Characterization of metabolic interrelationships and in silico phenotyping of lipoprotein particles using self-organizing maps[s]

Linda S. Kumpula,[1,*,†] **Sanna M. Mäkelä,**[1,†,§] **Ville-Petteri Mäkinen,***[,**,††] **Anna Karjalainen,**[§] **Johanna M. Liinamaa,**[§] **Kimmo Kaski,*** **Markku J. Savolainen,**[†,§] **Minna L. Hannuksela,**[†,§,§§] and **Mika Ala-Korpela**[2,†,§]

Department of Biomedical Engineering and Computational Science,* Helsinki University of Technology, Espoo, Finland; Computational Medicine Research Group,[†] Institute of Clinical Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, University of Oulu, Oulu, Finland; Department of Internal Medicine and Biocenter Oulu,[§] Clinical Research Center, University of Oulu, Oulu, Finland; Folkhälsan Institute of Genetics,** Folkhälsan Research Center, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland; Division of Nephrology, Department of Medicine,[††] Helsinki University Central Hospital, Helsinki, Finland; and Department of Clinical Chemistry,[§§] University of Oulu, Oulu, Finland

Abstract Plasma lipid concentrations cannot properly account for the complex interactions prevailing in lipoprotein (patho)physiology. Sequential ultracentrifugation (UCF) is the gold standard for physical lipoprotein isolations allowing for subsequent analyses of the molecular composition of the particles. Due to labor and cost issues, however, the UCF-based isolations are usually done only for VLDL, LDL, and HDL fractions; sometimes with the addition of intermediate density lipoprotein (IDL) particles and the fractionation of HDL into HDL$_2$ and HDL$_3$ (as done here; n = 302). We demonstrate via these data, with the lipoprotein lipid concentration and composition information combined, that the self-organizing map (SOM) analysis reveals a novel data-driven in silico phenotyping of lipoprotein metabolism beyond the experimentally available classifications. The SOM-based findings are biologically consistent with several well-known metabolic characteristics and also explain some apparent contradictions. The novelty is the inherent emergence of complex lipoprotein associations; e.g., the metabolic subgrouping of the associations between plasma LDL cholesterol concentrations and the structural subtypes of LDL particles. Importantly, lipoprotein concentrations cannot pinpoint lipoprotein phenotypes.[JLR] It would generally be beneficial to computationally enhance the UCF-based lipoprotein data as illustrated here. Particularly, the compositional variations within the lipoprotein particles appear to be a fundamental issue with metabolic and clinical corollaries.—Kumpula, L. S., S. M. Mäkelä, V-P. Mäkinen, A. Karjalainen, J. M. Liinamaa, K. Kaski, M. J. Savolainen, M. L. Hannuksela, and M. Ala-Korpela. **Characterization of metabolic interrelationships and in silico phenotyping of lipo-
protein particles using self-organizing maps. *J. Lipid Res.* **2010.** 51: 431–439.

Lipoprotein metabolism plays a key role in health and disease. Related measures, such as HDL and LDL cholesterol, are in common use to describe individuals' overall metabolic status and the potential risk for atherosclerosis and vascular complications. However, lipoprotein metabolism appears significantly more complex than just an interplay between HDL and LDL. Thus, information on lipoprotein subpopulations is presently needed to appreciate the various counteracting metabolic phenomena and also to more accurately assess the risk for various vascular outcomes (1–3).

It is well established that the liver plays a central role in the apolipoprotein B (apoB) particle cascade, i.e., the endogenous transport system of lipids to various tissues, by producing and secreting VLDL particles into the circulation (4–6). The hydrolysis of these particles by the lipoprotein lipase produces intermediate density lipoprotein (IDL) particles, some of which will further evolve to LDL. The counteracting lipid transport from the peripheral

Abbreviations: apoB, apolipoprotein B; CE, cholesterol ester; FC, free cholesterol; IDL, intermediate density lipoprotein; PL, phospholipid; SOM, self-organizing map; TG, triglyceride; UCF, ultracentrifugation.
[1] These authors contributed equally to this work.
[2] To whom correspondence should be addressed.
  e-mail: mika.ala-korpela@computationalmedicine.fi
[s] The online version of this article (available at http://www.jlr.org) contains supplementary data.

tissues is regulated by HDL that is often divided into two subpopulations, namely larger HDL_2 and smaller HDL_3 particles (7). Furthermore, the HDL metabolism and reverse cholesterol transport are related to the VLDL-IDL-LDL cascade via lipid transfer proteins (8, 9).

Substantial amounts of data exist on alterations in circulating lipoprotein concentrations, for example, in atherosclerosis and in metabolic disorders like diabetes and the metabolic syndrome that are familiar backgrounds for various vascular complications (10–13). The structural integrity and pertinent molecular composition of lipoprotein particles are the basis for the proper functioning of lipoprotein metabolism (14–16). However, the molecular composition as well as the metabolic and structural interrelationships between lipoprotein particles are often hampered due to limited measurement data available and the molecular complexity of lipoprotein metabolism. Physical isolation of lipoproteins by sequential ultracentrifugation (UCF) is common to measure plasma lipoprotein concentrations (17) but most studies are restricted regarding the analyzed subpopulations and detailed attention is rarely paid to the molecular composition of the isolated particles.

Here, we focus on an extensive set of UCF-based lipoprotein data in which the apoB particles were isolated as VLDL, IDL, and LDL fractions and the HDL particles separated to HDL_2 and HDL_3 (18–20). The use of UCF is tedious and expensive and thus, most often only the total HDL fraction is physically isolated and the IDL fraction might not be separately spun. However, the current data set provides an experimental extreme available for clinically oriented lipoprotein studies and was therefore considered optimal for assessing the capabilities of the self-organizing map (SOM) analysis to visualize and interpret lipoprotein metabolism. In fact, the SOM analysis enabled a holistic combination of plasma lipoprotein concentrations and the corresponding compositional features of the particles. Several well-known metabolic issues arose from the SOM analysis of these data per se. As a novel derivative, the analysis resulted in purely data-driven in silico lipoprotein phenotyping allowing detailed characterization of various compositional and metabolic details beyond the experimentally available classifications.

SELF-ORGANIZING MAP ANALYSIS

Background

The SOM is an unsupervised pattern recognition technique (21) that organizes the input data according to given similarity criteria. The end result is a two-dimensional map, where mutually similar input data profiles are placed next to each other and on which all the measures can be easily visualized and compared. The SOM analysis is currently one of the most popular neural network methods already recognized as an effective and advantageous tool to handle complex data in various areas [see, for example, refs (22–26)]. We have also recently implemented and developed the SOM analysis into a metabonomics framework with incorporated p-value statistics (26–28).

This new SOM approach was applied here to analyze and interpret the individual multivariate lipoprotein lipid data.
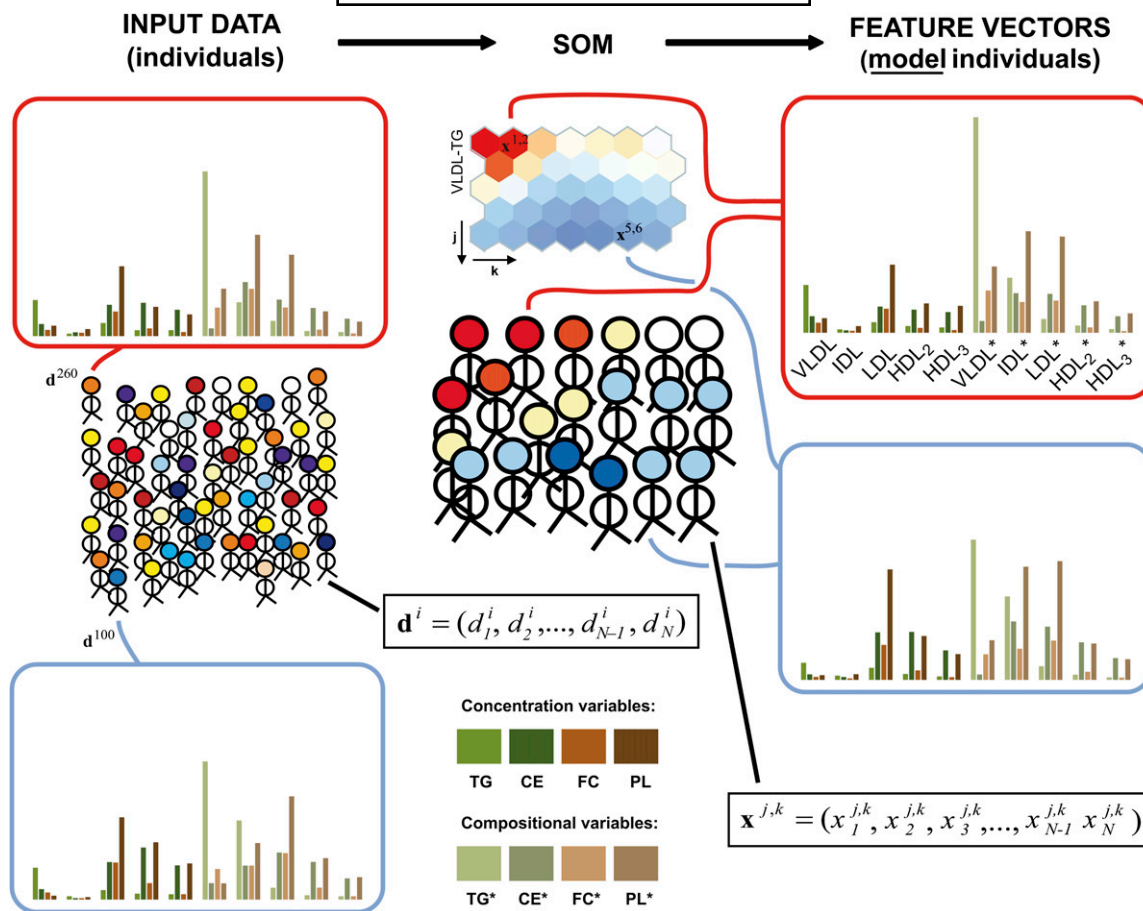
Logic

A characteristic feature of the SOMs is their ability to map nonlinear relations in multidimensional data sets into visually more approachable, typically two-dimensional planes of nodes. The overall concept of SOM analysis is illustrated in **Fig. 1**. The input data to the SOM from each case i, i.e., from each plasma sample in this particular application, contain a number of variables used to form a vector $\mathbf{d}^i = (d_1^i, d_2^i, ... d_{N-1}^i, d_N^i)$. The SOM algorithm (21, 26) then transforms the input data vectors into a two-dimensional map in which each node j,k (j goes over the rows and k over the columns, total of J rows and K columns) will be represented by a single feature vector $\mathbf{x}^{j,k} = (x_1^{j,k}, x_2^{j,k}, ..., x_{N-1}^{j,k}, x_N^{j,k})$ representing the original N dimensional space, i.e., the input data. After the self-organizing process, the point density of the feature vectors follows roughly the probability density of the data, thereby making SOM a valuable tool for detecting similarities and groupings in a data set. The training algorithm is rather simple (and also robust to missing values), and it is easy to visualize the resulting maps. The feature vectors of the neighboring nodes in the two-dimensional map are similar to each other and thereby, importantly, the individuals ending up in nodes close by are similar also in the original N dimensional space (21, 24, 27).

The visualization phase of the SOM analysis is two-fold: first, to look at potential constellations of nodes (feature vectors) formed that would describe similar individuals (groups) in the original variable space; second, to depict input (or other related) variables over the two-dimensional map in order to obtain a quick overview of their distribution and values in different nodes, i.e., in the case of each feature vector. In other words, each node describes a model individual, which, in turn, bares a link to the individuals specified in the original N dimensional space. The SOM algorithm, thus, offers the possibility to generate a form of average representations of model individuals along with identifying both metabolic and compositional characteristics and interrelationships out of multidimensional and complex lipoprotein data. Comparing the component planes of two or more variables in the two-dimensional map may provide insights into the dependencies between the variables and their potential similarities or dissimilarities for the various groups of model individuals. The use of color coding in the component planes is particularly helpful because clearly colored areas as well as correlated changes in the colors of different variables are visually easy to detect. Although it is difficult to exactly define groups in the organized map, subtle changes in colors are also good in indicating potentially diffuse borderline areas between various clusters (27).

Lipoprotein data

The lipoprotein lipid data represent complex metabolic conditions. The SOM analysis of these data revealed groupings of input data parameters that characterize and define

Fig. 1. A methodological overview and illustration of the overall concept of the self-organized map (SOM) analysis. An individual is described by N lipoprotein lipid variables ($\mathbf{d}^i$) as illustrated by the histograms for the two different individuals on the left. Based on the individual inputs, the SOM algorithm results in $j \cdot k$ model individuals $\mathbf{x}^{j,k}$, each having a definite location j, k in the map. All model individuals have the same amount of descriptive variables as the real individuals as illustrated by the histograms on the right. The feature vectors of the neighboring nodes (defining the model individuals and formed through the averaging process of the SOM algorithm) in the two-dimensional map are similar to each other and thereby, importantly, the individuals ending up close by in the map are also similar in the original N dimensional space. The map illustrated in the middle is for the VLDL-TG, one of the 40 variables used as an input for the current SOM analysis, also shown as the first bar in each histogram. The head-colors of the individuals and model individuals conceptually refer to the color-scale used in the component plane for the illustrated lipid variable; bluish colors refer to low and reddish colors to high values. Four lipid concentration measures per lipoprotein fraction were used, namely triglycerides (TG), phospholipids (PL), free cholesterol (FC), and cholesterol esters (CE), together with the corresponding compositional measures (marked with an asterisk) calculated by scaling the concentration measures with the protein amount in each fraction. VLDL, IDL, LDL, and HDL refer to very low, intermediate, low, and high density lipoprotein fractions, respectively.

model individuals according to both plasma lipid concentrations and lipoprotein particle compositions with a link to metabolic pathways. The component planes clarified the clustering of the data set, i.e., the grouping of model individuals into biochemically interpretable areas, for example, in which the concentration of VLDL triglycerides (TGs) is high but that of IDL-TGs is low. It is important to realize that the formation of the models is based solely on the experimental data and the self-organizing process of the SOM algorithm as illustrated in Fig. 1.

MATERIALS AND METHODS

Subjects

Biochemical lipoprotein lipid analyses were available from 233 individuals including 302 distinct lipid measurements (53% fe-

males; 47% males). For some individuals, more than one sample was included from separate blood collections with a typical time interval of 6 months. The study population consisted of heavy alcohol drinkers (40%) (19), hysterectomised postmenopausal women on estrogen replacement therapy (41%) (20), and apparently healthy control individuals (19%) (19), thereby representing a wide range of plasma lipoprotein lipid values. The phenotypic characteristics of the study population are discussed in supplementary information I.

Ethics statement

The study protocol was in accordance with the Declaration of Helsinki and approved by the Ethical Committee of the Northern Ostrobothnia Hospital District, Oulu, Finland, and written informed consent was obtained from all subjects.

Isolation and composition of lipoprotein fractions

The blood samples were drawn after an overnight fast of 12 h into EDTA-containing tubes. Plasma was separated by centrifuga-

tion at 1200–1500 *g* for 10–15 min at 4°C. The lipoprotein fractions were isolated from plasma by sequential UCF using density ranges of <1.006 g/ml for VLDL, 1.006–1.019 g/ml for IDL, 1.019–1.063 g/ml for LDL, 1.063–1.125 g/ml for HDL_2, and 1.125–1.210 g/ml for HDL_3 (18–20). The lipoprotein fractions were isolated from fresh plasma samples and the lipid and protein analyses were commenced immediately after isolation of each fraction. The concentrations of total cholesterol, cholesterol esters (CEs), TGs, phospholipids (PLs), and total protein in the isolated lipoprotein fractions were determined as described previously (18, 19) and expressed as mmol/l in plasma for lipids and mg/dl for proteins (**Table 1**).

SOM analysis

Four lipid concentration measures per lipoprotein fraction were used, namely TGs, PLs, free cholesterol (FC), and CEs, together with the corresponding compositional measures (marked with an asterisk) calculated by scaling the concentration measures with the total protein amount in each fraction (e.g., VLDL-PL* = VLDL-PL/ VLDL-protein). These latter parameters are approximations for the number of lipid molecules (~mol/g) in each lipoprotein particle and thereby provide a measure of the molecular composition of the physically isolated lipoprotein fractions. The concentration and compositional lipid parameters were deliberately used together in the SOM analysis to enable direct association of the concentration and structural information. However, the compositional inputs (e.g., VLDL-PL*) are not intuitive variables to interpret and therefore we have presented mass percentages (marked with two asterisks, e.g., VLDL-PL**) in the SOM component planes to demonstrate the compositional variability. Particle sizes for the VLDL, IDL, LDL, HDL_2, and HDL_3 fractions were estimated as previously described (16). Briefly, the number of lipid molecules in a particular lipoprotein particle was calculated on the basis of the experimental data. The known average volumes of the lipid and protein molecules were then used to calculate the average particle size.

The input data were scaled between −1 and 1 by rank-transformation for preventing unjustified domination of any of the variables and then normalized to smooth the distribution of individuals into the grid. We chose a 5 × 7 map of hexagonal units (resulting in 8.6 samples per unit) and a Gaussian neighborhood function. We also did several runs with different map sizes leading to essentially similar results, as expected, as SOM is known to be rather insensitive to choices of its size and other parameters (23). After the positions of the individuals on the SOM were computed, the map was colored according to the biochemical variables within different parts of the SOM with overall permutation estimations for the p-values for the statistical significance of the patterns (27). The null distributions from the permutation analysis were also the basis of the color scale in each component plane so that variables can be compared visually while maintaining the statistical interpretation. When interpreting the input variables, the p-value estimation is no longer strictly

valid. Consequently, q (instead of p) was used to denote the level of regional variability on the map (28). All the analyses were performed using in-house scripts in the MATLAB programming environment. An open source package, termed Melikerion (26, 27), for SOM analyses in the Matlab/Octave programming environment is freely available. After constituting the SOM, the main regions with differing metabolic features were chosen by visual examination to further analyses. Some individuals residing in the borderline areas were excluded to result in clearer lipoprotein phenotypes. All the analyses were performed on a laptop PC with an Intel Core2 Duo, 2.0 GHz processor, which trained a typical SOM and calculated the colorings in a few minutes.

RESULTS AND DISCUSSION

General aspects of the self-organizing map

Conventionally, correlation analysis is applied to study linear associations between two lipoprotein measures. We have illustrated this for the current data set in supplementary information II, supplementary Fig. II. The correlation patterns imply that at an individual level the lipoprotein particle structure is rather consistent, but within each main lipoprotein fraction substantial compositional variation takes place (a detailed discussion is available in supplementary information II). However, multiparametric and nonlinear relationships, inherent in lipoprotein metabolism, remain intrinsically undetected in simple correlation analyses between two variables. Hence, we will demonstrate how the SOM analysis can be used to extend beyond linear assumptions in a multiparametric manner.

The SOM analysis was performed using a combination of plasma lipoprotein lipid concentrations and compositional lipoprotein particle measures as inputs (for details, see Materials and Methods, SOM analysis). The SOM component planes for plasma concentration measures are illustrated in **Fig. 2** and for compositional particle measures in **Fig. 3**. The very same SOM analysis is the basis for all the component planes shown and thereby each of them can be directly compared, i.e., the distribution of the individuals is the same under every component plane. Now, logically, these component planes should also uncover the key linear associations evident in supplementary information II, supplementary Fig. II. Consequently, within each lipoprotein fraction, strong positive associations between plasma lipid concentrations (supplementary Fig. IIA) are clearly seen from the maps in Fig. 2; for example, the samples with highest concentrations of VLDL lipids are clus-

TABLE 1. The average plasma lipid and protein concentrations and mass percentage compositions of the lipoprotein particles isolated via sequential ultracentrifugation

	TG	C	FC	CE	PL	Protein	TG	C	FC	CE	PL	Protein
	mmol/l	mmol/l	mmol/l	mmol/l	mmol/l	mg/dl	Mass-%	Mass-%	Mass-%	Mass-%	Mass-%	Mass-%
VLDL (n≥294)	1.01 ± 1.58	0.50 ± 0.61	0.22 ± 0.32	0.28 ± 0.31	0.32 ± 0.43	19.7 ± 20.6	53 ± 6	17 ± 7	5 ± 2	12 ± 5	16 ± 4	14 ± 4
IDL (n≥256)	0.11 ± 0.07	0.25 ± 0.23	0.06 ± 0.07	0.19 ± 0.18	0.09 ± 0.08	6.45 ± 5.03	27 ± 11	37 ± 13	6 ± 3	31 ± 10	18 ± 6	18 ± 6
LDL (n≥300)	0.31 ± 0.14	3.39 ± 1.20	0.85 ± 0.33	2.54 ± 0.93	1.06 ± 0.42	89.0 ± 32.3	7 ± 3	49 ± 8	8 ± 2	41 ± 6	21 ± 4	23 ± 4
HDL_2 (n≥297)	0.17 ± 0.09	1.25 ± 0.60	0.23 ± 0.16	1.02 ± 0.47	1.00 ± 0.50	114.4 ± 44.3	6 ± 3	26 ± 5	3 ± 1	23 ± 4	27 ± 5	41 ± 6
HDL_3 (n≥278)	0.12 ± 0.08	0.75 ± 0.26	0.08 ± 0.06	0.68 ± 0.25	0.65 ± 0.19	137.0 ± 45.5	4 ± 3	19 ± 5	1 ± 1	18 ± 4	21 ± 4	56 ± 6

The values are given as mean ± SD. TG, triglycerides; C, cholesterol; FC, free cholesterol molecules; CE, cholesterol esters; PL, phospholipids; Protein, total proteins. The details for the lipoprotein isolation and lipid analysis are given in the Materials and Methods section.
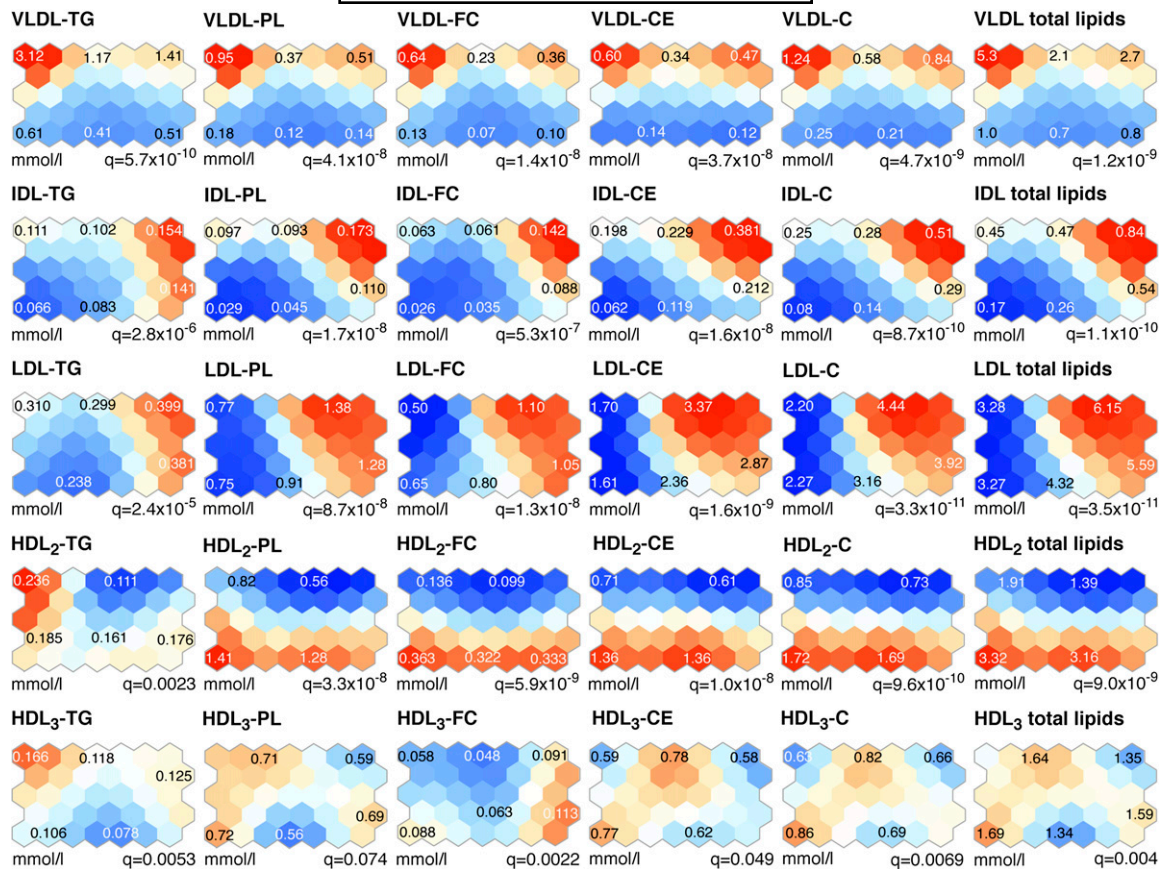
Fig. 2. Statistical colorings of the lipoprotein lipid concentration measures (in mmol/l) in the SOM analysis of the combined concentration and compositional lipoprotein variables. The coloring is according to the characteristics of the local residents within each hexagonal unit. The concentration levels are color-coded to visualize whether the concentration level is above (reddish), at (white) or below (bluish) the median of the variable. The numbers on selected units tell the local mean value for that particular region. The q-values below the plots indicate the probability of observing equivalent regional variability for random data (see Materials and Methods). Importantly, the very same SOM analysis is the basis for all the component planes shown (holds also for Fig. 3) and thereby each of them can be directly compared; i.e., the distribution of the individuals is the same under every component plane. The abbreviations are as given in the caption for Fig. 1.

tered in the northwest corner and those with high LDL lipid concentrations in the northeast and eastern areas of the map. The negative associations between VLDL and HDL_2 lipid concentrations are also clearly seen in Fig. 2 via the opposite colorings for the VLDL and HDL_2 component planes for PLs, FC, CEs, cholesterol, and total lipids.

Toward complex associations: LDL and HDL_2

In addition to the linear relationships (supplementary Fig. II), the SOM component planes shown in Figs. 2 and 3 are revealing further associations between the lipoprotein measures. For example, the SOM component planes in Fig. 2 give an explanation why the plasma concentrations of LDL lipids barely correlate with those of HDL_2 (supplementary Fig. IIA). The associations between LDL and HDL_2 lipid concentrations are complex in such a way that different metabolic models arose from the data in the SOM analysis. Each of these models correspond to various combinations of lipoprotein concentrations; those reflecting positive associations between LDL and HDL_2 are separated in the southeast and northwest areas of the SOM and those representing negative associations in the north-northeast and south-southwest areas of the SOM. These associations are pro-

nounced for the PLs, FC, CEs, cholesterol, and total lipids but somewhat different between the LDL-TG and HDL_2-TG. This is most likely an indication of a nonstructural role of TG molecules in these lipoprotein particles.

Thus, the SOM analysis enables an unsupervised discovery of multiple (nonlinear) associations; i.e., in composite data, a nonexistent linear correlation does not necessarily mean that the two measures would have no association. As noted above for the LDL and HDL_2 lipid concentrations, various metabolic pathways may exist that have distinct but different associations that, in the linear analysis, mix in such a manner that no clear common correlation is found.

High concentration of plasma HDL_2 is associated with two different lipoprotein phenotypes

The high plasma concentrations of HDL_2 lipids (the southern region of the SOM in Fig. 2) consistently associate with the low concentrations of plasma VLDL lipids (29) but differentiate into two subgroups with respect to LDL lipid concentrations. In the southeast region of the SOM, the high HDL_2 lipid concentrations relate to relatively high concentrations of plasma LDL lipids and large IDL, LDL, HDL_2, and HDL_3 particles together with small VLDL (Figs.
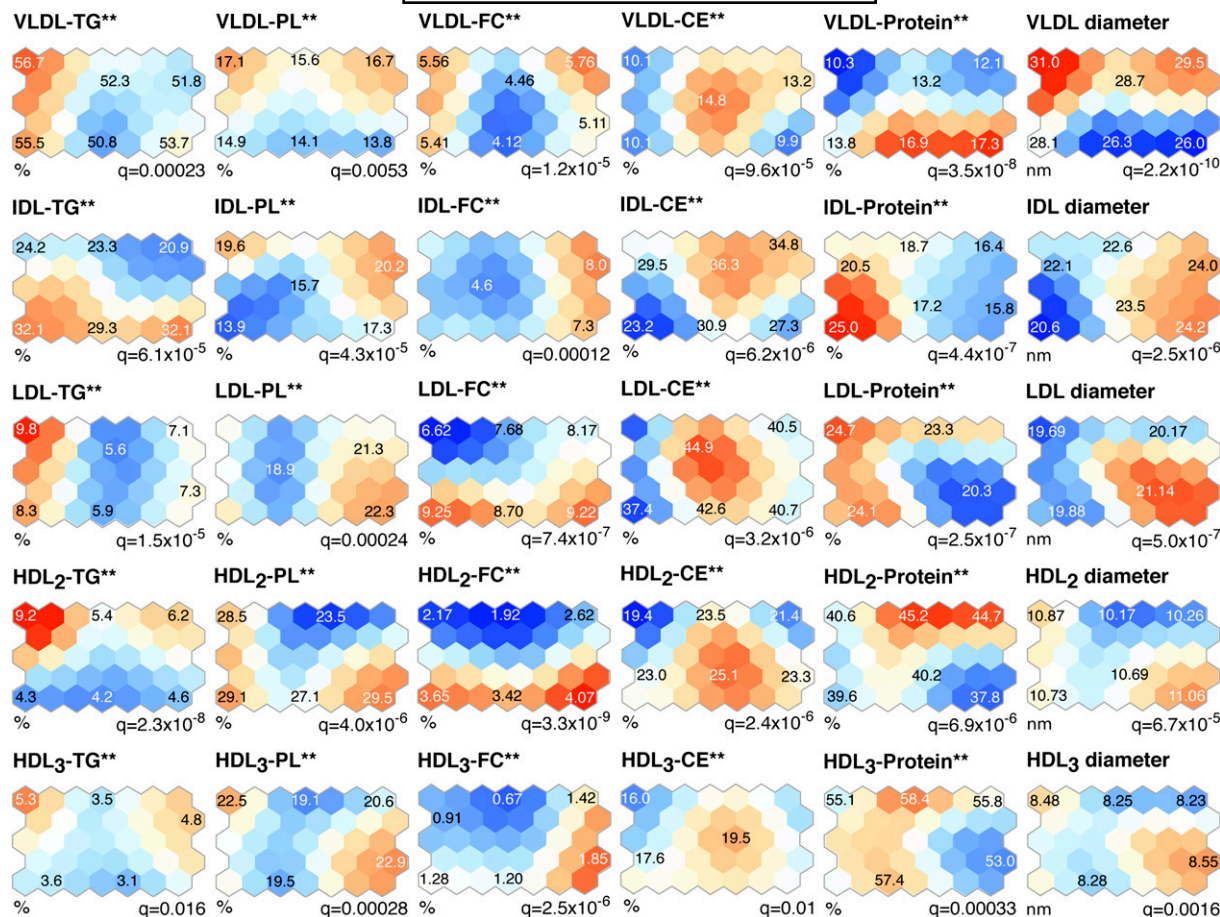
Fig. 3. Statistical colorings of the compositional lipoprotein measures in the SOM analysis of the combined concentration and compositional lipoprotein variables. For each lipoprotein fraction, the values are represented as mass percentages (**) for the lipids and protein or nm for the particle sizes. All other details are as given in the caption for Fig. 2. The abbreviations are as given in the caption for Fig. 1.

2, 3). The LDL, HDL_2, and HDL_3 particles are enriched in PLs and FC, whereas the VLDL and IDL particles are relatively PL-poor. In contrast, in the southwest region of the SOM, the high plasma HDL_2 lipid concentrations are associated with low concentrations of plasma lipids in all apoB-containing lipoprotein particles, i.e., VLDL, IDL, and LDL. All these apoB-particles are also relatively small. In general, it is evident from Fig. 3 that there are clear associations between the lipoprotein particle size and the PL as well as the protein content of the particles, higher amounts of PL and lower amounts of protein indicating larger particles.

Plasma LDL-C concentration and the structural subtypes of LDL particles

Quite contradictory results have been published regarding the association of plasma LDL-C concentrations with the composition and characteristics of LDL particles. It has been reported that small dense and large LDL particle distributions do not differ in plasma LDL-C concentration (30, 31). On the other hand, large LDL particles have been connected to higher LDL-C concentrations than the predominance of small LDL particles (32). The binding of LDL particles to the LDL receptor has been shown to be reduced with dense as well as large LDL compared with LDL with intermediate particle sizes (33), indicating that high LDL-C concentration in plasma could be associated with both small and large LDL particles. Generally, the amounts of LDL particles seem not to be able to account for the LDL-C concentration in plasma, because with the similar LDL-C concentrations, the LDL-particle numbers can differ (34, 35).

Our current findings via the SOM-based lipoprotein phenotyping indicate that the large LDL particles are associated with relatively high LDL-C (the southeast region of the SOM in Figs. 2 and 3) and the small dense LDL particles are mostly related to low plasma LDL-C concentrations (the western half of the SOM in Figs. 2 and 3). However, this association is not inclusive and there is a metabolic pathway in which the small dense LDL particles are related to high plasma LDL-C (the northeast region of the SOM in Figs. 2 and 3). The small LDL particles related to the high plasma LDL-C concentration seem to be rather TG-poor whereas the small LDL associated with low plasma LDL-C is enriched in TGs. Thus, the apparent contradictions noted above are most likely only reflections of different characteristics in the study populations. In fact, these associations are a good example of how simple (linear) correlation analysis is not able to reveal differently associated subgroups (nonlinearities) in the data. It is also notable that high concentration of plasma LDL-C associated
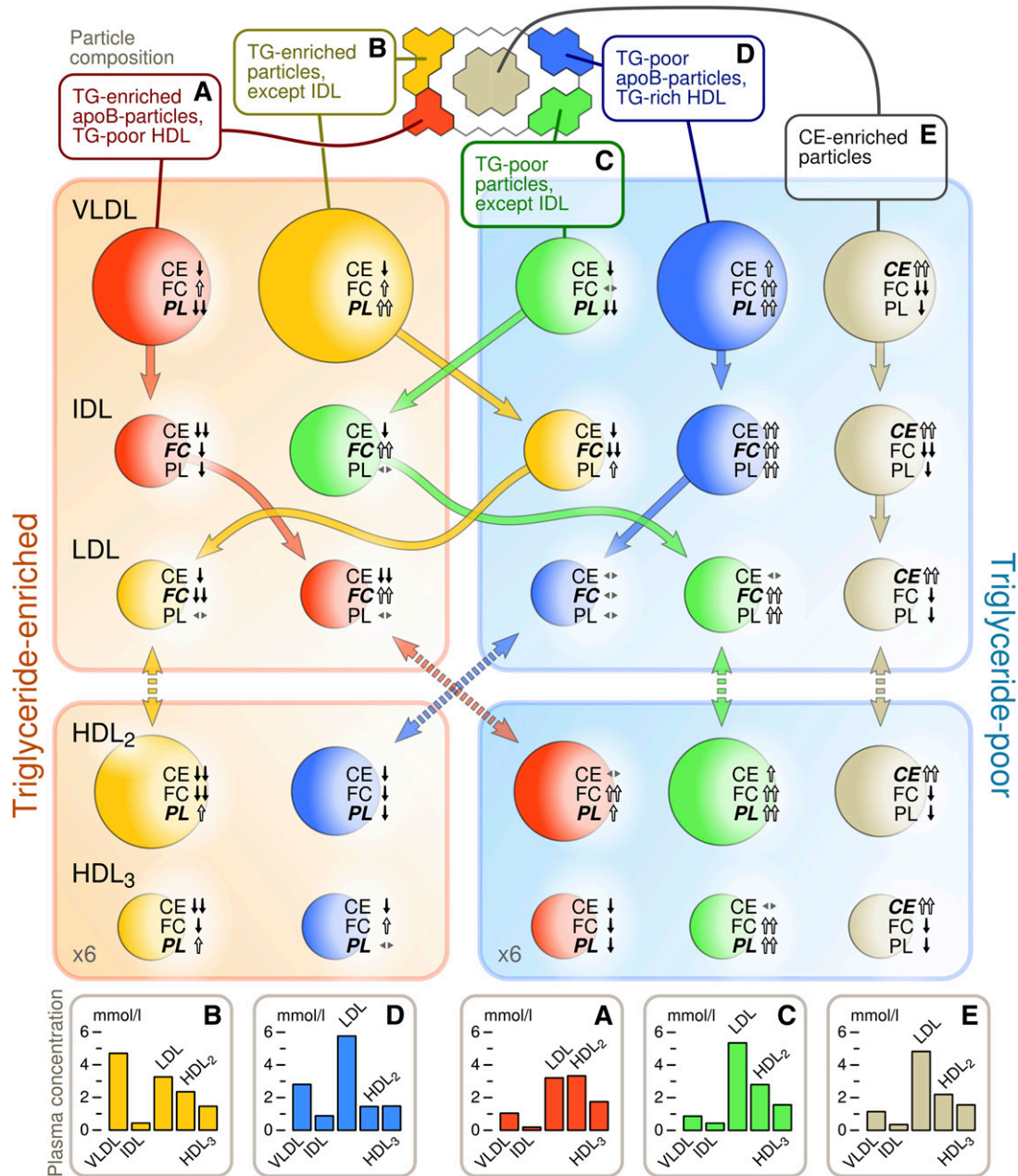
Fig. 4. A metabolic overview of the lipoprotein phenotypes arisen from the SOM analysis (illustrated in Figs. 2 and 3). The application of the SOM analysis to the combination of concentration and compositional lipoprotein data resulted in a novel perspective and also provided a subgrouping of the lipoprotein particles in each fraction, i.e., an in silico lipoprotein phenotyping beyond the experimental data. The term 'lipoprotein phenotype' is used here to denote a collection of lipoprotein subtypes for VLDL, IDL, LDL, HDL$_2$, and HDL$_3$ related to a particular plasma lipoprotein concentration profile and forming a metabolically connected entity. Five different phenotypes were discovered (A–E as marked and color-coded on the SOM), all with characteristic plasma concentration profiles (indicated in the bottom) as well as distinct compositional features (summarized on the top). The scale in the concentration profiles indicates the total plasma lipid concentrations of the lipoprotein fractions in mmol/l. The apoB-containing VLDL-IDL-LDL cascade is the principal route in the endogenous lipoprotein metabolism and relates primarily to the transport and hydrolysis of TG. Thus, the metabolic pathways of the lipoprotein phenotypes are organized here in two platforms, one for TG-enriched (on the left, with an orange background) and one for TG-poor particles (on the right, with a bluish background). The HDL particles are also accordingly divided into two groups with respect to their relative TG content. The solid color-coded arrows represent the metabolic pathways of the apoB-containing lipoprotein particles within each lipoprotein phenotype. The connections between apoB and HDL particles are indicated by the bidirectional dashed arrows. The sizes of apoB-containing particles as well as HDL particles are in scale although the sizes of the HDL particles are enlarged by a factor of 6. The relative contents of the various lipids in the lipoprotein particles are indicated by the up- and down-ward arrows. Structurally characteristic lipids in each particle are bolded. The abbreviations are as given in the caption for Fig. 1.

with high amount of small LDL particles is particularly risky with respect to cardiovascular disease (36).

Metabolic pathways of the lipoprotein phenotypes

The application of the SOM analysis enabled us to concomitantly assess various metabolic and compositional interrelationships between the experimentally isolated and characterized lipoprotein fractions. In fact, this appeared to be a novel viewpoint that also provided a detailed subgrouping of the lipoprotein particles within each fraction, i.e., a detailed in silico lipoprotein phenotyping beyond the experimental data. The term 'lipoprotein phenotype' is used here to denote a collection of lipoprotein subtypes for VLDL, IDL, LDL, HDL_2, and HDL_3 related to a particular plasma lipoprotein concentration profile and forming a metabolically connected entity. An overview of the lipoprotein phenotypes arisen from the SOM analysis is given in **Fig. 4**. The apoB-containing lipoprotein cascade, i.e., VLDL-IDL-LDL, is a key route in the endogenous lipoprotein metabolism and relates principally to the transport and hydrolysis of TG (6). Thus, in Fig. 4, the metabolic pathways of the lipoprotein phenotypes are organized in two platforms, one for TG-enriched and one for TG-poor particles (reflecting the SOM analysis shown in Fig. 3). The HDL particles are also accordingly divided into two groups with respect to their relative TG content. Five different lipoprotein phenotypes were discovered, all with characteristic plasma lipoprotein lipid concentration profiles, as well as distinct compositional and metabolic features. Some key findings will be highlighted and discussed below.

A lipoprotein phenotype reflecting characteristics of the metabolic syndrome

Figure 4 depicts lipoprotein phenotype B in which all the lipoprotein particles are enriched in TG except IDL being fairly TG-poor (see also the northwest region of the SOM in Fig. 3). The VLDL particles in this phenotype are large and also enriched in PL and FC. However, the IDL and LDL particles are quite small. Interestingly, although VLDL particles are FC-enriched, the corresponding IDL, LDL, HDL_2, and HDL_3 particles are FC-poor. The plasma concentration of VLDL total lipids is high but those of LDL and HDL_2 are low; the concentration of IDL total lipids is also somewhat elevated (see Fig. 4 and the northwest region of the SOM in Fig. 2). Consequently, the characteristics of lipoprotein phenotype B resemble those inherent for the metabolic syndrome (10).

These findings are also in line with studies showing that the delipidation of large VLDL can produce low levels of LDL (37) and that large VLDL is related to small dense LDL (32, 38, 39). In addition, the small dense LDL phenotype has been linked to increased production and decreased catabolism of VLDL particles (40). This is consistent with our findings here with respect to lipoprotein phenotype B in which the plasma concentration of VLDL is high and VLDL particles are large and TG-enriched, whereas the LDL lipid concentrations are low with the preponderance of small, TG-enriched LDL particles. It is also notable that the low percentage of FC in the IDL, LDL, and HDL parti-

cles of phenotype B may be a structural issue to enhance the oxidative susceptibility of these lipoproteins (41, 42).

Plasma lipoprotein concentrations do not predict lipoprotein phenotypes

Very similar plasma concentrations of VLDL and IDL in phenotypes A and E relate to significantly different composition of the VLDL as well as IDL particles in these phenotypes; the VLDL and IDL particles in phenotype A are TG-enriched and CE-poor, the situation being the opposite in the case of phenotype E. Also, high plasma LDL concentration, which is characteristic for phenotypes C, D, and E, relates to remarkable variations in the composition of the LDL particles between the phenotypes and, notably, even more profound differences in the composition as well as the size of the VLDL and IDL particles.

Lipoprotein metabolism is a complex crosstalk of various lipoprotein particles as well as enzymes and lipid transfer proteins. For example, during lipolysis of apoB-containing particles the phospholipid transfer protein increases the particle distribution of HDL toward HDL_2 subclasses (9, 43). On the other hand, the cholesteryl ester transfer protein mediates heteroexchange of CE and TG between HDL and apoB-containing lipoprotein particles, the transfer of TG being toward HDL and that of CE toward apoB-containing particles (8, 16). Therefore, it is not unexpected that the plasma lipoprotein concentrations alone can only give a limited view on the overall lipoprotein metabolism.

Rationale for the in silico lipoprotein phenotyping

Even though detailed data on lipoprotein particles would currently be preferred in cardiovascular research, the subpopulation analysis is usually based on particle size (e.g., using gradient gel electrophoresis or nuclear magnetic resonance spectroscopy) and therefore, the chemical composition of the particles remains unknown (7, 36). In analytical lipid biochemistry, sequential ultracentrifugation is the gold standard for physical lipoprotein isolation allowing for subsequent analyses of the molecular composition of the particles (18–20). However, the UCF-based lipoprotein work is most often restricted regarding the analyzed subfractions. This is because the finer the density ranges used for the isolation, the more tedious and expensive the analyses become (44). Consequently, it would generally be beneficial to computationally enhance the UCF-based lipoprotein data as illustrated in this work. In particular, deeper insight into the compositional variations in the lipoprotein particles appears a fundamental issue. The lipoprotein physiology and pathophysiology is about transfer and exchange of various lipid molecules between the lipoprotein particles and tissues. Thus, not only the concentration but also the quality of lipoprotein particles and the form of transportation does matter.**JLR**

REFERENCES

1. Stein, J. H., and P. E. McBride. 2006. Should advanced lipoprotein testing be used in clinical practice? *Nat. Clin. Pract. Cardiovasc. Med.* **3:** 640–641.
2. Arsenault, B. J., I. Lemieux, J. P. Després, N. J. Wareham, R. Luben, J. J. Kastelein, K. T. Khaw, and S. M. Boekholdt. 2007. Cholesterol

levels in small LDL particles predict the risk of coronary heart disease in the EPIC-Norfolk prospective population study. *Eur. Heart J.* **28:** 2770–2777.

3. Vekic, J., A. Topic, A. Zeljkovic, Z. Jelic-Ivanovic, and V. Spasojevic-Kalimanovska. 2007. LDL and HDL subclasses and their relationship with Framingham risk score in middle-aged Serbian population. *Clin. Biochem.* **40:** 310–316.

4. Adiels, M., S. O. Olofsson, M-R. Taskinen, and J. Boren. 2008. Overproduction of very low-density lipoproteins is the hallmark of the dyslipidemia in the metabolic syndrome. *Arterioscler. Thromb. Vasc. Biol.* **28:** 1225–1236.

5. Parhofer, K. G., and P. H. R. Barrett. 2006. What we have learned about VLDL and LDL metabolism from human kinetics studies. *J. Lipid Res.* **47:** 1620–1630.

6. Berneis, K. K., and R. M. Krauss. 2002. Metabolic origins and clinical significance of LDL heterogeneity. *J. Lipid Res.* **43:** 1363–1379.

7. Warnick, G. R., J. R. McNamara, C. N. Boggess, F. Clendenen, P. T. Williams, and C. C. Landolt. 2006. Polyacrylamide gradient gel electrophoresis of lipoprotein subclasses. *Clin. Lab. Med.* **26:** 803–846.

8. Qiu, X., A. Mistry, M. J. Ammirati, B. A. Chrunyk, R. W. Clark, Y. Cong, J. S. Culp, D. E. Danley, T. B. Freeman, K. F. Geoghegan, et al. 2007. Crystal structure of cholesteryl ester transfer protein reveals a long tunnel and four bound lipid molecules. *Nat. Struct. Mol. Biol.* **14:** 106–113.

9. Patsch, J. R., A. M. Gotto, Jr., T. Olivercrona, and S. Eisenberg. 1978. Formation of high density lipoprotein2-like particles during lipolysis of very low density lipoproteins in vitro. *Proc. Natl. Acad. Sci. USA.* **75:** 4519–4523.

10. Chapman, M. J. 2007. Metabolic syndrome and type 2 diabetes: lipid and physiological consequences. *Diab. Vasc. Dis. Res.* **4(Suppl 3):** S5–S8.

11. Sviridov, D., N. Mukhamedova, A. T. Remaley, J. Chin-Dusting, and P. Nestel. 2008. Antiatherogenic functionality of high density lipoprotein: how much versus how good. *J. Atheroscler. Thromb.* **15:** 52–62.

12. de Souza, J. A., C. Vindis, B. Hansel, A. Nègre-Salvayre, P. Therond, C. V. Serrano, Jr., S. Chantepie, R. Salvayre, E. Bruckert, M. J. Chapman, et al. 2008. Metabolic syndrome features small, apolipoprotein A-I-poor, triglyceride-rich HDL3 particles with defective anti-apoptotic activity. *Atherosclerosis.* **197:** 84–94.

13. Xiao, C., T. Watanabe, Y. Zhang, B. Trigatti, L. Szeto, P. W. Connelly, S. Marcovina, T. Vaisar, J. W. Heinecke, and G. F. Lewis. 2008. Enhanced cellular uptake of remnant high-density lipoprotein particles. A mechanism for high-density lipoprotein lowering in insulin resistance and hypertriglyceridemia. *Circ. Res.* **103:** 159–166.

14. Hevonoja, T., M. O. Pentikäinen, M. T. Hyvönen, P. T. Kovanen, and M. Ala-Korpela. 2000. Structure of low density lipoprotein (LDL) particles: basis for understanding molecular changes in modified LDL. *Biochim. Biophys. Acta.* **1488:** 189–210.

15. Mouritsen, O. G. 2005. Life - As a Matter of Fat. The Emerging Science of Lipidomics. Springer-Verlag, Berlin Heidelberg.

16. Kumpula, L. S., J. M. Kumpula, M-R. Taskinen, M. Jauhiainen, K. Kaski, and M. Ala-Korpela. 2008. Reconsideration of hydrophobic lipid distributions in lipoprotein particles. *Chem. Phys. Lipids.* **155:** 57–62.

17. Kahlon, T. S., L. A. Glines, and F. T. Lindgren. 1986. Analytic ultracentrifugation of plasma lipoproteins. *Methods Enzymol.* **129:** 26–45.

18. Liinamaa, M. J., M. L. Hannuksela, Y. A. Kesäniemi, and M. J. Savolainen. 1997. Altered transfer of cholesteryl esters and phospholipids in plasma from alcohol abusers. *Arterioscler. Thromb. Vasc. Biol.* **17:** 2940–2947.

19. Liinamaa, M. J., K. Kervinen, M. L. Hannuksela, Y. A. Kesäniemi, and M. J. Savolainen. 1997. Effect of apolipoprotein E phenotype on plasma lipids and lipoproteins in alcohol abusers. *Alcohol. Clin. Exp. Res.* **21:** 606–612.

20. Karjalainen, A., J. Heikkinen, M. J. Savolainen, A. C. Bäckström, and Y. A. Kesäniemi. 2000. Mechanisms regulating LDL metabolism in subjects on peroral and transdermal estrogen replacement therapy. *Arterioscler. Thromb. Vasc. Biol.* **20:** 1101–1106.

21. Kohonen, T. 2001. Self-Organizing Maps. Springer. Berlin.

22. Valkonen, V. P., M. Kolehmainen, H. M. Lakka, and J. T. Salonen. 2002. Insulin resistance syndrome revisited: application of self-organizing maps. *Int. J. Epidemiol.* **31:** 864–871.

23. Nikkilä, J., P. Törönen, S. Kaski, J. Venna, E. Castrén, and G. Wong. 2002. Analysis and visualization of gene expression data using self-organizing maps. *Neural Netw.* **15:** 953–966.

24. Suna, T., M. Hardey, J. Huhtinen, Y. Hiltunen, K. Kaski, J. Heikkonen, and M. Ala-Korpela. 2006. Self-organising map approach to individual profiles: age, sex and culture in internet dating. *Sociol. Res. Online.* 11: 31.03.2006.

25. Hyvönen, M. T., Y. Hiltunen, W. El-Deredy, T. Ojala, J. Vaara, P. T. Kovanen, and M. Ala-Korpela. 2001. Application of self-organizing maps in conformational analysis of lipids. *J. Am. Chem. Soc.* **123:** 810–816.

26. Mäkinen, V-P., P. Soininen, C. Forsblom, M. Parkkonen, P. Ingman, K. Kaski, P-H. Groop, and M. Ala-Korpela. 2008. ^{1}H NMR metabonomics approach to the disease continuum of diabetic complications and premature death. *Mol. Syst. Biol.* **4:** 167.

27. Mäkinen, V-P., C. Forsblom, L. M. Thorn, J. Wadén, D. Gordin, O. Heikkilä, K. Hietala, L. Kyllönen, J. Kytö, M. Rosengård-Bärlund, et al. 2008. Metabolic phenotypes, vascular complications and premature deaths in a population of 4,197 patients with type 1 diabetes. *Diabetes.* **57:** 2480–2487.

28. Tukiainen, T., T. Tynkkynen, V-P. Mäkinen, P. Jylänki, A. Kangas, J. Hokkanen, A. Vehtari, O. Gröhn, M. Hallikainen, H. Soininen, et al. 2008. A multi-metabolite analysis of serum by ^{1}H NMR spectroscopy: early systemic signs of Alzheimer's disease. *Biochem. Biophys. Res. Commun.* **375:** 356–361.

29. Yang, Y., B. Yan, M. Fu, Y. Xu, and Y. Tian. 2005. Relationship between plasma lipid concentrations and HDL subclasses. *Clin. Chim. Acta.* **354:** 49–58.

30. Katzel, L. I., R. M. Krauss, and A. P. Goldberg. 1994. Relations of plasma TG and HDL-C concentrations to body composition and plasma insulin levels are altered in men with small LDL particles. *Arterioscler. Thromb.* **14:** 1121–1128.

31. Kang, H. S., B. Gutin, P. Barbeau, M. S. Litaker, J. Allison, and N. A. Le. 2002. Low-density lipoprotein particle size, central obesity, cardiovascular fitness, and insulin resistance syndrome markers in obese youths. *Int. J. Obes. Relat. Metab. Disord.* **26:** 1030–1035.

32. Georgieva, A. M., M. M. van Greevenbroek, R. M. Krauss, M. C. Brouwers, V. M. Vermeulen, M. G. Robertus-Teunissen, C. J. van der Kallen, and T. W. de Bruin. 2004. Subclasses of low-density lipoprotein and very low-density lipoprotein in familial combined hyperlipidemia: relationship to multiple lipoprotein phenotype. *Arterioscler. Thromb. Vasc. Biol.* **24:** 744–749.

33. Nigon, F., P. Lesnik, M. Rouis, and M. J. Chapman. 1991. Discrete subspecies of human low density lipoproteins are heterogeneous in their interaction with the cellular LDL receptor. *J. Lipid Res.* **32:** 1741–1753.

34. Superko, H. R., and R. R. Gadesam. 2008. Is it LDL particle size or number that correlates with risk for cardiovascular disease? *Curr. Atheroscler. Rep.* **10:** 377–385.

35. Jeyarajah, E. J., W. C. Cromwell, and J. D. Otvos. 2006. Lipoprotein particle analysis by nuclear magnetic resonance spectroscopy. *Clin. Lab. Med.* **26:** 847–870.

36. Cromwell, W. C., and J. D. Otvos. 2004. Low-density lipoprotein particle number and risk for cardiovascular disease. *Curr. Atheroscler. Rep.* **6:** 381–387.

37. Packard, C. J., A. Munro, A. R. Lorimer, A. M. Gotto, and J. Shepherd. 1984. Metabolism of apolipoprotein B in large triglyceride-rich very low density lipoproteins of normal and hypertriglyceridemic subjects. *J. Clin. Invest.* **74:** 2178–2192.

38. Magkos, F., B. S. Mohammed, and B. Mittendorfer. 2008. Effect of obesity on the plasma lipoprotein subclass profile in normoglycemic and normolipidemic men and women. *Int. J. Obes. (Lond.).* **32:** 1655–1664.

39. Packard, C. J. 2003. Triacylglycerol-rich lipoproteins and the generation of small, dense low-density lipoprotein. *Biochem. Soc. Trans.* **31:** 1066–1069.

40. Verges, B. 2005. New insight into the pathophysiology of lipid abnormalities in type 2 diabetes. *Diabetes Metab.* **31:** 429–439.

41. Tribble, D. L., L. G. Holl, P. D. Wood, and R. M. Krauss. 1992. Variations in oxidative susceptibility among six low density lipoprotein subfractions of differing density and particle size. *Atherosclerosis.* **93:** 189–199.

42. Holvoet, P. 2008. Relations between metabolic syndrome, oxidative stress and inflammation and cardiovascular disease. *Verh. K. Acad. Geneeskd. Belg.* **70:** 193–219.

43. Jiang, X. C., C. Bruce, J. Mar, M. Lin, Y. Ji, O. L. Francone, and A. R. Tall. 1999. Targeted mutation of plasma phospholipid transfer protein gene markedly reduces high-density lipoprotein levels. *J. Clin. Invest.* **103:** 907–914.

44. Ala-Korpela, M., N. Lankinen, A. Salminen, T. Suna, P. Soininen, R. Laatikainen, P. Ingman, M. Jauhiainen, M-R. Taskinen, K. Héberger, et al. 2007. The inherent accuracy of ^{1}H NMR spectroscopy to quantify plasma lipoproteins is subclass dependent. *Atherosclerosis.* **190:** 352–358.