Characterization of systemic metabolic phenotypes associated with subclinical atherosclerosis

Peter Würtz, Pasi Soininen, Antti J. Kangas, Ville-Petteri Mäkinen, Per-Henrik Groop, Markku J. Savolainen, Markus Juonala, Jorma S. Viikari, Mika Kähönen, Terho Lehtimäki, Olli T. Raitakari, and Mika Ala-Korpela

Supplementary text, tables and figures:

Supplementary Discussion	
Supplementary Table 1	Metabolic phenotype characteristics in terms of lipoprotein subclasses, low-molecular-weight metabolites, and selected clinical and biochemical variables
Supplementary Table 2	Characteristics of the study population
Supplementary Figure 1	Metabolic phenotype characteristics in terms of lipoprotein subclasses, low-molecular-weight metabolites, and selected clinical and biochemical variables separately for men and women
Supplementary Figure 2	Self-organizing map analysis of the 2007 data subset (*n* = 2,123)

Supplementary Discussion

Characteristics of the metabolic phenotypes in terms of lipoprotein subclasses, low-molecular-weight metabolites, and selected clinical and biochemical variables are given in Supplementary Table 1. The lipoprotein subclasses were quantified in terms of total lipid concentrations [mmol/l]. Therefore the sums of the subclass measures do not directly correspond to the low-density or high-density lipoprotein cholesterol concentrations (LDL-C or HDL-C, respectively) given in Supplementary Table 2. The NMR-determined LDL-C does not contain the intermediate-density-lipoprotein cholesterol fraction which is part of the Friedewald estimated LDL-C,[1] and therefore the numerical values of the LDL-C shown in Supplementary Table 2 may appear low. The significance limit after Bonferroni-adjusting for 204 independent tests is $p < 0.00025$. Most of the *p*-values from the Kolmogorov-Smirnov tests were smaller than this limit and the number of tests to adjust for by the Holm-Bonferroni adjustment[2] was 20 independent tests making the *p*-value considered significant $p < 0.0025$. The metabolites with *non-significant p*-values according to this significance level are marked by asterisks in Supplementary Table 1. The significance correction applied here may still be considered conservative given the high intertrait correlation between lipoprotein subclasses.[3]

It is evident from Supplementary Table 1 that the features of the metabolic phenotypes illustrated in Fig. 3 are statistically significant characteristics; almost all of the metabolites serve as phenotype signatures. This holds for all of the lipoprotein subclasses and in general also for the low-molecular-weight metabolites with the exceptions of histidine, acetoacetate, and 3-hydroxybutyrate. The clinical and biochemical variables corroborate the view that the metabolic phenotypes are distinct. Nevertheless, the lacking statistical significance for many of the clinical and biochemical variables for phenotype D indicates that these measures are insufficient in their own right to indicate how this phenotype would be associated with cardiovascular risk. However, the results of this study indicate that this metabolic phenotype does not have significantly higher relative risk for subclinical atherosclerosis than that of the reference phenotype E. The detailed information obtained from the lipoprotein subclass and low-molecular-weight metabolite profiles for phenotype D indicates that, despite the elevated concentration of intermediate-density lipoprotein, large LDL and glycoproteins, this phenotype is not associated with subclinical atherosclerosis.

The characteristics of the study population of the ongoing Cardiovascular Risk in Young Finns Study are given in Supplementary Table 2 in terms of clinical and biochemical variables. This multi-centre study was designed to study the risk factors and precursors of cardiovascular diseases and their determinants in children and young adults. The baseline population in 1980 was selected to be representative of the Finnish population with individuals from both urban and rural areas.[4] Differences between the baseline population and the study population at follow-ups in 2001 and 2007 have been discussed earlier.[5] Individuals lost to follow-up were younger in both sexes than participants. Baseline body-mass index was higher in female non-participants than in participants. No statistically significant differences were seen in other risk factors.

Sex-stratified analysis of the phenotype characteristics are illustrated in Supplementary Fig. 1. The histograms shown for men and women were calculated on the basis of the SOM division depicted in Fig. 3, *i.e*., based on the SOM trained using the NMR spectral data from both sexes. In the case of lipoprotein subclass profiles the sex-stratified results are very similar to the combined analysis illustrated in Fig. 3. For the low-molecular-weight metabolite profiles the majority of the trends are also similar between men and women. For instance, the systemic concentrations of isoleucine and glycoproteins behave coherently for men and women in their associations with subclinical atherosclerosis. Nevertheless, more variation is present than in the case of the lipoprotein subclasses, as evident for, *e.g*., tyrosine. Similarly, there are also notable differences for the clinical and biochemical measures determined independently of NMR, *e.g*., in the case of the liver enzymes glutamyl transferase and alanine aminotransferase.

Results of the self-organizing map analysis for the 2007 data subset are shown in Supplementary Figure 2. The analysis was conducted with identical settings as used for calculating the SOM component planes shown in Fig. 2. Since the SOM-training algorithm for distributing the samples on the map was run separately for this data subset, the resulting component planes are visually different than those in Fig. 1. However, the analytical conclusions drawn from the analysis are identical: the component planes illustrate clear associations between the metabolic NMR profiles and the carotid IMT and carotid artery distensibility. Indeed, the statistical significance is retained almost as high as for the analysis on the complete data set. These results confirm that the combined analysis (Fig. 3) including 64% of the individuals twice, with serum samples collected and ultrasound conducted in 2001 and 2007, *i.e.*, 6-years apart, did not influence the results. Importantly, also the biological interpretations remain identical for the 2007 subset of samples and for the entire data set. For instance, a high prevalence of MetS associates with the region of high IMT, but the association is not exclusive. The region with high IMT and low prevalence of MetS is characterized by high levels of LDL-C as in the case of the entire data set.

Supplementary Table 1. Summary characteristics of the metabolic phenotypes.

Total lipid concentration in lipoprotein subclass [mmol/l]	**A**	**B**	**C**	**D**	**E**	**F**
Extremely large VLDL	0.080 (0.065 –0.11)	0.020 (0.011 –0.034)	0.024 (0.013 –0.037)	0.0097 (0.0013 –0.022)	0.0023 (0 –0.0066)	0.0046 (2.5e-4 –0.0098)
Very large VLDL	0.21 (0.17 –0.27)	0.067 (0.040 –0.10)	0.076 (0.047 –0.10)	0.042 (0.021 –0.073)	0.010 (6.0e-5 –0.021)	0.016 (4.6e-4 –0.032)
Large VLDL	0.76 (0.65 –0.90)	0.31 (0.20 –0.43)	0.34 (0.23 –0.44)	0.24 (0.12 –0.33)	0.078 (0.042 –0.13)	0.11 (0.067 –0.17)
Medium VLDL	1.2 (1.09–1.38)	0.68 (0.55–0.85)	0.70 (0.58–0.82)	0.57 (0.41–0.72)	0.29 (0.21–0.36)	0.37 (0.30–0.46)
Small VLDL	1.0 (0.89–1.1)	0.78 (0.67–0.88)	0.71 (0.64–0.79)	0.66 (0.56–0.81)	0.42 (0.33–0.50)	0.50 (0.42–0.57)
Very Small VLDL	0.63 (0.55–0.75)	0.64 (0.57–0.71)	0.52 (0.46–0.59)	0.62 (0.53–0.72)	0.45 (0.37–0.52)	0.44 (0.37–0.50)
IDL	1.4 (1.2–1.6)	1.4 (1.3–1.6)	1.2 (1.0–1.3)	1.5 (1.3–1.7)	1.1 (0.97–1.3)	1.0 (0.89–1.2)
Large LDL	1.7 (1.5–2.0)	1.8 (1.6–2.0)	1.5 (1.3–1.7)	1.8 (1.6–2.1)	1.3 (1.2–1.5)	1.3 (1.1–1.5)
Medium LDL	1.1 (0.92–1.3)	1.0 (0.92–1.2)	0.89 (0.78–1.0)	1.0 (0.87–1.2)	0.76 (0.65–0.87)	0.74 (0.63–0.87)
Small LDL	0.77 (0.66–0.88)	0.67 (0.59–0.75)	0.57 (0.50–0.65)	0.65 (0.55–0.77)	0.47 (0.40–0.54)	0.46 (0.39–0.54)
Very large HDL	0.31 (0.20–0.42)	0.28 (0.19–0.37)	0.20 (0.14–0.27)	0.51 (0.35–0.70)	0.53 (0.41–0.66)	0.27 (0.19–0.36)
Large HDL	0.42 (0.31–0.58)	0.63 (0.48–0.82)	0.45 (0.32–0.57)	1.32 (1.1–1.6)	1.1 (0.97–1.30)	0.65 (0.51–0.79)
Medium HDL	1.0* (0.93–1.2)	1.0 (0.93–1.2)	0.91 (0.82–0.99)	1.5 (1.3–1.7)	1.1 (0.98–1.2)	0.91 (0.82–1.0)
Small HDL	1.3 (1.2–1.4)	1.3 (1.2–1.4)	1.2 (1.2–1.3)	1.4 (1.3–1.6)	1.2 (1.1–1.3)	1.2 (1.1–1.2)

* indicates metabolite distributions with *non-significant p*-values

Low–molecular-weight metabolites [NMR signal area]	**A**	**B**	**C**	**D**	**E**	**F**
Alanine	31 (24–35)	27 (24–31)	28 (24–32)	28 (25–33)	24 (22–28)	25 (22–28)
Histidine	1.2* (1.1–1.3)	1.2* (1.1–1.3)	1.2* (1.1–1.3)	1.3 (1.2–1.5)	1.2 (1.1–1.3)	1.2 (1.1–1.3)
Isoleucine	4.6 (4.0–5.4)	2.6 (2.1–3.2)	2.9 (2.4–3.6)	2.1 (1.6–2.6)	1.5 (1.2–1.9)	2.0 (1.6–2.4)
Leucine	1.3	1.3	1.31	1.8*	1.2	1.3*

	(1.1–1.6)	(1.1–1.5)	(1.2–1.5)	(1.1–1.5)	(1.0–1.3)	(1.1–1.4)
Tyrosine	1.5 (1.3–1.7)	1.3 (1.1–1.5)	1.3 (1.1–1.5)	1.2 (0.96–1.5)	1.1 (0.96–1.3)	1.2 (1.0–1.4)
Valine	19 (17–22)	17 (15–19)	17 (15–19)	16 (13–19)	13 (12–16)	15* (14–17)
Acetoacetate	1.7* (1.2–2.2)	1.6* (1.2–2.2)	1.5 (1.1–2.0)	2.0 (1.4–2.8)	1.6* (1.1–2.4)	1.5 (1.1–2.2)
3-hydroxy-butyrate	1.2 (0.55–2.3)	2.9* (1.8–4.5)	3.1* (2.0–4.8)	2.7 (1.5–5.0)	3.1* (1.8–5.1)	3.8 (2.6–5.6)
Creatinine	1.6 (1.4–1.9)	1.5 (1.3–1.8)	1.6 (1.4–1.8)	1.5* (1.3–1.8)	1.4 (1.2–1.6)	1.4 (1.2–1.7)
Glycoproteins	113 (102–128)	74 (65–83)	69 (61–77)	84 (74–97)	55 (49–63)	51 (45–56)
Urea	68 (62–74)	58 (53–63)	56 (52–61)	64 (57–72)	49 (45–54)	50 (46–54)

* indicates metabolite distributions with *non-significant p*-values

Clinical and biochemical variables	**A**	**B**	**C**	**D**	**E**	**F**
Body-mass index [kg/m^2]	29.0 (26–31.9)	26.5 (24.1–29.4)	26.5 (24.1–29.4)	24.2 (21.9–26.5)	22.7 (20.8–24.9)	24.2 (22.1–26.9)
Systolic blood pressure [mmHg]	127 (117–136)	121 (113–131)	121 (113–129)	118* (109–127)	113 (106–122)	115 (107–125)
Glucose [mmol/l]	5.4 (5.0–5.7)	5.3 (4.9–5.6)	5.3 (5.0–5.6)	5.0 (4.7–5.4)	5.0 (4.7–5.3)	5.0 (4.8–5.4)
Insulin [mU/l]	12 (8.0–18)	8.0 (5.0–11)	8.0 (6.0–12)	6.8* (4.6–9.0)	5.0 (3.5–7.0)	6.0 (4.0–8.8)
Oxidized LDL [U/l]	107 (87–125)	98 (82–115)	87 (75–102)	85* (69–102)	72 (60–85)	71 (59–84)
Glutamyl transferase [U/l]	44 (26–67)	24 (17–37)	24 (17–36)	19* (13–29)	15 (12–21)	17 (13–25)
Alanine amino-transferase [U/l]	26 (16–40)	17 (12–25)	17 (12–26.3)	13* (10–18)	11 (9–15)	13* (10–19)
Plasminogen activator inhi-bitor-1 [ng/ml]	9.2 (5.4–14)	5.3 (3.2–8.5)	5.6 (3.6–9.8)	3.2 (2.1–5.3)	2.7 (2.0–3.9)	3.7* (2.5–5.7)
C–reactive protein [mmol/l]	1.4 (0.66–2.8)	1.0 (0.48–2.2)	0.89* (0.42–2.0)	0.97* (0.41–2.3)	0.57 (0.28–1.3)	0.61 (0.30–1.5)

* indicates metabolite distributions with *non-significant p*-values

All values are median (25th – 75th percentiles). For each variable the distribution of a given phenotype was compared with that of all the other phenotypes combined using two-tailed Kolmogorov–Smirnov test. Note that in most cases the variable distribution for a specific metabolic phenotype was significantly different from the distribution of the combined set of the other five metabolic phenotypes at the Holm–Bonferroni significance level of $p < 0.0025$. **The metabolite distributions with *non-significant p*-values are marked by an asterisk (*).**

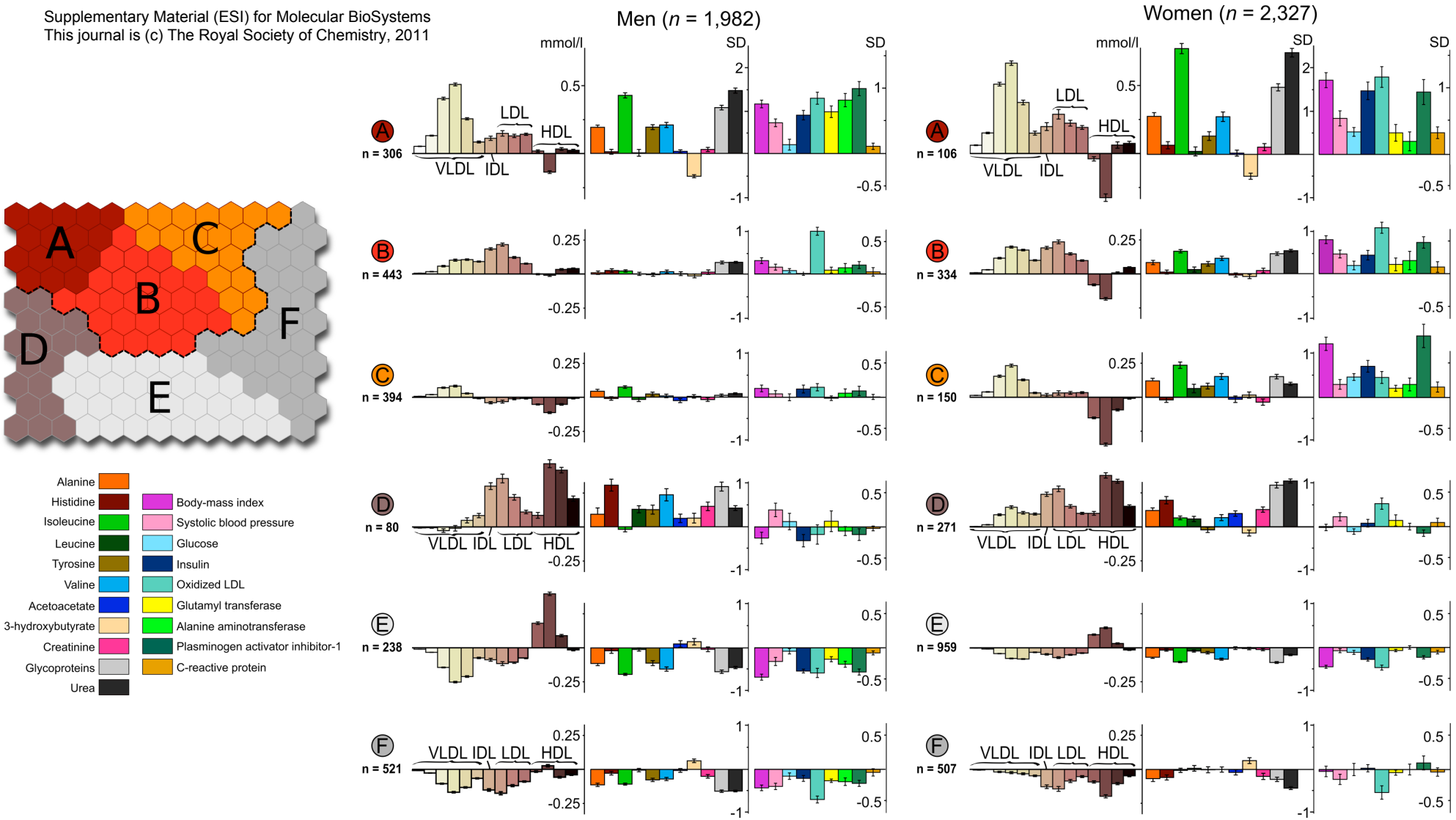
Supplementary Table 2. Characteristics of the study population.

	All	Men	Women
Age [years]	34.7 ± 5.8 (*n* = 4,309)	34.6 ± 5.8 (*n* = 1,982)	34.7 ± 5.8 (*n* = 2,327)
Body mass index [kg/m^2]	25.5 ± 4.6 (*n* = 4,267)	26.2 ± 4.2 (*n* = 1,968)	24.9 ± 4.9 (*n* = 2,299)
Systolic blood pressure [mmHg]	119 ± 14 (*n* = 4,274)	124 ± 13 (*n* = 1,970)	115 ± 13 (*n* = 2,304)
Serum triglycerides [mmol/l]	1.2 (0. 87–1.6) (*n* = 4,263)	1.4 (1.0–1.9) (*n* = 1,953)	1.0 (0.80–1.4) (*n* = 2,310)
LDL-C [mmol/l]	2.0 ± 0.56 (*n* = 4,263)	2.1 ± 0.56 (*n* = 1,953)	1.9 ± 0.53 (*n* = 2,310)
HDL-C [mmol/l]	1.6 ± 0.38 (*n* = 4,263)	1.5 ± 0.32 (*n* = 1,953)	1.7 ± 0.39 (*n* = 2,310)
Glucose [mmol/l]	5.1 (4.8–5.5) (*n* = 4,309)	5.3 (5.0–5.6) (*n* = 1,982)	5.0 (4.7–5.3) (*n* = 2,327)
Insulin [mU/l]	6.9 (4.5–10) (*n* = 4,307)	7.0 (4.6–10) (*n* = 1,980)	6.6 (4.3–10) (*n* = 2,327)
Carotid IMT [mm]	0.60 ± 0.097 (*n* = 4,271)	0.62 ± 0.11 (*n* = 1,964)	0.59 ± 0.088 (*n* = 2,307)
Carotid artery distensibility [%/10 mmHg]	2.26 ± 0.71 (*n* = 4,257)	2.09 ± 0.63 (*n* = 1,956)	2.41 ± 0.74 (*n* = 2,301)
Metabolic syndrome [%]	16.5 (*n* = 4,243)	21.4 (*n* = 1,947)	12.5 (*n* = 2,296)
Oxidized LDL [U/l]	83 ± 25 (*n* = 2,181)	89 ± 26 (*n* = 1,007)	79 ± 23 (*n* = 1,174)
Glutamyl transferase [U/l]	19 (14–31) (*n* = 4,302)	27 (18–43) (*n* = 1,978)	15 (12–22) (*n* = 2,324)
Alanine transferase [U/l]	14 (10–22) (*n* = 2,123)	19 (14–29) (*n* = 974)	11 (9.0–15) (*n* = 1,149)
Plasminogen activator inhibitor-1 [ng/ml]	4.0 (2.5–7.0) (*n* = 2,121)	4.80 (3.0–8.5) (*n* = 973)	3.3 (2.2–5.6) (*n* = 1,148)
C-reactive protein [mmol/l]	0.78 (0.36–1.8) (*n* = 4,309)	0.71 (0.34–1.5) (*n* = 1,982)	0.89 (0.39–2.2) (*n* = 2,327)

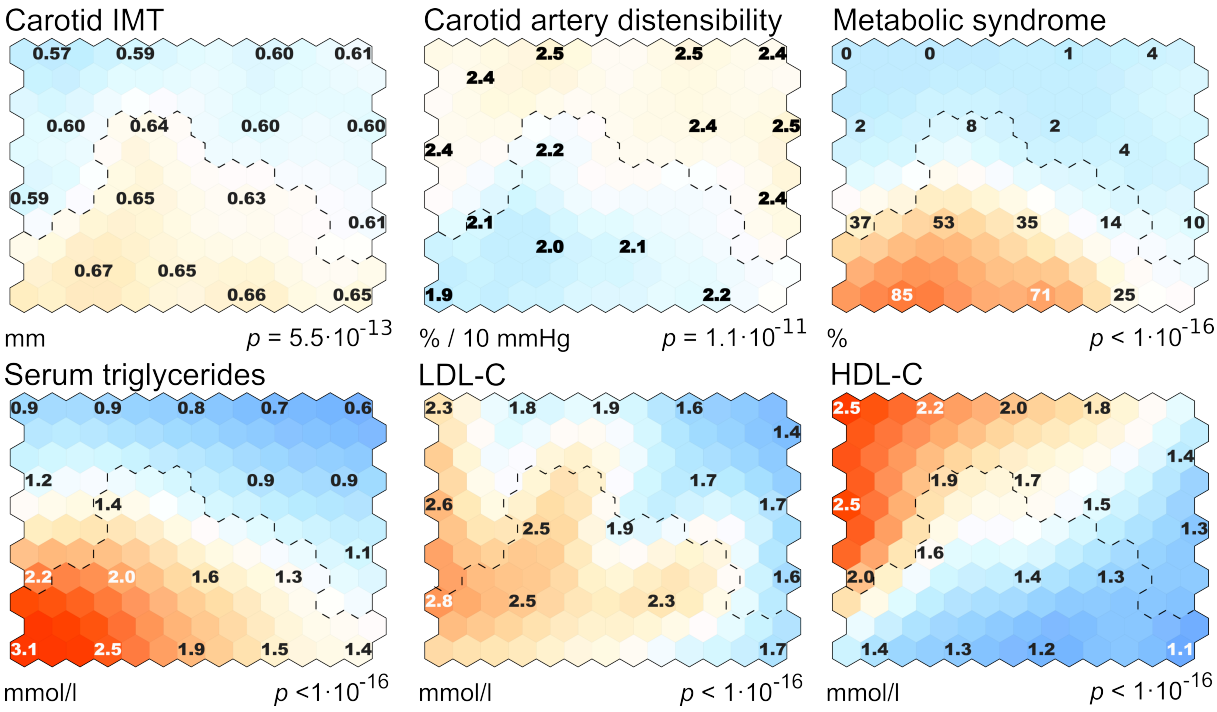
Values are indicated as mean ± SD for normal distributions and median (25^{th} – 75^{th} percentile) for skewed distributions (glucose, insulin, serum triglycerides, glutamyl transferase, alanine transferase, plasminogen activator inhibitor-1, and C-reactive protein). *n* indicates the number of measurements available for the study.

$p < 10^{-6}$ for all two-tailed Kolmogorov-Smirnov comparisons between men and women, except for age and insulin where $p > 0.1$.

LDL-C and HDL-C refer to NMR-determined low- and high-density lipoprotein cholesterol, respectively.



Supplementary Figure 1. Characteristics of the metabolic phenotypes for men and women. Lipoprotein subclass concentrations (in mmol/l), low-molecular-weight metabolites (in units of standard deviation) and non-NMR variables (in units of standard deviation) are shown in the histograms as deviation from the median value for all men and all women. Error bars indicate standard error. The histogram profiles of the sex-stratified phenotypes were calculated based on the SOM analysis depicted in Fig. 3.



Supplementary Figure 2. Self-organizing map analysis of the 2007 data subset. Statistical colorings of selected clinical and lipid variables in the SOM analysis of the serum NMR metabonomics data for 2,123 samples from the 27-year follow-up survey of the Cardiovascular Risk in Young Finns Study conducted in 2007. The SOM analysis positions the samples so that spectral differences between adjacent samples are minimized. A given serum sample is in the same place in each component plane. The values in the component planes are visualized to show whether the unit values are above (red), at (white) or below (blue) the median of the variable. Numbers on selected units indicate the local mean value for that particular region. The SOM is divided into high and low risk phenotypes indicated by the dashed line based on the carotid IMT median. When comparing the component planes to Fig. 2 it should be kept in mind that the organizations in different SOM analysis are independent and thereby preclude direct visual comparison of the results.

References

1. J. Niemi, V.-P. Mäkinen, J. Heikkonen, L. Tenkanen, Y. Hiltunen, M. L. Hannuksela, M. Jauhiainen, C. Forsblom, M.-R. Taskinen, Y. A. Kesäniemi, M. Savolainen, K. Kaski, P.-H. Groop, P. T. Kovanen, M. and Ala-Korpela, *Ann Med.,* 2009, **41**, 451–461.
2. S. Holm, *Scand. J. Statist*., 1979, **6**, 65–70.
3. D.I. Chasman, G. Paré, S, Mora, J.C. Hopewell, G. Peloso, R. Clarke, L.A Cupples, A Hamsten, S Kathiresan, A Mälarstig, J.M. Ordovas, S. Ripatti, A.N. Parker, J.P. Miletich, and P.M. Ridker, *PLoS Genet.*, 2009, **5**, e1000730.
4. O.T. Raitakari, M. Juonala, M. Kähönen, L. Taittonen, T. Laitinen, N. Mäki-Torkko, M.J. Järvisalo, M. Uhari, E. Jokinen, T. Rönnemaa, H.K. Åkerblom, and J.S. Viikari, *JAMA,* 2003, **290**, 2277–2283.
5. J. Koskinen, C.G. Magnussen, L. Taittonen, L. Räsänen, V. Mikkilä, T. Laitinen, T. Rönnemaa, M. Kähönen, J.S. Viikari, O.T. Raitakari, and M. Juonala, *Circulation,* 2009, **120**, 229–236.