

## SUPPLEMENTAL MATERIAL

### **Associations between adipose tissue volume and small molecules in plasma and urine among asymptomatic subjects from the general population**

Lerina Otto, Kathrin Budde, Gabi Kastenmüller, Anne Kaul, Uwe Völker, Henry Völzke, Jerzy Adamski, Jens P. Kühn, Jan Krumsiek, Anna Artati, Matthias Nauck, Nele Friedrich, Maik Pietzner

#### **Material and Methods**

##### *Metabolomics Measurements*

Non-targeted metabolomics analysis for metabolic profiling was conducted at the Genome Analysis Center, Helmholtz Zentrum München. Two separate LC-MS/MS analytical methods were used as previously published, i.e. in positive and in negative ionization modes, were used to detect a broad metabolite panel (1). In this study, samples were divided into two sets according to the biological matrices of the samples, i.e. plasma and urine. On the day of extraction, samples were thawed on ice. A 100µL of the sample were pipetted into a 2mL 96-well plate. In addition to study samples, a human pooled reference plasma sample (Seralab, West Sussex, United Kingdom) and another pooled reference matrix of each sample set (Seralab, West Sussex, United Kingdom) were extracted and placed in 1 and 6 wells, respectively, of the 96-well plate. These samples served as technical replicates throughout the data set to assess process variability. Beside those samples, 100µL of water was extracted as samples and placed in 6 wells of the 96-well plate to serve as process blanks. Protein was precipitated and the metabolites were extracted with 475µL methanol, containing four recovery standards to monitor the extraction efficiency. After centrifugation, the supernatant was split into 4 aliquots of 100µL each onto two 96-well microplates. The first 2 aliquots were used for LC-MS/MS analysis in positive and negative electrospray ionization mode. Two further aliquots were kept as a reserve. The extracts were dried on a TurboVap 96 (Zymark, Sotax, Lörrach, Germany). Prior to LC-MS/MS in positive ion mode, the samples were reconstituted with 0.1% formic acid (50µl for plasma, 100µl for urine). Whereas samples analyzed in negative ion mode were reconstituted with 6.5mM ammonium bicarbonate (50µl for plasma, 100µl for urine), pH 8.0. Reconstitution solvents for both ionization modes contained internal standards that allowed monitoring of instrument performance and also served as retention reference markers. To minimize human error, liquid handling was performed on a Hamilton Microlab STAR robot (Hamilton Bonaduz AG, Bonaduz, Switzerland). LC-MS/MS analysis was performed on a linear ion trap LTQ XL mass spectrometer (Thermo Fisher Scientific GmbH, Dreieich, Germany) coupled with a Waters Acquity UPLC system (Waters GmbH, Eschborn, Germany). Two separate columns (2.1 x 100 mm Waters BEH C18, 1.7 µm particle-size) were used either for acidic (solvent A: 0.1% formic acid in water, solvent B: 0.1% formic acid in methanol) and or for basic (A: 6.5mM ammonium bicarbonate, pH 8.0, B: 6.5mM ammonium bicarbonate in 95%

methanol) mobile phase conditions, optimized for positive and negative electrospray ionization, respectively. After injection of the sample extracts, the columns were developed in a gradient of 99.5% A to 98% B over an 11 min run time at 350 $\mu$ L/min flow rate. The eluent flow was directly run through the ESI source of the LTQ XL mass spectrometer. The mass spectrometer analysis alternated between MS and data-dependent MS/MS scans using dynamic exclusion and the scan range was from 80-1000 m/z. Metabolites were identified by Metabolon, Inc. from the LC-MS/MS data by automated multiparametric comparison with a proprietary library, containing retention times, m/z ratios, and related adduct/ fragment spectra (2). Identification criteria for the detected metabolites are described in Evans *et al.*(1). Quality control methods and normalization of metabolite levels are explained in detail in the supplement.

#### Metabolomics Measurements: Quality Control and Normalization of Metabolite Levels

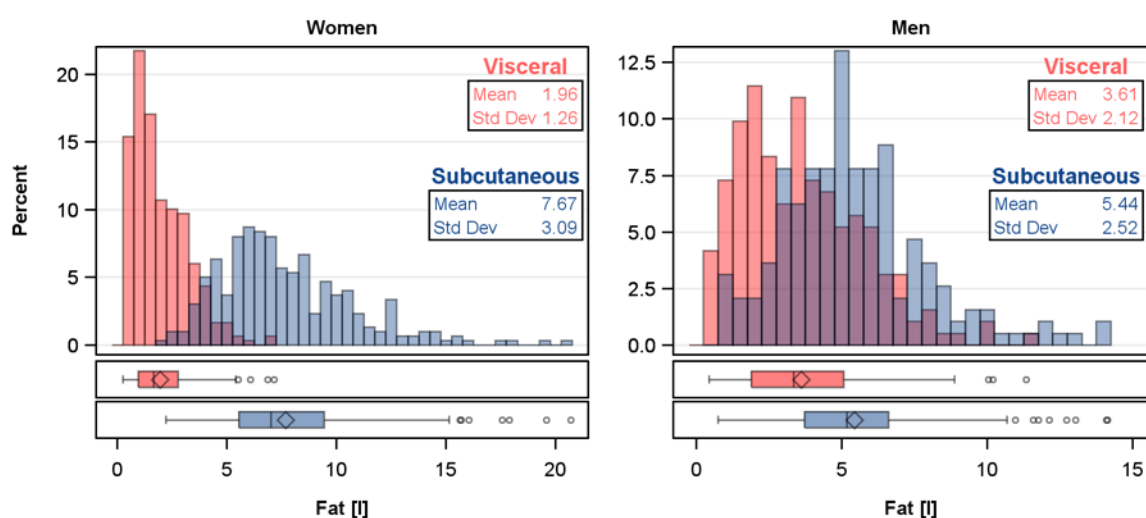
To correct for daily variations of platform performance, the raw ion count of each metabolite was rescaled by the respective median value of the run day. Valid estimation of the median was ensured by keeping only metabolites with at least three measured values on more than the half of the run days. This procedure resulted in 475 and 558 metabolites for plasma and urine, respectively, available for the present analysis. 263 metabolites were measured in both bio fluids. We chose probabilistic quotient normalization (PQN) (3) to account for diurnal variation of urine samples, since this procedure was shown to be superior to the common creatinine scaling. For this purpose we calculated a mean-pseudo-spectrum depending on metabolites with measurements for all participants (131 urine metabolites). Subsequently, we calculated a dilution factor as the median quotient between the reference spectrum and each sample. Of note, urine creatinine and the estimated dilution factor were highly correlated ( $r=0.91$ ,  $p<0.001$ ) within the present study sample. Afterwards all metabolite levels were log<sub>2</sub>-transformed. Separately for plasma and urine samples we performed multivariate outlier detection using an algorithm proposed by Filzmoser *et al.* (4) as implemented in the *pcout* function within the R package *mvoutlier*. The algorithm provides an outlier score for each sample based on a weighted combination of location and scatter estimations using principle component analysis and the Mahalanobis distance on a robustly scaled data matrix. The default parameters were used for the identification process, except the critical value for the location outliers was set to 4, as it corresponds to a 4 SD exclusion criteria. The minimum score was used as cut-off for outlier identification. As a result 13 and 8 samples from plasma and urine were excluded, respectively. Metabolite levels were log<sub>2</sub>-transformed showing a clear improvement towards normality. To avoid artificial results in linear regression analysis from single outlier, additional univariate outliers for each metabolite were excluded based on the common mean plus 3SD rule.

#### Calculation of Gaussian graphical models

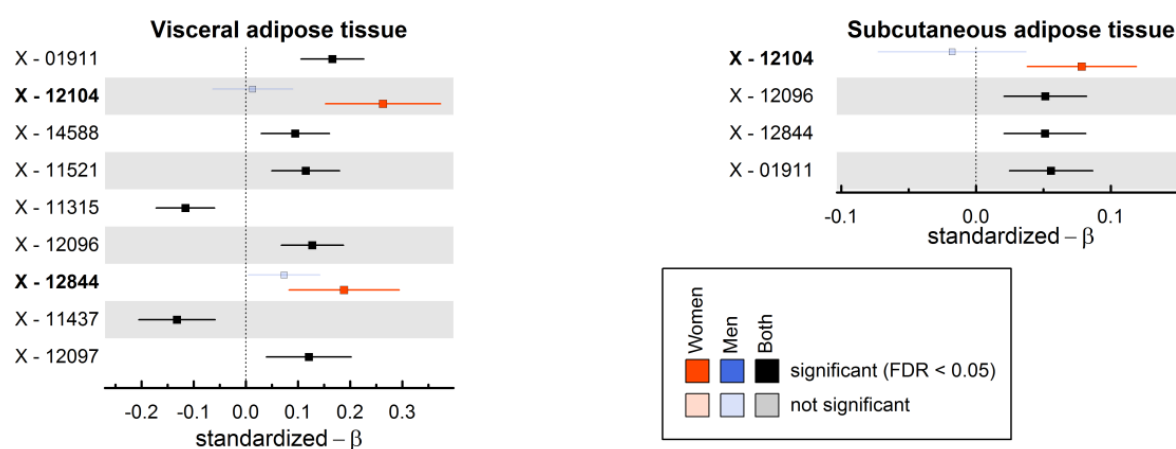
To guide interpretation of our metabolome-wide association results, we calculated Gaussian graphical models (GGMs) for each of the bio fluids because of their ability to mirror

physiological dependencies between metabolites. The procedure was extensively discussed previously (5-7). Briefly, GGMs are based on partial correlations, which represent the correlation between two metabolites correcting for all remaining metabolites. We additionally included age, sex and BMI as covariates to account for major confounding factors. Edges in the GGMs were declared significant if both partial and Pearson correlation were significant at  $\alpha = 0.05$  after Bonferroni correction for all possible edges (correcting for  $\binom{p}{2}$  tests, where  $p$  is the number of metabolites). Since GGM calculation requires a full data matrix, imputation of missing values was performed. The influence of the imputation was minimized by exclusion of all metabolites with more than 20% missing values, resulting in 263 and 399 metabolites in plasma and urine, respectively. Assuming missing values mainly due to low concentrations of metabolites, the distribution of each metabolite on a run day could be estimated as a left-censored log-normal distribution prior normalization. Hence, we reconstructed these distributions based on maximum likelihood estimation and sampled missing values from the censored part of the distribution. This procedure was only applied for metabolites with at least ten observations on the specific run day. Remaining missing values were imputed by multiple chained equations using the R package mice. Both approaches are expected to rather lower (truncated sampling) or maintain (mice, using predictive mean matching) correlations between metabolites than falsely increasing them. The final GGM comprised 732 nodes and 930 edges. Subsequently, results from linear regression analyses were mapped on the graph to visually inspect altered clusters of metabolites.

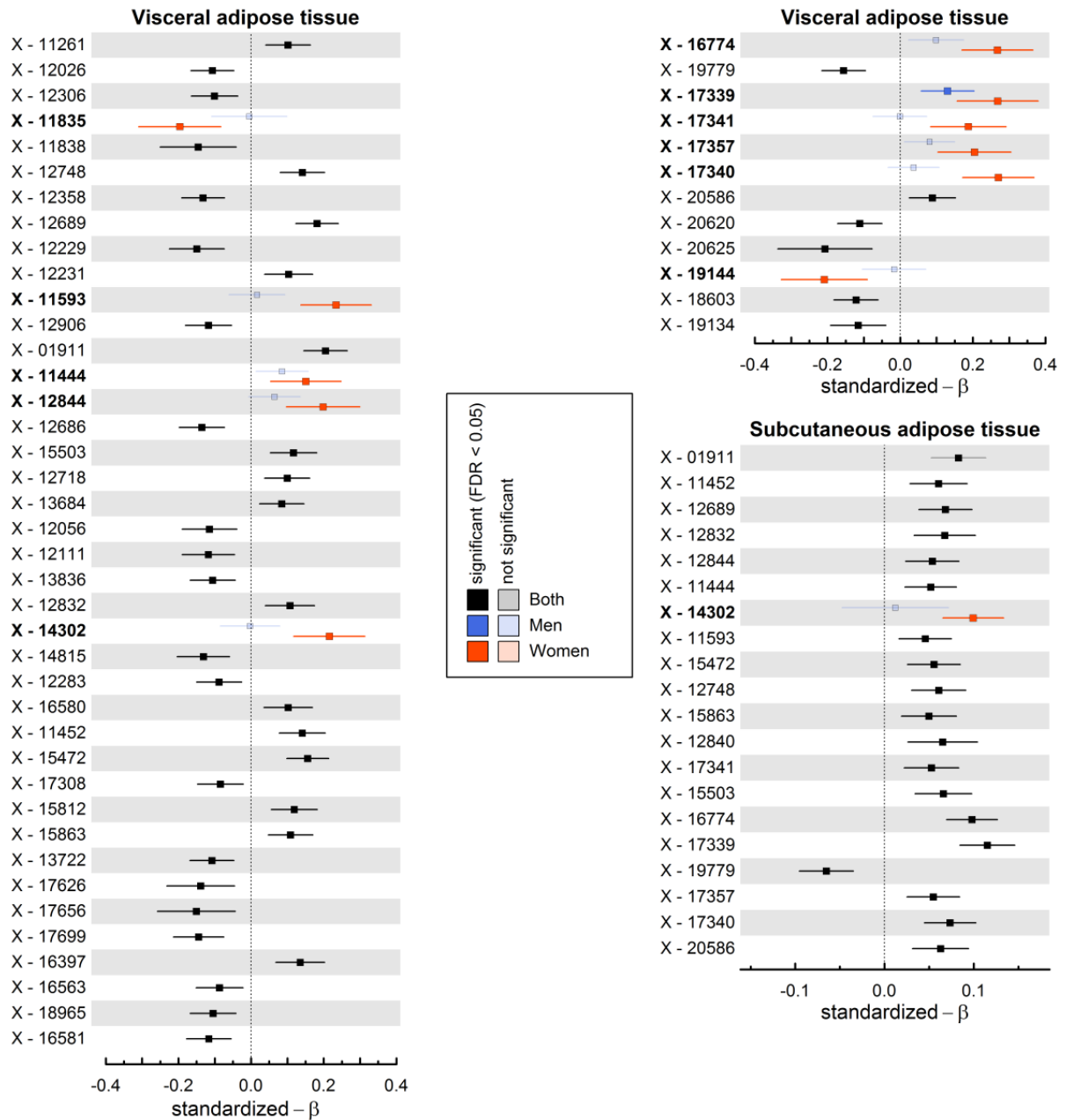
## Figures



**Figure S1** Histograms and boxplots of visceral and subcutaneous fat by sex.



**Figure S2** Standardized  $\beta$ -estimates from linear regression analysis with the amount of visceral (VAT; left panel) or subcutaneous (SAT; right panel) as exposure and unknown plasma metabolites as outcome conducting either the whole population (black), only men (blue) or women (orange). Displayed are only metabolites which were significant (controlling the false discovery rate (FDR) at 5%) in at least one of the subsets (indicated by darker colors). Metabolites printed in bold showed a nominal significant ( $p < 0.05$ ) interaction term between VAT or SAT and sex.



**Figure S3** Standardized  $\beta$ -estimates from linear regression analysis with the amount of visceral (VAT; left panel) or subcutaneous (SAT; right panel) as exposure and unknown urine metabolites as outcome conducting either the whole population (black), only men (blue) or women (orange). Displayed are only metabolites which were significant (controlling the false discovery rate (FDR) at 5%) in at least one of the subsets (indicated by darker colors). Metabolites printed in bold showed a nominal significant ( $p < 0.05$ ) interaction term between VAT or SAT and sex.

## References

1. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Analytical chemistry*. 2009;81(16):6656-67.
2. Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, Guo L, et al. Analysis of the adult human plasma metabolome. *Pharmacogenomics*. 2008;9(4):383-97.
3. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in <sup>1</sup>H NMR metabonomics. *Analytical chemistry*. 2006;78(13):4281-90.
4. Filzmoser P, Maronna R, Werner M. Outlier identification in high dimensions. *Computational Statistics and Data Analysis*. 2008;52(3):1694-711.
5. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol*. 2011;5:21.
6. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46(6):543-50.
7. Do KT, Kastenmuller G, Mook-Kanamori DO, Yousri NA, Theis FJ, Suhre K, et al. Network-based approach for analyzing intra- and interfluid metabolite associations in human blood, urine, and saliva. *J Proteome Res*. 2015;14(2):1183-94.