Ontology term enrichment for metabolite cluster gene modules

René Pool[1,2,3,4,*], Harmen H.M. Draisma[1,2,3,*], Mohammed El-Kebir[5,6,*], Jan Bert van Klinken[3,7], Ko Willems van Dijk[3,7], **Cornelia M. van Duijn[3,4,8], Christian Gieger[3,9],** Peter A.C. 't Hoen[3,7], Gunnar W. Klau[5,6], **Nicholas G. Martin[3,10],** Mark I. McCarthy[3,11], Michael Kobl[3,9], Andres Metspalu[3,12], P. Eline Slagboom[3,4,13], **Tim D. Spector[3,14], Konstantin Strauch[3,9,15],** Dorret I. Boomsma[1,2,3,4] and Aaron Isaacs[3,8]

* These authors contributed equally; 1) VU University Amsterdam, Faculty of Psychology and Education, dept. Biological Psychology, Amsterdam, The Netherlands; 2) The EMGO+ Institute for Health and Care Research, Amsterdam, The Netherlands; 3) ENGAGE: European Network of Genomic And Genetic Epidemiology; 4) BBMRI-NL: Infrastructure for the Application of Metabolomics Technology in Epidemiology (RP4), The Netherlands; 5) Centrum Wiskunde & Informatica (CWI), Life Sciences Amsterdam, The Netherlands; 6) VU University Amsterdam, The Centre for Integrative Bioinformatics VU (IBIVU), Amsterdam, The Netherlands; 7) Leiden University Medical Center, Department of Human Genetics, Leiden, The Netherlands; 8) Erasmus Medical Center, Department of Epidemiology, Rotterdam, The Netherlands; 9) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 10) Queensland Institute of Medical Research, Genetic Epidemiology, Brisbane, Australia; 11) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 12) University of Tartu, Estonian Genome Center, Tartu, Estonia; 13) Leiden University Medical Center, Department of Molecular Epidemiology, Leiden, The Netherlands; 14) King's College London, Department of Twin Research & Genetic Epidemiology, London, United Kingdom; 15) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany

Introduction

We conducted genome-wide association and meta-analyses for 163 metabolites as detected using the Biocrates AbsoluteIDQ p150 sample preparation kit. These analyses only provide the associations between a single metabolite and all single nucleotide polymorphisms in the genotype data. Also, it is expected that there is considerable pleiotropy (the same gene influencing the levels of multiple metabolites).

Objective

To highlight connections among metabolites, we sought to identify sets of genes (gene modules) shared across different metabolites.

Methods

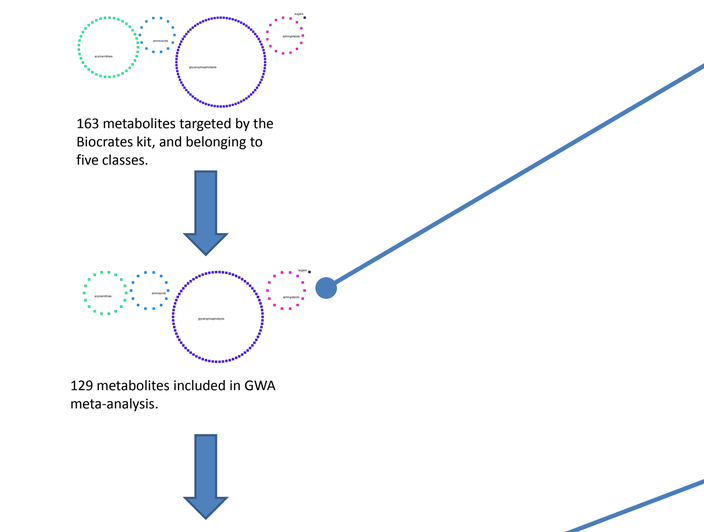

Figure 1. Numbers of metabolites retained at each step.

A flowchart of the analysis is depicted in Figure 2. In brief, gene-based P-values were computed on the basis of the results from the meta-analyses of the genome-wide association results for each metabolite. For this, an adaptation of the VEGAS (VErsatile Gene-based Analysis Software) [Liu et al., Am J Hum Genet 2010] algorithm was used.

We used the Benyamini-Hochberg method to determine which associations between genes and metabolites were statistically significant. The associated genes for each metabolite were denoted as "gene sets".

Then, the overlap in gene sets across different metabolites was quantified by computing the matrix of Jaccard index values for each possible pair-wise metabolite combination. The resulting similarity matrix was used as input for a hierarchical clustering analysis using the average linkage algorithm. The "dynamic tree cut" method was used to cut the resulting dendrogram in order to define clusters of metabolites sharing the same significantly associated genes ("gene sets").

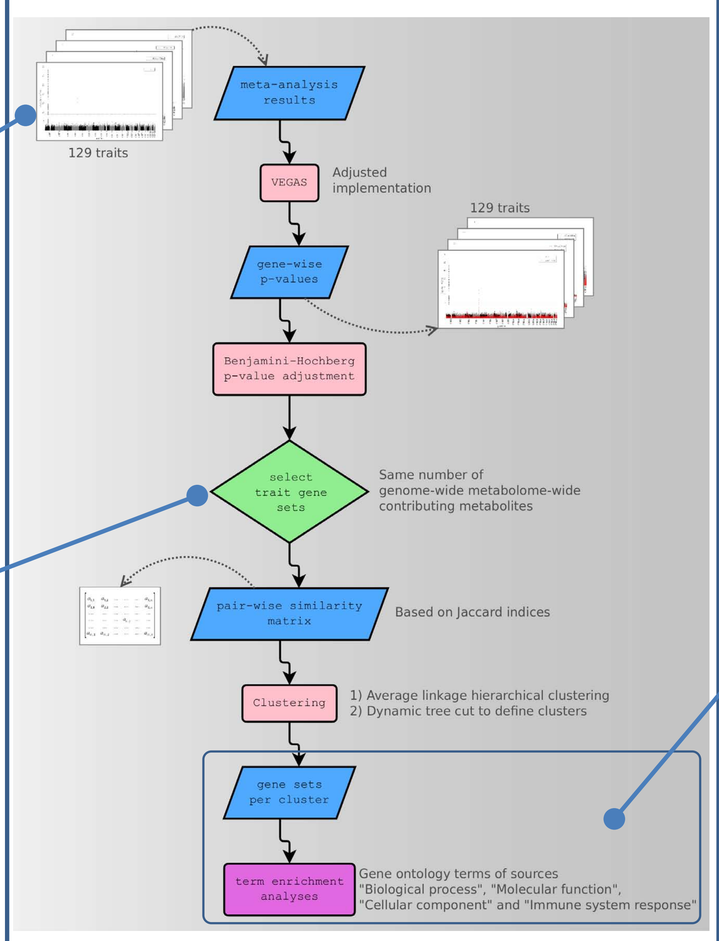

Figure 2. Flowchart of analysis. Bulleted lines connect corresponding steps depicted in this Figure and in Figures 1 or 3.

Finally, the enrichment of the genes within the found clusters for ontology terms in four categories ("Biological process", "Molecular function", "Cellular component", "Immune system response") was assessed.

Results

Ten (10) clusters of metabolites with overlapping gene sets (gene modules) were found. Examples of the first three clusters are given in Figures 3A–C.

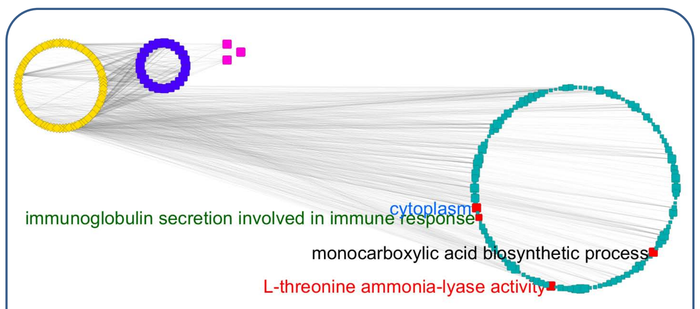

Figure 3A. Cluster 1

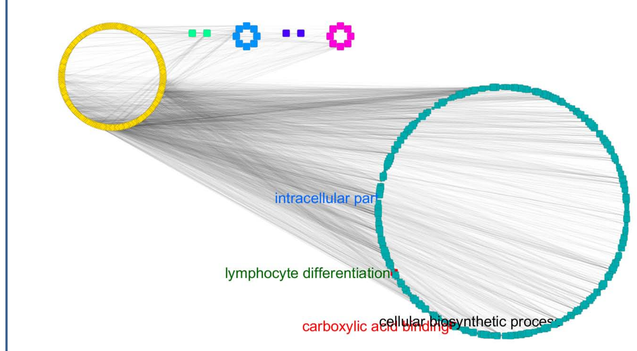

Figure 3B. Cluster 2

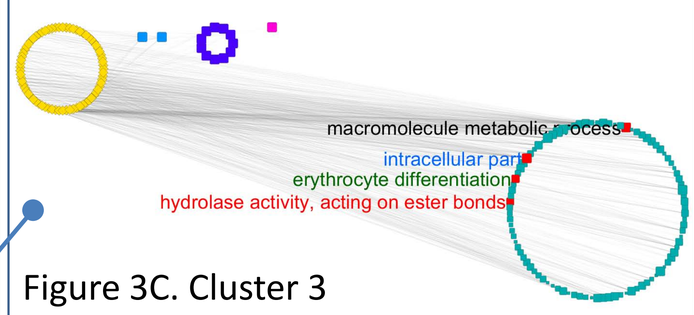

Figure 3C. Cluster 3

Figure 3. Examples of found clusters. Ontology terms with lowest enrichment P-values are shown explicitly.

Conclusions and future work

- We have developed and used a method to elucidate, *de novo*, clusters of metabolites and their associated genes.
- The developed method represents one way to functionally follow up the results of genome-wide association studies based on metabolomics.
- The biological significance of the found clusters still needs to be addressed in future studies.

This research was supported through funds from The European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE Consortium, grant agreement HEALTH-F4-2007- 201413.