

# High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism†

Pasi Soininen,<sup>‡,ab</sup> Antti J. Kangas,<sup>‡,b</sup> Peter Würtz,<sup>bc</sup> Taru Tukiainen,<sup>b</sup> Tuulia Tynkkynen,<sup>a</sup> Reino Laatikainen,<sup>a</sup> Marjo-Riitta Järvelin,<sup>b,def</sup> Mika Kähönen,<sup>g</sup> Terho Lehtimäki,<sup>h</sup> Jorma Viikari,<sup>i</sup> Olli T. Raitakari,<sup>c</sup> Markku J. Savolainen<sup>bj</sup> and Mika Ala-Korpela<sup>\*abj</sup>

Received 27th May 2009, Accepted 23rd July 2009  
First published as an Advance Article on the web 30th July 2009  
DOI: 10.1039/b910205a

A high-throughput proton (<sup>1</sup>H) nuclear magnetic resonance (NMR) metabonomics approach is introduced to characterise systemic metabolic phenotypes. The methodology combines two molecular windows that contain the majority of the metabolic information available by <sup>1</sup>H NMR from native serum, *e.g.* serum lipids, lipoprotein subclasses as well as various low-molecular-weight metabolites. The experimentation is robotics-controlled and fully automated with a capacity of about 150–180 samples in 24 h. To the best of our knowledge, the presented set-up is unique in the sense of experimental high-throughput, cost-effectiveness, and automated multi-metabolic data analyses. As an example, we demonstrate that the NMR data as such reveal associations between systemic metabolic phenotypes and the metabolic syndrome ( $n = 4407$ ). The high-throughput of up to 50 000 serum samples per year is also paving the way for this technology in large-scale clinical and epidemiological studies. In contradiction to single ‘biomarkers’, the application of

this holistic NMR approach and the integrated computational methods provides a data-driven systems biology approach to biomedical research.

## Introduction

Metabonomics is an -omics approach to identify and monitor metabolic phenotypes with respect to various synergistic factors such as environment, life-style, diet as well as potential pathophysiological processes.<sup>1–5</sup> Metabonomics offers supplementary information to other -omics areas by representing the far end in the chain of phenomena from gene expression to systemic metabolism.<sup>6,7</sup> Mass spectrometry and proton (<sup>1</sup>H) nuclear magnetic resonance (NMR) spectroscopy have become the two key experimental technologies in the field.<sup>8</sup>

Proton NMR offers high analytical reproducibility and provides specific quantitative data in a non-selective manner.<sup>9,10</sup> Thus, it is well suited for probing continuous multi-metabolic variations with respect to potential systemic effectors, for example, in the grey zone of atherosclerosis pathophysiology between health and disease.<sup>11,12</sup> While the biochemistry of serum is reflected by the atherosclerotic processes in the arterial wall (and *vice versa*), the biological heterogeneity as well as the slow development and progression of pathological conditions make the borderline between health and disease inherently indistinct. It is in these types of situations in which single ‘biomarkers’ are currently being recognised to fail in describing the complex molecular foundations of diseases and where the metabonomics approaches have a remarkable potential to provide cost-effective solutions *via* high-throughput analytics and advanced computational methods.<sup>1–15</sup>

Here, we describe a new <sup>1</sup>H NMR spectroscopy protocol for high-throughput metabonomics of serum to characterise individual systemic metabolic phenotypes. Blood serum is the primary body fluid connected to systemic metabolism and is therefore the natural choice for studies related to vascular and systemic diseases.<sup>1,4,8–12</sup> Proton NMR *per se* allows fast and reliable detection of a large number of metabolites. However, the molecular variety and multiple environments, particularly in serum, partly hamper the molecular identification and quantification. To overcome this issue, we have adopted an approach based on two molecular windows, LIPO and LMWM. The LIPO window represents a conventional water-suppressed <sup>1</sup>H NMR spectrum of serum providing mainly information on lipoprotein lipids and subclasses, while the LMWM window is acquired in such a way that majority of the broad signals,

<sup>a</sup>NMR Metabonomics Laboratory, Laboratory of Chemistry, Department of Biosciences, University of Kuopio, Kuopio, Finland

<sup>b</sup>Computational Medicine Research Group, Institute of Clinical Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, P.O. Box 5000, FI-90014 University of Oulu, Oulu, Finland. E-mail: mika.alakorpela@computationalmedicine.fi; Tel: +358 50 35 35 457; Web: <http://www.computationalmedicine.fi/>

<sup>c</sup>Department of Clinical Physiology and the Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and Turku University Central Hospital, Turku, Finland

<sup>d</sup>National Institute of Health and Wellbeing, Department of Child and Adolescent Health, Oulu, Finland

<sup>e</sup>Department of Epidemiology and Public Health, Imperial College London, London, UK

<sup>f</sup>Institute of Health Sciences and Biocenter Oulu, University of Oulu, Oulu, Finland

<sup>g</sup>Department of Clinical Physiology, Tampere University Hospital and University of Tampere, Tampere, Finland

<sup>h</sup>Department of Clinical Chemistry, Tampere University Hospital and University of Tampere, Tampere, Finland

<sup>i</sup>Department of Medicine, University of Turku and Turku University Central Hospital, Turku, Finland

<sup>j</sup>Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, University of Oulu, Oulu, Finland

† Authors' contributions: P. S., A. J. K., P. W., J. V., O. T. R., and M. A. K. conceived and designed the study; P. S., A. J. K., and M. A. K. designed the serum NMR platform; P. S., A. J. K., P. W., T. Tu., T. Ty., R. L., and M. A. K. took part in analysing the spectroscopic data; M. R. J., M. K., T. L., J. V., O. T. R., and M. J. S. collected and interpreted clinical and biochemical data; all authors interpreted the results; P. S., A. J. K., P. W., and M. A. K. wrote the paper. All authors read and approved the final manuscript.

‡ Contributed equally to this work.

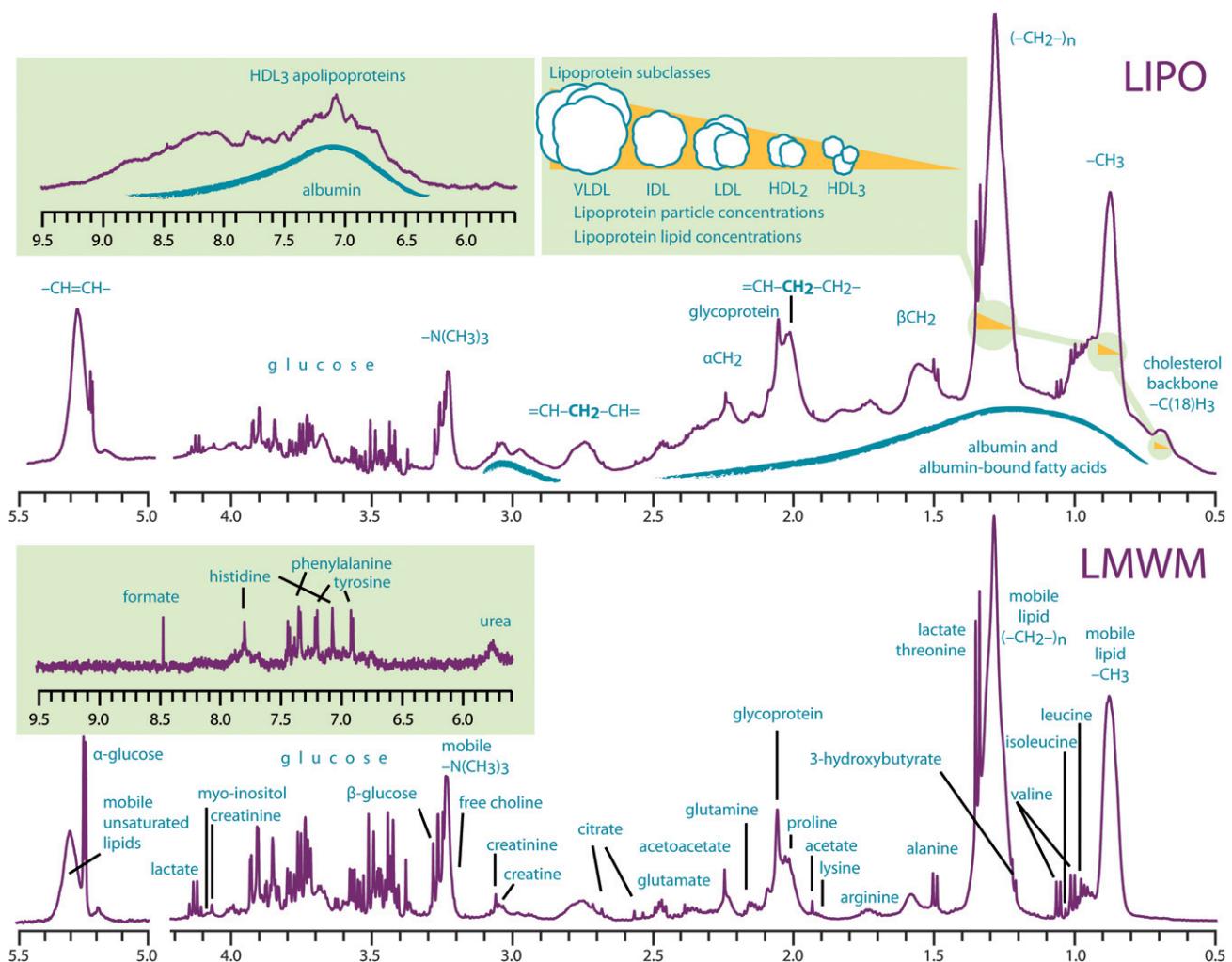
characterising the LIPO window, are not visible and the detection of the low-molecular-weight metabolites is considerably improved. These combined data represent the majority of the metabolic information available by  $^1\text{H}$  NMR metabolomics of native serum.<sup>1,4,9,10</sup> The spectral characteristics and the metabolic contents of these molecular windows are illustrated in Fig. 1. The protocol to be presented was established, optimised, and tested in our laboratory during approximately six months. In addition to all the testing and optimisation, we have now run around 12 000 actual study samples from a few epidemiological studies.

## Experimental

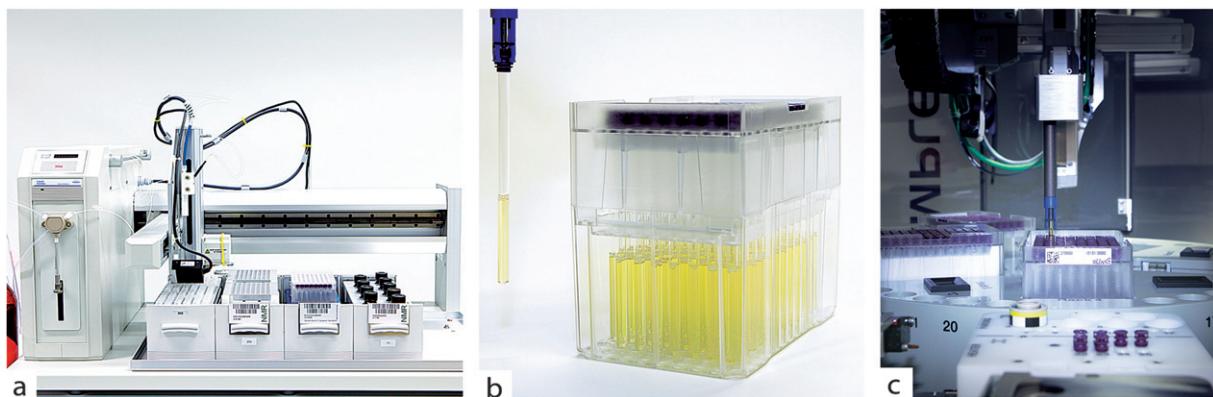
### $^1\text{H}$ NMR spectroscopy – instrumentation

Some of the key elements in the high-throughput serum NMR metabolomics instrumentation are depicted in Fig. 2. The buffer allocation and subsequent serum mixing is performed by a Gilson 215

Liquid Handler (Fig. 2a). The prepared samples are stored in 96-tube racks (Fig. 2b) ready to be inserted into one of the five well-plate positions in the SampleJet<sup>TM</sup> (Bruker BioSpin GmbH, Germany) sample changer (Fig. 2c) placed on top of the superconducting magnet inside which the actual NMR measurements take place. The robotic sample changer includes a cooling unit, which keeps the samples awaiting the measurements at a refrigerator temperature ( $+6\text{ }^\circ\text{C}$ ), and a preheating unit, which is used to warm up the sample just before the measurement. The samples are preheated to  $0.5\text{ }^\circ\text{C}$  above the physiological measurement temperature of  $37\text{ }^\circ\text{C}$  since some heat is lost during the sample transfer into the NMR probehead inside the magnet. Preheating is essential to reduce the time needed for temperature stabilisation and thereby one of the key issues in enabling a high-throughput for serum samples. Prior to starting the actual measurements, the (preheated) sample is kept idle inside the probehead at least for 2 min to ensure temperature stabilisation. In our set-up, the NMR data are measured using a Bruker AVANCE III spectrometer operating at 500.36 MHz ( $^1\text{H}$  observation frequency;



**Fig. 1** The NMR spectral characteristics and the metabolic contents of the two molecular windows – LIPO and LMWM. The LIPO window is dominated by broad signals arising from macromolecules, mainly lipoprotein lipids and albumin. Despite the broad overall characteristics and heavy overlap of the resonances, appropriate data analyses provide abundant information on lipoprotein particles as indicated by the inset illustrating the lipoprotein subclasses. In the LMWM window a pulse sequence that suppresses the macromolecule signals is applied, thus, enhancing the detection of smaller solutes. The residual water peak region (4.2–5.0 ppm) in the LIPO and LMWM windows is not shown. Abbreviations used: HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein.



**Fig. 2** Some of the key technical devices for high-throughput serum NMR metabolomics. (a) A Gilson 215 Liquid Handler for automatic sample preparation. (b) Prepared serum samples in a 96-tube rack and a 5 mm outer-diameter SampleJet NMR tube with serum (300 µl of original serum + 300 µl of the NMR buffer). (c) View inside the SampleJet™ (Bruker BioSpin GmbH, Germany) sample changer showing the well-plate positions to place the sample racks. The robotic sample changer includes a cooling unit as well as a preheating unit. The actual experimentation takes place in an NMR probehead inside the superconducting magnet.

11.74 T) and equipped with an inverse selective SEI probehead including an automatic tuning and matching unit and a *z*-axis gradient coil for automated shimming. A BTO-2000 thermocouple serves for temperature stabilisation at the level of approximately 0.01 °C at the sample. Notably, very stable and high-performance electronics is also a prerequisite for some of the new implemented concepts, including metabolite quantification without per sample chemical referencing or double tube systems.<sup>16</sup>

### Sample storage and preparation

The serum samples are stored in a freezer at –80 °C. Before sample preparation, the frozen samples are first slowly thawed in a refrigerator (+4 °C) overnight. The samples are then mixed gently and spun in a centrifuge at 3400 *g* to remove possible precipitate. Aliquots of each sample (300 µl) are mixed with 300 µl of sodium phosphate buffer (75 mM Na<sub>2</sub>HPO<sub>4</sub> in 80%/20% H<sub>2</sub>O/D<sub>2</sub>O, pH 7.4; including also 0.08% sodium 3-(trimethylsilyl)propionate-2,2,3,3-d<sub>4</sub> and 0.04% sodium azide). The sample preparation is done automatically with a Gilson Liquid Handler 215 (Fig. 2a) to 5 mm outer-diameter SampleJet NMR tubes (Fig. 2b). In the automatic preparation procedure, 300 µl of buffer is first transferred to the NMR tubes. After that, 300 µl of serum is added to the buffer and the resulting solution is mixed thoroughly by aspirating three times. Slow aspiration and mixing is required due to the high viscosity of serum and its high susceptibility to foam. In this way, the preparation of 96 samples for the NMR experimentation takes approximately 2 h.

### The molecular windows

A typical LIPO window together with the key signal assignments is illustrated on the top of Fig. 1. The LIPO window represents a conventional <sup>1</sup>H NMR spectrum of human serum showing broad overlapping resonances arising mainly from different lipid molecules in various lipoprotein particles.<sup>9,10</sup> The LIPO data are now recorded with 80k data points after 4 dummy scans using 8 transients acquired with an automatically calibrated 90° pulse and applying a Bruker noesypresat pulse sequence with mixing time of 10 ms and irradiation field of 25 Hz to suppress the water peak. The acquisition time is 2.7 s and the relaxation delay 3.0 s. The 90° pulse

is calibrated automatically for each sample. A constant receiver gain setting is used for all the samples.

The LMWM data are acquired with such spectrometer settings (using a *T*<sub>2</sub>-relaxation-filtered pulse sequence) that suppress most of the broad macromolecule and lipoprotein lipid signals and in that way enhance the detection of rapidly tumbling smaller solutes. A characteristic LMWM spectrum is shown at the bottom of Fig. 1. The LMWM data are recorded with 64k data points using 24 (or 16) transients acquired after 4 steady state scans with a Bruker 1D CPMG pulse sequence with water peak suppression and a 78 ms *T*<sub>2</sub>-filter with a fixed echo delay of 403 µs to minimise diffusion and *J*-modulation effects. The acquisition time is 3.3 s and the relaxation delay 3.0 s.

Both LIPO and LMWM data are processed and phase corrected in an automated fashion. Prior to Fourier transformations to spectra, the measured free induction decays for both LIPO and LMWM windows are zero-filled to 128k data points and then multiplied with an exponential window function with a 1.0 Hz line broadening. In total, the experimental time needed to record and process the LIPO and LMWM windows for one serum sample is less than 9 min (or around 8 min if 16 transients are acquired for the LMWM data), and the current set-up enables measurement of up to 180 serum samples per 24 h.

### Data storage, analyses and visualisation

Applications of metabolomics in epidemiological studies with thousands of samples generate vast amounts of data and require particular attention to the data storage, access and analyses. Our approach is an in-house designed database integrating all data in a similar and combinable fashion in all studies. The basis of the data analysis infrastructure is a centralised SQL-database running in a dedicated web-server. All the spectra and clinical metadata as well as the results of the data analyses are stored in the database and are thus available regardless of time and place. An application interface has also been created to make it possible to access the database programmatically. The idea is to have client software to operate through this interface and perform various data analyses automatically for large data sets and submit the results back to the server *via* the same application interface. At this point, for example, a first-stage

data analysis for LIPO and LMWM spectra for 1000 serum samples, including the self-organising map (SOM) analysis and visualisation of the clinical metadata, as well as lipoprotein lipid and metabolite quantification can well be achieved in a working day provided that the data have already been incorporated into the database.

The spectroscopic data can be seen either as a holistic profile for statistical classification of the sample (in terms of, *e.g.*, metabolic phenotypes, risk assessment, or diagnostics) or as a quantitative source for particular metabolites (*e.g.* triglycerides or glucose). In most of the published metabolomics applications the former approach has been taken.<sup>1,8,17</sup> It is quite a natural choice and makes the statistical analyses rather straightforward since there is no need to focus on individual resonances and their quantification. Instead, the strategy is to commence relating the whole spectral pattern to the biochemical and/or clinical condition of interest with the aid of statistical and chemometric data analyses.<sup>1,8,17</sup>

We usually start the data analysis and visualisation by applying the SOM analysis to the combined experimental NMR data from the two molecular windows. The SOM is an unsupervised pattern recognition technique that organises the input data according to data-driven similarity criteria. The end result is a two-dimensional map where, in this context, mutually similar NMR spectra of serum (as representatives of the individual metabolic phenotypes) are placed next to each other and on which all the clinical and biochemical measures can easily be visualised and compared.<sup>4,5,18</sup> We have recently developed and implemented the SOM analysis into a metabolomics framework with incorporated *p*-value statistics.<sup>4,5</sup>

In the long run, we believe that bringing metabolomics to clinical and epidemiological use necessitates a more specific metabolite approach than the one currently provided by the chemometric approaches analysing the entire spectral data.<sup>1,8,17,19</sup> In the case of lipoprotein lipid quantification from the LIPO window, we are implementing automated regression models with which the computation of the measures is instant (since the analysis with trained models is non-iterative).<sup>20</sup> The method of choice to quantify the low-

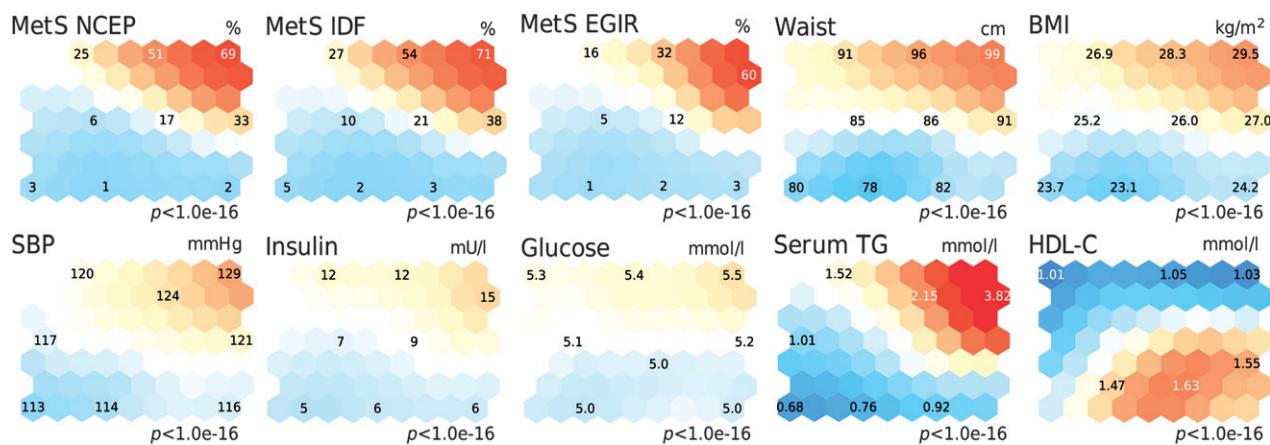
molecular-weight metabolites (as well as the residual lipoprotein lipid resonances) from the LMWM spectra is iterative lineshape fitting analysis.<sup>9,21</sup>

### A clinical application

We have previously presented an NMR metabolomics approach to study the disease continuum of diabetic complications and premature death in the case of 613 patients with type-1 diabetes<sup>4</sup> and also applied serum NMR metabolomics to study the early systemic signs of Alzheimer's disease.<sup>5</sup> However, in these studies earlier generation experimental protocols and equipment were used with mainly manual experimentation (the LIPO and LMWM measurements for one serum sample taking about 30 min in total). Here we present our first preliminary results of applying the presented high-throughput serum NMR metabolomics methodology in an extensive epidemiological study. Notably, the NMR experimentation as well as the data analysis were performed during approximately four weeks.

Fig. 3 shows the SOM analysis of the LIPO and LMWM spectra together with visualisation of some associated biochemical and clinical measures for 4407 serum samples from the Cardiovascular Risk in Young Finns Study.<sup>22</sup> It is notable that the data analysis *via* the SOM was based solely on the NMR spectral data and thereby provides purely data-driven metabolic phenotyping. The NMR-independent clinical and biochemical variables were used only to elucidate and interpret the observed structure of the SOM.

The component planes in Fig. 3 relate to clinically defined metabolic syndrome (MetS), a combination of disorders that together increase the risk of developing cardiovascular disease and diabetes.<sup>22</sup> The distribution of MetS in Fig. 3 indicates that the metabolic phenotypes characterising the northeast corner of the SOM relate to a high preponderance of MetS (as indicated by the three different clinical definitions of MetS).<sup>22</sup> The other seven measures shown in Fig. 3 are all components of the clinical definitions of MetS, namely, waist circumference, body mass index



**Fig. 3** Statistical colourings of selected clinical and biochemical variables in the SOM analysis of the combined LIPO and LMWM molecular windows from  $^1\text{H}$  NMR spectroscopy of 4407 serum samples (corresponding to 8814 spectra, see Fig. 1) from the Cardiovascular Risk in Young Finns Study.<sup>22</sup> The SOM analysis positions the samples so that the multi-metabolite differences in the metabolic phenotypes between nearby samples are minimised. A given serum sample (person) is in exactly the same place in each of the component planes. The colouring is according to the corresponding measures of the local residents within each hexagonal unit. The values in each component plane are colour-coded to visualise whether the value is above (reddish), at (white) or below (bluish) the median of the variable. The numbers on selected units tell the local mean value for that particular region. MetS refers to the metabolic syndrome with NCEP, IDF and EGIR being three different clinical definitions of the condition, namely by the National Cholesterol Education Program (2005 revised criteria), by the International Diabetes Federation, and by the European Group for the Study of Insulin Resistance, respectively.<sup>22</sup> BMI refers to body mass index, SBP to systolic blood pressure, TG to triglycerides, and HDL-C to high-density lipoprotein cholesterol.

(BMI), systolic blood pressure (SBP), insulin, glucose, serum triglycerides (TG) and high-density lipoprotein cholesterol (HDL-C). In the case of the first six, the MetS is associated with high values and for the HDL-C with low values of the corresponding measures. This indicates that the metabolic data revealed by serum NMR metabonomics bear a clear link to the (patho)-physiological pathways connected to the metabolic syndrome. This is by no means a surprise since, for example, glucose, serum TG, and HDL-C are known to be quantifiable from the LIPO window.<sup>9,10</sup> Overall, these first preliminary results, obtained using the newly established high-throughput NMR metabonomics set-up and the integrated data analysis, give further support to the applications of serum NMR metabonomics in clinical and epidemiological research.

## Conclusion

The data storage and automated spectral analyses are under continuous development. The aim is to maximise the number of metabolic measures that can be identified and automatically quantified. The set-up and analyses are very cost-effective in comparison to other, either traditional or systems biology, approaches on serum analytics. It is anticipated that the presented set-up paves the way for clinical and epidemiological serum NMR metabonomics.

## Acknowledgements

This work has been supported by the Academy of Finland Research Funding (J. V., O. T. R., M. J. S.), the Academy of Finland SALVE programme for 2009–2012 (M. R. J., M. J. S., M. A. K.), the Emil Aaltonen Foundation (M. K., T. L.), the Finnish Cardiovascular Research Foundation (J. V., O. T. R., M. J. S.), the Finnish Cultural Foundation (J. V., O. T. R.), the Finnish Foundation for Alcohol Studies (M. J. S.), the Sigrid Jusélius Foundation (M. J. S.), the Social Insurance Institution of Finland (J. V., O. T. R.), and the Tampere University Hospital Medical Fund (M. K., T. L.).

## Notes and references

- 1 M. Ala-Korpela, *Expert Rev. Mol. Diagn.*, 2007, **7**, 761–773.
- 2 E. Holmes, I. D. Wilson and J. K. Nicholson, *Cell*, 2008, **134**, 714–717.
- 3 J. K. Nicholson and J. C. Lindon, *Nature*, 2008, **455**, 1054–1056.
- 4 V.-P. Mäkinen, P. Soininen, C. Forsblom, M. Parkkonen, P. Ingman, K. Kaski, P.-H. Groop and M. Ala-Korpela, *Mol. Syst. Biol.*, 2008, **4**, 167.
- 5 T. Tukiainen, T. Tynkkynen, V.-P. Mäkinen, P. Jylänki, A. Kangas, J. Hokkanen, A. Vehtari, O. Gröhn, M. Hallikainen, H. Soininen, M. Kivipelto, P.-H. Groop, K. Kaski, R. Laatikainen, P. Soininen, T. Pirttilä and M. Ala-Korpela, *Biochem. Biophys. Res. Commun.*, 2008, **375**, 356–361.
- 6 J. K. Nicholson, *Mol. Syst. Biol.*, 2006, **2**, 52.
- 7 J. L. Griffin and A. Vidal-Puig, *Physiol. Genomics*, 2008, **34**, 1–5.
- 8 J. C. Lindon, J. K. Nicholson and E. Holmes, *The Handbook of Metabonomics and Metabolomics*, 2007, Elsevier, Amsterdam.
- 9 M. Ala-Korpela, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1995, **27**, 475–554.
- 10 M. Ala-Korpela, *Clin. Chem. Lab. Med.*, 2008, **46**, 27–42.
- 11 M. Ala-Korpela, P. Sipola and K. Kaski, *Ann. Med.*, 2006, **38**, 322–336.
- 12 T. Suna, A. Salminen, P. Soininen, R. Laatikainen, P. Ingman, S. Mäkelä, M. J. Savolainen, M. L. Hannuksela, M. Jauhainen, M.-R. Taskinen, K. Kaski and M. Ala-Korpela, *NMR Biomed.*, 2007, **20**, 658–672.
- 13 T. Y. Wong, G. Liew, R. J. Tapp, M. I. Schmidt, J. J. Wang, P. Mitchell, R. Klein, B. E. Klein, P. Zimmet and J. Shaw, *Lancet*, 2008, **371**, 736–743.
- 14 P. Hunter, *EMBO Rep.*, 2009, **10**, 20–23.
- 15 M. Basson, *Nature*, 2008, **451**, 903.
- 16 G. Wider and L. Dreier, *J. Am. Chem. Soc.*, 2006, **128**, 2571–2576.
- 17 J. Trygg, E. Holmes and T. Lundstedt, *J. Proteome Res.*, 2007, **6**, 469–479.
- 18 T. Kohonen, *Self-Organizing Maps*, 1995, Springer-Verlag, Heidelberg.
- 19 K. S. Opstad, C. Ladroue, B. A. Bell, J. R. Griffiths and F. A. Howe, *NMR Biomed.*, 2007, **20**, 763–770.
- 20 A. Vehtari, V.-P. Mäkinen, P. Soininen, P. Ingman, S. M. Mäkelä, M. J. Savolainen, M. L. Hannuksela, K. Kaski and M. Ala-Korpela, *BMC Bioinf.*, 2007, **8**(Suppl 2), S8.
- 21 P. Soininen, J. Haarala, J. Vepsäläinen, M. Niemitz and R. Laatikainen, *Anal. Chim. Acta*, 2005, **542**, 178–185.
- 22 N. Mattsson, T. Rönnemaa, M. Juonala, J. S. Viikari, E. Jokinen, N. Nutri-Kähönen, M. Kähönen, T. Laitinen and O. T. Raitakari, *Eur. Heart J.*, 2008, **29**, 784–791.