Metabolic Phenotypes, Vascular Complications and Premature Deaths in a Population of 4,197 Patients with Type 1 Diabetes

Ville-Petteri Mäkinen1,2,3*, Carol Forsblom2,3, Lena M Thorn2,3, Johan Wadén2,3, Daniel Gordin2,3, Outi Heikkilä2,3, Kustaa Hietala2,3, Laura Kyllönen2,3, Janne Kytö2,3, Milla Rosengård-Bärlund2,3, Markku Saraheimo2,3, Nina Tolonen2,3, Maija Parkkonen2,3, Kimmo Kaski1, Milla Ala-Korpela1,2,3 and Per-Henrik Groop2,3*; on behalf of the FinnDiane Study Group

1Department of Biomedical Engineering and Computational Science, Helsinki University of Technology, Finland; 2Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Finland; 3Division of Nephrology, Department of Medicine, Helsinki University Hospital, Finland

*Corresponding authors

Ville-Petteri Mäkinen MSc (Tech)
Department of Biomedical Engineering and Computational Science,
Computational Medicine Research Group,
Helsinki University of Technology,
ville-petteri.makinen@computationalmedicine.fi

Per-Henrik Groop MD, DMSc, Docent
Folkhälsan Institute of Genetics,
Folkhälsan Research Center, Biomedicum Helsinki,
P.O. Box 63, FI-00014, University of Helsinki, Finland
per-henrik.groop@helsinki.fi

Received 7 March 2008 and accepted 22 May 2008.

Additional information for this article can be found in an online appendix at http://diabetes.diabetesjournals.org.
**Objective:** Poor glycemic control, elevated triglycerides and albuminuria are associated with vascular complications in diabetes. However, few studies have investigated combined associations between metabolic markers, diabetic kidney disease, retinopathy, hypertension, obesity and mortality. Here the goal was to reveal previously undetected association patterns between clinical diagnoses and biochemistry in the FinnDiane dataset.

**Research Design and Methods:** At baseline, clinical records and serum and 24h-urine samples of 2,173 men and 2,024 women with type 1 diabetes were collected. The data were analyzed by the self-organizing map (SOM), which is an unsupervised pattern recognition algorithm that produces a two-dimensional layout of the patients based on their multi-variate biochemical profiles. At follow-up, the results were compared against all-cause mortality during 6.5 years (295 deaths).

**Results:** The highest mortality was associated with advanced kidney disease. Other risk factors included i) a profile of insulin resistance, abdominal obesity, high cholesterol, triglycerides and low HDL$_2$C, and ii) high adiponectin and high LDLC for older patients. The highest population adjusted risk of death was 10.1-fold (95% CI 7.3-13.1) for men and 10.7-fold (7.9-13.7) for women. Non-significant risk was observed for a profile with good glycemic control and high HDL$_2$C, and for a low cholesterol profile with a short diabetes duration.

**Conclusions:** The SOM analysis enabled detailed risk estimates, described the associations between known risk factors and complications, and uncovered statistical patterns difficult to detect by classical methods. The results also suggest that diabetes *per se*, without an adverse metabolic phenotype, does not contribute to increased mortality.
Patients with type 1 diabetes are susceptible to severe microvascular complications such as proliferative retinopathy and chronic kidney disease, which are often accompanied by cardiovascular disease and premature death (1,2). Currently, the risk assessment and diagnostics rely on the urine albumin excretion, serum creatinine and lipid profile (3,4). In many cases, however, the biochemical measurements are treated as independent factors without explicit attention to the overall metabolic imbalance behind the complications. While the risk factors for cardiovascular disease and diabetic complications have been verified statistically in large clinical studies (5-7), the overall picture on the mutual relationships and their relevance for risk assessment remains fragmented.

The metabolic syndrome (8) is one attempt to describe the co-occurrence of vascular complications and insulin resistance, but so far its applicability to type 1 diabetes and exact definition remain controversial (9,10). Moreover, gradually developing conditions, such as cardiovascular disease, do not present a physiologically clear border between health and disease, so quantitative risk assessment tools are needed to augment and even replace discrete differential categorizations (11). Hence we are developing new approaches to attain more accurate phenotypes without excessive cost, both through high-throughput analytics (12,13) and computational methods (14,15).

In this work, we develop an analysis framework based on the self-organizing map (SOM) and statistically verified visualizations (13,16,17) for a large clinical study. Our goal is to characterize typical phenotypes (or metabolic profiles) that can be associated with high or low mortality during several years of follow-up. We do not question the role of albuminuria as the primary marker of increased risk, but we aim for a more comprehensive metabolic description of the conditions that manifest as albumin excretion. We show how numerous biochemical variables from serum and urine can be integrated under one statistical model while maintaining interpretability of the results, and we also demonstrate how non-linear multivariate statistics can reveal complex phenotypes that are difficult to detect by classical methods.

Research Design and Methods

Patients with type 1 diabetes were recruited by the Finnish Diabetic Nephropathy Study (FinnDiane), a nationwide multi-center effort to identify genetic and clinical risk factors for diabetic nephropathy. Diagnostic criteria for type 1 diabetes included age of onset below 35 years and transition to insulin treatment within a year of onset. Four patients were excluded due to insufficient biochemical data. The design was cross-sectional (n = 4,197), but with longitudinal records of albuminuria and clinical events before baseline, and with all-cause mortality data available after an average of 6.5 years of follow-up from baseline (25,714 patient years). The study protocol was in accordance with the Declaration of Helsinki, and approved by the local ethics committee in each of the participating centers.

Biochemical data came both from centrally organized measurements (90% of values) and from local health care centers and hospitals (10%). When both were available, the centrally measured value was used. The
pattern of missing values was regular (Online Appendix 1), but no significant sampling bias was detected.

**Clinical Definitions**

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. Vitality was obtained from the national registry maintained by the Population Register Center of Finland.

The classification of renal status was made centrally according to urinary albumin excretion rate (AER) in at least two out of three consecutive overnight or 24h-urine samples. Patients on renal replacement therapy (dialysis or transplantation) were classified as having end-stage renal disease (ESRD). Absence of nephropathy was defined as normoalbuminuria (AER <20 µg/min or <30 mg/24h), while overt kidney disease (DKD) was obtained by pooling macroalbuminuria (AER ≥200 µg/min or ≥300 mg/24h) and ESRD. The intermediary range was defined as microalbuminuria (20 ≤AER <200 µg/min or 30 ≤AER <300 mg/24h). Timed albumin excretion from the latest 24h-urine collection for each patient was also available, and it was used as a biochemical variable in parallel with the longitudinal records of AER.

The metabolic syndrome (MetS) was defined as a score of ≥3 according to the modified NCEP ATP III criteria (10,13,18), where every patient with type 1 diabetes has an initial score of one for hyperglycemia. Diabetic retinopathy (DRP) was defined present if a patient had undergone laser treatment of the retina. Macrovascular disease was defined as a pooled end-point of coronary heart disease, myocardial infarction, stroke and peripheral vascular disease. Blood pressure was measured twice with 2 min intervals in the sitting position after a 10 min rest.

**Biochemical Measurements**

Serum lipid and lipoprotein concentrations were measured from fasting blood samples at the research laboratory of Helsinki University Central Hospital, Division of Cardiology, Finland. Total cholesterol and triglycerides were determined enzymatically using an autoanalyzer (Cobas Mira or Mira Plus; ABX Diagnostics, Montpellier, France). Total HDL- and HDL₃-cholesterol were determined enzymatically using an assay reader (HTS 7000 Plus Bio; Perkin Elmer, Wellesley, MA, USA). HDL₂-cholesterol was calculated by subtracting HDL₃-cholesterol from total HDL-cholesterol. LDL-cholesterol was calculated according to the Friedewald formula. Serum apolipoprotein A-I, A-II and B concentrations were determined by immunoassays (Orion Diagnostica, Espoo, Finland). Serum and 24h-urine creatinine (enzymatic), 24h-urine albumin (immunoturbidimetric) and C-reactive protein (radioimmunoassay), and C-peptide (radioimmunoassay) were quantified at the Helsinki University Central Hospital laboratory, Finland. Adiponectin and mannan binding lectin were measured as previously described (19,20). 24h-urine urea (enzymatic), and Na and K (ion selective electrode) were measured on a Cobas Integra analyzer (Hoffmann-La Roche, Basel, Switzerland) by Medix Laboratories, Finland. Hemoglobin A₁c was determined by standardized assays at local health care centers and hospitals.
Statistical Analyses
To characterize metabolic profiles and compare them to the current clinical classifications, a self-organizing map (SOM) was constructed based on the biochemical variables only. Before analysis, variables that were highly correlated (e.g. 24h-urine creatinine and urea) were pruned if considered biologically redundant, which left 14 variables in the final training set (Online Appendix 1). The input data was rank-transformed (shifted and scaled rank indices from -1 to 1) for both sexes separately to ensure that each variable had an identical value distribution irrespective of gender-specific differences in absolute values. Simultaneously, the impact of outliers and extreme values was attenuated. Missing values were ignored in the transformation, but imputed from the preprocessed data by least-squares linear regression before analysis.

The SOM is an unsupervised pattern recognition method (13,16) and was employed here for automated comparisons between the multi-variate biochemical profiles. The resulting map layout of patients becomes such that those who have similar profiles are as close to each other as possible, whereas those who have different profiles are placed far apart on the map (Online Appendix 2). Here, a 7 x 10 grid of hexagonal map units was chosen with a Gaussian neighborhood function (60 patients per unit). The SOM algorithm was initialized based on the two first principal components of the input data and finished by the batch training procedure.

After the patients' positions were computed, the map was colored according to the patients' characteristics (clinical variables) within different regions (13). To verify that the results were statistically reliable, 10,000 random colorings were computed by permuting the data values to obtain empirical P-values, for each of the variables separately. The null distributions from the permutation analysis were also the basis of the color scale in each figure so that the categorical and continuous variables, possibly with some data missing, could be compared visually while maintaining the statistical interpretation. In addition, 95% confidence intervals were obtained by bootstrapping and are listed in the text when available. The aforementioned techniques were also applied to the biochemical variables, but since they were the inputs for the SOM, the null hypothesis was no longer valid. The respective P-values and confidence intervals were thus omitted.

Relative risk of dying (Fig. 4) was estimated by first computing the colorings for age and for the frequency of deaths observed on each map unit, and adjusting the estimated mortality against the expected mortality in the respective age segment in the Finnish population, for men and women separately, during the follow-up period. The unadjusted death percentages and follow-up times are available in Online Appendix 3. The population history was collected by Statistics Finland (URL:http://www.stat.fi).

The SOM analysis was performed with the MeliKerion software package for the Matlab/Octave programming environment (Online Appendix 2). A web-based interface for the software is also available (URL:http://www.computationalmedicine.fi/software). In-house scripts were used for the data preprocessing and additional analyses. The computational analyses for the study took approximately 2 hours on a workstation with a single core 3.2 GHz CPU.

Results
Initial analyses showed that the SOM was heavily influenced by the physiological differences between men and women. To avoid confounding effects, the gender-specific features were eliminated by rank transformation (see Methods). Figure 1 illustrates the gender distribution on the SOM of the preprocessed data, and shows no spatial separation between men and women.

The patient's locations on the map were determined by their similarity to the SOM profiles. The profiles, in turn, were derived from the average properties of the local population in the corresponding neighborhood. During the training process, this reciprocal relationship stabilized to a set of models, a number of which are depicted in Figure 1. For instance, patients in the south-west corner (row 7, column 1) were characterized by a profile with good glycemic control (low hemoglobin A1c), high HDL cholesterol, low apolipoprotein B and low C-reactive protein. Patients in the opposite corner in the north-east (2,10) were summarized by a profile of high levels of triglycerides and low HDL cholesterol.

By examining several profiles from various parts of the map the global features begin to emerge. For example, serum creatinine is higher on the northern regions (1,1; 1,6; 2,10) if compared with the south (7,1; 7,10) and apolipoprotein A-I is higher in the west (1,1; 7,1) compared with the east (2,10; 7,10). With this many variables, though, the bar profiles are not the most visually effective way to investigate the global properties, and thus the map coloring approach was applied to better reveal the statistical patterns.

**Vascular Complications and Premature Death**

The patients that are located on a given hexagonal unit determine the color for the respective area of the SOM (13). Figure 2 depicts the spatial distributions of clinical categories for the male patients. Not surprisingly, the highest 47% (38-57%) 10-year mortality is observed near the highest DKD prevalence (84%, 80-89%), and is coupled with a history of macrovascular complications (Fig. 2A,B,C). Women have lower mortality (36%, 26-46%) and DKD prevalence (71%, 65-77%), but the coupling with macrovascular complications remains (Online Appendix 4). The highest percentages of DRP coincide with DKD and high mortality in the north with 84% (79-88%) prevalence for men and 77% (72-83%) for women; on the southern half less than one third of patients have DRP (Fig. 2D). The MetS does not predict death accurately, since the highest percentages (men 88%, 83-93%; women 86%, 81-91%) occur in the north-eastern corner (Fig. 2E), away from the areas with the highest mortality. A detailed breakdown of the DKD grades and the MetS scores is available in Online Appendix 3.

Age and diabetes duration (Fig. 2F,G) contribute to the development of diabetic complications, but do not fully explain the deaths: although the patients on the north-western regions of the map are older (men 44.3, 42.9-45.9; women 43.2, 41.4-45.1 years), they have average mortality. The same regions are associated with lower insulin doses (men 0.62, 0.59-0.65; women 0.60, 0.57-0.63 IU/kg) and lower hemoglobin A1c, especially towards the south (Fig. 2H,I). C-peptide levels are negligible (<10 pmol/l) in every map region (Fig. 2J).
Systolic blood pressure is highest in the north (men 147, 145-151; women 142, 139-146 mmHg), and is linked to high mortality (Fig. 2K,L). Patients in the same regions are on medication, with up to 93% (90-96%) of men and 85% (81-90%) of women under antihypertensive treatment (Fig. 2M). Abdominal obesity is most prevalent in the north-eastern parts of the map (Fig. 2N,O), with wide waist (men 96, 94-98; women 87, 85-89 cm) and large BMI (men 27.0, 26.5-27.5; women 26.6, 26.1-27.3 kg/m²). Neither of the anthropometric markers predict death accurately.

Biochemical Features of Diabetic Complications

The complex nature of lipoproteins and lipids is explicitly revealed by the SOM (Fig. 3A-D, Online Appendix 5). The MetS region in the north-east is characterized by the highest triglyceride concentration (men 2.7, women 2.2 mmol/l) and high LDL and total cholesterol, but low HDL cholesterol (men 0.23, women 0.32 mmol/l). Closer to the western corner, HDL is restored (men 0.49, women 0.67 mmol/l) with elevated LDL (men 3.9, women 3.7 mmol/l). Patients on the southern half have lower triglycerides and cholesterol, except for HDL that exhibits an ascending east-west pattern. Lipid lowering treatment is more common in the north, but fails to explain the changes in lipoprotein fractions (Fig. 3E).

The number of VLDL/IDL/LDL particles (apolipoprotein B in Fig. 3G) is the highest in the MetS corner (men 123, women 126 mg/dl), with an overall pattern similar to cholesterol and triglycerides. Apolipoprotein A-I is related to HDL cholesterol, as expected (Fig. 3F).

24h-urine albumin (Fig. 3H) is consistent with the DKD classification (Fig. 2C), and the highest serum creatinine concentrations (men 221, women 162 µmol/l) are found near the high-mortality regions (Fig. 1A,3H). This advanced DKD phenotype in the north is also associated with decreased 24h-urine creatinine (men 11.0, women 8.2 mmol), elevated serum potassium (men 4.9, women 4.7 mmol/l) and decreased 24h-urine potassium (men 50, women 42 mmol) in Figure 3J-L.

The highest adiponectin levels can be observed in the north and north-west (men 20, women 26 mg/l). Albuminuria coincides with high adiponectin only on the western quadrant; the MetS corner in the north-east shows a relative reduction in adiponectin (Fig. 3H), but higher serum C-reactive protein levels (men 4.3, women 7.6 mg/l) in Figure 3N. Mannan-binding lectin concentrations exhibit no clear patterns (Fig. 3O).

High-Risk Metabolic Phenotypes

The observations on clinical and biochemical characteristics were summarized by five model phenotypes. The five models do not represent patient clusters; they show a condensed version of the colorings in Figures 2-3 (see also Online Appendix 6). Figure 4A depicts a phenotype with low A1C, high HDL cholesterol and low triglyceride concentration – with very few complications. Accordingly, the population adjusted risk of premature death remains non-significant. In the north, total cholesterol, apolipoprotein B and 24h-urine albumin are higher, along with higher age in Figure 4B. The risk of premature death is 3.8-fold (2.7-5.0) for men and 3.9-fold (2.4-5.7) for women on the map units (2,2) and (1,2).
DKD is the defining factor for a high-risk phenotype that is characterized by high serum creatinine, albuminuria and elevated adiponectin (Fig. 4C). For men, the 10-year mortality peaks in this region (Fig. 2A) but, after adjusting by age, the 9.1-fold (7.3-10.4) risk is no longer the highest on the map. For women, the highest 10.7-fold (7.9-13.7) risk coincides with mortality (Online Appendix 4), but closer to the north-east corner.

The highest 10.1-fold (7.3-13.1) risk of death in male patients is observed at the MetS corner in Figure 4D. By definition, the phenotype is characterized by wide waist, high triglycerides and low HDL₂ cholesterol. Although the prevalence of DKD is high in these patients, they exhibit the highest total cholesterol and apolipoprotein B concentrations, in contrast to the lower values in Figure 4C.

Diabetes duration is the shortest in the south-eastern regions, where mortality is low and relative risk non-significant (Fig. 4E). Nevertheless, the phenotype shares low HDL₂ cholesterol with the north-east corner, which contributes to the observed prevalence of the MetS. The two corners are separated by the triangle of hemoglobin A₁c, total cholesterol and triglycerides, which are elevated in the north.

A number of patients in the high-risk zone (Fig. 4F) have normoalbuminuria despite otherwise adverse biochemical profiles. Aside from 24h-urine albumin and serum creatinine, anti-hypertensive treatment is the most significant discriminating factor, with 23% of normoalbuminuric men (14% of women), but over 90% of those with DKD ($P = 1.9 \times 10^{-12}$) and women have 11 years shorter duration ($P = 2.3 \times 10^{-16}$) than the “metabolic peers” with DKD.

Discussion

The life expectancy for a patient with type 1 diabetes depends heavily on the development of kidney disease and the concurrent incidence of macrovascular events (1,7,21,22). Our analyses with the self-organizing map (SOM) confirmed this, and provided carefully adjusted estimates for the relative risk of premature death. In addition, we also revealed direct associations between the metabolic features and several clinical outcomes.

Urine albumin is biologically highly variable, which has raised questions about its role as an early marker of complications (23). Also, there is continuous discussion on the exact cutoffs that define the clinical DKD categories, which in turn have significant impact on the treatment decisions at an individual level (24). Could albuminuria be supplemented by other sources of information? Much work is being concentrated at finding new biomarkers for kidney disease and some candidates have been found (25-27), although many of them are compared against existing albuminuria and hence do not directly lead to better accuracy in sub-clinical risk assessment.

Genetic susceptibility has also been investigated (28-32), but conclusive results are still lacking, and it is uncertain whether genetic testing will have clinical applications in the near future. Considering the slow development of these complications, the substantial individual and environmental variation, and difficulties in exact phenotyping, it may be too optimistic to
expect any single factor to be decisive in the traditional reductionist fashion (33). Our attention has therefore shifted back to the known biomarkers and risk factors, but with a more comprehensive approach.

Multivariate exploratory analysis has been previously applied to dyslipidemia, the MetS and type 2 diabetes (34-36), but most of these studies were focused on the associations between variables (linear factorization) rather than on the individuals. Here, we described the dataset by five model phenotypes after examining the diversity of the patients' biochemical profiles and combining the results with existing clinical knowledge. The phenotypes do not represent patient clusters since that would imply detectable boundaries between the metabolic states (see Online Appendix 6). Instead, the data reflect different stages of gradual damage in a continuous and time-dependent manner, which prevents clear categorizations in a cross-sectional population-based study.

The SOM has been applied previously in the analysis of spectroscopic data (13,14), in molecular conformation analysis (37), in studies of gene-metabolite interactions (38) and in multivariate assessment of insulin resistance (17), among others. While the sensitivity was demonstrated also in this study, the non-linear algorithm may also lead to the typical problems of multi-variate modeling such as overfitting and false positives (39). Here, the number of samples was large compared to the number of explanatory variables (4,197 patients vs. 14 variables), which already makes the model less prone to exaggerate weak results. Modest sized maps with smoothed estimates were chosen and, to be absolutely sure, only the biochemical variables were used as inputs; this way the statistical significance of the observed variability for the clinical features on the SOM could be verified.

The SOM is an indeterministic method in the sense that a small change in the input data may change the shapes of observed patterns significantly. We therefore recommend testing with different map sizes and subsets of the input data to determine what are the biologically relevant interpretations. The problem is not so much in the robustness of the algorithm, but with the fact that non-biological effects such as sampling bias, differences in data collection methods and laboratory protocols over time may produce data-driven artifacts in clinical studies with a long history.

The highest relative risk of premature death (10.1-fold for men and 10.7-fold for women) was associated with a metabolic profile that shared features from DKD (high 24h-urine albumin and serum creatinine), and from the MetS (high triglycerides, low HDL cholesterol and wide waist). Albuminuria was the common risk factor, as expected, but the two sides were also distinguished by adiponectin, which was higher on the map units dominated by microvascular complications (north-west vs. north-east in Figs. 2 and 3), and C-reactive protein, which was higher in the MetS corner.

Earlier studies already showed that female sex and renal insufficiency are associated with higher adiponectin levels (40) and that DKD is associated with increased C-reactive protein levels (41). Similar results have been obtained with respect to microangiopathy (19,42) and long diabetes duration (43). Unlike C-reactive protein, adiponectin is considered to be inversely related to obesity and dyslipidemia and low levels have been linked to cardiovascular disease (44,45).
considering only the correlation between the MetS and DKD one could conclude that this is not true in type 1 diabetes. The SOM analysis, however, was able to show that the link between adiponectin and the MetS (or abdominal obesity and low HDL cholesterol) is, in fact, present, although concealed by the overall increase of adiponectin due to diabetic complications. Also C-reactive protein seemed to have a more complex role than classical analyses would suggest: the highest levels were not observed for DKD specifically, but for patients with high scores of the MetS.

The incomplete overlap between DKD and MetS was also visible from the lipid-centered perspective. The north-west and north-east quadrants of the colorings in Figures 2 and 3 indicated that the lipid profile of a typical patient with advanced DKD was located between a high HDL phenotype with microvascular complications, and a high triglyceride-low HDL phenotype with the MetS (Fig. 4B-D). Survival bias (ESRD+MetS under-represented) may have been the cause for the reduction in cholesterol and triglyceride concentrations in Figure 4C, since mortality was high. Altered nutrition and other effects due to ESRD are other likely causes, but the data are insufficient for a final conclusion.

The diverse picture of different lipid profiles in various manifestations of microvascular and macrovascular diseases sheds light on the heterogeneous results from previous studies (46-49). For instance, the role of HDL metabolism in the development of DKD has not yet been determined conclusively; this study suggests that low HDL\textsubscript{2} cholesterol is coupled with abdominal obesity and increased insulin requirement, but not necessarily with persistent albuminuria. On the other hand, high HDL\textsubscript{2} cholesterol may represent a marker of good glycemic control and a low prevalence of complications at higher age (Fig. 2 south-west and western regions).

The clinical material was extensive from the cross-sectional phase, but the exact causes of death were not available for the prospective part. Nevertheless, the population registry in Finland is highly accurate and the vitality status itself was reliable. From previous studies it is known that most premature deaths related to type 1 diabetes are caused by cardiovascular events (2,9), which is plausible also in this study: a history of macrovascular disease coincided with the high-mortality regions (Fig. 2A,B).

On one hand, a clear relationship between diabetic kidney disease and mortality exists but, on the other hand, there seems to be a dual nature to the vascular complications, with a quartet of poor glycemic control, adverse lipid profile, abdominal obesity and insulin resistance on one corner (50), and a high-HDL, high adiponectin, leaner phenotype with microvascular complications on the other. When these two collide, the risk of death peaks according to our present and previous results (13). Many of these observations have already been reported, but here all the data was viewed within a single statistical framework that not only described the known risk factors in their metabolic context, but also showed additional details that would have been laborious to detect by classical methods. We therefore expect that the multi-variate approach, as applied here, will enable more rapid and thorough investigations of large clinical datasets, give grounds to novel hypotheses of the pathology of diabetic complications, and open new perspectives into the complex interactions between metabolic risk factors.
ACKNOWLEDGMENTS

For a complete listing of the FinnDiane Study Group, please see Online Appendix 7. The study was supported by grants from the Folkhälsan Research Foundation, the Jenny and Antti Wihuri Foundation, the Wilhelm and Else Stockmann Foundation, the Liv och Hälsa Foundation and the European Commission (QLG2-CT-2001-01669, LSHB-CT-2006-037681). The work was also supported by the Center of Excellence Program of the Academy of Finland (VPM, KK, MAK).

AUTHOR CONTRIBUTIONS

VPM, KK, MAK and PHG conceived and designed the study. VPM programmed the visualization software, analyzed the data and wrote the first draft. VPM, MAK, CF and PHG edited the final manuscript. CF, LT, JW, DG, OH, KH, LK, JK, MRB, MS, NT, MP and PHG collected the clinical and biochemical data.
References


18. NCEP: National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) 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, 2002


33. Loscalzo J, Kohane I, Barabasi A: Human disease classification in the postgenomic era: a
Metabolic Phenotypes in Type 1 Diabetes


**Figure 1.** Multivariate metabolic profiles of patients with type 1 diabetes.

The figure depicts the distribution of males and females on the self-organizing map (SOM) that was constructed based on the listed biochemical variables. The self-organizing neural network algorithm places those patients that have similar biochemical profiles close to each other, and those that have differing profiles far apart on the map. The bar plots illustrate the averaged profile for patients that reside on a given hexagonal region. The closed circles depict sets of 10 men and the open circles sets of 10 women; this was done to avoid excessive clutter from individual markers for each patient.
Figure 2. SOM colorings of clinical features for men with type 1 diabetes.

The map in Figure 1 can be colored according to the characteristics of the local residents within each hexagonal unit. The color scale indicates the deviation from population mean with respect to the random fluctuations that could be expected by chance. The numbers on selected units tell the local prevalence (binary variables) or mean value (continuous variables) for that particular region. For plot A, which illustrates time-adjusted mortality, the random fluctuations could not be estimated using the standard procedure, hence the pseudo-colors are different to avoid direct comparisons with the rest of the colorings. The P-values below the plots indicate the probability of observing equivalent regional variability for random data. *The MetS included variables that were also SOM inputs, hence the P-value is only suggestive. Colorings for women are available in Online Appendix 4.
**Figure 3.** Colorings of biochemical variables on the SOM.

The map colorings were produced with the same procedure as in Figure 2. However, empirical $P$-values are not available for the biochemical variables that were included in the SOM training data. Colorings for women are available in Online Appendix 5.
Figure 4. Metabolic phenotypes and risk of premature death.

The relative risk of death for men and women was estimated against the expected gender-specific mortality in Finland. A-E: Five model phenotypes were constructed based on observations from Figures 1-3. The models do not represent distinct clusters in the dataset, but they summarize the characteristics of patients around the highlighted area in order to make the discussion in relation to Figures 2-3 easier. F: A high-risk region was highlighted for detailed comparisons of DKD categories (results listed in text).