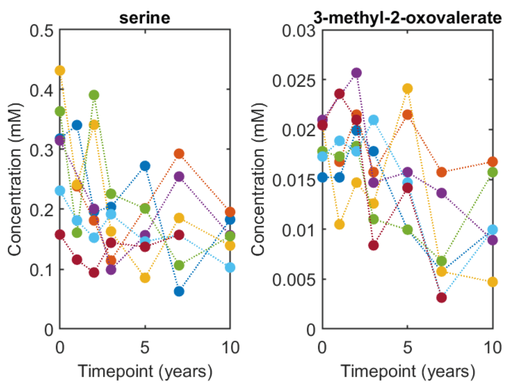
[1]University of East Anglia, Norwich Research Park, NR4 7TJ, United Kingdom
[2]Mestrelab Research SL, c/ Feliciano Barrera, 9B-Bajo, 15706, Santiago de Compostela, Spain
[3]Centro de Investigación Mestrelab (CIM), Av. Barcelona 7, 15706, Santiago de Compostela, Spain
*Authors for correspondence

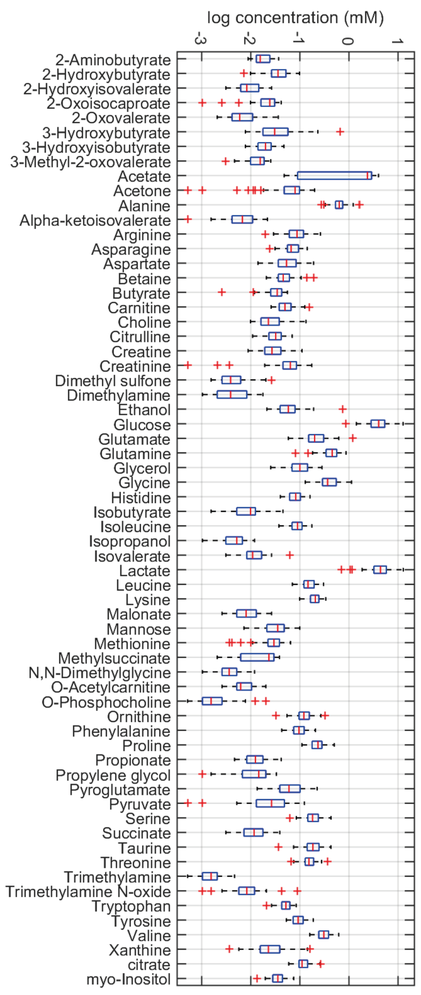
OVERVIEW

Understanding how diet may influence the course of arthritic disease is of interest to patients and medical practitioners. We use samples and data from multiple cohorts to examine the relationships of endogenous and exogenous metabolites with disease onset and progression, aiming to improve treatments and aid in prevention. Presented here are initial results from high-resolution 600 MHz NMR analysis of serum samples obtained from 7 patients over a study period of 10 years.

Metabolite concentration data

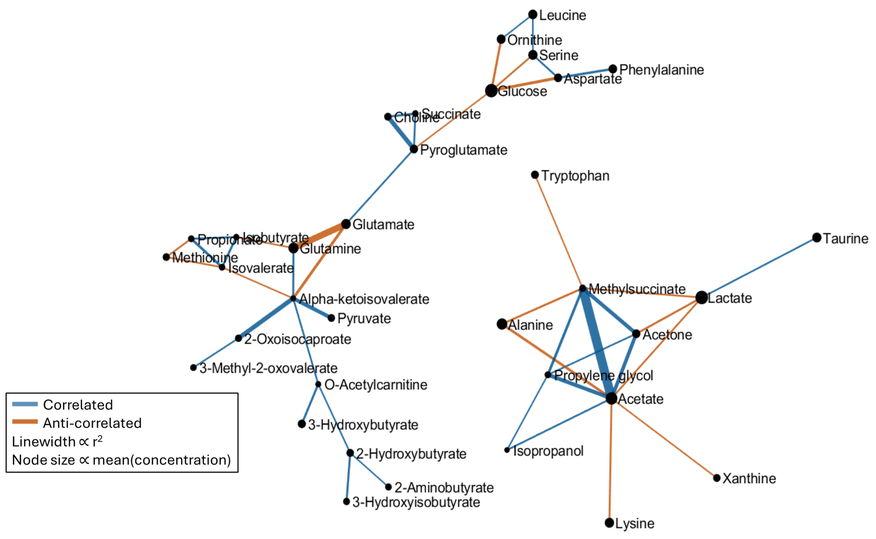
The concentration values obtained for each annotated metabolite are summarised in the boxplot. Several of these show interesting effects, for example **serine** and **3-methyl-2-oxovalerate**, both of which trend downwards in most patients over the duration of the study. Serine is a non-essential amino acid involved in the synthesis of purines and pyrimidines, and many proteins and enzymes. 3-methyl-2-oxovalerate is produced by the breakdown of isoleucine, related to protein intake, and has a role in various metabolic processes.





Relationships between metabolites

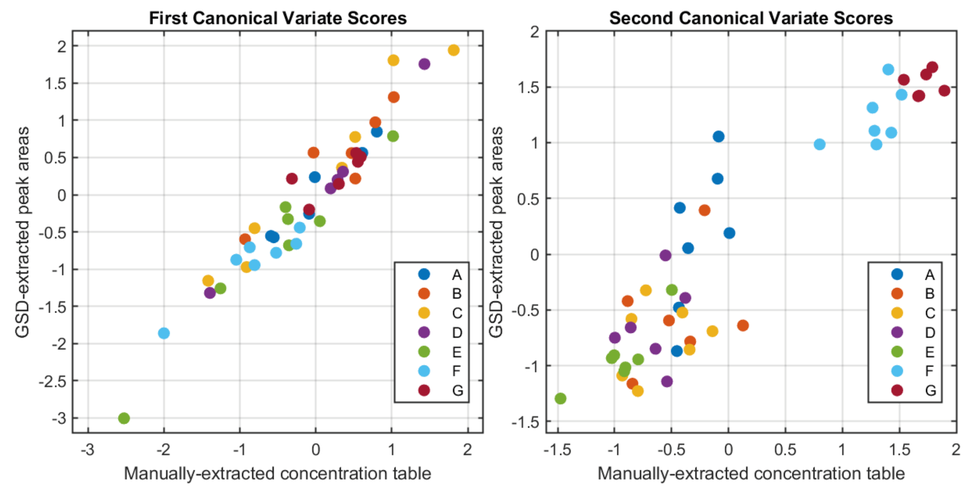
Both serine and 3-methyl-2-oxovalerate are found to be involved in a network obtained by graphing the metabolite concentration inter-correlations (r-values with $p<0.001$, multiple test-corrected). A second, unconnected network is obtained at the same significance level. The networks broadly map onto the dendrogram built by hierarchical clustering applied to the full matrix of r-squared values. This is shown in the heatmap alongside.



Data Extraction

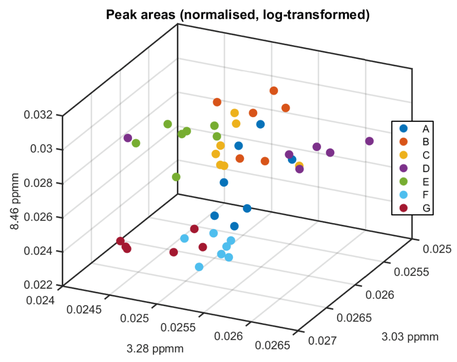
Two data tables were extracted from the raw spectra: the first was prepared with researcher oversight using Chenomx software and contains **concentrations of 64 annotated metabolites**. The second dataset comprises **areas of 39 peaks** common to all spectra, extracted using a fully automated process comprising global spectral deconvolution using Mnova software, followed by density-based clustering applied to the peak chemical shifts.

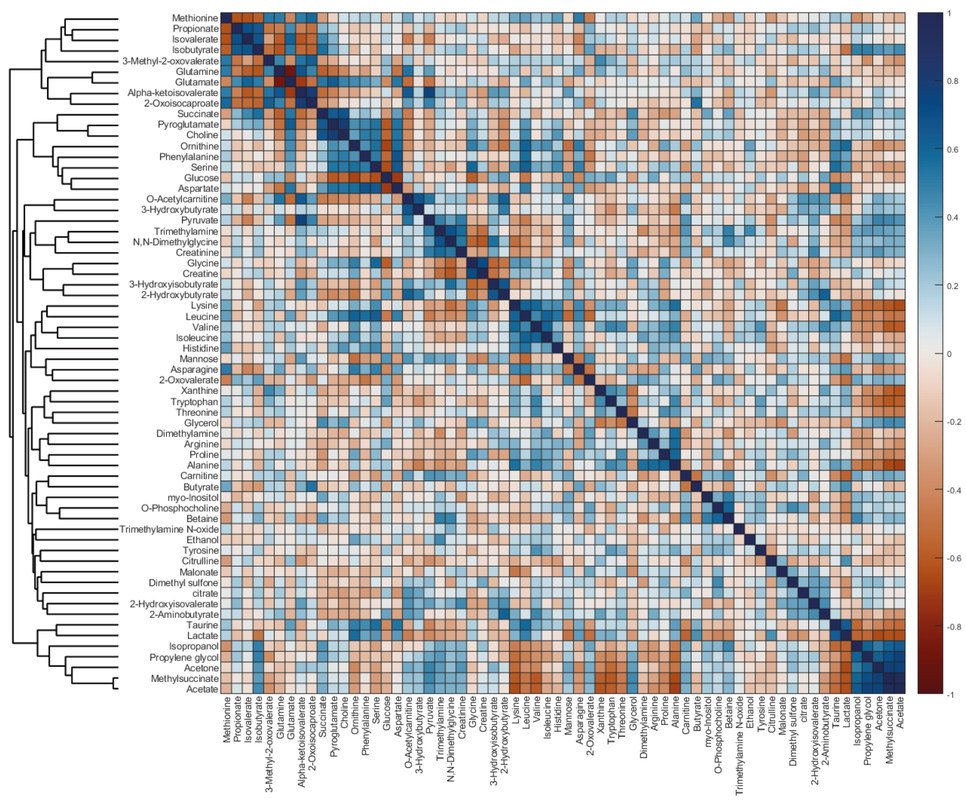
Canonical correlation analysis (CCA)



Canonical correlation analysis (CCA) shows that there is much shared information content in the two extracted data tables. The first and second canonical variates (from principal component subsets with cumulative variance of >90% in each case) are significantly correlated.

Notably, a major source of variance in both datasets is information that distinguishes patients from one another. This is seen from the symbol colour coding on the CCA scores plot; it is also evident in many of the individual table variables. It is illustrated here for three selected peak areas from the automatically extracted dataset.





Acknowledgements:
V.M.R. thanks the creation of the Mestrelab Research Center (CIM) subsidized by the Axencia Galega de Innovación. Operation funded by the Xunta de Galicia, through the business aid program for the creation and integration of new business research centers 001_IN853D_2022.