



Quantitative high-throughput metabolomics: a new era in epidemiology and genetics

Mika Ala-Korpela, Antti J Kangas and Pasi Soininen

MUSINGS

Quantitative high-throughput metabolomics: a new era in epidemiology and genetics

Mika Ala-Korpela^{1,2*}, Antti J Kangas¹ and Pasi Soininen^{1,2}

Metabolites in body fluids reflect multiple biochemical processes and pathways relevant to health and disease. Comprehensive approaches to gain insights into metabolic variation and diseases, such as metabolic phenotyping, have become increasingly popular over recent years [1-3]. These developments have been driven by mass spectrometry (MS) and proton nuclear magnetic resonance (NMR) spectroscopy as the two key experimental technologies. On the basis of findings from multiple disciplines, it has been envisaged that metabolic phenotyping will eventually lead to holistic risk assessment for various diseases [4].

Inevitability of the health-disease continuum

In the field of metabolomics, the complexity of the data generated has directed analyses towards spectral chemometrics and black-and-white thinking, the main goal often being to separate individuals who are 'healthy' from those who are 'diseased'. These kinds of simple classification approaches, often also using spectroscopic measures that are non-specific in molecular terms, are not optimal for deriving metabolic understanding in epidemiology and genomic medicine. They also poorly reflect the continuity of biological processes and states (Figure 1). Even though single biomarkers and diagnostic thresholds are necessary criteria in current clinical practice, it will be pivotal that future medicine builds on our growing understanding of disease etiology. Importantly, we should inherently accept the biologically inevitable metabolic and disease continuum instead of being hampered by the apparently unattainable black-and-white diagnostics. Common disorders, being multi-genic, are essentially quantitative traits [5], and ultimately we cannot hide from this fundamental feature of nature.

Quantification rules

Techniques that can quantify large numbers of metabolites, representing multiple metabolic pathways in systemic metabolism, are particularly relevant for the risk assessment of metabolic disorders, such as diabetes and vascular diseases. Currently, NMR-based applications are the only approaches that can offer fully automated and highly reproducible high-throughput experimentation in a very cost-effective manner [6,7]. Although the variety of molecules measurable by MS cannot be matched by NMR, the per-sample costs using MS still tend to be high, and the quantitative throughput is limited. Recent MS applications, however, show appealing progress in both throughput and molecular identifications [3], suggesting that, in the near future, MS and NMR will most likely be used as complementary technologies in large-scale epidemiology. Combined with genome-wide and gene expression data at the population level, comprehensive metabolic information has started to trigger detailed systems-level findings [8,9]. This novel line of multiomics research is anticipated to grow rapidly and to allow a more thorough molecular understanding of biochemical pathways and disease pathologies. Yet metabolomics will be truly useful in epidemiology or in genetic studies only if quantitative data on specific, identified metabolites are available.

Specificity is power

Broadly speaking, epidemiological research and genome-wide association studies (GWASs) aim to discover associations in order to generate biological hypotheses. Conventional thinking holds that the number of people is the primary adjustable variable for increasing the statistical power to detect variants of a given effect in a GWAS. However, effect sizes depend on phenotypic definitions and will get stronger as one moves closer to the biochemical source. This implies that 'missing heritability' is not simply a reflection of what cannot be found by a common genetic variant association; it relates fundamentally to the biological and molecular rationale of the trait. We have recently demonstrated this in a GWAS on (only) 8,330 Finnish individuals for a range of 216 serum metabolic measures quantified by our NMR

*Correspondence: mika.ala-korpela@computationalmedicine.fi
¹Computational Medicine, Institute of Clinical Medicine, Faculty of Medicine, University of Oulu, PO Box 5000, FI-90014 Oulu, Finland
Full list of author information is available at the end of the article

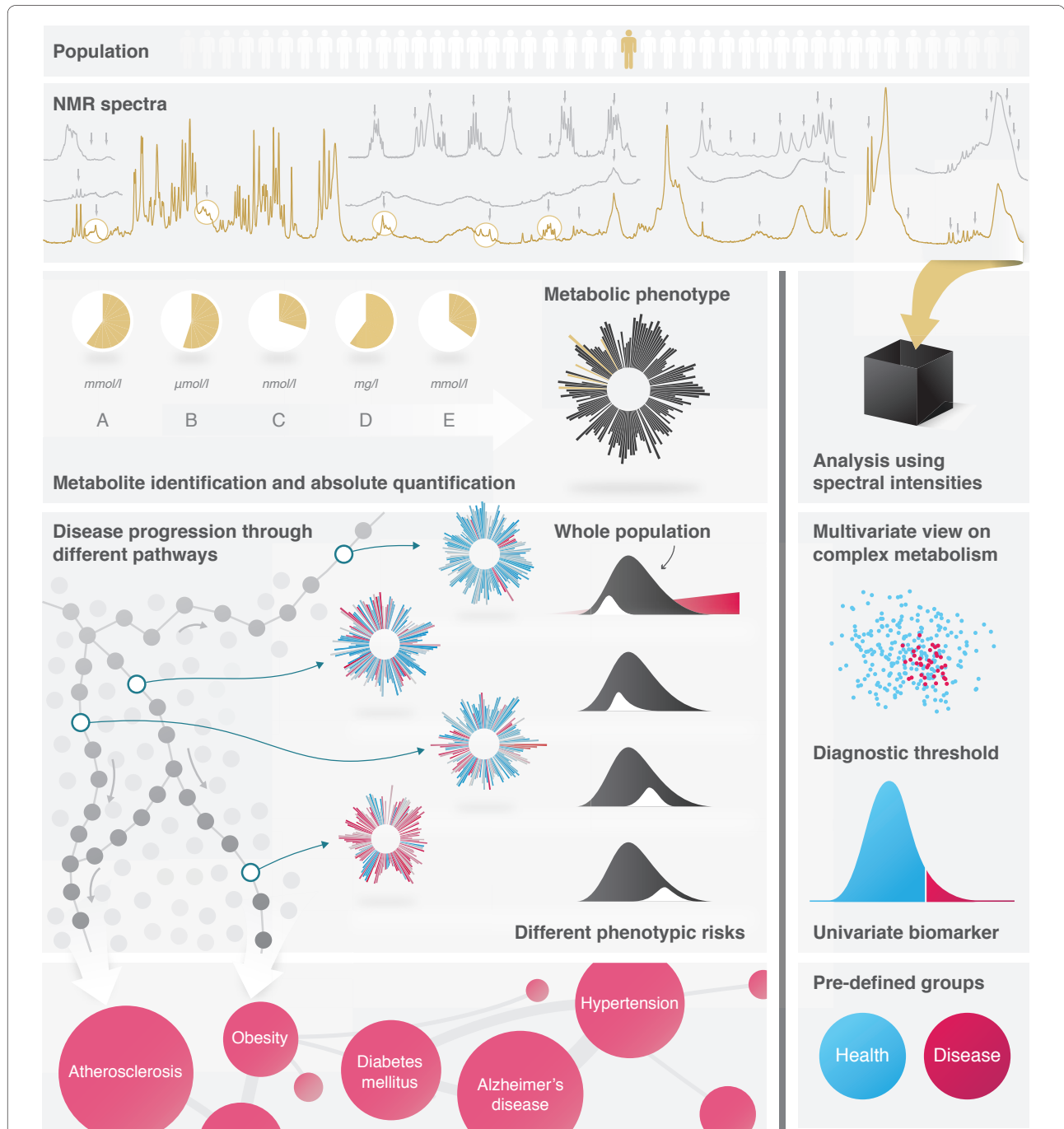


Figure 1. Quantitative metabolic phenotyping for continuous pathway modeling versus spectral-based black-and-white diagnostic classification. **Right:** A common metabolomics approach is to use the spectral data directly in a chemometrics model that explores the overall differences between individuals expected to belong to different diagnostic groups. This approach is not optimal for epidemiological or genetic research in which metabolic and disease continuum should be appreciated, because common disorders, being multigenic, are fundamentally quantitative traits. Therefore, the real data (strongly overlapping metabolic characteristics) do not match with the pre-defined groups (health versus disease). **Left:** New high-throughput methodologies involve sophisticated automation, including absolute quantification of identified metabolites. This provides new opportunities to understand disease etiologies and to handle disease risks and diagnostics as truly continuous multivariate phenomena. When these kinds of approaches mature and extensive datasets accumulate, it is anticipated that characteristic metabolic phenotypes for various disease-related pathways can be identified. This would allow overall assessment of individual health status and disease risks. Here, from the spectral NMR data of an individual, metabolites are identified and quantified in a fully automated manner, resulting in a comprehensive metabolic phenotype. Different pathways, which predispose to metabolic disorders in a distinctive way, have characteristic (time-dependent) metabolic signatures with specific risk distributions. In real life, metabolic disorders are interrelated and rarely exist in isolation.

metabolomics platform [7]. We identified 31 loci associated with one or more metabolic measures at a genome-wide significance level, including seven newly identified loci for low-molecular-weight metabolites and four for serum lipoprotein and lipid measures. Using the same Finnish individuals, we also performed an association analysis of the 95 genetic loci known to affect serum cholesterol and triglyceride levels [10]. Our analysis included comprehensive data on lipids and lipoprotein subclasses, obtained via the NMR metabolomics platform, and four enzymatic lipid traits. For 30 of the 95 loci, we identified new metabolic or genetic associations. In the majority of the loci, the strongest association was to a more specific metabolite measure than the total lipids measured enzymatically. Interestingly, in four loci, the smallest high-density lipoprotein (HDL) measures showed effects opposite to the larger ones, a finding that indicates distinct metabolic characteristics for small and large HDL particles as also previously indicated by the gene co-expression findings in circulating leukocytes [8]. Thus, the findings feature considerable diversity in association patterns for the loci originally identified through associations with enzymatic total lipid measures, and they reveal association profiles of far greater resolution than those from routine clinical lipid measures [10]. Therefore, not unexpectedly, metabolic measures of pathway specificity (such as HDL subclasses) can provide far better insights into biological processes than common clinical measures representing merely a sum of multiple biochemical components (such as HDL cholesterol).

One-for-all goes multiple

Increased etiological understanding and good biomarkers for disease prediction and prevention could facilitate clinical progress and translational medicine. However, in typical epidemiological studies, the findings, although statistically significant at a population level, reflect only weak relationships between metabolism and demographic or clinical measures, and therefore do not provide a sound basis for individual prediction models. Metabolic measures close to the underlying molecular pathways are needed to increase the accuracy of such modeling. Metabolomics approaches can intrinsically provide holistic molecular perspectives and thereby lead to better representations of disease progression, especially if we stop using univariate cut-offs for diagnostics and start handling disease risk as multivariate continuous dimensions. Combined with new bioinformatics schemes that include biological justification, comprehensive metabolic phenotyping has the potential to provide globally useful solicitous societal solutions with better individual well-being together with more efficiently spent health budgets.

It is clearly understood that typical clinical outcomes of metabolic diseases, like infarctions in coronary heart

disease, occur as a result of life-long effects of multiple molecular pathways. If characteristic combinations of these pathways exist, and if they are distinct in relation to metabolic disorders, a pathway-specific identification and risk assessment of individuals might be possible at the population level (Figure 1). The potential translation from the current 'one model for all situations' to a new clinical practice incorporating pathways and metabolic phenotypes requires better understanding of the etiologies of metabolic diseases and their interplay. This invokes the need for quantitative metabolomics for the masses.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We acknowledge financial support from the Academy of Finland, the Responding to Public Health Challenges Research Program of the Academy of Finland, the Finnish Foundation for Cardiovascular Research, the Jenny and Antti Wihuri Foundation, the Finnish Diabetes Research Foundation, and the Strategic Research Funding from the University of Oulu.

Author details

¹Computational Medicine, Institute of Clinical Medicine, Faculty of Medicine, University of Oulu, PO Box 5000, FI-90014 Oulu, Finland. ²NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, PO Box 1627, FI-70211, Kuopio, Finland.

Published: 30 April 2012

References

1. Mäkinen V-P, Soininen P, Forsblom C, Parkkonen M, Ingman P, Kaski K, Groop P-H, FinnDiane Study Group, Ala-Korpela M: **¹H NMR metabolomics approach to the disease continuum of diabetic complications and premature death.** *Mol Syst Biol* 2008, **4**:167.
2. Madsen R, Lundstedt T, Trygg J: **Chemometrics in metabolomics - a review in human disease diagnosis.** *Anal Chim Acta* 2010, **659**:23-33.
3. Suhre K, Shin S-Y, Petersen A-K, Mohny RP, Meredith D, Wägele B, Altmaier E, CARDIoGRAM, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes H-W, Milburn MV, Prehn C, Raffler J, Ried JS, Römisch-Margl W, Samani NJ, Small KS, Wichmann H-E, Zhai G, Illig T, et al: **Human metabolic individuality in biomedical and pharmaceutical research.** *Nature* 2011, **477**:54-60.
4. Holmes E, Wilson ID, Nicholson JK: **Metabolic phenotyping in health and disease.** *Cell* 2008, **134**:714-717.
5. Plomin R, Haworth CMA, Davis OSP: **Common disorders are quantitative traits.** *Nat Rev Genet* 2009, **10**:872-878.
6. Soininen P, Kangas AJ, Würtz P, Tukiainen T, Tynkkynen T, Laatikainen R, Järvelin M-R, Kähönen M, Lehtimäki T, Viikari J, Raitakari OT, Savolainen MJ, Ala-Korpela M: **High-throughput serum NMR metabolomics for cost-effective holistic studies on systemic metabolism.** *Analyst* 2009, **134**:1781-1785.
7. Kettunen J, Tukiainen T, Sarin A-P, Ortega-Alonso A, Tikkanen E, Lyytikäinen L-P, Kangas AJ, Soininen P, Würtz P, Silander K, Dick DM, Rose RJ, Savolainen MJ, Viikari J, Kähönen M, Lehtimäki T, Pietiläinen KH, Inouye M, McCarthy MI, Jula A, Eriksson J, Raitakari OT, Salomaa V, Kaprio J, Järvelin M-R, Peltonen L, Perola M, Freimer NB, Ala-Korpela M, Palotie A, et al: **Genome-wide association study identifies multiple loci influencing human serum metabolite levels.** *Nat Genet* 2012, **44**:269-276.
8. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, Hämäläinen E, Jousilahti P, Kangas AJ, Männistö S, Savolainen MJ, Jula A, Leiviskä J, Palotie A, Salomaa V, Perola M, Ala-Korpela M, Peltonen L: **Metabonomic, transcriptomic, and genomic variation of a population cohort.** *Mol Syst Biol* 2010, **6**:441.
9. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, Holm H, Sanna S, Kavousi M, Baumeister SE, Coin LJ, Deng G, Gieger C, Heard-Costa

- NL, Hottenga J-J, Kühnel B, Kumar V, Lagou V, Liang L, Luan J, Vidal PM, Mateo Leach I, O'Reilly PF, Peden JF, Rahmioglu N, Soininen P, Speliotes EK, Yuan X, Thorleifsson G, Alizadeh BZ, *et al.*: **Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma.** *Nat Genet* 2011, **43**:1131-1138.
10. Tukiainen T, Kettunen J, Kangas AJ, Lyytikäinen LP, Soininen P, Sarin AP, Tikkanen E, O'Reilly PF, Savolainen MJ, Kaski K, Pouta A, Jula A, Lehtimäki T, Kähönen M, Viikari J, Taskinen MR, Jauhiainen M, Eriksson JG, Raitakari O, Salomaa V, Järvelin MR, Perola M, Palotie A, Ala-Korpela M, Ripatti S: **Detailed**

metabolic and genetic characterization reveals new associations for 30 known lipid loci. *Hum Mol Genet* 2012, **21**:1444-1455.

doi:10.1186/gm335

Cite this article as: Ala-Korpela M, *et al.*: Quantitative high-throughput metabolomics: a new era in epidemiology and genetics. *Genome Medicine* 2012, **4**:36.