Title: Automated pathway and reaction prediction facilitates *in silico* identification of unknown metabolites in human cohort studies

Author: Jan D. Quell Werner Römisch-Margl Marco Colombo Jan Krumsiek Anne M. Evans Robert Mohney Veikko Salomaa Ulf de Faire Leif C. Groop Felix Agakov Helen C. Looker Paul McKeigue Helen M. Colhoun Gabi Kastenmüller

PII: S1570-0232(17)30568-8
DOI: http://dx.doi.org/doi:10.1016/j.jchromb.2017.04.002
Reference: CHROMB 20537

To appear in: *Journal of Chromatography B*

Received date: 1-12-2016
Revised date: 22-3-2017
Accepted date: 1-4-2017

Please cite this article as: J.D. Quell, W. R*ddoto*misch-Margl, M. Colombo, J. Krumsiek, A.M. Evans, R. Mohney, V. Salomaa, U. de Faire, L.C. Groop, F. Agakov, H.C. Looker, P. McKeigue, H.M. Colhoun, G. Kastenm*ddotu*ller, Automated pathway and reaction prediction facilitates *in silico* identification of unknown metabolites in human cohort studies, *Journal of Chromatography B* (2017), http://dx.doi.org/10.1016/j.jchromb.2017.04.002

Automated pathway and reaction prediction facilitates *in silico* identification of unknown metabolites in human cohort studies

Jan D. Quell[a,b], Werner Römisch-Margl[b], Marco Colombo[c], Jan Krumsiek[d], Anne M. Evans[e], Robert Mohney[e], Veikko Salomaa[f], Ulf de Faire[g], Leif C. Groop[h], Felix Agakov[i], Helen C. Looker[j], Paul McKeigue[c], Helen M. Colhoun[j], Gabi Kastenmüller[b,*]

Affiliations:
[a] Genome-oriented Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
[b] Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[c] Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, UK
[d] Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
[e] Metabolon Inc., Durham, North Carolina, USA
[f] Department of Health, National Institute for Health and Welfare, Finland
[g] Unit of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
[h] Department of Clinical Sciences - Diabetes and Endocrinology, Lund University, Malmö, Sweden
[i] Pharmatics Limited, Edinburgh, UK
[j] Population Health Sciences, University of Dundee, Dundee, UK
[*] corresponding author

E-Mail:
g.kastenmueller@helmholtz-muenchen.de

Abstract

Identification of metabolites in non-targeted metabolomics continues to be a bottleneck in metabolomics studies in large human cohorts. Unidentified metabolites frequently emerge in the results of association studies linking metabolite levels to, for example, clinical phenotypes. For further analyses these unknown metabolites must be identified. Current approaches utilize chemical information, such as spectral details and fragmentation characteristics to determine components of unknown metabolites. Here, we propose a systems biology model exploiting the internal correlation structure of metabolite levels in combination with existing biochemical and genetic information to characterize properties of unknown molecules.

Levels of 758 metabolites (439 known, 319 unknown) in human blood samples of 2279 subjects were