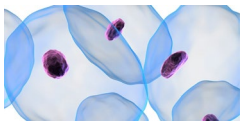
Laboratory of Systems and Synthetic Biology
www.wur.nl\ssb
www.systemsbiology.nl



Integration and Analysis of Metabolomics and Clinical data using Gaussian Graphical Modelling

Supervisors: Edoardo Saccenti

Contacts: edoardo.saccenti@wur.nl

Type of thesis: Computational

Required competences: Basic statistics, Basic knowledge of R and possibly Python

Acquired competences: Network analysis, Graphical models, Multivariate statistics, analysis of (metabol)omics data, Phenotyping of human metabolism,

Date: 2-10-2019

Description

Omics data facilitate the gain of novel insights into the pathophysiology of diseases and, consequently, their diagnosis, treatment, and prevention. To this end, omics data are integrated with other data types, e.g., clinical, phenotypic, and demographic parameters of categorical or continuous nature.

Scope of the thesis is to integrate complex data comprising clinical, demographic, and one-dimensional ^{1}H nuclear magnetic resonance metabolic variables from a large cohort of healthy subjects with the aim of identifying markers of possible latent disease status.

This can accomplished using network approaches like the so called Gaussian graphical models.

Graphical models are a unifying framework for describing the statistical relationships between large collections of random variables. In graphical models, the dependencies between different variables, the so-called nodes or vertices, are represented as edges in a network. Thus, two nodes are connected by an edge only if their association or interaction cannot be explained by any other node in the graphical model. Consequently, probabilistic graphical models eliminate spurious associations between

variables and can potentially reveal new associations adjusted for all other variables in the dataset.

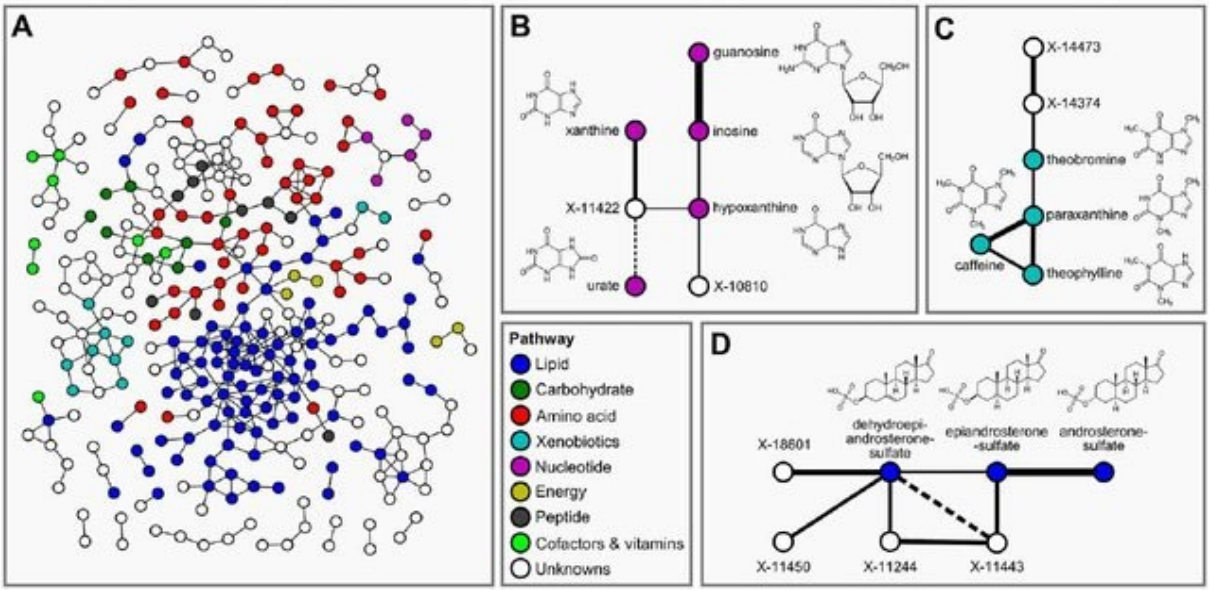
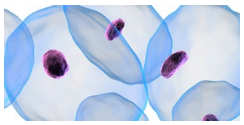


Figure 1 Example of Gaussian Graphical Modelling: GGMs embed unknown metabolites into their biochemical context. A: Complete network presentation of partial correlations that are significantly different from zero at a = 0.05 after Bonferroni correction. The unknown metabolites are spread over the entire network and are involved in various metabolic pathways. B–D: Selected high-scoring sub-networks. We observe that GGM edges directly correspond to chemical reactions which alter specific chemical groups (e.g. carbonyl groups and methyl groups). Solid lines denote positive partial correlation. Dashed lines indicate negative partial correlations. Line widths represent partial correlation strengths. From: doi:10.1371/journal.pgen.1003005.g003

At the end of this project the student will have gain knowledge in (bio)logical network analysis, integration of different data type, (advanced) multivariate statistics, NMR spectroscopy, analysis of large metabolomics data sets.

This project is collaboration with the Centre for Nuclear Magnetic Resonance of the University of Florence (http://www.cerm.unifi.it). At the end of the thesis, it may be possible to arrange an Internship in Florence.



References

1. Altenbuchinger, M., H.U. Zacharias, S. Solbrig, A. Schäfer, M. Büyüközkan, et al., *A multi-source data integration approach reveals novel associations between metabolites and renal outcomes in the German Chronic Kidney Disease study.* Scientific Reports, **2019**. 9(1): p. 13954.
2. Vignoli, A., L. Tenori, C. Luchinat and E. Saccenti, *Age and Sex Effects on Plasma Metabolite Association Networks in Healthy Subjects.* Journal of Proteome Research, **2018**. 17(1): p. 97-107.
3. Saccenti, E., M. Suarez-Diez, C. Luchinat, C. Santucci and L. Tenori, *Probabilistic Networks of Blood Metabolites in Healthy Subjects As Indicators of Latent Cardiovascular Risk.* Journal of Proteome Research, **2014**. 14(2): p. 1101--1111.
4. Hawe, J.S., F.J. Theis and M. Heinig, *Inferring Interaction Networks From Multi-Omics Data.* Frontiers in genetics, **2019**. 10: p. 535-535.
5. Højsgaard, S., D. Edwards and S. Lauritzen, *Gaussian graphical models*, in *Graphical Models with R*. 2012, Springer. p. 77-116.