1H NMR metabonomics approach to the disease continuum of diabetic complications and premature death

Ville-Petteri Mäkinen1,2,3, Pasi Soininen4, Carol Forsblom5,3, Maija Parkkonen2,3, Petri Ingman5, Kimmo Kaski1, Per-Henrik Groop2,3,* and Mika Ala-Korpela1,*, on behalf of the FinnDiane Study Group

1 Computational Medicine Research Group, Laboratory of Computational Engineering, Systems Biology and Bioinformation Technology, Helsinki University of Technology, PO Box 9203, Helsinki FI-02015 HUT, Finland. Tel.: + 35894514833; E-mail: mika.ala-korpela@hut.fi
2 FinnDiane Study Group, Folkhålsan Research Center, Folkhålsan Institute of Genetics, Biomedicum Helsinki, University of Helsinki, PO Box 63, Helsinki FI-00014, Finland. Tel.: + 358919125452; Fax: + 358919125459; E-mail: per-henrik.groop@helsinki.fi and Mika Ala-Korpela, Computational Medicine Research Group, Laboratory of Computational Engineering, Systems Biology and Bioinformation Technology, Helsinki University of Technology, PO Box 9203, Helsinki FI-02015 HUT, Finland. Tel.: + 358503535457; Fax: + 35894514833; E-mail: mika.ala-korpela@hut.fi
3 Division of Nephrology, Department of Medicine, Helsinki University Hospital, Finland, 4 Laboratory of Chemistry, Department of Biosciences, University of Kuopio, Finland and 5 Instrument Centre, Department of Chemistry, University of Turku, Finland

* Correspondence: Per-Henrik Groop, FinnDiane Study Group, Folkhålsan Research Center, Folkhålsan Institute of Genetics, Biomedicum Helsinki, University of Helsinki, PO Box 63, Helsinki FI-00014, Finland. Tel.: + 358919125459; Fax: + 358919125452; E-mail: per-henrik.groop@helsinki.fi and Mika Ala-Korpela, Computational Medicine Research Group, Laboratory of Computational Engineering, Systems Biology and Bioinformation Technology, Helsinki University of Technology, PO Box 9203, Helsinki FI-02015 HUT, Finland. Tel.: + 358503535457; Fax: + 35894514833; E-mail: mika.ala-korpela@hut.fi

Received 8.10.07; accepted 5.12.07

Subtle metabolic changes precede and accompany chronic vascular complications, which are the primary causes of premature death in diabetes. To obtain a multimetabolite characterization of these high-risk individuals, we measured proton nuclear magnetic resonance (1H NMR) data from the serum of 613 patients with type I diabetes and a diverse spread of complications. We developed a new metabonomics framework to visualize and interpret the data and to link the metabolic profiles to the underlying diagnostic and biochemical variables. Our results indicate complex interactions between diabetic kidney disease, insulin resistance and the metabolic syndrome. We illustrate how a single 1H NMR protocol is able to identify the polydiagnostic metabolite manifold of type I diabetes and how its alterations translate to clinical phenotypes, clustering of micro- and macrovascular complications, and mortality during several years of follow-up. This work demonstrates the diffuse nature of complex vascular diseases and the limitations of single diagnostic biomarkers. However, it also promises cost-effective solutions through high-throughput analytics and advanced computational methods, as applied here in a case that is representative of the real clinical situation.

Molecular Systems Biology 12 February 2008; doi:10.1038/msb4100205

Subject Categories: bioinformatics; molecular biology of disease

Keywords: 1H NMR spectroscopy; biomarkers; metabonomics; serum; type I diabetes

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. This licence does not permit commercial exploitation or the creation of derivative works without specific permission.

Introduction

Type I diabetes is caused by an autoimmune reaction against the insulin-producing pancreatic β-cells and subsequent disturbance of normal blood glucose metabolism. Insulin replacement therapy cures the acute symptoms, but is not able to match the natural response to rising or falling glucose levels. This persistent metabolic imbalance is linked to high incidence of vascular complications such as diabetic kidney disease (DKD) (Finne et al., 2005), diabetic retinal disease (DRD) (Roy et al., 2004) and macrovascular diseases (MVDs) (Libby et al., 2005), all of which are co-occurring in vulnerable patients (Groop et al., 2005; Thorn et al., 2005; Ala-Korpela, 2007). The diagnosis, risk assessment and treatment of these conditions are currently determined by a number of biochemical and clinical variables, although none of these are conclusive on its own (Soedamah-Muthu et al., 2004; Stadler et al., 2006). Furthermore, the simultaneous clustering of complications and metabolic risk factors has not been studied by high-throughput analytical techniques that could reveal the multidimensional metabolic state of an individual more effectively.

The standard differential diagnostics in medicine may not be sufficient in detecting complex perturbations of biological systems (Zenker et al., 2007). Conditions such as insulin resistance and atherosclerosis stem from nonlinear interactive pathways between the genes (Hakonarson et al., 2007), gene expression (Sieberts and Schadt, 2007), metabolic environment (Goodacre, 2007) and the symbiotic microflora (Martin et al., 2007). To pinpoint the nodes and their roles in the disease...
networks requires a large number of samples with multi-dimensional quantitative data—a direct consequence of the curse of dimensionality. The genome-wide association studies have shown that this can be achieved at the DNA level (Frayling, 2007; Wellcome Trust Case Control Consortium, 2007). However, for personalized risk assessment and treatment the genetic approach is limited, as it does not take into account the dynamic environment, unlike the metabolomics approach (Nicholson and Wilson, 2003; Clayton et al, 2006), which has gained popularity as analytical technologies are evolving (Nicholson, 2006; Ala-Korpela, 2007; Salek et al, 2007).

Diabetic complications pose a difficult challenge to public health care, as populations grow older and life style becomes more sedentary and energy-rich (Reunanen et al, 2000). For this reason, we are aiming at new screening methods and metabolic characterization tools to find the vulnerable patients at an early stage when preventive treatment is still effective (Tenenbaum et al, 2004; Gross et al, 2005). Mass spectrometry and proton (1H) nuclear magnetic resonance (NMR) spectroscopy are the two key experimental methods in the area of ‘global biochemistry’ (Fernie et al, 2004). 1H NMR, in particular, is advantageous for screening, as it can efficiently extract detailed molecular information on a large number of metabolites in various biofluids (Tang et al, 2004; Beckonert et al, 2007; Ala-Korpela, 2008). The earliest experiments with plasma have already demonstrated this in type II diabetes (Nicholson et al, 1984). Recently, we have shown that DKD can be detected by 1H NMR of serum (Mäkinen et al, 2006) and that the metabolic syndrome (MetS) can be distinguished by multivariate methods and 1H NMR spectroscopy (Suna et al, 2007). A similar approach has been applied to cardiovascular disease, but with limited success (Brindle et al, 2002; Kirschenslohr et al, 2006). The metabolic changes in type II diabetes have also been studied by chromatographic methods (Yang et al, 2004; Wang et al, 2006). Animal models have provided encouraging results and further justification for the metabolomic NMR approach (Williams et al, 2005; Clayton et al, 2006; Salek et al, 2007), but more experience from human populations is needed (Griffin and Nicholls, 2006; Ala-Korpela, 2007).

In this study, the emphasis is on the metabolic continuum that underlies the slow and often elusive development of chronic complications. We focus first on DKD due to its high significance in the treatment and prognosis of diabetic patients (Gross et al, 2005). Our main goal, however, is to extract a metabolite manifold that highlights not only DKD, but also other important clinical and biochemical characteristics and their complex relationships (Fernie et al, 2004). The combination of 1H NMR of serum and metabolomic mapping provides the necessary insight: neural network analysis and statistically verified visualizations of both the spectroscopic and clinical data will not only help decision making in clinical environment, but will also increase the knowledge of multifactorial disease states that are difficult to pinpoint by reductionist approaches (Sams-Dodd, 2005; Weckwerth and Morgenthal, 2005; Loscalzo et al, 2007). The new source of information can then be used in personalized risk assessment as a cost-effective high-throughput alternative to a collection of specific biomarker assays (Lindon et al, 2006; Ala-Korpela, 2008).

**Results**

**Molecular windows to metabolism**

We obtained serum samples from the FinnDiane study to measure two molecular windows for 613 patients with type I diabetes. A typical 1H NMR spectrum of human serum is characterized by broad resonances from the lipid molecules of lipoprotein particles, such as the –CH3 group of triglycerides, cholesterol compounds and phospholipids (Figure 1C and D). This so-called lipoprotein lipid (LIPO) window is a complex mixture of the aforementioned lipid signals, serum albumin and albumin-bound fatty acid resonances across the aliphatic region, and the less intense signals from smaller molecules such as creatinine, lactate and glucose (Ala-Korpela, 1995, 2008).

To reveal the resonances from smaller molecules, the spectrometer settings can be altered to suppress most of the broad resonances while still enabling the detection of the more mobile low molecular weight molecules (LMWM). The LMWM window is dominated by the numerous glucose resonances between 3.1 and 3.9 p.p.m., although some lipid signals still remain (Figure 1B and E). The spectral shapes from both windows share a common axis of chemical shift and are superimposable, except for the intensity scaling constant. For example, lactate creates a strong doublet signal around 1.28 p.p.m. in the LMWM window, but only small shapes on top of the wider lipid and albumin resonances in the LIPO window. On the other hand, most of the molecules with the –CH3 group contribute to the prominent signal around 0.8 p.p.m. in the LIPO window, but only the more mobile species can be detected in the LMWM window.

**Spectral profile of type I diabetes**

A self-organizing map (SOM) (Kohonen, 2000) was constructed from the 1H NMR data (Figure 1A). The SOM was the result of reducing the 613 experimental spectra into 9 x 9 = 81 representative spectral models, each of which was assigned to a unique hexagonal unit on the map grid. Subsequently, a best-matching model was determined for each experiment, thus each patient had a best-matching unit or a ‘place of residence’ on the map. Localized similarity is the fundamental idea behind the SOM, that is, neighboring units or the patients therein are more similar to each other than those from the opposite sides of the map. In this case, similarity was defined by the arithmetic multidimensional difference between the two spectral models; thus, any two neighbors shared more metabolic characteristics (their spectra looked the same) than two randomly picked patients, on average.

As the SOM is analogous to a geographic map in all but the way the patients’ coordinates are assigned, it is possible to use ordinary demographic methods to visualize the properties of patients in different metabolic neighborhoods (Supplementary data 1). Here, we started by coloring the units based on the percentage of DKD patients within a local population (Figure 1A). The highest value of 70% can be seen on the western edge of the map, on the unit at row 5 and column 1 (5,1). The unit at (2,9) near the northeast corner has the lowest percentage of DKD (16%), and is located far from (5,1). In fact,
when looking at the overall coloring, the DKD patients are clustered on the western side, whereas the patients with fewer complications are concentrated on the northeast corner of the map.

The typical spectral profile of DKD was examined by comparing the spectral models at (5,1) and (2,9) to see if any of the metabolite resonances differed. Each of the spectral plots, such as Figure 1C, consists of three components. First, the mean model over all data is depicted as a solid black line to serve as a constant reference to which the spectral models can be compared. The second curve alongside the reference is the unit-specific spectral model, which was split into orange or blue segments depending on where the model exceeded or was less intense than the reference. As the first two curves are close to each other in terms of absolute intensity, a third curve that depicts the proportional differences is helpful in revealing any significant changes in intensity. In Figure 1C for instance, the third curve was drawn below the two absolute intensity curves.

![Figure 1](image-url)
and painted similarly to the unit-specific model. The lipoprotein lipid resonances at around 0.81, 1.23 and 1.95 p.p.m. show reduced values from the mean, whereas the albumin background is increased. In Figure 1E, the two creatinine peaks at 2.98 and 3.99 p.p.m. are higher for the DKD region (5,1). However, looking at just two map locations is not enough for accurate interpretation; a more global perspective is required.

Diabetic kidney disease, the metabolic syndrome and mortality

A majority of the patients in this study had either micro- (22%) or macroalbuminuria (37%), the spatial distributions of which are revealed by class-specific colorings in Figure 2A and B. The macroalbuminuric group (clinically diagnosed with DKD) is concentrated on the western side of the map ($P=1.2 \times 10^{-8}$), whereas the diagnostically intermediate microalbuminuric group does not form any statistically significant pattern ($P=0.077$). While the DKD status was determined by urine albumin excretion, the map was constructed solely based on the $^1$H NMR spectra of serum, thus illustrating the systemic biological connection between the two biofluids.

All-cause mortality in Figure 2C ($P=0.00057$) was estimated based on 8.2 ± 0.6 years of follow-up and scaled to the percentage of deaths in a decade (number of deaths per 1000 patient years). As expected, there is a clear connection between DKD and increased mortality, and the highest value of 25% is observed at (7,1) close to the highest DKD percentage at (5,1). Additional details are available in Figure 1 in Supplementary data 2.

The MetS (NCEP, 2002; Eckel et al., 2005) represents a binary classification according to a clinical scoring system (a score of 3 or more is considered positive) that combines several components of insulin resistance and obesity (Figure 2D–H). Patients with the lowest score 1 reside on the northern part of the SOM with a 42% ($P=1.1 \times 10^{-5}$) occupancy at (1,8), and those with a score above 3 are tightly concentrated on the southwestern corner, with hardly any overlap with the first group ($P=1.8 \times 10^{-8}$ for score 4, $P=1.3 \times 10^{-7}$ for score 5).

The SOM colorings indicate strong associations between DKD, mortality and the MetS, but with subtle differences. For instance, the first two MetS categories split the normoalbuminuric northeastern side, rather than spread evenly to mirror the DKD group (Figure 2A, B, D and E). The highest percentage of DKD at (5,1) does not coincide with the highest MetS scores at (9,1). Interestingly, a history of macrovascular complications in these patients seem to be related more to the MetS than to DKD (Figure 2I), although the numbers are too small for statistical significance ($P=0.035$ for the MVD pattern). Finally, the highest 10-year mortality of 25% is observed at (7,1) in Figure 2C, where the MetS and DKD overlap the most.

Confounding factors and treatments

The colorings for age (mean ± s.d. 40 ± 11 years), type I diabetes duration (27 ± 10 years) and gender (311 males, 302 females) show only minor spatial clustering and weak statistical significance (Figure 3A–C). Furthermore, the observed patterns show little similarity to DKD, the MetS or MVD in Figure 2, suggesting that the major chronological and physiological determinants of risk do not confound the biochemical characteristics.

Most of the patients were on medication or had undergone laser treatment for DRD. Antihypertensive treatment ($P=2.7 \times 10^{-5}$) is most common (up to 75%) in those areas on the western side of the map (Figure 3G), which have a high percentage of DKD in Figure 2B and elevated blood pressure in Figure 3D and E, as expected. Furthermore, the same areas have a high proportion of DRD ($P=0.00027$), although the pattern is more widely dispersed on the southwestern half of the SOM (Figure 3H). Lastly, patients with the highest MetS scores have also the widest waist (94 cm) in Figure 3F and the highest percentage (30%) of lipid-lowering treatment (Figure 3I).

Biochemical backdrop of diabetic complications

To create a more comprehensive metabolic picture than that in Figure 1, we colored the map according to estimates from regression models and spectral features that quantify key biochemical variables directly from the $^1$H NMR spectra (Mäkinen et al., 2006). The previously used null hypothesis of no dependence between the map and a target variable could
not be used here, as both the map and the coloring were derived from the spectra. The dynamic range of statistical fluctuations was nevertheless estimated to determine a suitable color scale (Supplementary data 1).

Triglyceride concentration is a part of the MetS definition, and the highest unit-specific value (3.6 mmol/l) can therefore be seen at the southwest corner of the map, where also the MetS is most severe (Figure 4A). Total serum cholesterol is only partially linked to triglycerides, as it produces an ascending north-south pattern on the SOM (Figure 4B). Nevertheless, the highest value (6.0 mmol/l) coincides with that of triglycerides at (1,9). HDL\textsubscript{2} cholesterol exhibits a more complicated pattern (Figure 4C), with the highest value (0.64 mmol/l) located near the southeast corner, and the lower values (0.41–0.47 mmol/l) located on the western side.

Creatinine singlets at 2.98 and 3.99 p.p.m. and urea around 5.68 p.p.m. have a strong association with DKD (Figure 2B) and produce similar colorings of approximately doubled values on the western side of the SOM as compared with the eastern side (Figure 4D and E). Furthermore, lactate at 1.28 and 4.05 p.p.m. and acetate at 1.86 p.p.m. follow the same trend (Figure 4G and H), but with patterns closer to the MetS. Serum albumin is higher on the eastern half of the map (Figure 4F), which was already evident in the typical spectral model of DKD in Figure 1C and D.

The effect of the glucose resonances between 3.1 and 3.9 p.p.m. was suppressed when constructing the SOM, but nevertheless the remaining doublet from \(\alpha\)-glucose around 5.19 p.p.m. was enough to separate patients who had high blood glucose at the time of sample collection (Figure 4I). Although the daily variations are large in type I diabetic patients, high glucose values do partially overlap with the MetS and other complications on the southwestern half.

Validation by standard biochemistry

Standard non-NMR measurements of a number of metabolites were also available for validating the NMR-derived results with methodologically independent data (Table 1 in Supplementary data 2). In addition, the metabolite manifold from the spectra was supplemented with biomolecular data that could not be detected by NMR to confirm biologically relevant observations.

Figure 5A–C depicts the three most important lipid variables (see also the NMR-derived colorings in Figure 4A–C), with highly statistically significant patterns for triglycerides (\(P=3.1 \times 10^{-19}\)), total cholesterol (\(P=2.1 \times 10^{-11}\)) and HDL\textsubscript{2} cholesterol (\(P=2.5 \times 10^{-6}\)). HDL\textsubscript{2}C (\(P=4.3 \times 10^{-6}\)) has a pattern similar to HDL\textsubscript{2}C in Figure 5F. The highest concentration of ApoB (117 mg/dl) coincides with the highest triglycerides and
cholesterol in the southwest corner at (9,1) and appears to be elevated if either of the two is higher than average ($P = 1.1 \times 10^{-12}$ for ApoB).

Creatinine is easily detectable by $^1$H NMR as two singlets at 2.98 and 3.99 p.p.m.; accordingly, the non-NMR measurement in Figure 5D closely matches the NMR-derived creatinine (and biologically correlated urea) in Figure 4D and E ($P = 9.2 \times 10^{-7}$). The 24-h urine albumin in Figure 5I is indicative of persistent albuminuria by definition ($P = 4.6 \times 10^{-10}$), and albuminuria is tightly linked to the metabolite manifold from NMR, as was evident in Figure 2A and B.

Patients in the southwest corner have significantly higher weight-adjusted insulin doses ($P = 0.00053$), but, despite that, the worst glycemic control, which is indicated by HbA$\text{c}$ of 9.7% ($P = 2.9 \times 10^{-8}$). Corresponding patterns were previously observed for the MetS (Figure 2F–H) and NMR-derived lactate, acetate and glucose (Figure 4G–I), which can be seen as an indication of insulin resistance.

**Summary of clinical and metabolic characteristics**

In the final stage, we merged map units into larger districts and collected the regional characteristics into tabular format to create a summary of the metabolic characteristics (Figure 6; Table 3 in Supplementary data 2). For instance, the southwest district (Figure 6A) is populated by patients with high relative risk due to DKD and the MetS (7.8 versus 2.0–2.1), compared...
with the districts in the north and northeast (Figure 6E and F). Biochemically, these groups differ significantly: triglycerides (2.8 versus 0.84 mmol/l), cholesterol (5.8 versus 4.5–4.9 mmol/l), serum creatinine (116 versus 87–88 µmol/l) and 24 h urine albumin (356 versus 13–16 mg) are high, whereas HDL-subfractions are low in the MetS district. Patients in the MetS corner have also poor glycemic control (HbA1c 9.2 versus 8.2%) and larger waist circumference (94 versus 81–83 cm).

The northwest corner is characterized by high susceptibility to microvascular complications, as DKD (58%) and DRD (58%) are common but MetS (42%), MVD (10%) and 10-year mortality (17%) are closer to average. Furthermore, there is a notable difference in HDL$_2$C when comparing the northwest corner and the southeast corner of the SOM (0.45 versus 0.60 mmol/l), which is also visible in Figure 4C. On the other hand, total cholesterol is higher in the southeast (4.9 versus 5.7 mmol/l), so the difference in the ratio of HDL$_2$ to total cholesterol is less pronounced.

The two districts with favorable phenotypes are similar if complications alone are considered, but some minor biochemical differences can be observed (Figure 6E and F). Although both districts share relatively high HDL$_2$C concentrations, ApoA$_1$, ApoB, HDL$_2$C and total cholesterol are lower in the northwestern district. The districts are also different with respect to serum albumin (Figure 4F). Triglycerides and glycemic control (HbA1c 8.2%) exhibit no clear differences, as was seen also in the spectral estimates of triglycerides in Figure 4A.

**Discussion**

Diabetes is associated with increased incidence of vascular complications and premature aging (Morris et al., 2001; Rönntack et al., 2005; Pambianco et al., 2006). Here, we characterized the metabolic background of adverse clinical phenotypes within this high-risk population by $^1$H NMR spectroscopy of serum. Specifically, we were able to identify differences and similarities in the biochemical patterns of DKD, insulin resistance, DRD, macrovascular complications and all-cause mortality in 613 type I diabetic patients.

The $^1$H NMR analyses were targeted at two molecular windows simultaneously. The LIPO window carries information particularly on lipoprotein lipids and albumin, whereas the LMWM window contains signals from smaller metabolites such as creatinine and glucose (Mäkinen et al., 2006; Alakorpela, 2008). Lipoprotein levels are altered in type 1 diabetes in the presence and during the development of complications (Chaturvedi et al., 2001; Lyons et al., 2004; Thomas et al., 2006), but they alone cannot capture all aspects of the diabetic condition. It is therefore plausible that the inclusion of the LMWM window as an integral part of the biochemical analysis is crucial for polydiagnostic applications that are targeted at several concurrent disease mechanisms (Kell, 2006).

The raw metabolite manifold obtained from the two molecular windows was not usable as such, but multivariate computational analysis was required to transform the spectral data into an accessible form of information. We chose the SOM (Kohonen, 2000; Suna et al., 2007) instead of linear decomposition methods such as principal component analysis (PCA), since the SOM algorithm preserves the spectral shapes and summarizes the data via a small number of spectral models that are directly relatable to individual samples (Figure 1). The decompositions, on the other hand, are focused on the variables and on the reduction of the data into abstract multidimensional linear spaces that may not have a direct connection to the observed NMR resonances.

PCA and partial least squares methods have been established as the standard pattern recognition methods in the field of metabonomics (Dieterle et al., 2006; Trygg et al., 2007). However, the SOM analysis has been applied, for instance, in spectroscopic classification problems (Beckonert et al., 2003; Lavine et al., 2004; Suna et al., 2007), in molecular conformation analysis (Hyvönen et al., 2001) and in studies of gene–metabolite interactions (Hirai et al., 2004). The last case is an example of complementary application of both PCA and SOM. Some studies have also compared SOM with other methods (Mangiameli et al., 1996; Giraudel and Lek, 2001; Astel et al., 2007) and, in most cases, the use of SOM produced additional insight compared with PCA (see also Supplementary data 1).

The ambiguity in model selection and the danger of overfitting are the main drawbacks of neural network algorithms (Lampinen and Kostiainen, 1999). Here, the statistical significance of the observed patterns was addressed from a practical point of view, with emphasis on the ability of $^1$H NMR to extract relevant metabolic information from a serum sample. The question of overtraining was not relevant in this context, as we were not using the SOM as a predictive tool but included only the spectra as inputs. Thus, we were able to integrate different sources of data in a user-friendly fashion by stochastically normalized map colorings, without giving up statistical reliability.

We chose not to perform supervised feature extraction or selection before the SOM analysis. This may have led to suboptimal predictive performance for a particular clinical endpoint but, on the other hand, the results reflect the intrinsic metabolic information content of the spectra as such, and the complex dependencies between the various bodies of data. For instance, we have previously shown that a linear regression model with nonlinear feature extraction can indicate the albumin excretion categories from the $^1$H NMR spectra of serum (Mäkinen et al., 2006), but this model has less utility in nondiabetic populations. The SOM presented here, however, is not methodologically dependent on albuminuria and thus easier to relate to the general population.

Omitting explicit feature selection does not guarantee optimal results, but other options would require detailed methodological treatment that is not yet available, and might complicate the interpretations of the original spectral shapes. DKD, for instance, is a discrete classification based on one biomarker (urine albumin), and while it makes a perfect test case for analytical techniques (Mäkinen et al., 2006), a systems biology approach to diabetic complications should not aim to predict albuminuria, but to distinguish the multifactorial disease states, even if not all of them are associated with overt albuminuria (Caramori et al., 2006). Put differently, a predictive model will only work in the clinical practice if the phenotype to be predicted is accurate (Loscalzo et al., 2007).

Despite the shortcomings, albumin excretion is an excellent indicator of progressing kidney damage after the early stages of
the disease. Albuminuria starts to present itself within the first 15 years of diabetes (Gross et al, 2005; Caramori et al, 2006); patients included in this work had a long diabetes duration (26 years on average), so the normoalbuninuric subgroup has a low risk of ever developing DKD. Consequently, they represent a generally healthier subset of diabetic patients, especially when considering that low-grade albuminuria is associated with macrovascular complications, even outside type I diabetes (Gerstein et al, 2001). Still, normoalbuninuria does not preclude the cardiovascular risk factors: a significant portion of patients in the map region with dyslipidemia and insulin resistance (Figure 6A) had normal (17%) or intermediate (20%) albumin excretion as opposed to macroalbuminuria (63%).

The metabolic differences between most normo- and macroalbuminuric (DKD) patients were evident in the 1H NMR spectra of serum. DKD was associated with elevated triglycerides, lower HDL cholesterol and decreased albumin in the LIPO window. Other studies have also reported the connection between albuminuria and triglycerides (Chaturvedi et al, 2001; Jenkins et al, 2003; Thomas et al, 2006), but the exact role of HDL metabolism remains unclear. Serum creatinine and urea are two waste products that are normally excreted by the kidneys and, accordingly, the LMWM window revealed elevated values for the macroalbuminuric group, although none of the patients had end-stage renal disease. Crucially, the microalbuminuric patients did not clearly identify with either the low-risk or high-risk regions of the SOM, again highlighting the limited usefulness of any single biomarker in complex disease environments (Caramori et al, 2005).

The MetS and DKD are strongly associated in the Finnish type I diabetic population (Thorn et al, 2005). Our results from the metabonomic analysis were similar: the SOM regions with patients that have a detectable loss in kidney function (i.e., elevated creatinine and urea, decreased serum albumin) overlapped with insulin resistance and related problems in glucose metabolism (dyslipidemia, high insulin dose, high HbA1c, elevated lactate and acetate and high fasting glucose) (Krentz et al, 1991; Lovejoy et al, 1992; Avogaro et al, 1996; Choi et al, 2002). Interestingly, the neighborhood with the most severe insulin resistance did not coincide with the highest values of creatinine and urea, which suggests that there is a subtle systematic difference between the two clinical conditions. Furthermore, the highest mortality was observed on the intersection of the two defects. Unfortunately, the data set size in this study was insufficient for any definitive conclusions, but we have made similar observations with the non-NMR data for the full FinnDiane population (Mäkinen et al, in preparation), so these results nevertheless support a multifactorial approach to the study of pathophysiology (Loscalzo et al, 2007).

The discrepancy between the MetS and DKD could not be explained by nonbiochemical factors. For instance, although lipid treatment was most common in the MetS neighborhood, this group of patients still had the highest triglyceride concentration. This suggests that DKD with lower triglycerides is not a product of lipid-lowering treatment alone. In contrast, the highest percentage of antihypertensive treatment coincided with the highest blood pressure and DKD, but not with the MetS, which in turn suggests that the MetS with lower albumin excretion is not just a product of decreased albuminuria due to blood pressure medication (Thomas and Atkins, 2006). Furthermore, there was little regional gender difference in the MetS-DKD half of the SOM, so the wide waist circumference in the MetS neighborhood did not represent a gender bias. Hence, our results did not match some of the previous findings on statistical gender groupings in 1H NMR experiments of plasma (Kirschenlohr et al, 2006).

Macrovascular events are the primary targets of intervention, as they are the most common cause of premature death in type I diabetic patients (Libby et al, 2005; Stadler et al, 2006). In this respect, the difference between insulin resistance and kidney function becomes significant; the SOM neighborhood with the highest percentage of MVD was also the one with the strongest insulin resistance and the highest MetS scores, which agrees well with previous studies (Sierra-Johnson et al, 2006). The result is only suggestive, as the numbers were low and we had only cross-sectional data on MVD available, but this finding nevertheless demonstrates the sensitivity of the metabonomics approach to MVD vulnerability, as opposed to the albuminuria classification alone. Undoubtedly, triglyceride concentration is the most significant serum biomarker (Davis et al, 2007; Pambianco et al, 2007), and its role may be even further emphasized here, as triglycerides produce pronounced 1H NMR resonances.

DRD was the most common (51%) complication among the study population. Map regions with DKD or insulin resistance hallmarks had the highest percentages, but DRD did not specifically associate with either of the two. The connection to mortality was less obvious due to the large number of DRD cases, which agrees with previous observations (Torffvit et al, 2005).

Low concentration of HDL cholesterol was the common, albeit not exclusive, denominator for all complications, which suggests that HDL2 has a protective effect, or is a marker of a favorable phenotype with respect to both micro- and macrovascular complications (Cutri et al, 2006). This is in contrast to many observations regarding type I diabetes (Chaturvedi et al, 2001; Jenkins et al, 2003; Thomas et al, 2006), but the disconnect may be due to the statistical models that were used in these studies. There is a negative correlation between HDL cholesterol and serum triglycerides, so any linear model that includes both the two is likely to emphasize one at the expense of the other. Furthermore, men and women have different concentrations, so gender bias is likely to influence the results. To resolve the controversy, studies that use modern multivariate pattern recognition techniques in addition to the classical medical statistics should be performed on large population-based cohorts.

The biological heterogeneity of diabetic complications make the borderline between health and disease ambiguous, as is the case with many other slow pathophysiological processes. Our work illustrates this fundamental diagnostic challenge: DKD, DRD, the MetS and MVD shared much of the same biochemical basis, but nevertheless did not conclusively define each other. Even though the patients in this study were carefully selected to represent clinically relevant phenotypes, the metabolic landscape remained diffuse.
This work is, to our knowledge, the first metabolomics study on premature death and vascular complications in a large human cohort. We used only serum to characterize the patients, and yet the high-risk metabolic features were easily observable. This is an encouraging result with respect to general applicability as, unlike type I diabetes, urine albumin (or any other single biomarker) does not have an equally critical role in type II diabetes, let alone in the non-diabetic population. Furthermore, our application of $^1$H NMR metabolomics and statistical visualizations may improve the tracking of patients’ progress in the diabetic disease continuum in a way not attainable by traditional approaches. Hence, it may become possible to re-route the multimetabolite path of a vulnerable patient away from adverse clinical endpoints and towards a more favorable phenotype before it is too late (Ala-Korpela et al., 2006).

**Materials and methods**

**Study population**

Patients with type I diabetes were recruited by the Finnish Diabetic Nephropathy Study (FinnDiane), which is a nationwide multicenter effort to identify genetic and clinical risk factors for DKD. The study protocol was in accordance with the Declaration of Helsinki and approved by the local ethics committee in each of the participating health care centers. Diagnostic criteria for type I diabetes included age of onset below 35, the transition to permanent insulin treatment within a year from onset and C-peptide negativity.

Data on medication, cardiovascular status and diabetic complications were registered by a standardized questionnaire, which was completed by the patient’s attending physician according to the medical file. Death certificates were obtained from the national registry maintained by the Population Register Centre of Finland. The average follow-up time was 8.2 ± 0.6 years (4972 patient years in total).

The classification of renal status was made centrally according to urinary albumin excretion rate (AER) in at least two out of three successive overnight urine samples. Absence of kidney disease was defined as normoalbuminuria (AER < 20 μg/min), while the presence of overt kidney disease was defined as macroalbuminuria (AER ≥ 200 μg/min). The intermediary range is referred simply as microalbuminuria (20 μg/min ≤ AER < 200 μg/min). Albumin from 24-h urine samples was also available, and it was used as a biochemical variable in parallel with the longitudinal records of albuminuria.

The MetS scores were calculated according to the NCEP ATP III recommendations (NCEP, 2002), with every type I diabetic patient having a base score of 1 for hyperglycemia (Table 2 in Supplementary data 2). DRD was defined present if a patient had undergone laser eye treatment. MVD was defined as a pooled composite of coronary heart disease, acute myocardial infarction, stroke and peripheral vascular disease. The events were pooled to have more cases for the statistical analyses. Of the 54 cases, 40 had coronary heart disease, 24 had acute myocardial infarction, 12 had undergone coronary bypass, 15 had cerebral stroke and 17 had undergone peripheral vascular bypass.

Patients for the $^1$H NMR experiments were chosen based on the renal status. First, a random subset of macroalbuminuric patients was selected from the FinnDiane clinical database with preference for those individuals with genetic and clinical information already available. Next, sex- and age-matched peers were chosen from the normo- and microalbuminuric group, with preference for normoalbuminuria and availability of information. In total, 613 patients were included, of which 251 (41%) were normoalbuminuric, 137 (22%) had microalbuminuria and 225 (37%) had macroalbuminuria. Patients with end-stage renal disease or kidney transplant were excluded. The study set does not reflect the population-based cross-section, except within the macroalbuminuric group.

$^1$H NMR spectroscopy

$^1$H NMR data from serum samples at 37 °C were recorded on a Bruker AVANCE spectrometer with a field strength of 500.13 MHz. The reference substance (sodium 3-trimethylsilyl[2,2,3,3-d$_4$]propionate (TSP)) 40 mmol/l, MnSO$_4$ 0.6 mmol/l in 99.8% D$_2$O) was placed coaxially into the NMR sample container (o.d. 5 mm, contains 430 μl of serum) in a separate tube (o.d. 1.7 mm, supported by a Teflon adapter). This double-tube system was chosen to avoid mixing of the sample fluid and reference substance, which would make the absolute metabolite quantification less reliable (Ala-Korpela, 1995, 2008; Ma¨kinen et al., 2006).

The $^1$H NMR experiments were targeted at two different molecular windows: LIPO and LMWM. For the LIPO spectra, 128 transients were recorded with a 90° flip angle, a 6.2 s acquisition time and a 0.1 s relaxation delay. The LMWM data were collected with a standard one-dimensional Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence with a 325 ms T$_1$-filter and a fixed 400 μs echo delay to eliminate diffusion and J-modulation effects. Forty-eight transients were collected after 16 dummy scans with a 6.2 s acquisition time and an 8.7 s relaxation delay. Water suppression was not used. The free induction decays (FID) with 65 536 data points were zero-filled and multiplied by an exponential window function with a 1.0 Hz line-broadening for the LIPO spectra and 0.5 Hz line-broadening for the LMWM spectra. The preprocessing of the FIDs and subsequent Fourier transformations were performed on the PERCH software platform (PERCH Solutions Ltd, Kuopio, Finland).

Metabolite intensities in each spectrum were scaled according to the area of the respective TSP reference signal at 0 ppm. The TSP area was obtained by first subtracting the linear baseline between −0.1 and 0.1 ppm., and then integrating over the remainder. In the LMWM window, the effects of the water peak around 4.6 ppm were removed by fitting a Lorenzian tail on the aliphatic side and a piece-wise polynomial curve on the aromatic side, and then subtracting the tails. Furthermore, a minor piece-wise linear correction was applied in both windows, mainly affecting the region between 3 and 5 ppm. Peaks in the LMWM spectra were aligned by first estimating the shift offsets at selected locations, then linearly interpolating the offsets where direct peak detection was difficult and finally re-mapping the frequency axis according to the estimated shift offsets. The same correction was performed also for the LIPO spectra, although the effects were small due to wider line shapes.

Self-organizing map and statistical analysis

Before constructing the SOM, the spectra were truncated to 0–3.3 ppm (LIPO) and 0.7–5.8 ppm (LMWM), and the chemical shift resolution was reduced to 0.003 ppm (LIPO) and 0.001 ppm (LMWM). Data points between 4.2 and 5.0 ppm were omitted and the intensities between 3.22 and 3.38 ppm were given only 0.1% weight in the SOM algorithm to attenuate the effect of glucose. The respective intensity units for the two molecular windows were adjusted such that the total variance of the data points within the LIPO window was equal to that in the LMWM window. This was done to ensure that both windows would have comparable effect to the SOM structure while preserving the relative intensities between individual experiments.

We chose a 9 × 9 hexagonal sheet of map units with a Gaussian neighborhood function for the analysis (7.6 samples per unit). The SOM was initialized based on the two first principal components of the input data and finished by the batch-training algorithm, with a 3.4% topological error.

After the positions of the study subjects on the SOM were computed, the map was colored according to the demographic properties (i.e., clinical and biochemical variables) within different parts of the SOM (Supplementary data 1). To verify statistical significance, the variability for the colorings was estimated by first computing the squared deviations from the global average for each map unit, and then adding these unit-specific values to obtain a single descriptive statistic. Put differently, the ‘bumpiness’ of a coloring was expressed as one
Numerical value. Some bumps will occur by change alone, and to estimate the expected null distribution, the target variable (DKD status for instance) was shuffled randomly several times, and at each round, the descriptive statistic was recomputed. The probability of getting the observed statistic (or a more extreme value) from the random distribution determined the statistical significance. P. The null distributions from the permutation analysis were also the basis of the color scale in each figure so that categorical and continuous variables, possibly with some data missing, can be compared visually while maintaining the statistical interpretation.

Mortality in Figure 2C was estimated by first computing the coloring for the percentage of deaths observed on each map unit, and then normalizing these values by the average unit-specific follow-up time in decades. Note that this is equivalent to the number of deaths in 1000 patient years. Only patients that had not died (91%) were used for the follow-up time estimation to avoid bias due to deaths. The unadjusted event percentages and follow-up times are available in Figure 1 in Supplementary data 2. The discrepancy between Figures 2C and 6A (max 25 versus 26%) stems from different formulations: the former was obtained from the smoothed map estimates for the number of deaths and follow-up period, whereas the latter was obtained from point estimates within the resident patient subgroup. Relative risk for early death in Figure 6 was obtained as the ratio of observed deaths and the mortality in the corresponding age segments of the entire Finnish population within the follow-up period (data from Statistics Finland).

In addition to clinical variables, numerous quantities were estimated computationally from the spectral data. Easily distinguishable peaks in the LMWM window such as urea around 5.68 p.p.m. and acetate at 1.86 p.p.m. were quantified by direct peak integration. The wide baseline signal from serum albumin was estimated by fitting a polynomial curve at selected locations in the LIPO window and computing the respective curve area. Lastly, semilinear regression modeling was applied to quantities that were also available via non-NMR methods (Mäkinen et al., 2006). All analyses were performed using the SOM Toolbox 2.0 for the Matlab environment (URL: http://www.cis.hut.fi/projects/somtoolbox/), and in-house scripts for subsequent quantification, map coloring and permutation analyses.

Supplementary information
Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

Acknowledgements
The skilled technical assistance by Antti Niinikoski and Taru Tukiainen is gratefully acknowledged. The study was supported by grants from the Folkhälsoin Research Foundation, Samfundet Folkhälsan, the Jenny and Antti Wihuri Foundation and the Graduate School of Electrical and Communications Engineering at Helsinki University of Technology. This work was also supported by the Centre of Excellence Program of the Academy of Finland (KK, MAK). For a complete listing of the FinnDiane Study Group, please see Supplementary data 4.

References
Ala-Korpela M (1995) 1H NMR spectroscopy of human blood plasma. Progr Nucl Magn Reson Spectr 27: 475–554
Ala-Korpela M (2007) The potential role of body fluid 1H NMR metabonomics as a prognostic and diagnostic tool. Expert Rev Mol Diagn 7: 71–77
Ala-Korpela M (2008) Critical evaluation of 1H NMR metabonomics of serum as a methodology for disease risk assessment and diagnostics. Clin Chem Lab Med 46: 27–42
Ala-Korpela M, Sipola P, Kaski K (2006) Characterization and molecular detection of atherosclerosis by magnetic resonance—potential tools for individual risk assessment and diagnostics. Ann Med 38: 322–336
Astell A, Tsakovski S, Barbieri P, Simeonov V (2007) Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. Water Res 41: 4566–4578
Avogaro A, Crepaldi C, Miola M, Maran A, Pengo V, Tiengo A, del PS (1996) High blood ketone body concentration in type 2 non-insulin dependent diabetic patients. J Endocrinol Invest 19: 99–105
Becknort E, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC, Nicholson JK (2007) Metabolic profiling, metabonomic and molecular procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat Protoc 2: 2692–2703
Becknort E, Monnerjahn J, Bonk U, Leibfritz D (2003) Visualizing metabolic changes in breast-cancer tissue using 1H NMR spectroscopy and self-organizing maps. NMR Biomed 16: 11
Brindle J, Antti H, Holmes E, Tranter G, Nicholson J, Bethell H, Clarke S, Schofield P, McKeiligin E, Mosedale D, Grainger D (2002) Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H NMR-based metabonomics. Nat Med 8: 1439–1444
Caramori M, Fioretto P, Mauer M (2000) The need for early predictors of diabetic nephropathy risk: is albumin excretion rate sufficient? Diabetes 49: 1399–1408
Caramori M, Fioretto P, Mauer M (2006) Enhancing the predictive value of urinary albumin for diabetic nephropathy. J Am Soc Nephrol 17: 339–342
Chaturvedi N, Fuller J, Taskinen M (2001) Differing associations of lipid and lipoprotein disturbances with the macrovascular and microvascular complications of type 1 diabetes. Diabetes Care 24: 2071–2077
Choi CS, Kim YB, Lee FN, Zabolotny JM, Kahn BB, Youn JH (2002) Lactate induces insulin resistance in skeletal muscle by suppressing glycolysis and impairing insulin signaling. Am J Physiol Endocrinol Metab 283: E233–E240
Clayton T, Lindon J, Cloarec O, Antti H, Charuel C, Hanton G, Provost J, Le NJ, Baker D, Walley R, Everett J, Nicholson J (2006) Pharmaco-metabonomic phenotyping and personalized drug treatment. Nature 440: 1073–1077
Cutri B, Hime N, Nicholls S (2006) High-density lipoproteins: an emerging target in the prevention of cardiovascular disease. Cell Res 16: 799–808
Davis T, Bruce D, Davis W (2007) Prevalence and prognostic implications of the metabolic syndrome in community-based patients with type 1 diabetes: the Fremantle Diabetes Study. Diabetes Res Clin Pract 78: 414–417
Dieterle F, Ross A, Schlotterbeck G, Senn H (2006) Metabolite projection analysis for fast identification of metabolites in metabonomics. Application in an amidarone study. Anal Chem 78: 3551–3561
Eckel R, Grundy S, Zimmet P (2005) The metabolic syndrome. Lancet 365: 1415–1428
Fernie A, Trethewey R, Krotzky A, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol 5: 763–769
Finne P, Reunanen A, Stenman S, Groop P, Groenhagen-Riska C (2005) Incidence of end-stage renal disease in patients with type 1 diabetes: the Fremantle Diabetes Study. J Am Soc Nephrol 16: 3551–3561
Finnar A, Gretehuy K, Krotzky A, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol 5: 763–769
Finne P, Reunanen A, Stenman S, Groop P, Grönhagen-Riska C (2005) Incidence of end-stage renal disease in patients with type 1 diabetes. JAMA 294: 1782–1787
Frayling TM (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nat Rev Genet 8: 657–662
Gerstein H, Mann J, Qiong Y, Zinman B, Dinneen S, Hoogwerf B, Hallén J, Young J, Rashkow A, Joyce C, Nawaz S, Yusuf S (2001) Albutamurina and risk of cardiovascular events, death, and heart failure in diabetic and nondiabetic individuals. JAMA 286: 421–436
Giraudel J, Lek S (2001) A comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. Water Res 41: 4566–4578
Goodacre R (2007) Metabolomics of a Superorganism. J Nutr 137: 2595–266S
Griffin J, Nicholls A (2006) Metabolomics as a functional genomic tool for understanding lipid dysfunction in diabetes, obesity and related disorders. Pharmacogenomics 7: 1095–1107

Groop P, Forsblom C, Thomas M (2005) Mechanisms of disease: pathway-selective insulin resistance and microvascular complications of diabetes. Nat Clin Pract Endocrinol Metab 1: 100–110

Gross J, de AM, Silveiro S, Canani L, Caramoni M, Zelmanovitz T (2005) Diabetic nephropathy: diagnosis, prevention, and treatment. Diabetes Care 28: 164–176

Hakonarson H, Grant SF, Bradford JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skrabalan R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SC, Rappaport EF, Orange JS, Monos DS et al. (2007) A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 448: 591–594

Hirai M, Yano M, Goodeowde D, Kanaya S, Kimura T, Awazuahara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proc Natl Acad Sci USA 101: 10205–10210

Hyvönen MT, Hiltunen Y, El-Deredy W, Ojala T, Vaara J, Kovanesen PT, Ala-Korpela M (2001) Application of self-organizing maps in conformational analysis of lipids. J Am Chem Soc 123: 810–816

Jenkins A, Lyons T, Zheng D, Otvos J, Lackland D, McGee D, Garvey W, Klein R, Dr. (2003) Serum lipoproteins in the diabetes control and complications trial/epidemiology of diabetes intervention and complications cohort: associations with gender and glycaemia. Diabetes Care 26: 810–818

Kell D (2006) Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug Disc Today 11: 1085–1092

Kirschenvlohr H, Griffin J, Clarke S, Rhydwen R, Grace A, Schofield P, Brindle K, Metcalfe J (2006) Proton NMR analysis of plasma is a weak predictor of coronary artery disease. Nat Med 12: 705–710

Kohonen T (2000) Self-organizing Maps. New York, USA: Springer-Verlag and Heidelberg

Krentz A, Singh B, Nattrass M (1991) Impaired glucose tolerance is characterized by multiple abnormalities in the regulation of intermediary metabolism. Diabet Med 8: 848–854

Lampinen J, Kostiainen T (1999) Overtraining and model selection with the self-organizing map. Neural Networks 3: 1911–1915

Lavine B, Davidson C, Westover D (2004) Spectral pattern recognition using self-organizing maps. J Chem Inf Comput Sci 44: 1056–1064

Libby P, Nathan D, Abraham K, Brunzelli J, Fradkin J, Haffner S, Huseh W, Rewers M, Robert B, Savage P, Skralatos S, Wassef M, Rabadan-Diehl C (2005) Report of the national heart, lung, and blood institute—national institute of diabetes and kidney disease working group of cardiovascular complications of type 1 diabetes mellitus. Circulation 111: 3489–3493

Lindon J, Holmes E, Nicholson J (2006) Metabolomics techniques and applications to pharmaceutical research and development. Pharm Res 23: 1075–1088

Loscalzo J, Kohane I, Bladner P (2007) Human disease classification and subclasses in the DCCT/EDIC cohort. Invest Ophthalmol Vis Sci 45: 910–918

Mäkinen VP, Soininen P, Forsblom C, Parkkonen M, Ingman P, Kaski K, Groop PH, Ala-Korpela M (2006) Diagnosing diabetic nephropathy by 1H NMR metabolomics of serum. Magn Reson Mater Physiol 19: 281–296

Mangiameli M, Chen S, West D (1996) A comparison of SOM neural network and hierarchical clustering methods. Eur J Op Res 93: 402–417

Martin FPI, Dumas ME, Wang Y, Legido-Quigley C, Yap IKS, Tang H, Zirah S, Murphy GM, Cloarec O, Linden JC, Sprenger N, Fay LB, Kochhar S, van Bladeren P, Holmes E, Nicholson JK (2007) A top-down systems biology view of microbiome-mammalian metabolic interactions in a mouse model. Mol Syst Biol 3: 112

Morrish N, Wang S, Stevens L, Fuller J, Keen H, the WMSGT (2001) Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes. Diabetologia 44: S14–S21

National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) (2002) Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. Circulation 106: 3142–3421

Nicholson JK (2006) Global systems biology, personalized medicine and molecular epidemiology. Mol Syst Biol 2: 52

Nicholson JK, O’Flynn MP, Sadler PJ, Macleod AF, Juul SM, Sønksen PH (1994) Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. Biochim Biophys Acta 1217: 365–375

Nicholson JK, Wilson ID (2003) Understanding ‘global’ systems biology: metabolomics and the continuum of metabolism. Nat Rev Drug Discov 2: 668–676

Pambianco G, Costacou T, Ellis D, Becker D, Klein R, Orchard T (2006) The 30-year natural history of type 1 diabetes complications. Diabetes 55: 1463–1469

Pambianco G, Costacou T, Orchard T (2007) The prediction of major outcomes of type 1 diabetes: a 12 year prospective evaluation of three separate definitions of the metabolic syndrome, and their components and estimated glucose disposal rate: the Pittsburgh Epidemiology of Diabetes Complications Study experience. Diabetes Care 30: 1248–1254

Reunanen A, Kangas T, Martikainen J, Klaukka T (2000) Nationwide survey of comorbidity, use, and costs of all medications in Finnish diabetic individuals. Diabetes Care 23: 1265–1271

Rönnback M, Fagerudd J, Forsblom C, Petterson-Fernholm K, Reunanen A, Groop P (2005) Finnish Diabetic Nephropathy (FinnDiane) Study Group: altered age-related blood pressure pattern in type 1 diabetes. Circulation 110: 1076–1082

Roy M, Klein R, O’Colmain B, Klein B, Moss S, Kenpen J (2004) The prevalence of diabetic retinopathy among adult type 1 diabetic persons in the United States. Arch Ophthalmol 122: 546–551

Salek R, Maguire M, Bentley E, Rubtsov D, Hough T, Chessman M, Nunez D, Sweatman B, Haselden J, Cox R, Conner S, Griffin J (2007) A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat and human. Physiol Genomics 29: 99–108

Sams-Dodd F (2005) Target-based drug discovery: is something wrong? Drug Disc Today 10: 139–147

Sieberts SK, Schadt EE (2007) Moving toward a systems genetics view of disease. Mamm Genome 18: 389–401

Sierra-Johnson J, Johnson B, Allison T, Bailey K, Schwartz G, Turner S, Jhang M, Bjerke J, Sams-Dodd F (2005) Target-based drug discovery: is something wrong? Drug Disc Today 10: 139–147

Siegert S, Schadt EE (2007) Moving toward a systems genetics view of disease. Mamm Genome 18: 389–401

Sierra-Johnson J, Johnson B, Allison T, Bailey K, Schwartz G, Turner S (2006) Correspondence between the Adult Treatment Panel III criteria for metabolic syndrome and insulin resistance. Diabetes Care 29: 668–672

Soedamah-Muthu S, Chaturvedi N, Toeller M, Ferriss B, Reboldi P, Michel G, Manes C, Fuller J (2004) Eurodiab prospective complications study group risk factors for coronary heart disease in type 1 diabetic patients in Europe: the Eurodiab Prospective Complications Study. Diabetes Care 27: 530–537

Stadler M, Auinger M, Anderwald C, Küstenbauer T, Kramar R, Feinböck C, Insigler K, Kronenberg F, Prager R (2006) Long-term mortality and incidence of renal dialysis and transplantation in type 1 diabetes mellitus. J Clin Endocrinol Metab 91: 3814–3820

Suna T, Salminen A, Soininen P, Laatikainen R, Ingman P, Mäkelä S, Savolainen M, Hannuksela M, Jauhiainen M, Taskinen M, Kaski K, Ala-Korpela M (2007) 1H NMR metabolomics of plasma

© 2008 EMBO and Nature Publishing Group Molecular Systems Biology 2008 11
lipoprotein subclasses: elucidation of metabolic clustering by self-organising maps. NMR Biomed 20: 658–672
Tang H, Wang Y, Nicholson JK, Lindon JC (2004) Use of relaxation-edited one-dimensional and two-dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. Anal Biochem 325: 260–272
Tenenbaum A, Motro M, Schwammenthal E, Fisman E (2004) Macrovascular complications of metabolic syndrome: an early intervention is imperative. Int J Cardiol 97: 167–172
Thomas M, Atkins R (2006) Blood pressure lowering for the prevention and treatment of diabetic kidney disease. Drugs 66: 2213–2234
Thomas M, Rosengård-Bärlund M, Mills V, Rönnback M, Thomas S, Forsblom C, Cooper M, Taskinen M, Viberti G, Groop P (2006) Serum lipids and the progression of nephropathy in type 1 diabetes. Diabetes Care 29: 317–322
Thorén L, Forsblom C, Fagerudd J, Thomas M, Petterson-Fernholm K, Saraheimo M, Wadén J, Rönnback M, Rosengård-Bärlund M, Björkестen CG, Taskinen M, Groop P (2005) Metabolic syndrome in type 1 diabetes: association with diabetic nephropathy and glycemic control (the FinnDiane Study). Diabetes Care 28: 2019–2024
Torffvit O, Lövestam-Adrian M, Agardh E, Agardh C (2005) Nephropathy, but not retinopathy, is associated with the development of heart disease in type 1 diabetes: a 12-year observation study of 462 patients. Diabet Med 22: 723–729
Trygg J, Holmes E, Lundstedt T (2007) Chemometrics in metabonomics. J Proteome Res 6: 469–479
Wang T, Gona P, Larson M, Tofler G, Levy D, Newton-Cheh C, Jacques P, Rifai N, Selhub J, Robins S, Benjamin E, D’Agostino R, Vasan R (2006) Multiple biomarkers for the prediction of first major cardiovascular events and death. N Engl J Med 355: 2631–2639
Weckwerth W, Morgenthal K (2005) Metabolomics: from pattern recognition to biological interpretation. Drug Discov Today 10: 1551–1558
Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678
Williams R, Lenz E, Lowden J, Rantalainen M, Wilson I (2005) The metabolomics of aging and development in the rat: an investigation into the effect of age on the profile of endogenous metabolites in the urine of male rats using 1H NMR and HPLC-TOF MS. Mol Biosyst 1: 166–175
Yang J, Xu G, Hong Q, Liebich H, Lutz K, Schmülling R, Wahl H (2004) Discrimination of type 2 diabetic patients from healthy controls by using metabolomics method based on their serum fatty acid profiles. J Chromatogr B 813: 53–58
Zenker S, Rubin J, Clermont G (2007) From inverse problems in mathematical physiology to quantitative differential diagnoses. PLoS Comp Biol 3: e204

Molecular Systems Biology is an open-access journal published by European Molecular Biology Organization and Nature Publishing Group.
This article is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Licence.