Gender-dependent associations of metabolite profiles and body fat distribution in a healthy population with central obesity: towards metabolomics diagnostics


Published in:
Omics

DOI:
10.1089/omi.2012.0062

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Gender-Dependent Associations of Metabolite Profiles and Body Fat Distribution in a Healthy Population with Central Obesity: Towards Metabolomics Diagnostics

Ewa Szyman’ska,1,2* Jildau Bouwman,1,3* Katrin Strassburg,1,4 Jacques Vervoort,1,5 Antti J. Kangas,6 Pasi Soininen,6,7 Mika Ala-Korpela,6,8 Johan Westerhuis,1,2 John P.M. van Duynhoven,1,9,10 David J. Mela,10 Ian A. Macdonald,11 Rob J. Vreeken,1,4 Age K. Smilde,1,2 and Doris M. Jacobs1,10

Abstract

Obesity is a risk factor for cardiovascular diseases and type 2 diabetes especially when the fat is accumulated to central depots. Novel biomarkers are crucial to develop diagnostics for obesity and related metabolic disorders. We evaluated the associations between metabolite profiles (136 lipid components, 12 lipoprotein subclasses, 17 low-molecular-weight metabolites, 12 clinical markers) and 28 phenotype parameters (including different body fat distribution parameters such as android (A), gynoid (G), abdominal visceral (VAT), subcutaneous (SAT) fat) in 215 plasma/serum samples from healthy overweight men (n = 32) and women (n = 83) with central obesity. (Partial) correlation analysis and partial least squares (PLS) regression analysis showed that only specific metabolites were associated to A:G ratio, VAT, and SAT, respectively. These association patterns were gender dependent. For example, insulin, cholesterol, VLDL, and certain triacylglycerols (TG 54:1-3) correlated to VAT in women, while in men VAT was associated with TG 50:1-5, TG 55:1, phosphatidylcholine (PC 32:0), and VLDL (X)L. Moreover, multiple regression analysis revealed that waist circumference and total fat were sufficient to predict VAT and SAT in women. In contrast, only VAT but not SAT could be predicted in men and only when plasma metabolites were included, with PC 32:0 being most strongly associated with VAT. These findings collectively highlight the potential of metabolomics in obesity and that gender differences need to be taken into account for novel biomarker and diagnostic discovery for obesity and metabolic disorders.

Introduction

The accumulation of additional fat in central and intra-abdominal depots increases the risk of cardiovascular disease (CVD) and type 2 diabetes mellitus. However, not all overweight/obese individuals develop cardiovascular or metabolic diseases. A substantial portion of overweight/obese subjects are without any of the metabolic abnormalities associated with insulin resistance, while insulin resistance and associated metabolic abnormalities are not uncommon in normal-weight subjects (Reaven, 2011). Much effort has been made to relate susceptibility to CVD to clinical measures of adiposity (Wormser et al., 2011). The difficulty in finding correlations between fat distribution and CVD may be rooted in the uncertainty that people with a high risk profile develop disease, in the heterogeneity of the disease, and in the diverse nature of its association with body fat distributions (BFDs).
Gender differences in body fat distributions are apparent and may explain differences in the prevalence of CVD between men and women. In general, men accumulate excessive fat in the abdominal region, whereas women are more susceptible to the accumulation of fat in the gynoid (gluteofemoral) region. Although increase in gynoid adipose tissue also correlates to CVD risk in obese women (mainly due to the general increase in weight), high storage of fat in the gynoid region relative to total fat is negatively related to CVD risk (Wiklund et al., 2008). Excessive accumulation of fat in the abdominal region has strongly been associated with metabolic alterations such as disturbed plasma lipoprotein profiles, hyperinsulinemia, insulin resistance, and glucose intolerance (Lemieux et al., 1994).

The adipose tissue in the abdominal region can be further divided into intra-abdominal (visceral) and subcutaneous fat, which are dissimilar in more than just their localization. For example, excess visceral fat accumulation, in contrast to subcutaneous fat accumulation, appears to be the main component of abdominal obesity and has been suggested to be an important correlate of gender difference in CVD risk (Lemieux et al., 1994).

Numerous studies have demonstrated that specific regional BFD parameters are clearly associated with metabolic complications (Despres et al., 1990; Williams, 2004). However, these evaluations were generally limited to a number of relevant blood biomarkers such as triacylglycerols (TG), high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, very-low-density lipoprotein (VLDL) cholesterol, insulin, and glucose. The armory of metabolomics technology available nowadays offers a broader view on metabolism. For instance, lipidomics provides detailed information on more than 90 lipid components in plasma and has been demonstrated to be useful in identifying lipid components that are linked to metabolic and cardiovascular disorders (de Mello et al., 2009; Graessler et al., 2009; Quehenberger and Dennis, 2011). Recently, untargeted metabolite profiling using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC-Q-TOF MS) has identified three lysophosphatidylcholines (LPC) as potential plasma markers and confirmed eight known metabolites for overweight/obesity in men (Kim et al., 2010). Moreover, lipoprotein subclasses and particles have been considered to play a critical role in the assessment of coronary heart disease (Ala-Korpela, 2008; Despres et al., 1990; Superko, 2001).

Central obesity, as measured by waist-to-hip ratio, has been shown to be associated with modifications in lipoprotein distribution and composition indicative of CVD (James et al., 1997). In another study, the visceral fat area was positively correlated with certain VLDL and LDL subclasses and negatively with large and medium HDL subclasses in male obese subjects (Okazaki et al., 2005). From the current evidence it is apparent that an in-depth view of metabolism will help to describe relationships to phenotypic parameters in more detail and to better understand their underlying metabolic processes.

In the current study we systematically investigated relationships between multiple blood parameters and phenotypic characteristics (including body fat distribution parameters) among 215 samples from healthy overweight male (n = 32) and female (n = 83) individuals (including one or two samples per subject). In total, 177 metabolites were measured in blood using lipidomics (parameters (p) = 136), lipoprotein subclass profiling (p = 12), NMR-based low-molecular-weight-metabolite (LMWM) profiling (p = 17), and clinical chemistry (p = 12). The phenotypes were described by typical phenotypic parameters and diverse BFD parameters as measured by magnetic resonance imaging (MRI) (intra-abdominal and subcutaneous fat) and dual-energy X-ray absorptiometry (DEXA) (android and gynoid fat). We calculated (partial) correlations between single metabolites and the different BFD parameters in order to estimate the importance of each metabolite with respect to a certain BFD parameter. Considering that a combination of certain metabolites rather than a single metabolite may be linked to a physiological endpoint, we also identified sets of metabolites that are needed to describe best a specific BFD parameter. In addition, we estimated the added value of metabolomics analysis in the presence of phenotypic and clinical chemistry data. Therefore, we quantified the incremental gain in the prediction of the certain fat distribution parameters with different measures from anthropometry, clinical chemistry, and metabolomics analysis.

Materials and Methods

Subjects

Study subjects were 83 women and 32 men with abdominal obesity, waist circumference of over 80 cm for women or 94 cm for men, age 18–55 years, and only pre-menopausal women. Apart from the increased waist circumference and overweight or obese status, they were apparently healthy (no known cardiovascular, respiratory, neurological, or metabolic disease) and nonsmoking. Although currently healthy, many subjects had evidence of elevated cardiovascular risk: 14% of the women and 48% of the men would be classified with metabolic syndrome (MetS) according to inclusion criteria of World Health Organization (WHO). Subjects were not taking any medications, did not follow any weight loss or medically prescribed diet, were not suffering from renal and/or liver disease, and did not have an abnormal lipid profile at the screening visit. They participated in a double-blinded, randomized, parallel nutritional intervention trial that was conducted at the University of Nottingham. The experimental protocol allowing for metabolomics analysis was approved by the University of Nottingham Medical School Ethics Committee. In this study, subjects consumed twice a day either a drink (250 mL) containing a catechin-enriched green tea extract (daily dose: 600 mg) or a green tea flavored placebo drink (without any of the active ingredients) over 12 weeks.

Blood collection

Blood samples were taken after an overnight fasting period at baseline (start of the experiment) and after 12 weeks of intervention from an arterialized venous cannula. Blood samples were placed in EDTA-treated or plain tubes, centrifuged to yield plasma or serum, and stored at −80°C until analysis. In total, 166 and 64 blood samples from women and men, respectively, were collected and analyzed.

Phenotype parameters

The BMI (weight/height²) was calculated from measured body weight and height. Six circumferences were measured,
including upper arm, waist, hip, proximal thigh, middle thigh, and distal thigh. The waist/hip ratio was calculated from the waist and hip circumferences. The total body fat (based on estimated body density) and five regional measures (biceps, triceps, subscapular, suprailiac, and thigh) were derived from skinfold measurement (Durnin and Womersley, 1974). The lean body mass (LBM) was calculated by using the following equation: \( \text{LBM} = \frac{\text{weight} - \text{bodyfat}}{100} \). The total body, android, and gynoid fat contents were measured by dual energy X-ray absorptiometry (DEXA) (Kiebzak et al., 2000) and the android/gynoid ratio was calculated. The visceral (VAT) and subcutaneous (SAT) fat content were determined by magnetic resonance imaging (MRI) of 16 transaxial scans reaching at the fourth and fifth lumbar interspace (Hu et al., 2011). Basal systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate (HR) were measured twice (with a 2-minute interval) using an automated BP-monitor after a 10-minute resting period in a sitting position.

**Clinical parameters**

Fasted blood samples were analyzed for glucose (YSI 2300 analyzer; YSI Inc, Yellow Springs, Ohio), free fatty acids (FFA) (NEFA HR kit; Wako Chemicals, Neuss, Germany), glycerol (Sigma kit F6428; Sigma-Aldrich, St Louis, MO, USA), 3-hydroxybutyrate (Williamson et al., 1974), insulin (Siemens TKNX radioimmunoassay kit; Siemens, Erlangen, Germany), creatinine, triacylglycerols (total TG), high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, total cholesterol (Olympus AU automated analyzer; Olympus UK Ltd. Southall, UK), epinephrine, and norepinephrine (Forster and MacDonald, 1999). The total cholesterol/ HDL cholesterol ratios were calculated. Insulin resistance (IR) was calculated by the homeostasis model assessment (HOMA) using the following equation:

\( IR = \frac{\text{fasting insulin} (\mu \text{IU/mL}) \times \text{fasting glucose} (\text{mmol/L})}{22.5} \) (Matthews et al., 1985).

**NMR-based metabolite profiling**

NMR samples (3-mm tubes) were prepared by adding 100 μL phosphate buffer solution [0.1 M, NaH2PO4/Na2HPO4, pH 7.4 containing 4 mM 3-(trimethylsilyl)propionic acid-d4 sodium salt (TSP), 10 μL sodium azide (4%), and 20 μL deuterium oxide (D2O)], and 100 μL D2O to 100 μL serum. The buffer allocation and subsequent serum mixing was performed manually in a gentle fashion.

One-dimensional 1H NMR spectra were acquired on a Bruker Avance III 600 NMR spectrometer operating at 600.13 MHz and equipped with a cryoprobe and a sample changer (Bruker BioSpin GmbH, Germany) for sample delivery. For the lipoprotein (LIFO) profiles, a ‘noesyppgrld’ pulse sequence was used with a mixing time of 10 msec to suppress the water peak. The acquisition time was 2.7 sec and the relaxation delay 4.0 sec. The spectra were acquired with 4 dummy scans, 64 transients and a sweep width of 18028 Hz. For the low-molecular weight metabolites (LMWM) profiles, a cpmg pulse sequence with water suppression was used to suppress water and most of the broad macromolecule and lipoprotein lipid signals. A 38.4 msec T2-filter with a fixed echo delay of 300 μsec was applied to minimize diffusion and J-modulation effects. The acquisition time was 3.3 sec and the relaxation delay 4.0 sec. The spectra were acquired with 4 dummy scans, 64 transients, and a sweep width of 12019 Hz.

Samples of the same volunteer were measured in a randomized manner. Both spectra were acquired at a temperature of 310 K and with 65536 data points. The temperature was equilibrated for 5 min before data acquisition. The 90° pulse was calibrated automatically for each sample. A constant receiver gain of 57 was used for all samples.

The spectra were automatically corrected for phase and baseline distortions using Topspin 1.3 software (Bruker Analytik Rheinstetten, Germany). Zero-filling and an exponential window function with a line-broadening factor of 1.0 Hz were applied to the free induction decay (FID) prior to Fourier transformation. The spectra were referenced to TSP \( (\delta \text{H} \text{C} = 0 \text{ppm}) \). Total lipids in 12 lipoprotein subclasses were quantified using a calibrated regression model (Ala-Korpela, 2008; Kettunen et al., 2012; Soininen et al., 2009; van den Berg et al., 2009). The subclasses were defined via high-performance liquid chromatography (Okazaki et al., 2005) and are as follows: very large and large VLDL (average particle diameter 58.8 nm, (X)L-VLDL) medium VLDL (44.5 nm, M-VLDL), small VLDL (36.8 nm, S-VLDL), very small VLDL (31.3 nm, XS-VLDL), IDL (28.6 nm), large LDL (25.5 nm, L-LDL), medium LDL (23.0 nm, M-LDL), small LDL (18.7 nm, S-LDL), very large HDL (14.3 nm, VL-HDL), large HDL (12.1 nm, L-HDL), medium HDL (10.9 nm, M-HDL), and small HDL (8.7 nm, S-HDL).

13 LMWM (i.e., glucose, valine, isoleucine, leucine, alanine, tyrosine, phenylalanine, histidine, 3-hydroxybutyrate, lactate, acetoacetate, acetate, and creatinine) were quantified from the LMWM spectra using lineshape fitting analysis with the PERCH NMR software platform (PERCH Solution Ltd., Kuopio, Finland) (Mierisova and Ala-Korpela 2001; Soininen et al., 2005).

**Lipid profile**

The UPLC-MS Lipidome Platform was developed, validated, and applied in this study at the Demonstration and Competition Laboratory (DCL) of the Netherlands Metabolomics Centre, situated at Leiden University in The Netherlands. This lipidome platform was based on the protocol given by Hu et al., (2008) and Castro-Perez et al., (2011) and was described in detail in the previous study by Jansen et al., (2011) as well as in Supplementary Box S1. (See Supplementary Box S1 at www.liertpub.com/omi). With this platform, 136 lipids species were measured in plasma samples. They are cholesterol esters (CE), di- and triacylglycerols (DG and TG), lysophosphatidylethanolamines (LPE), phosphatidylethanolamines (PE), lysophosphatidylcholines (LPC), phosphatidylcholines (PC), and sphingomyelins (SM). These groups also include ether lipids [i.e., plasmalogens (PE-O, LPC-O, PC-O, LPC-O)]. A complete list of the detected metabolites is given in Supplementary Table S1. For 215 samples (out of analyzed 230, 94%), all data including phenotype parameters, clinical parameters,
NMR-based metabolite profiles, lipids profiles were available and only those are included in statistical analysis. For 15 samples (out of analyzed 230, 6%), a complete data set was not available due to problems with data acquisition or processing. The data sets contained 157 samples from 83 women and 58 samples from 32 men. All statistical analysis was performed on nontransformed data, because log- or square-root transformation did not affect the current results.

Data analysis strategy

Our data analysis strategy (Fig. 1) pursued two main goals, namely (i) to determine robust correlations of metabolites with BFD, (ii) to assess the added value of metabolomics analysis in describing BFD. Initially, an exploratory analysis of phenotype and metabolite data sets was performed where main trends in each data set, and associations between two data sets and gender differences were studied. Next, due to observed gender differences, data sets were divided by gender and used in extensive analysis of associations of BFD parameters and metabolites separately for women and men. Because of the different size of the data sets available for women ($n=157$) and men ($n=58$), not only correlations but also their statistical significance were taken into account when comparing the associations between women and men. Analysis of associations was composed of three types of analyses:

**Fig. 1.** Data analysis strategy.
analysis of bivariate associations, analysis of multivariate associations, and multiple regression analysis including different sets of phenotype and metabolic variables. Bivariate associations were combined with multivariate associations into association patterns to determine the best set of metabolites for each BFD parameter. Bivariate associations independent on other phenotype parameters were studied to select associations that are specific only for a given BFD parameter. Finally, the added value of metabolomics analysis was assessed by performing multiple regression analysis with variable selection on several data sets encompassing different groups of variables, such as phenotype parameters, clinical markers, and metabolites. The best set of variables for each BFD parameter was evaluated on an independent set of samples (these were kept aside) and used to assess the added value of metabolomics analysis in describing each BFD parameter.

Statistical analyses were performed in the Matlab 2010a (The Mathworks Inc., Natick, MA, USA) using Statistics Toolbox (The Mathworks Inc.) and in-house written routines, partly based on the PLS Toolbox (Eigenvector Research, USA). Permutation tests have been performed on the LISA-SARA Dutch super-computer.

Exploratory analysis and gender differences

Principal component analysis (PCA) was performed separately on the phenotype data set and metabolic data set in order to identify major trends in each. Data sets were autoscaled before PCA analysis (Fig. 2A and 2B). The resulting score plots were color coded according to gender. Canonical correlation analysis (CCA) was performed on PCA scores to examine the link between the two data sets (Fig. 2C). (Massart et al., 1997; van den Berg et al., 2009). PCA scores from the 4 (phenotype data set) and 10 (metabolic data set) first PC components were used in this analysis. Results of the CCA analysis were validated according to procedure presented by van den Berg et al., 2009.

Furthermore, gender differences in the levels of physiological parameters and metabolites were assessed by comparison of basic statistical parameters (mean, standard deviation (SD), median, minimum, and maximum). Radar plots were used to visualize gender differences in parameters (Fig. 3). The median for men was used as the reference value (Szymanska et al., 2010). Mann-Whitney tests were used to test whether the gender differences were statistically significant at \( \alpha < 0.05 \), corrected for multiple comparisons with False Discovery Rate (FDR) by Benjamini-Hochberg at a significance threshold \( q < 0.05 \).

Correlations with BFD: (Independent) bivariate associations

Spearman correlation coefficients were calculated to assess the bivariate associations of certain BFD parameters and all metabolites separately for women and men. The BFD parameters included android fat (A), gynoid fat (G), A:G ratio, VAT, SAT, waist circumference (waist), and lean body mass (LBM). By this analysis, an overview of all associations of those BFD parameters was obtained. Additionally, in order to identify associations that were independent on other phenotype parameters, partial correlations were calculated also with Spearman correlation coefficients. In this analysis, the bivariate associations of A:G ratio, visceral fat, and subcutaneous fat were adjusted by one of selected phenotype parameters (possible covariate) at a time (Fig. 1). Phenotype parameters used for adjustment were: age, diastolic blood pressure (DBP), heart rate (HR), total fat, lean body mass (LBM), waist circumference, A:G ratio (for associations of VAT and subcutaneous fat), VAT (for associations of A:G ratio and SAT), and SAT (for associations of A:G ratio and visceral fat).

The statistical significance (p value) of all obtained Spearman correlation coefficients was assessed by both large-sample approximations and 2000 permutations tests (Fujikoshi et al., 2010; van den Berg et al., 2009). The significance of bivariate associations (p values) is dependent on data set size, so usually lower for men (n = 58) than women (n = 157) as given in Supplementary Table S2. FDR correction was performed per BFD parameter (i.e., for 177 p-values of correlations between a BFD parameter and 177 metabolites/clinical parameters). In general, all bivariate associations with an
absolute Spearman correlation coefficient higher than 0.22 (for women) and 0.38 (for men) were statistically valid after FDR correction.

Correlations with BFD: Association patterns

Partial least squares (PLS) regression analysis was used to identify associations of specific BFD parameters with a metabolic data set (all metabolites at once) (Fig. 1). The BFD parameters (selected from phenotype data set for this analysis) were A:G ratio, VAT, and SAT. All metabolites \((p = 177)\) were included. The analysis was run separately for women and men and with a double-cross validation scheme (7-fold single cross validation and 8-fold double cross validation), including 5-step variable selection (van Velzen et al., 2008). Root mean square error of cross validation (RMSECV) was

FIG. 3. Radar plots of median levels of selected phenotypic and metabolic variables. (A) Selected phenotype variables; (B) selected lipids; (C) selected lipoproteins; and (D) selected clinical parameters. *Statistically significant gender difference. Radar plots of all variables are presented in Supplementary Fig. S1.

metabolic data set (all metabolites at once) (Fig. 1). The BFD parameters (selected from phenotype data set for this analysis) were A:G ratio, VAT, and SAT. All metabolites \((p = 177)\) were included. The analysis was run separately for women and men and with a double-cross validation scheme (7-fold single cross validation and 8-fold double cross validation), including 5-step variable selection (van Velzen et al., 2008). Root mean square error of cross validation (RMSECV) was

FIG. 4. Gender-dependent association patterns of BFD parameters; (A) A:G ratio in women; (B) A:G ratio in men; (C) intra-abdominal fat (VAT) in women; (D) VAT in men; (E) subcutaneous fat (SAT) in women; (F) SAT in men. Each association pattern is composed of 177 markers representing metabolic variables: orange squares, clinical parameters, green crosses, low molecular weight metabolites, blue circles, lipid metabolites, and green triangles, lipoproteins. Strength of single association of each metabolic variable with phenotype parameter is presented on horizontal axis and expressed as value of Spearman correlation coefficient. The significance of contribution of each metabolic variable to multivariate association of phenotype parameter with all metabolites simultaneously is presented on vertical axis and expressed as significance of contribution to multivariate association based on rank products in PLS regression models. Regions of plots marked a–f contain metabolic variables with similar associations with phenotype parameter and are used in interpretation. For clarity of the figures, only the most important metabolic variables are labeled.
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>A:G Ratio</th>
<th>VAT</th>
<th>SAT</th>
<th>Waist</th>
<th>LBM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPC O-16:1</td>
<td></td>
<td></td>
<td>-ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPC 18:2</td>
<td></td>
<td></td>
<td>-adht</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPC 22:6</td>
<td></td>
<td></td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 32: (1:2)</td>
<td>ah</td>
<td></td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC O-34:1</td>
<td>ah</td>
<td></td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 34:4</td>
<td>ah</td>
<td></td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC O-36:3</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 36:3</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 38:3</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>SM d18:1/16:1</td>
<td>ah</td>
<td></td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>DG 36:3</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 44:(0:2)</td>
<td></td>
<td></td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 46:(1:2)</td>
<td></td>
<td></td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 48:(0:3)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 48:4</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 50:1</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 50:(2:3)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 50:4</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 50:5</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 51:1</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 51:3</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 52:(2:3)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 52:1, TG 53:2</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 54:(1:4)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 54:(5:7)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 55:1, TG 56:(6:7, 1:2)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 56:5</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 56:8</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 58:2</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TG 59:2</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>VLDL(XL-S)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>VLDL (XS)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>HDL (XL-L)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>isoleucine</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>valine</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>TOTAL TAG</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>Chol:HDL ratio</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>Insulin</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>HOMA-IR</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>A:G Ratio</th>
<th>VAT</th>
<th>SAT</th>
<th>Waist</th>
<th>LBM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CholE 22:6</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPC 16:0, O-16:1</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPC 16:1</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPC 18:(0:2)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPC 20:4</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPC 20:5</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>LPE 18:0</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 32:0</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 32:(1:2)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 34:0</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC O-34:2</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC O-34:3</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 36:6</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC O-36:5</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC O-36:(2:3)</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 38:3</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 38:7</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PC 40:8</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
<tr>
<td>PE O-36:5</td>
<td>ah</td>
<td>ah</td>
<td></td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
<td>ah</td>
</tr>
</tbody>
</table>

(continued)
Table 1. (Continued)

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>A</th>
<th>G</th>
<th>A:G Ratio</th>
<th>VAT</th>
<th>SAT</th>
<th>Waist</th>
<th>LBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE 38:6</td>
<td></td>
<td></td>
<td>+adhlwsv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE O-38:5</td>
<td></td>
<td></td>
<td>+dhtrsv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM d18:1/14,17,18,20,21:0</td>
<td></td>
<td></td>
<td>-adhlwsv</td>
<td>+adhlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM d18:0/24:0</td>
<td></td>
<td></td>
<td>+dhtrsv</td>
<td></td>
<td></td>
<td></td>
<td>+dhro</td>
</tr>
<tr>
<td>SM d18:0/19:1</td>
<td></td>
<td></td>
<td>-dhlsv</td>
<td>+adhlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 46:0, 48:(0:2)</td>
<td></td>
<td></td>
<td>+adhlwsv</td>
<td>+adhlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 50:0</td>
<td></td>
<td></td>
<td>++adhlsv</td>
<td>+adhtlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 50:1</td>
<td></td>
<td></td>
<td>+adhlwsv</td>
<td>++adhtlwsv</td>
<td>+tr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 50:2</td>
<td></td>
<td></td>
<td>+adhlwsv</td>
<td>++adhtlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 51:1</td>
<td></td>
<td></td>
<td>++adhlwsv</td>
<td>++adhtlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 52:1</td>
<td></td>
<td></td>
<td>+adhlwsv</td>
<td>++adhtlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 53:2</td>
<td></td>
<td></td>
<td>+adhlwsv</td>
<td>++adhtlwsv</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 54:1(2, 6:7)</td>
<td></td>
<td></td>
<td>+adhts</td>
<td>+adhtlwrs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 55:1</td>
<td></td>
<td></td>
<td>+adhts</td>
<td>++adhtlwrs</td>
<td>+atr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG 56:1(3), TG 58:3, TG 59:2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLDL (XL-S)</td>
<td></td>
<td></td>
<td>+adhtlwrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDL (L)</td>
<td></td>
<td></td>
<td>++dhtr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDL, LDL (M-S)</td>
<td></td>
<td></td>
<td>+adhlwsv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDL(XL)</td>
<td></td>
<td></td>
<td>-dhtrsv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alanine</td>
<td></td>
<td></td>
<td>+hts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>isoleucine</td>
<td></td>
<td></td>
<td>+dhlsv</td>
<td>+adhtlwrs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tyrosine</td>
<td></td>
<td></td>
<td>+hvs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>creatinine</td>
<td></td>
<td></td>
<td>-adhlwsv</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholesterol, Chol:HDL ratio</td>
<td></td>
<td></td>
<td>+dhtr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td></td>
<td>+adhlwsv</td>
<td>-adhtlwrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFA</td>
<td></td>
<td></td>
<td>+adhro</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL TG</td>
<td></td>
<td></td>
<td>++adhtlrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insulin</td>
<td></td>
<td></td>
<td>++adhlwrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOMA-IR</td>
<td></td>
<td></td>
<td>++adhlwrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlations are measured with Spearman correlation coefficient: + weak positive correlations with 0.4 > R > 0.3 with p < 1*10^-6 for women and p < 0.03 for men, ++ strong positive correlations with R > 0.4 with p < 1*10^-7 for women and p < 2*10^-3 for men, - weak negative correlations with 0.4 > R < 0.3 with p < 1*10^-4 for women and p < 0.05 for men, and - - strong negative correlations with R < - 0.4 with p < 1*10^-7 for women and p < 2*10^-3 for men. Listed correlations were controlled by other phenotype parameters in partial correlation analysis and independence of each correlation is indicated by letters.

Correlations independent of: a, age; d, DSP; h, HR; t, total fat; l, LBM; r, A:G ratio; s, SAT; v, VAT; and w, waist; (e.g., in women correlation of SAT with LPC 18:2 is independent on age, DSP, HR, but is dependent on total fat, LBM, waist, A:G ratio, and VAT).

A, android fat; CholE, cholesterol esters; Chol:HDL ratio, ratio of total cholesterol and HDL-cholesterol; DG, diacylglyceride; FFA, free fatty acids; G, gynoid fat; HDL, high density lipoprotein; LBM, lean body mass; LDL, low density lipoprotein; LPC, lysophosphatidylcholine; LPE, lysophosphatidyletanolamine; PC, phosphatidylcholine; PE, phosphatidylethanolamine; SAT, subcutaneous fat; SM, sphingomyelin; TG, triacylglyceride; Total TG, total triacylglycerides; VAT, visceral fat; VLDL, very low density lipoprotein; waist, waist circumference.

used as a diagnostic statistic to assess model quality as well as to assess its statistical significance in comparison with the results of 3000 permutation tests run for each model. Contribution of each metabolite to the obtained model was evaluated on the basis of its rank product and its statistical significance (p value) by relating it to rank products obtained by 3000 permutation tests run for this model.

To determine the best set of metabolites that is associated to A:G ratio, VAT, and SAT, respectively, the p values from PLS regression analysis were plotted against the Spearman correlation coefficients from bivariate associations (Fig. 4). Each of these association patterns is composed of 177 metabolites. For clarity of Figure 4, not all metabolites are labeled. (However, all metabolites with absolute values of Spearman correlation coefficients higher than 0.3 are listed in Table 1.) In each plot of Figure 4, six regions (labeled “a”-“f”) can be distinguished. Region “a” is comprised of metabolites with strong bivariate correlations (absolute value of Spearman correlation coefficient higher than 0.4) and statistically significant contribution to multivariate association (significant at p < 0.05). This group of metabolites gains in importance when the high order interaction between all metabolites is included and thus belongs to the optimal set of metabolites related to this BFD parameter. In contrast, the region “b” is comprised of metabolites with strong bivariate correlations but an insignificant contribution to multivariate associations. Strong bivariate associations of these metabolites with phenotype parameter lose relevance when other variables are taken into consideration, and thus provide the same (redundant) information about the phenotype parameter. The regions “c” and “e” are comprised of metabolites with weak or very weak bivariate correlations with phenotype parameters, but gain in importance when multivariate associations are taken into account, meaning that they only emerge when interactions with other metabolites are present. Metabolites from the regions “d” and “f” have weak or very weak associations with phenotype parameter and they also have insignificant contributions to multiple associations. These metabolites do not contribute to the association to the phenotype parameter. In brief, in this article, metabolites from regions “a” and “c” were considered as best set of metabolites associated with the respective phenotype parameter.
Multiple linear regression analyses were used to assess the strength of associations of one of three BFD parameters: A:G ratio, VAT, and SAT, with different sets of phenotype and metabolic variables (Fig. 1). Four different sets including phenotype and metabolic parameters were tested:

- a. set 1: age, total fat, LBM, waist circumference, DBP and HR (p = 6)—the set of the basic phenotypic and cardiovascular parameters
- b. set 2: age and all variables from the phenotype data set, excluding android fat, gynoid fat, % VAT, and % SAT (p = 24)—the extensive set of the phenotypic, cardiovascular, and body fat distribution parameters
- c. set 3: all variables of set 2 plus clinical parameters from the metabolic data set (p = 36)—the set containing set 2 with clinical parameters
- d. set 4: all variables of set 2 and all variables from the metabolic data set (p = 201)—the set containing set 3 with metabolomics parameters

Multiple linear regression analyses were performed separately for each BFD parameter (A:G ratio, VAT, and SAT) and for each gender. They consisted of three different types of analyses: multiple linear regression analyses with stepwise variable selection from four different sets of variables (i.e., set 1, set 2, set 3, and set 4), simple linear regression analysis with only one variable at a time (i.e., separate analysis for each of 201 variables), and final multiple regression analyses with variables selected based on the results from the two previous analyses. First and second types of analyses were performed on a calibration set (containing 75% of all samples) and were used to select the best variables from each of four sets for each phenotype parameter. The following criteria for variables selection were used: both the relevance score obtained during the first analysis had to be above 20% as well as the percentage of explained variance, and significance of standardized regression coefficient of models (the second analysis) had to be higher than 0.1. The only exception to these criteria was the selection of variables from set 1 for regression models of VAT for men. In this case, none of variables met criterion of R^2 > 0.1 and all variables meeting the criterion of the relevance score > 20% were selected for the final model. The final multiple regression analysis was performed on independent test set (containing 25% of all samples) and was used to select the overall best set of variables for each phenotype parameter.

In the first analysis, multiple linear regression analysis with forward stepwise variable selection was performed within a double-cross validation scheme with 7-fold single-cross validation and 8-fold double-cross validation (Smit et al., 2007) (in-house written routine available on request). The analysis was repeated 30 times with different assignments of samples into validation, calibration, and test sets. In the first cross validation loop, initial conditions of analysis were optimized by selecting one optimal initial variable to which other variables were added in further steps of stepwise variable selection (forward selection). In the stepwise variable selection, inclusion and exclusion criteria were set at p value = 0.05. In the second cross validation loop, performance of the regression model optimized in the first cross validation loop was assessed by a test set giving Root-Mean-Square Error of Cross-Validation (RMSECV). The number of times that each variable was present in multiple linear regression models (models optimized in the first cross validation loop of performed linear regression analysis) was calculated and referred to as the relevance score. A relevance score of 100% means that the variable was always selected in stepwise variable selection and is present in all models optimized in the single-cross validation loop.

In the second analysis, simple linear regression analyses were performed for each parameter: A:G ratio, VAT, and SAT, including one independent variable at a time (separate analysis for each of 201 variables tested). Percent of explained variance (R^2), standardized regression coefficients and their p values were used to assess importance of each variable in describing: A:G ratio, VAT, and SAT.

In the third analysis, final multiple regression models for A:G ratio, VAT, and SAT were obtained for an independent set of samples, including subsets of variables from sets 1–4 selected in the first and the second analysis. Performance of each regression model was assessed by R^2 value adjusted by number of variables included in this model.

Results

Subject characteristics

In the current study, all 83 women and 32 men were apparently healthy yet had increased waist circumference and an overweight or obese status. All participants were included regardless of whether they followed the green tea extract treatment or placebo treatment, because the green tea extract had no statistically significant effect on any plasma lipids (Jansen et al., 2011), lipoproteins, small molecular-weight metabolites, or phenotype parameters (internal communication). For each subject, parameters derived from two blood samples and two measures of phenotype parameters collected before and after treatment were taken into account. By including both samples and measures of each subject, more robust associations between phenotype parameters and metabolic parameters could be obtained, because high intra-subject variation in metabolic profiles (variation between before and after treatment samples from the same subject) not related to phenotype are taken into account when phenotype-metabolite associations are estimated. The characteristics of the subjects are presented in detail in Supplementary Table S1. Complete data sets were available for 215 (out of 230) samples/measures.

Biological parameter selection

All phenotype parameters (28 parameters) were tested on associations with metabolic parameters. However, the main focus was given to associations of parameters describing body fat distribution, especially A:G ratio, VAT, and SAT, because they have been associated with different metabolic activities related to cardiovascular risk (Thomas et al., 2011).

Gender differences

Results of principal component analysis showed major trends in the phenotype data set and in the metabolic data set (Fig. 2A, Fig. 2B, and on studies.nmcdsp.org/06E6-P). Women and men could clearly be separated by means of their phenotype. PC1 and PC2 explained more than 60% of the variation. In contrast, the main trend in the metabolite profiles
was only partially related to a gender difference. PC1 and PC2 explained only 39% of the variation.

Canonical correlation analysis revealed a strong significant association of 0.85 between the phenotype and metabolic data sets (Fig. 2C). The association between these data sets was gender dependent (i.e., canonical correlation analysis provided evidence of a systematic difference between women and men) in both datasets. This result necessitated separate analyses of phenotype–metabolite associations for women and men. In addition, gender differences based on the median concentrations of single parameters were determined to interpret the phenotype–metabolite associations. Clear gender-related differences were found in global and regional body fat distribution parameters (Fig. 3, Supplementary Table S1, and Supplementary Fig. S1) and most of them were statistically significant (22 out of 28 phenotype parameters were statistically significant). Male subjects had higher levels of insulin, HOMA-IR, total TG, epinephrine, cholesterol/HDL–cholesterol ratio, VLDL’s, creatinine, and all three branched-chain amino acids (BCAA: valine, isoleucine, and leucine). In comparison, female subjects had higher levels of FFA (total), glycerol, HDL-cholesterol, HDL’s (very large and large HDL), acetate, and 3-hydroxybutyrate. From the lipid profiles, it became apparent that men had generally higher levels of LPCs and TGs, whereas women showed higher levels of certain SMs and PCs.

**Correlations with BFD: (Independent) bivariate associations**

Spearman correlation coefficients between several BFD parameters and all metabolites were calculated separately for women and men. Table 1 lists all metabolites with absolute values of Spearman correlation coefficients higher than 0.3. In general, different associations were found in women and men and only a limited number of associations were common. In women, more and generally stronger correlations were observed than in men and their statistical significance is higher due to bigger size of data set for women (n = 157) than for men (n = 58). Overall, similar correlations were found to android fat, A:G ratio, VAT, waist circumference, and LBM in women, while the correlations in men were more specific to each BFD parameter. For example, insulin and HOMA-IR were strongly positively correlated to android fat, A:G ratio, VAT, waist circumference, and LBM in women, while these clinical markers were only correlated to android fat in men. Similarly, total TG was positively correlated to android fat, A:G ratio, visceral fat, and waist circumference in women, while it was only correlated to visceral fat in men. Other examples can be seen in Table 1.

Furthermore, we tested whether the bivariate correlations were specific or dependent on certain influencing factors such as age, total fat, and waist circumference. In Table 1, the letters indicate the phenotype parameter on which the correlation is independent. For example, in women the bivariate correlation of SAT with LPC 18:2 is independent on age, DSP, HR, but is dependent on total fat, LBM, waist circumference, A:G ratio, and VAT. Overall, most correlations were dependent on waist circumference in women. In contrast, the correlations found in men were mostly independent on waist circumference and other factors, meaning that these associations were unique and specific for a given pair of phenotype and metabolite. It is also notable that the correlations to A:G ratio were more specific than to android fat in women, possibly resulting from negative and positive correlations to gynoid fat and android fat, respectively, that were at the borderline of being statistically significant.

**Correlations with BFD: Association patterns**

Bivariate and multivariate phenotype–metabolite associations were determined with Spearman correlation analysis and PLS regression analysis, respectively, as described in the Material and Methods section. The results for both analyses were combined in a single plot of an association pattern to determine the best set of metabolites that is associated to A:G ratio, VAT, and SAT, respectively. In brief, we have considered metabolites with bivariate associations stronger than R = 0.3 and with statistically significant contribution to multivariate associations (p < 0.05) as members of the best set (regions ‘a’ and ‘c’ in Fig. 4; more details in Methods section).

In women, the A:G ratio could best be described by certain triacylglycerols (TG 48:0-3, 50:1-4, TG 52:2-3), total TG, PC 38:3, VLDL (M and (X)L), isoleucine, insulin, and HOMA-IR. These associations were positive and mostly determined by the android fat. A similar set of metabolites was found for VAT. This set of metabolites included in addition positive correlations of cholesterol, VLDL (S) and TG 54:2-3 and negative correlations of HDL (L, XL). On the other hand, PC 38:3 and TG 50:1-4 did not significantly contribute to the association pattern of VAT. In comparison, the association pattern of SAT was substantially different from those of the A:G ratio and VAT. It was mostly determined by the negative correlations of LPC 18:2-3, LPC 0-16:1, and HDL (XL) and the positive correlation of cholesterol:HDL ratio.

In men, the optimal set of metabolites that is associated to the A:G ratio included certain negative correlations with phosphatidylecholines (PC O-34:2-3, PC O-36:2-3), and positive correlations with triacylglycerols (TG 51:1, TG 51:3), sphingomyelins (SM d18:1/17:0, SM d18:1/20:1), and alanine. These associations were typical for the A:G ratio and not reflected in associations to android or gynoid fat. A specific association pattern was also observed for VAT including total TG, TG 50:1-5, TG 55:1, PC 32:0, and VLDL ((X)L). In comparison to women, cholesterol, insulin, HOMA-IR, and HDL were not correlated to VAT, but to android fat in men. No relevant set of metabolites could be determined for SAT in men.

**Added value of metabolomics analysis**

The performances of multiple regression models of A:G ratio, VAT, and SAT are shown in Table 2. For each BFD parameter, four gender-specific models were established based on independent set of samples and parameters selected from four different sets of phenotype parameters and metabolites (set 1, set 2, set 3, and set 4) as described in Methods section. All multiple regression models were obtained on independent set of samples and with strict variable selection criteria to avoid selection bias. More details are presented in Supplementary Tables S2 and S3.
In women, the multiple regression models based on set 1 or 2 (set 1 included the basic phenotypic and cardiovascular parameters \( p < 0.05 \), and set 2 included all phenotype parameters \( p = 0.1 \) ) had similar performance when compared to models based on set 3 and 4 (which also included clinical and metabolomic parameters, respectively). For our study, this means that phenotype parameters in women were able to predict best the A:G ratio, VAT, and SAT, and that metabolomics and clinical parameters had no added value in describing those body fat distribution parameters. More specifically, the waist circumference, together with LBM and total fat, were predictive for the A:G ratio and for VAT and SAT, respectively.

In men, statistically significant multiple regression models were only obtained for VAT and only when including clinical parameters and metabolomics data (sets 3 and 4). Interestingly, the VAT model including metabolomics data clearly outperformed the other models that did not include metabolomics data (increase from \( R^2 = 0.48 \) for set 3 to \( R^2 = 0.78 \) for set 4 selection; see Table 2 for explanation). Four parameters, namely PC 32:0, acetate, insulin, and HOMA-IR were predictive for VAT with a performance of 78%, suggesting that these four parameters together were the most powerful predictors of VAT in men. The predictability of the metabolites based on both the relevance scores and the significance levels were 60%, 69.6%, 28.8%, and 27.5% for PC 32:0, acetate, insulin, and HOMA-IR, respectively (see Supplementary Table 3 for significance levels). Insulin and HOMA-IR did not have strong bivariate or multivariate associations with VAT and had low relevance scores. Their presence in the regression model of set 4 can be explained by two influential subjects (two outliers in Supplementary Fig. S2). One group of subjects had low acetate concentrations, which correlated with a wide range of VAT sized, whereas the other group had high acetate concentrations that correlated only to low VAT sized. This nonlinear relationship was only observed for acetate and VAT in men, and the underlying reason is unclear. We speculate that subjects with high acetate levels have a relatively higher lipid turnover in VAT (small VAT size), considering that acetate is a substrate/product in lipid turnover (Wajchenberg, 2000). An alternative explanation may be related to a differential microbiome in men with different VAT size (Flint, 2011).

In the current study, we identified the most important sets of metabolites associated with A:G ratio, VAT, and SAT, which may be related to differential risk in cardiovascular disease (CVD) and type 2 diabetes mellitus. In addition, we assessed the added value of metabolomics analysis in the presence of other parameters including BFD parameters and clinical markers. The particular strength of our study relied on detailed measures of BFD (including MRI and DEXA measurements) and plasma/serum metabolites (205 different parameters in total) as well as on a relatively large sample size (215 samples in total). Furthermore, we used a sophisticated data analysis strategy including careful and strict validation procedures.

**Table 2.** \( R^2 \) of Multiple Linear Regression Models Adjusted for Variables.

<table>
<thead>
<tr>
<th>Set/ Parameter</th>
<th>Women ( (n = 39) )</th>
<th>Men ( (n = 15) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A:G ratio</td>
<td>VAT</td>
</tr>
<tr>
<td>Set 1 selection</td>
<td>0.420*</td>
<td>0.498*</td>
</tr>
<tr>
<td>Set 2 selection</td>
<td>0.445*</td>
<td>0.492*</td>
</tr>
<tr>
<td>Set 3 selection</td>
<td>0.447*</td>
<td>0.522*</td>
</tr>
<tr>
<td>Set 4 selection</td>
<td>0.451*</td>
<td>0.526*</td>
</tr>
</tbody>
</table>

The number of variables were selected by stepwise approach from four different subsets of data (set 1–4) and included in multiple regression models of A:G ratio, VAT, and SAT separately for women and men. Statistically significant models for each BFD parameter \( (p < 0.05) \) are indicated by *.

\( ^{\ast} \) Selection of variables from set 1 composed of age, total fat, LBM, DBP, HR, and waist circumference.

\( ^{\ast} \) Selection of variables from set 1 composed of age and variables of phenotype data set excluding: android fat, gynoid fat, \%SAT, and \%VAT.

\( ^{\ast} \) Selection of variables from set 3 composed of set 2 and clinical parameters.

\( ^{\ast} \) Selection of variables from set 3 composed of set 2 and clinical parameters.

\( ^{\ast} \) Selection of variables from set 4 composed of set 3, lipidomics, lipoproteins, and LMWM variables.

\( ^{\ast} \) Selection of variables from set 4 composed of set 3, lipidomics, lipoproteins, and LMWM variables.

\( ^{\ast} \) Set composed of LBM and waist circumference.

\( ^{\ast} \) Set composed of total fat and waist circumference.

\( ^{\ast} \) Set composed of total fat and waist circumference.

\( ^{\ast} \) Set composed of proximal thigh circumference and distal thigh circumference.

\( ^{\ast} \) Set composed of total fat, proximal thigh circumference, triceps circumference, PC O-34:2, and PC O-36:5.

\( ^{\ast} \) Set composed of PC 32:0, acetate, insulin, and HOMA-IR.
Gender differences

Significant differences in metabolite profiles and phenotypic parameters were observed between men and women, and thus a separate analysis increased the statistical power. Most of these differences are in agreement with earlier findings (Seidell et al., 1991; Wajchenberg, 2000). For instance, the higher gynoid fat levels in women and the higher A-G ratios in men go along with the higher prevalence of female ‘pears’ and male ‘apples’, respectively (Wajchenberg, 2000). The gender differences in clinical markers generally reflect a more favorable metabolic risk profile for women, including lower levels of insulin and triacylglycerols and higher levels of HDL-cholesterol (Lemieux et al., 1994; Seidell et al., 1991). In addition, women have lower levels of branched-chain amino acids that have been shown to be positively associated with insulin resistance (Adeva et al., 2011).

The gender differences observed by metabolomics analysis generally corroborate evidence from the literature. The higher levels of most sphingomyelins in women, which has also been found in another study (Mittelstrass et al., 2011), may be related to differences in insulin resistance (Blachnio-Zabiełska et al., 2011; Straczkowski et al., 2004). For instance, ceramides are produced in the sphingomyelin signaling pathway by de novo synthesis or by sphingomyelin hydrolysis. In adipose tissue, ceramides have been shown to correlate negatively with HOMA-IR and higher activity levels of enzymes involved in de novo sphingomyelin production, and lower activity levels of enzymes involved in sphingomyelin hydrolysis have been found in obese subjects (Blachnio-Zabiełska et al., 2011). In muscle, ceramides have been shown to correlate negatively with insulin sensitivity independent of obesity and positively to the enzyme involved in sphingomyelin hydrolysis (Straczkowski et al., 2004). This shows the importance of the sphingomyelin pathway in insulin sensitivity and the complexity of the regulation of this pathway. Insulin may also cause the higher level of PC 38:3 (possibly corresponding to PC(18:0-20:3) (Mahadevappa and Holub, 1984) in men, as some studies have shown a relationship between the type of fatty acids in phosphatidylcholines and insulin sensitivity (Clöre et al., 1998).

Correlations with BFD

Our systematic correlation analysis revealed significant differences in the association patterns between the BFD parameters. Generally, SAT is a well-studied tissue due to its easy accessibility. However, it is not the most interesting tissue, because SAT correlates less strongly with metabolic risks factors than VAT, as has been demonstrated in the Framingham heart study (Fox et al., 2007). In addition, VAT has been shown to correlate with insulin resistance, whereas SAT is protective against insulin resistance, which has been shown by analysis of VAT and SAT in insulin-resistant and insulin-sensitive obese subjects (women and men) by adjusting the relation between SAT and insulin for BMI and VAT (McLaughlin et al., 2011). In our study, insulin and HOMA-IR were positively correlated to VAT, but not to SAT, corroborating the literature observations. Furthermore, the VLDL subclasses were positively associated to VAT which is in agreement with previous studies showing negative correlations of VLDL with insulin resistance (Garvey et al., 2003) and positive correlations with VAT in men (Okazaki et al., 2005).

Interestingly, VLDL and TG consistently showed similar correlations, probably because TGs are mainly transported by VLDL particles in blood (Vance and Vance, 2008). In particular, several TGs, including TG 50:1, TG 50:2, TG 53:2, TG 54:1 and TG 54:7, were clearly associated to VAT and not to SAT, suggesting that SAT is less sensitive to lipolytic stimuli (Ibrahim, 2010; Williams, 2004). Hormone-sensitive lipase has a preference for triacylglycerols with longer chain fatty acids as substrate (Haemmerle et al., 2002) and VAT has been shown to be highly lipolytic (Ibrahim, 2010; Williams, 2004). In contrast to VLDL, large (XL-L) HDL subclasses have generally been shown to correlate negatively with VAT (Lemieux et al., 1994; Okazaki et al., 2005) and positively with insulin resistance (Garvey et al., 2003). In our study, we found weak correlations between HDL (XL-L) subclasses and VAT. Remarkably, these relations with VAT were only clearly observed in women and largely coincided with those found for android fat, A:G ratio, waist circumference, and LBMI. In men, however, a similar association pattern (correlations between lipoproteins and BFD) was only found for android fat with the exception of VLDL (M, XL-L) and certain TGs that were clearly associated to VAT and not to other BFD. This indicates that metabolite profiles in men were more specific to certain BFD parameters when compared to women, suggesting that the underlying metabolic processes in women were similar, while in men they seemed to depend on specific fat deposits. Please note, however, that the number of men with metabolic syndrome was higher in our study (48% men and 14% women). Therefore, the tissue specificity in men may be confounded by effects resulting from metabolic syndrome.

Interestingly, our metabolomics approach revealed correlations that have not been reported before and mostly consisted of TGs and phosphatidylcholines (including PC-O, PC, LPC, and LPC-O). In a recent study, certain LPCs (14:0, 18:0, and 18:1) have been correlated to obesity (Kim et al., 2010). However, none of these LPCs were associated with the BFD parameters in our study. In our study, SAT in women was negatively correlated to PCs (PC O-36:3; PC O-34:1) and LPCs (LPC O-16:1, LPC 18:2-3) that are downstream metabolites of PCs. Similarly, negative correlations between SAT and PCs (PC, PC-O) have been found in a study with 24 monozygotic twins (Pietiläinen et al., 2007). In this study, however, these lipid components were also negatively correlated to VAT and involved different species. In our study, the respective PCs had a low number of double bonds, whereas in the study by Pietiläinen and associates, the respective PCs had a high number of double bonds. The discrepancy may be explained by the higher number of subjects in our study as well as differences in the regulation of PC production pathways. There are two different pathways for the formation of PCs with an enzyme specificity depending on the fatty acid species (Pynn et al., 2011).

More specifically, a positive correlation between PC 32:0 and VAT was observed in men, which may result from the relative high levels in saturated fatty acids in obese subjects (Phinney, 2005). Moreover, certain unsaturated plasmalogens (PC-O 34:1-3, PC O-36:2,3,5) were negatively correlated to A:G ratio in men and to SAT in women. Plasmalogens influence the fluidity of lipids (Pietiläinen et al., 2011), which may be of importance considering the cell stress in adipocytes of obese subjects. Earlier relations between obesity and plasmalogens have shown that plasmalogens are higher in SAT in
obese subjects (Pietiläinen et al., 2011) and that plasma levels of unsaturated plasmalogens are lower in hypertensive sub-
jects (Graessler et al., 2009).

**Added value of metabolomics analysis**

Our multiple regression models could explain up to 78% of
the variation in A:G ratio, VAT, and SAT. The models ob-
tained for women and men were different in their statistical
significance (probably related to the larger number of women
in the study). In females, the waist circumference and LBM
already gave reasonable estimates for the prediction of A:G
ratio, and waist circumference and total fat were predictive
for VAT and SAT. Clinical markers nicely complemented
these predictors. Metabolites obtained by metabolomics
analysis (lipidomics, lipoprotein profiling, and LMWM pro-
file) did not contribute to the improvement of the models.
In comparison to women, phenotypic parameters as well as
clinical markers did not result in predictive models for the
A:G ratio, VAT, and SAT in men. In particular, the waist cir-
cumference was only a weak or even no predictor for these
BFD parameters in men, which is in contrast to literature
(Klein et al., 2007). This may be explained by the higher av-
erage muscle mass in men, possibly obscuring the relations
between phenotypic and BFD parameters. Only when meta-
bolomics data were included, the prediction of the models
could be significantly improved. The model for VAT was
significant and the model for A:G ratio was on the borderline
of significance. Remarkably, the model of VAT excluded
waist circumference, BMI, and related phenotype parameters.
Specific phosphatidylcholine and triacylglycerol components
were important determinants in both models and in the bi-
ivariate associations. Phosphatidylcholine may be a source
of fatty acids for triacylglycerols (van der Veen et al., 2012),
which may explain why they often co-occur in the models. PC
32:0 was included in the model of VAT, whereas PC-O 36:5
and PC O-34:2 were included in the model of A:G ratio,
possibly indicating tissue and fatty acid specific enzyme ac-
tivity (Pynn et al., 2011; Raclo et al., 2001). For example, en-
zymes converting esterified fatty acids are known to vary
between tissues (Kotronen et al., 2010). PC 32:0 is a new po-
tential indicator of VAT size and abdominal obesity, as PC
32:0 has been shown earlier to be related to obesity (Öberbach
et al., 2011). It has also been shown that PC levels in plasma
are related to hepatic steatosis (van der Veen et al., 2012).
The main enzyme responsible for PC clearance from the blood is
lecithin-cholesterol acyltransferase (LCAT), an enzyme with a
substrate prevalence for certain PC species (Subbaiah and
Monshizadegan, 1988). A positive relation between LCAT
and obesity has been described (Sutherland et al., 1979) and
LCAT-null mice have been shown to be protected from diet-
duced obesity (Li et al., 2011). Therefore, the relative high
level of LCAT in obese subjects may be responsible for
clearance of specific PC species. In addition, relative high
levels of PCs in adipose tissues compared to other tissues may
explain the relation of PCs to BFD (Kotronen et al., 2010).

**Conclusion**

In summary, our findings demonstrate significant gender-
related differences in the metabolite patterns associated to
various BFD parameters, which is in line with a recent meta-
analysis that revealed sexual dimorphism in the genetic basis
of fat distribution (Heid et al., 2010). The consistent associa-
tion patterns observed in women suggest that VAT, android
fat, A:G ratio, and waist circumference are mainly related to
well-known parameters, such as insulin, VLDL, HDL, and TG.
Interestingly, lipidomic analysis identified certain phospha-
tidylcholine species that were negatively correlated to SAT. In
comparison to women, the metabolite patterns were more
depot-specific in men. Multiple regression analysis revealed a
predictive metabolite, namely PC 32:0, for VAT in healthy
overweight men. Plasma PCs are related to storage of TG,
which may explain why PC in plasma can be predictive for
VAT. Considering that TGs and PCs are involved in obesity-
related disease, PC32:0 may be an early marker for diabetes
and CVD. However, future studies still need to confirm the
predictability of PC32:0 in similar populations. In addition,
lipidomic analysis in fat tissues should provide further evi-
dence of the underlying metabolism.

**Acknowledgments**

Robert van der Berg (GSK Biologicals) is acknowledged for
providing matlab codes for canonical correlation analysis.
This project was (co)funded by the Netherlands Metabo-
lemics (NMC) which is part of the Netherlands Genomics
Initiative/Netherlands Organisation for Scientific Research.

**Author contributions:** Design and supervision of the study:
DM, JPMvD. Design and supervision of the original study:
IM, DJM. Phenotypic and clinical measurements: IM, DJM.
Lipidomics profiling: KS, RJV. Lipoprotein profiling: JV, AJK,
PS, MAK. Data analysis: ES, JW, AKS. Biological interpreta-
tion: JB. The article was written by ES, JB, and DMJ.

**Author Disclosure Statement**

The authors declare that no competing financial interests
exist.

**References**

resistance and the metabolism of branched-chain amino acids

Ala-Korpela M. (2008). Critical evaluation of 1H NMR metabo-
lomics of serum as a methodology for disease risk assessment

Ceramide metabolism is affected by obesity and diabetes in

dylcholine fatty acids and insulin sensitivity in normal hu-

plasma ceramides, inflammation and insulin resistance: As-
sociation with serum IL-6 concentration in patients with cor-

Despres JP, Moorjani S, Lupien PJ, et al. (1990). Regional dis-
tribution of body fat, plasma lipoproteins, and cardiovascular

Durnin JV, and Womersley J. (1974). Body fat assessed from total
body density and its estimation from skinfold thickness: Mea-
surements on 481 men and women aged from 16 to 72

Flint HJ. (2011). Obesity and the gut microbiota. J Clin Gastro-
enterol 45, S128–S132.


Address correspondence to:
Doris M. Jacobs
Unilever R&D
Olivier van Noortlaan 120
3130 AC Vlaardingen
The Netherlands

E-mail: doris.jacobs@unilever.com