

# Familial Resemblance for Serum Metabolite Concentrations

Harmen H. M. Draisma,<sup>1,2</sup> Marian Beekman,<sup>3,4</sup> René Pool,<sup>1,2</sup> Gert-Jan B. van Ommen,<sup>5</sup>  
Anika A. M. Vaarhorst,<sup>3</sup> Anton J. M. de Craen,<sup>6</sup> Gonneke Willemsen,<sup>1</sup> P. Eline Slagboom,<sup>3,4</sup>  
and Dorret I. Boomsma<sup>1,2</sup>

<sup>1</sup>Department of Biological Psychology, Faculty of Psychology and Education, VU University Amsterdam, The Netherlands

<sup>2</sup>The EMGO Institute for Health and Care Research, Amsterdam, The Netherlands

<sup>3</sup>Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

<sup>4</sup>Netherlands Consortium for Healthy Ageing, Leiden University Medical Center, Leiden, The Netherlands

<sup>5</sup>Department of Human Genetics, Center for Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

<sup>6</sup>Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands

Metabolomics is the comprehensive study of metabolites, which are the substrates, intermediate, and end products of cellular metabolism. The heritability of the concentrations of circulating metabolites bears relevance for evaluating their suitability as biomarkers for disease. We report aspects of familial resemblance for the concentrations in human serum of more than 100 metabolites, measured using a targeted metabolomics platform. Age- and sex-corrected monozygotic twin correlations, midparent–offspring regression coefficients, and spouse correlations in subjects from two independent cohorts (Netherlands Twin Register and Leiden Longevity Study) were estimated for each metabolite. In the Netherlands Twin Register subjects, who were largely fasting, we found significant monozygotic twin correlations for 121 out of 123 metabolites. Heritability was confirmed by midparent–offspring regression. For most detected metabolites, the correlations between spouses were considerably lower than those between twins, indicating a contribution of genetic effects to familial resemblance. Remarkably high heritability was observed for free carnitine (monozygotic twin correlation 0.66), for the amino acids serine (monozygotic twin correlation 0.77) and threonine (monozygotic twin correlation 0.64), and for phosphatidylcholine acyl-alkyl C40:3 (monozygotic twin correlation 0.77). For octenoylcarnitine, a consistent point estimate of approximately 0.50 was found for the spouse correlations in the two cohorts as well as for the monozygotic twin correlation, suggesting that familiarity for this metabolite is explained by shared environment. We conclude that for the majority of metabolites targeted by the used metabolomics platform, the familial resemblance of serum concentrations is largely genetic. Our results contribute to the knowledge of the heritability of fasting serum metabolite concentrations, which is relevant for biomarker research.

■ **Keywords:** monozygotic twins, spouse correlations, midparent–offspring resemblance, Biocrates platform

Metabolites are the small molecules that represent the substrates, intermediates, or end products of cellular metabolism. There is considerable interest in the familial clustering of circulating metabolite concentrations, because this is informative of the likelihood that a particular metabolite may serve as a biomarker for disease (Nicholson et al., 2011). For instance, if the level of a particular metabolite is completely heritable, then it is unlikely that this metabolite will be suitable as a biomarker for diseases that are primarily driven by environmental factors. In addition, the assessment of the heritability of metabolite levels

through modeling of family data may help to explain the heritability of metabolic diseases characterized by altered metabolite levels (Shah et al., 2009). The heritability of

RECEIVED 9 July 2013; ACCEPTED 17 July 2013. First published online 28 August 2013.

ADDRESS FOR CORRESPONDENCE: Harmen Draisma, Department of Biological Psychology, Faculty of Psychology and Education, VU University Amsterdam, Van der Boerhorststraat 1, 1081 BT Amsterdam, The Netherlands. E-mail: h.h.m.draisma@vu.nl

metabolite levels is also of interest because it harbors the contribution of all genetic variants that influence individual differences in metabolite concentration (Kettunen et al., 2012; Nicholson et al., 2011) and thus sets an upper limit to the contribution of genetic variants.

In this study, we address several aspects of familial resemblance for metabolite concentrations as measured using metabolomics techniques in human serum. Metabolomics is the study of ideally all metabolites (small molecules, typically <1 kDa) as can be found in a specimen, tissue, organ, or complete organism (Dunn et al., 2011). The aim of metabolomics is to obtain a holistic overview (snapshot) of cellular metabolism both quantitatively and qualitatively. With respect to other types of 'omics' such as transcriptomics, metabolomics is particularly interesting because metabolites are relatively close to the observable phenotype (e.g., disease). On the other hand, metabolites are still sufficiently close to the genome to provide enhanced statistical power to detect genetic effects on a phenotype compared with directly linking genotype and phenotype itself (Gieger et al., 2008; Illig et al., 2010). Metabolomics uses analytical chemical techniques such as proton nuclear magnetic resonance spectroscopy ( $^1\text{H}$  NMR) or mass spectrometry to obtain information about the identity and abundance of metabolites in a particular sample at much higher resolution (e.g., about the individual triglycerides) than would be possible using classical clinical chemistry measures (such as enzymatically determined total triglyceride level). Metabolomics analyses can be classified into targeted and untargeted (global) approaches (Griffiths et al., 2010). The targeted approaches are used to obtain good quantification of known metabolites, whereas the global approaches are used to obtain a broad overview of all metabolites (both known and unknown) that are present in a particular sample. However, given the often widely differing physicochemical characteristics of the metabolites from different metabolite classes that are present in a biological sample, no single analytical technique is able to detect and quantify all metabolites. In the current article, we describe various aspects of familiarity for serum metabolite concentrations as determined using a targeted metabolomics platform that combines sample preprocessing using the Biocrates AbsoluteIDQ p150 kit (Biocrates Life Sciences AG, Innsbruck, Austria) with metabolite detection by tandem mass spectrometry. The Biocrates sample preparation kit allows simultaneous quantification of 163 metabolites belonging to various metabolite classes that are part of key metabolic pathways (for more information, see e.g., Berg et al., 2012; Vance & Vance, 2008; Zhai et al., 2010). In brief, acylcarnitines are acyl esters of carnitine, which allows the transport of fatty acids over the mitochondrial membrane for beta-oxidation. Amino acids are building blocks for protein synthesis and several amino acids can also be used as energy sources during gluconeogenesis. The phosphatidylcholines are important building blocks of cell membranes

and of the outer membrane of lipoprotein particles, as well as the precursors of lysophosphatidylcholines and sphingomyelins. Lysophosphatidylcholines in blood are formed from phosphatidylcholines by the liver enzyme lecithin-cholesterol transferase. These pro-inflammatory lipids are found in oxidized low-density lipoprotein (LDL) particles (Wu et al., 1998). Sphingomyelins are very long-chain structural analogues of phosphatidylcholines that are important for signal transduction. Finally, the Biocrates AbsoluteIDQ p150 kit allows for the quantification of hexose, which is ~90–95% glucose (Goek et al., 2012) and this is an important source of energy for the body. Indeed, metabolites as detected using this sample preprocessing kit have been associated previously with various diseases and other conditions (see, e.g., Floegel et al., 2013; He et al., 2012; Jourdan et al., 2012; Mittelstrass et al., 2011; Wang-Sattler et al., 2012; Xu et al., 2013; Yu et al., 2012).

Several authors report familial resemblance for metabolites detected in blood using metabolomics platforms. For instance, Menni et al. computed monozygotic (MZ) and dizygotic (DZ) twin correlations for 11 (four acylcarnitines, six phosphatidylcholines, and one sphingomyelin) and estimated heritability for nine serum metabolites that showed significant association with dietary variables in female twins from the TwinsUK cohort (Menni et al., 2013). These authors used as the basis for their study metabolomics data obtained using the same Biocrates sample preprocessing kit as we do in the current study, and also employed mass spectrometry to detect metabolites. Twins from the TwinsUK cohort were also included in the study by Nicholson et al. (2011), who employed a longitudinal study design to estimate familiarity (in their study, heritability plus shared environmental effects) for metabolites as detected in plasma by  $^1\text{H}$  NMR. Alul et al. (2013) computed the heritability of metabolite concentrations as determined in dried blood spots obtained from heel stick in neonatal MZ and DZ twins. The authors included hormones, enzymes, acylcarnitines, and amino acids (47 analytes in total) as detected in routine newborn screening in their study; the amino acids and acylcarnitines were measured by a mass spectrometry-based platform. Next to the heritability of the individual biomolecules, these authors also report heritability for eight amino acid ratios and nine acylcarnitine ratios. In families burdened with premature cardiovascular disease, Shah et al. (2009) performed heritability analysis of 66 metabolites (including 37 acylcarnitines, 15 amino acids, 9 free fatty acids, conventional analytes, and ketones) detected in plasma using a targeted mass spectrometry-based metabolomics platform. In the context of a large genome-wide association study for metabolite levels as determined in serum by  $^1\text{H}$  NMR, MZ and DZ twin correlations, and heritability were computed for 216 metabolic variables (measured or computationally derived 'single' metabolites, or selected ratios among these single metabolites; Kettunen et al., 2012; van Dongen et al., 2012).

Of all metabolites targeted by the Biocrates sample pre-processing kit used in the current study, the familiarity of concentrations in blood has been reported in particular for amino acids and acylcarnitines. However, the lipids (sphingomyelins, phosphatidylcholines, and lysophosphatidylcholines) that are targeted by the Biocrates kit are relatively underrepresented in previous studies. Here, we describe aspects of familiarity for circulating metabolite concentrations that have not yet been addressed in previous metabolomics-based studies. First, we present MZ twin correlations for all metabolites detected in subjects originating from the Netherlands Twin Register (NTR), who were mostly fasting. The resemblance of MZ twins, expressed for example in correlations, gives an upper limit for the heritability of human traits (Falconer & Mackay, 1996). If there is no shared environmental influence, or assortative mating in the parental generations, the correlation in MZ pairs is the most precise estimate of heritability.

Next, we estimated midparent–offspring regression coefficients for all metabolites detected in serum samples from a small subgroup of NTR participants. If there is significant assortative mating, the midparent–offspring regression coefficient provides a direct estimate for the heritability of a trait in the parental generation (Falconer & Mackay, 1996), again as long as there is no contribution of shared environment.

Finally, we report the correlations between spouses for all metabolites detected in NTR and in the Leiden Longevity Study (LLS), which is a study of long-lived siblings and their offspring. Spouse correlations have not yet been described for metabolomics-based metabolic variables. Non-zero correlation between mates (spouses) can be due to assortative mating, spousal interaction, or due to the effect of a shared environment (Di Castelnuovo et al., 2009; van Grootheest et al., 2008). From a practical point of view, significant spouse correlations suggest that interventions aiming at reduction of metabolic risk factors should address both members of a marital couple rather than only one of the spouses (Di Castelnuovo et al., 2009). Significant resemblance for metabolite levels between spouses who share a household may be an indication that common environment ('C') contributes to familiarity. Hence, in the current study, we are able to obtain an estimate for the contribution of C to MZ correlations by comparing MZ correlations with spouse correlations. We further investigated the contribution of C to the familiarity of metabolite levels by looking at the effect of self-reported marriage duration on spousal resemblance.

We find considerable MZ twin correlations for most metabolites, and generally lower spouse correlations in both investigated cohorts. Our findings for the MZ twin correlations are of particular interest to obtain a better understanding of the genetic and biochemical underpinnings of individual differences observed during fasting.

## Methods

### Participants

Participants included in the MZ correlation and parent–offspring analyses, and part of the spouse pairs in this study belong to a subgroup of NTR participants who are included in NTR-Biobank studies (Willemsen et al., 2010). Roughly 50% of adult NTR participants (twins, their parents, spouses, and siblings) are part of NTR-Biobank projects and inclusion was not based on selection of phenotype. Venous blood was drawn in the morning from participants after overnight fasting. An attempt was made to bleed fertile women on the same day (day 3) of their cycle and women who took oral contraceptives in their pill-free week. The zygosity of MZ twins was confirmed by DNA markers. Data on the relationship duration of spouse pairs at the time of blood sampling were reported in NTR questionnaires. Ethical approval was obtained from the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam.

Metabolomics data obtained for the calculation of parent–offspring regression coefficients and spouse correlations (total 1,372 participants from the first NTR-Biobank project; Willemsen et al., 2010) and for computing the MZ twin correlations (total 480 participants from the second NTR-Biobank project; Willemsen et al., 2013) were obtained in two different measurement batches. Data from seven participants and one participant were identified as outliers in the metabolomics data originating from the first and second measurement batch, respectively, and excluded from further analysis. In the remaining data from the first batch, there were 281 complete spouse pairs who also had data on age, sex, and fasting status, and these were all included in the current study. Of these 281 spouse pairs, 70 pairs were also parents of one or two children. There were 10 families in which two children participated; from these families one child was selected at random. The resulting 70 parent–offspring trios were used to estimate midparent–offspring regression coefficients. In the metabolomics data from the second measurement batch after outlier removal, 181 complete MZ twin pairs were available for analysis, and these were all included in the study. Ten MZ twins were also offspring in the parent–offspring sample.

Details of the design and recruitment procedure for LLS have been provided elsewhere (Schoenmaker et al., 2006; Westendorp et al., 2009). Briefly, LLS was designed to identify phenotypic and genetic markers of longevity. To this end, long-lived (nonagenarian) siblings from in total 421 families together with their offspring and the partners of the offspring were recruited. The offspring were included because these also have a higher propensity to reach old age, and the partners were included as similarly aged controls. In the current study, the offspring and their partners were included. For LLS, blood samples were drawn for 266 participants after overnight fasting and for 388 non-fasting

participants. All samples were processed within 2 hours and the serum samples were stored at  $-80^{\circ}\text{C}$  until the time of analysis. From the 656 eligible participants with metabolomics data, one was identified as an outlier and was removed.

### Metabolomic Measurements

The Biocrates metabolomics platform has been validated extensively and complies with 21 CFR (Code of Federal Regulations) Part 11, indicating reproducibility within a given error range (Gieger et al., 2008). Metabolite measurements were carried out by flow injection analysis coupled to tandem mass spectrometry (MS/MS) at the Metabolomic Facility of the Genome Analysis Centre at the Helmholtz Centre in Munich, Germany, as described previously (Goek et al., 2012; Illig et al., 2010; Mittelstrass et al., 2011; Römisch-Margl et al., 2012). Assessment of metabolomics measures was done in three batches: one for NTR parents and offspring, one for LLS participants, and one for the MZ twins. Each batch had additional unrelated participants as well and data preprocessing was done on all data.

### Metabolomics Data Preprocessing

We excluded from further analysis data for metabolites with a coefficient of variation over the measurements of a pooled quality control sample larger than 25%, or with a median value below the limit of detection. Outlying data points and samples were identified as described previously (Goek et al., 2012). Missing values were imputed using multivariate imputation by chained equations as implemented in the ‘mice’ package (Van Buuren & Groothuis-Oudshoorn, 2011) in the R statistical language and programming environment (R Core Team, 2012). The imputed values were normalized by natural logarithm transformation. Throughout the article, lipids detected in the study samples are denoted as follows: acylcarnitines (Cx:y), sphingomyelins (SMx:y) and sphingomyelin-derivatives [SM(OH)x:y]; and glycerophospholipids (PC). Glycerophospholipids are differentiated with respect to the presence of ester (a) and ether (e) bonds in the glycerol moiety, where two letters (aa = diacyl, ae = acyl-alkyl) denote that two glycerol positions are bound to a fatty acid residue, while a single letter (a = acyl) indicates the presence of a single fatty acid residue. Lipid side chain composition is abbreviated as Cx:y, where *x* denotes the number of carbon atoms in the side chain and *y* the number of double bonds (Menni et al., 2013). The full list of metabolites detected in subjects from each cohort (NTR or LLS) included in the current study is provided in the Appendix and in Supplementary Table S1.

### Statistical Analysis

All analyses were carried out in R. Three measures of familial resemblance for metabolite concentration values were studied: MZ twin correlations, midparent–offspring regression, and spouse correlations. For all analyses, the effects of age

and sex were regressed out of the raw metabolite levels and the residuals were used for further analysis. For the spouse correlation analyses in the LLS cohort, fasting state was an additional binary covariate when computing residuals. MZ twin and spouse correlations were estimated as the Pearson correlations between the residuals for both twins or spouses. Parent–offspring regressions for all metabolic variables were estimated by simple linear regression analysis of the values for the offspring on the mean values for the parents. A *p* value of .05 was adopted as the threshold for nominal significance in this study. In the NTR MZ twins and in the NTR and LLS spouses, the multiple testing–corrected significance threshold was defined as  $(0.05/\text{Meffli})$ , with Meffli being equal to the estimated number of independent tests using the method of Li and Ji (2005). We assessed the concordance between the three measures of familial resemblance that were studied as follows. For each set of analyses, we investigated for how many metabolic variables the 95% confidence intervals of the correlation or regression coefficients overlapped. Also, for MZ twin correlations, we investigated how many metabolites were significant and compared the estimates of heritability with the estimates from midparent–offspring regression. As the number of MZ pairs is about twice as large as the number of parent–offspring pairs, for this comparison we adopted a liberal multiple testing–corrected significance threshold for the parent–offspring regressions (see the Discussion section). We used simple linear regression analysis to estimate the association between absolute metabolite level differences within pairs (spouses or MZ twins) and relationship duration (for spouses) or age (for MZ twin pairs). When computing the association of absolute within-spouse metabolic variable levels with self-reported relationship duration, we used the average of the self-reported relationship duration values over both spouses. The classification of Robinson et al. (1949) was adopted to categorize familiality estimates as low (0–30%), moderate (31–60%), or high (>60%).

## Results

Characteristics of participants for each subgroup are given in Table 1. In both the NTR MZ twin sample and the NTR parent–offspring sample, 123 metabolites passed quality control. In the LLS spouses, 120 metabolites passed quality control (see Table 2). A full list of all metabolites included in each of the three groups of participants (LLS, NTR parents and offspring, and NTR MZ twins) and their quartile values is given in Supplementary Table S1. All NTR MZ twins were over 18 years of age (see Table 1) with the NTR offspring group being the youngest on average. For NTR and LLS data, the Meffli values equaled 44 and 42, respectively.

### Familial Resemblance

**MZ twins.** In the MZ twin group, men had higher values for all acylcarnitines and for hexose, and for most amino

**TABLE 1****Basic Participant Characteristics**

Subgroup	N	Age in years (mean $\pm$ SD)	Minimum age	Maximum age	Female (%)
NTR spouses	281 pairs	59.8 $\pm$ 6.2	43	81	50
NTR parents	70 fathers and mothers	59.2 $\pm$ 6.2	43	76	50
NTR offspring	70 offspring	30.9 $\pm$ 4.6	22	42	60
NTR MZ twins	181 pairs	32.6 $\pm$ 11.5	19	68	69
LLS spouses	327 pairs	63.0 $\pm$ 6.6	39	82	50

Note: NTR: Netherlands Twin Register; LLS: Leiden Longevity Study.

**TABLE 2****For All Possible Pair-Wise Combinations of Analysis Types: Numbers of Metabolites for Which the 95% Confidence Intervals Obtained in Both Types of Analyses Overlap<sup>a</sup>**

	MZ correlation	Parent–offspring regression	Spouse correlation NTR	Spouse correlation LLS
MZ correlation	(123)			
Parent–offspring regression	100	(123)		
Spouse correlation NTR	34	122	(123)	
Spouse correlation LLS	21	108	105	(120)

Note: <sup>a</sup>Numbers on the diagonal indicate the numbers of metabolites that passed quality control in each sample (NTR MZ twin pairs, NTR parent–offspring trios, and NTR and LLS spouse pairs).

acids and lysophosphatidylcholines as indicated by the negative regression coefficients for female sex when computing residuals of metabolite levels (see Supplementary Table S2 for the point estimates). Values were lower in men for most phosphatidylcholines and sphingomyelins. Serum values of hexose, of all acylcarnitines and sphingomyelins and of most amino acids, phosphatidylcholines, and lysophosphatidylcholines increased with age. Significant Pearson correlations between MZ twins were observed for 121 measured metabolites (listed in the Appendix and in Supplementary Table S2) that passed quality control. Point estimates and 95% confidence intervals for the MZ twin correlations for all metabolites that passed quality control are given in Figure 1 and in the Appendix. Supplementary Table S2 also provides *p* values for the correlations. The mean MZ twin correlation across all metabolites was equal to 0.53 (range 0.21–0.77). The highest correlation was observed for serine and for PC ae C40:3 (point estimate for both metabolites equal to 0.77); the lowest correlation was observed for histidine and for SM C26:1 (point estimate for both metabolites equal to 0.21).

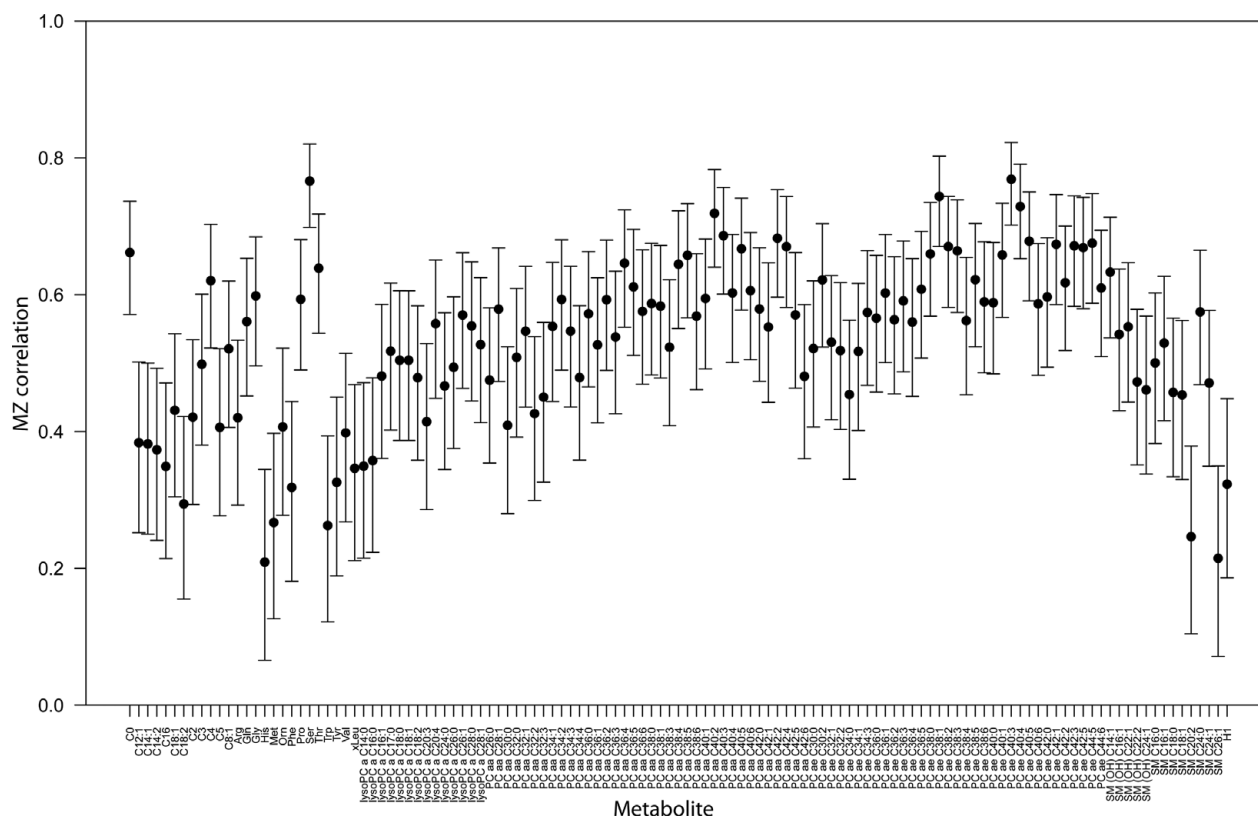
In MZ pairs, we also investigated if there was an association of absolute within-MZ pair differences in metabolite levels with age of the twins. Such associations would indicate that MZ twins become less (or possibly more) alike as they age. Only for PC ae C42:5 was found a nominally significant positive association ( $\beta = 0.002$ ; *p* value = .03; *df* = 179, see Supplementary Table S6).

**Midparent–offspring regression.** Similar to the MZ twins, in the offspring included in the parent–offspring sample, serum values of hexose, of all acylcarnitines and sphingomyelins, and of most phosphatidylcholines and lysophosphatidylcholines, increased with age (see Table S3; see the results of the spouse correlation analysis in NTR for the

effects of age and sex in the parental generation). However, in contrast to the MZ twins, the concentrations of most amino acids decreased with age in the offspring included in the parent–offspring sample, which might be due to the slightly different age and sex distributions in both groups of subjects. Also, in the offspring included in the parent–offspring sample, serum values of hexose, of most amino acids, of most lysophosphatidylcholines, and of all acylcarnitines were higher in males than in females. The values of almost all phosphatidylcholines and sphingomyelins were lower in males. Point estimates and associated 95% confidence intervals for the parent–offspring regression estimates for all metabolites are given in Figure 2 and in Table S3. The mean midparent–offspring regression coefficient across all metabolites was equal to 0.31 (range -0.13–0.85).

**Spouse correlations.** In the NTR spouses, in both men and women the levels of all acylcarnitines (except C3 in male spouses) increased with age (see Table S4). There was more heterogeneity in the age effects among the amino acids, lysophosphatidylcholines, and phosphatidylcholines (both among the metabolites and between the sexes). The levels of almost all sphingomyelins and hexose increased with age in both males and females. In NTR, significant Pearson correlations between the metabolite levels of spouses were observed for 74 metabolites (listed in Table S4). The mean spouse correlation across all metabolites was equal to 0.24 (range 0.02–0.54). Figure 3(a) displays the correlation point estimates as well as the boundaries of the 95% confidence intervals for the 123 metabolites that passed quality control for NTR.

In the LLS spouses, the levels of hexose and all acylcarnitines increased with age in both sexes (see Table S7). Also, the levels of most amino acids except


**FIGURE 1**

Monozygotic twin correlations in NTR for all 123 metabolites that passed quality control. For the explanation of metabolite names, see the Appendix and Supplementary Table S1.

glycine, histidine, serine, threonine, and tryptophan increased with age in both sexes. The levels of most phosphatidylcholines, lysophosphatidylcholines, and sphingomyelins decreased with age. In LLS, 58 metabolites (listed in Table S7) displayed significant spouse correlations. The mean spouse correlation across all metabolites was equal to 0.18 (range -0.06–0.51). Correlation point estimates and 95% confidence intervals are given for the LLS spouses in Figure 3(b).

#### Comparison of Correlations in Pairs of Different Genetic Relatedness

Table 2 indicates for MZ correlations, midparent–offspring regressions, and spouse-pair correlations, how many metabolic variables the 95% confidence intervals overlapped. For most pair-wise combinations of these measures, the 95% confidence intervals overlapped for the majority of metabolites. However, this overlap was considerably lower between the MZ correlations and the spouse correlation estimates in both NTR and LLS.

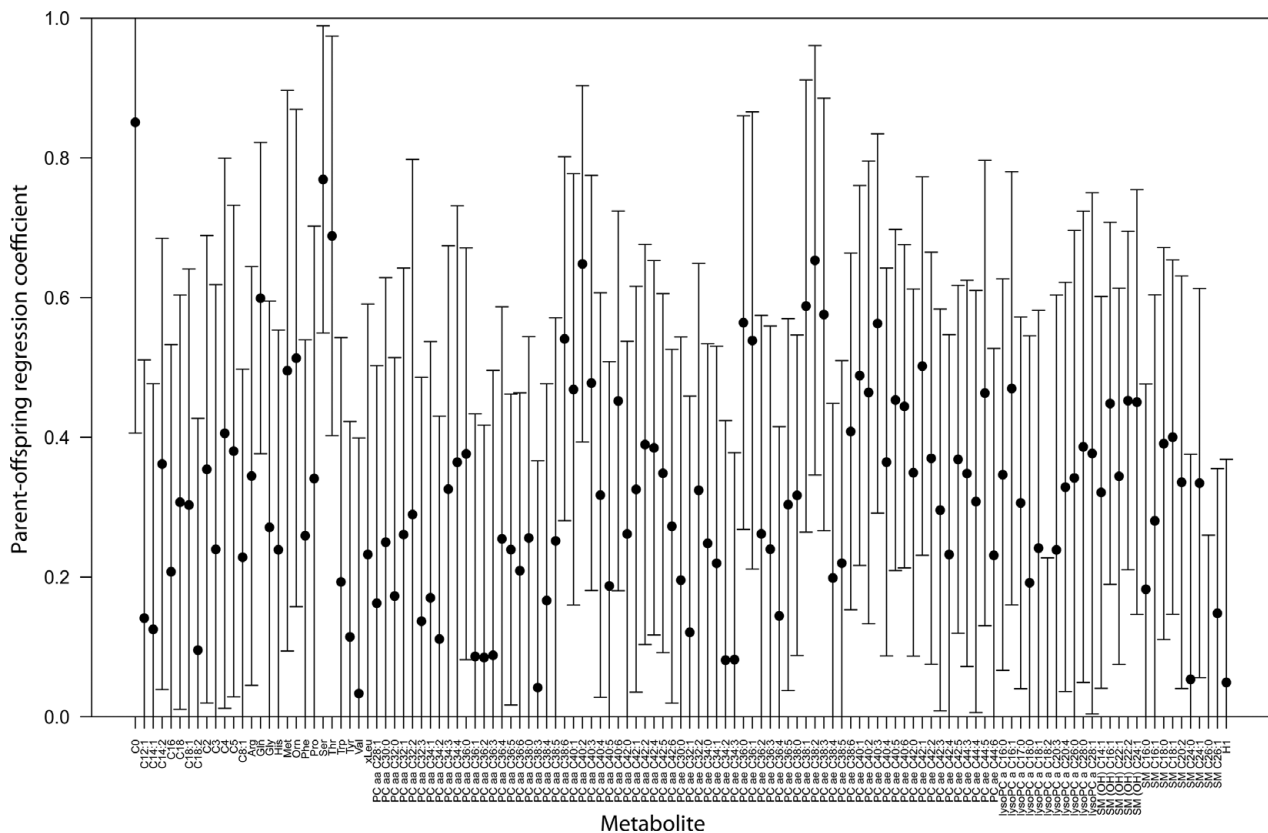
#### Association With Relationship Duration

For the majority of spouses in NTR, there were data regarding their self-reported relationship duration ( $N = 150$  pairs). We assessed whether there were associations between relationship duration and absolute within-spouse pair dif-

ferences in levels for each metabolite. At the nominal significance level, we found indications for such association for three single metabolites (C18:2; lysoPC a C18:2; SM (OH) C16:1; see Supplementary Table S5) that displayed significant spouse correlations; for all three metabolites, the association of relationship duration with within-pair differences was negative.

#### Discussion

The current study investigated familial resemblance for serum metabolite concentrations as measured using a well-established targeted metabolomics platform. Moderate-to-high MZ twin correlations were observed for most metabolites (mean 0.53; range 0.21–0.77), providing upper limits of heritability. Spouse correlations in two independent cohorts were generally lower and ranged from low to moderate. Therefore, these results suggest a substantial contribution of genetic factors to individual differences in serum metabolite concentrations. Consistent with previous reports of MZ twin correlations and heritability of blood metabolite concentrations (Alul et al., 2013; Kettunen et al., 2012; Nicholson et al., 2011; Shah et al., 2009), there was considerable heterogeneity in the MZ correlations across all metabolites. We observed more variation in MZ correlation estimates among the acylcarnitines, amino acids, and

**FIGURE 2**

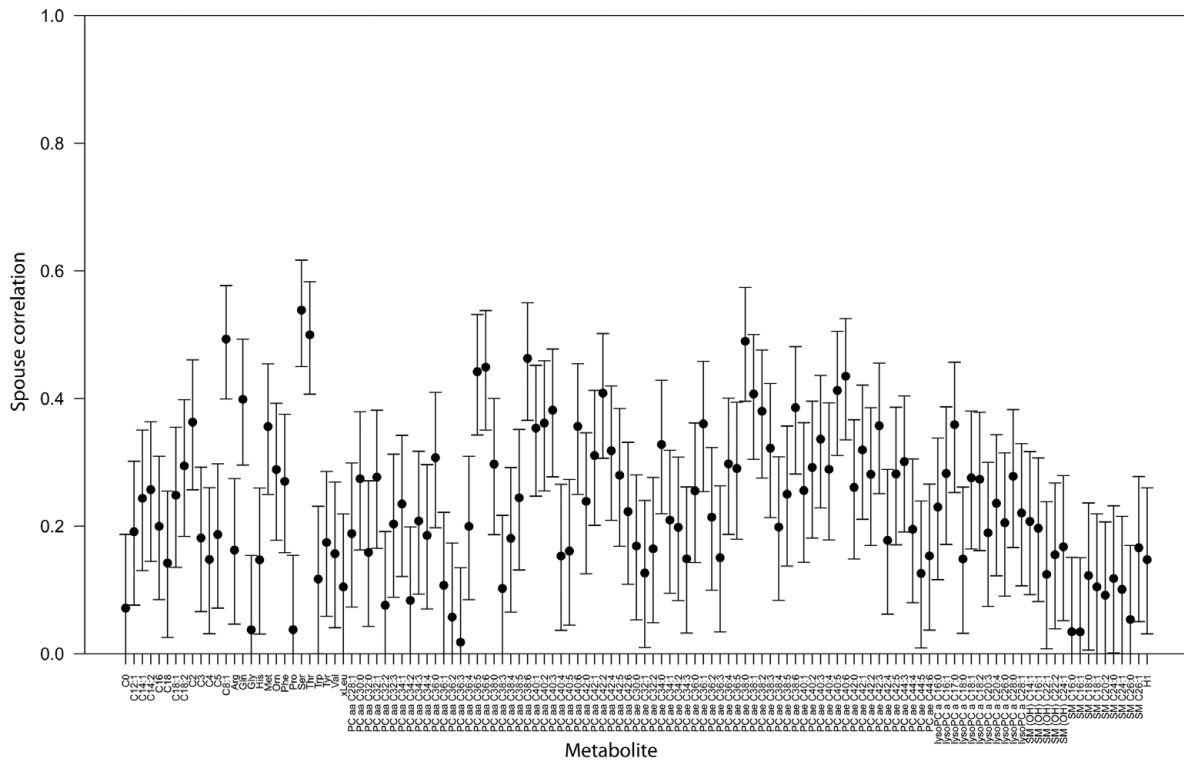
Midparent–offspring regression coefficients for all metabolites. For the explanation of metabolite names, see the Appendix and Supplementary Table S1.

sphingomyelins compared with the lysophosphatidylcholines and phosphatidylcholines.

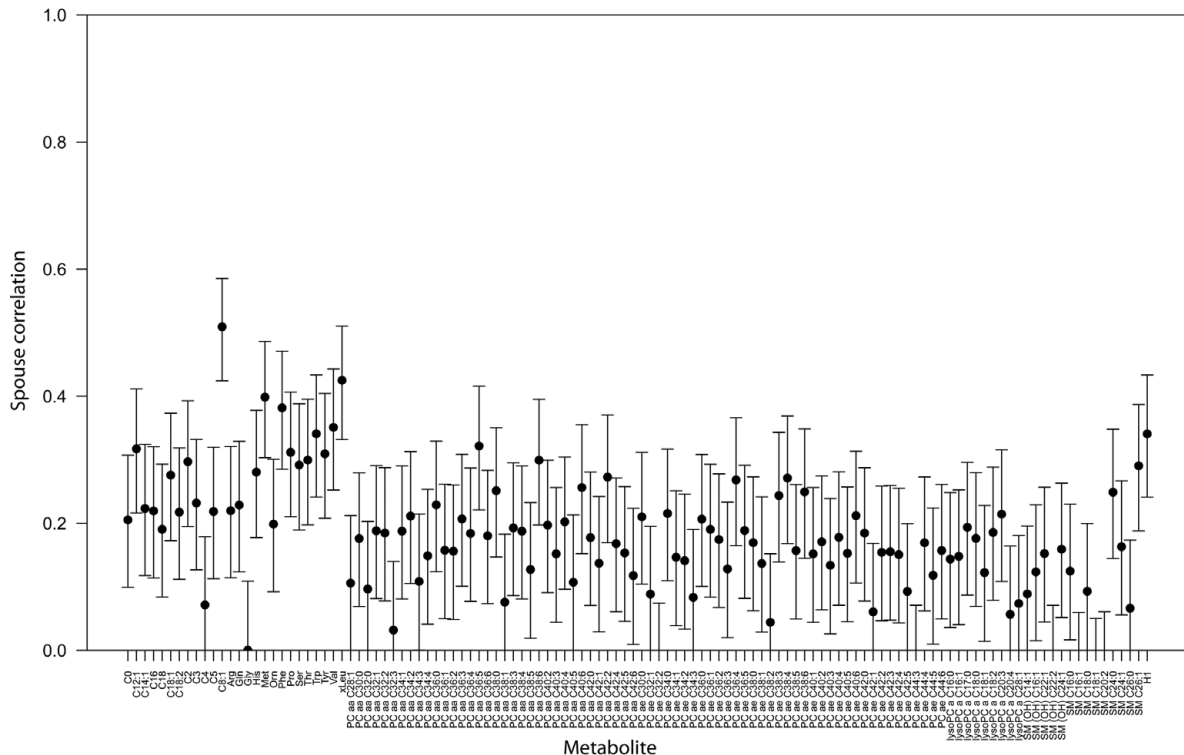
In particular, the estimates of MZ twin correlations for the acylcarnitines and amino acids can be compared with estimates of heritability obtained in previous studies for these metabolites. Supplementary Table S8 gives the point estimates for the MZ correlations together with the 95% confidence intervals for all metabolites targeted by the Biocrates AbsoluteIDQ p150 kit, along with the estimates of MZ correlations (Kettunen et al., 2012; Menni et al., 2013), heritability (Alul et al., 2013; Shah et al., 2009), or ‘familiality’ (Nicholson et al., 2011) (in the study by Nicholson et al., ‘familiality’ was defined as heritability plus shared environmental influences) in five previous studies that investigated familial aspects of metabolite concentrations as determined by metabolomics platforms. In the discussion that follows, we will denote as ‘close’ or ‘consistent’ those point estimates from these previous studies that are within the 95% confidence interval of the MZ correlation observed in the current work (a similar comparison was made by Nicholson et al. in their study; Nicholson et al., 2011).

All 12 (acyl)carnitines that were included in the MZ twin correlation analysis in our study were also included in the metabolomic newborn screen used in the study by Alul

et al. (2013), except for C14:2 (see Supplementary Table S8). However, this latter acylcarnitine was included in the study by Shah and colleagues (2009), who report heritability for in total 10 acylcarnitines for which we report MZ correlations. In line with the findings of Alul et al., in general we find lower heritability for medium- and long-chain acylcarnitines (more than eight carbon atoms) compared with the short-chain acylcarnitines. This is opposite to the pattern in the heritability observed by Shah and colleagues in families burdened with premature cardiovascular disease: compared with our MZ correlation estimates and the heritability estimates observed by Alul et al., the heritability estimates observed by Shah et al. are generally at the upper end or (much) higher, whereas the estimates for the short- and medium-chain species are similar or even lower. Remarkably, for C8:1 our estimate of the MZ correlation is (much) higher than the heritability estimates obtained by Alul et al. and Shah et al., but close to the MZ correlation observed by Menni et al. (2013). By comparison with the spouse correlations in NTR for this acylcarnitine, it can be seen that the high MZ correlations in our study might have been due in part to a substantial contribution of shared environment; this notion is supported by the considerably lower DZ correlation in the study by Menni et al. (2013).



(a)



(b)

**FIGURE 3**

Spouse correlations for all metabolites that passed quality control in NTR (a) and LLS (b). For the explanation of metabolite names, see the Appendix and Supplementary Table S1.



Aspects of familial resemblance for the amino acids threonine (MZ correlation 0.64) and tryptophan (MZ correlation 0.26) have not been reported in the five previous studies with which we compare our results. The estimates obtained in the current study might therefore be the first reported for these metabolites. Results for tyrosine, valine, and total leucine (leucine/isoleucine) are reported in four previous studies (Alul et al., 2013; Kettunen et al., 2012; Nicholson et al., 2011; Shah et al., 2009). Our estimate (0.33) of the MZ correlation for tyrosine is close to that observed by Kettunen et al. (2012), and to the heritability estimates of 0.36 and 0.38 obtained by Alul et al. (2013) and Shah et al. (2009). However, Nicholson et al. (2011) find a familiarity for this amino acid more than twice as high as our MZ correlation. The estimates of heritability obtained in the current study for the branched-chain amino acids (valine and total leucine) are rather consistent with those in previous reports. Our point estimate for the MZ correlation of valine (0.40) is close to the estimates in three of the four studies (range 0.43–0.44), but Alul et al. found almost zero heritability. The MZ correlation in the current study of 0.35 for total leucine is consistent with ‘familiarity’ (Nicholson et al., 0.35) or heritability (Shah et al., 0.39) reported by two previous studies, but slightly lower than the MZ correlation estimate of 0.50 obtained by Kettunen et al. and much higher than the heritability observed in the study by Alul et al. Remarkably, Alul and colleagues report almost zero heritability for this and a number of other metabolites that show considerably higher familiarity in other studies, including the current study. The MZ correlations for methionine (comparison with Alul et al. and Shah et al.), ornithine, and proline (comparison with Shah et al.) are close to the heritability estimates obtained in previous studies, suggesting that for these amino acids the MZ correlations provide an accurate estimate of heritability. However, the MZ correlations for glycine and serine obtained in the current study are higher than the MZ correlations or heritability observed by previous authors; our estimate for serine (0.77) is more than three times as high as the heritability calculated by Shah and colleagues. As we also find a considerable spouse correlation for serine in the NTR participants included in the current study, the relatively high MZ correlation may be due to the effects of shared environment. On the other hand, for glycine we find very low spouse correlations in samples from both cohorts (NTR and LLS) included in the current study, suggesting that the relatively high MZ correlation (0.60) that we find does not necessarily overestimate the heritability for this metabolite in the currently investigated subjects. The MZ correlations for the remaining amino acids are somewhat less consistent with those in earlier reports. Our estimate for arginine is almost twice as large as the heritability found by Alul et al., but approximately half the heritability observed by Shah and colleagues. For glutamine, our MZ correlation estimate is close to that obtained by Kettunen and colleagues, but lower than the heritability in the study by Shah

et al. For histidine, the MZ correlation estimate is lower than the heritability estimated by Shah and colleagues, and lower than the MZ correlation in the study of Kettunen and colleagues. We speculate, therefore, that this metabolite was measured with higher precision in those previous studies. Our MZ correlation estimate for phenylalanine is consistent with that of Kettunen et al., but slightly lower than the heritability in the study by Shah and colleagues and much higher than in the study by Alul and colleagues.

Menni et al. (2013) calculated MZ correlations for three diacyl-phosphatidylcholines and three acyl-alkylphosphatidylcholines that were also included in our MZ correlation analyses. For all six phosphatidylcholines, our MZ correlation estimates are higher than those observed by Menni and colleagues. The only sphingomyelin for which estimates of familial resemblance have been reported previously is SM C26:1, also in the study by Menni et al., who found a higher MZ correlation than we did. The relatively low congruence between the estimates obtained in both studies may be due to differences in age and sex distributions across the study subjects: the mean age in the study of Menni et al. was 58.5 years, and their study included females only.

In most studies that estimate heritability for complex traits, MZ twins are included and DZ twins serve as ‘controls’ to test for the contribution of shared environment. Most heritability studies of metabolite concentrations that employed the classical twin design in adult twins did not report evidence for common environment (C) shared by family members, although it should be recognized that these prior studies had too little power to obtain such evidence (Posthuma & Boomsma, 2000). In this study, we estimate upper limits of heritability for each metabolite in the MZ twins. If we are willing to assume that for metabolite levels in adults, the parents and their offspring do not share C we can compare heritability estimates from MZ twins to heritability based on the midparent–offspring regression. The midparent–offspring regression has the advantage that it gives an estimate for heritability that is unaffected by assortative mating. Therefore, it is interesting to investigate whether metabolites displaying significant midparent–offspring regression also display significant MZ correlation. For this comparison, because of the differences in sample size, we adopted a different significance threshold for the parent–offspring regression coefficients compared with the MZ twin correlations. This threshold was determined on the basis of the following power calculation using the G\*Power software (v. 3.1.7; Faul et al., 2009). Given a multiple testing–corrected significance threshold of  $p = (.05/44) = 1.14E-3$  with 44 being the estimated number of independent tests in the metabolomics data and a sample size of 181 MZ twin pairs, we would be able to detect an MZ twin correlation of 0.30 with 80% power. To detect a parent–offspring regression coefficient of minimally this magnitude with 80% power in the parent–offspring sample

comprising 70 parent–offspring pairs (one child with mean of parents), a significance threshold equal to 0.041 should be adopted. Based on this liberal multiple testing–corrected threshold, midparent–offspring regression coefficients are significant for 57 of the 123 metabolites that passed quality control in NTR (Supplementary Table S2). The average midparent–offspring regression coefficient over these 57 metabolites would be equal to 0.44 (range 0.24–0.85). Of these metabolites, 55 are also significant in the MZ correlation analysis. This considerable overlap in the number of statistically significant metabolites in both MZ correlation and parent–offspring regression analysis is consistent with the notion that both measures are estimates of heritability under the assumption of no shared environment between parents and their adult offspring.

To the best of our knowledge, spouse correlations have not yet been reported for metabolomics-based traits. In a systematic review, significant but low (upper limit of 95% confidence interval, maximal 0.10) spouse correlations were identified for the classical cardiovascular risk factors total and LDL cholesterol and total triglycerides (Di Castelnuovo et al., 2009). The meta-analysis estimate for total triglycerides can be compared with the spouse correlation estimates for the diacylphosphatidylcholines in the current study; for instance, because lipids from both classes can be formed from the intermediate molecule phosphatidic acid during biosynthesis (Vance & Vance, 2008). In concordance with the meta-analysis result, in the current study we found only low-to-moderate spouse correlations for most phosphatidylcholines, although these were lower in LLS than in NTR generally. Heritability studies for triglycerides have been reviewed by Snieder et al. (1999). In one study by Knoblauch et al. (1997), 100 MZ pairs and 72 DZ twin pairs were included with age characteristics (average age 33.0; *SD* 14.0; range 15–69) close to the age characteristics of the MZ twins in the current study. The heritability estimate of 0.66 for fasting total triglycerides obtained in the study by Knoblauch et al. is in line with the average MZ correlation of 0.59 we observe for the phosphatidylcholines.

The differences in the spouse correlations observed between the NTR and LLS subjects may be due to the intrinsic differences between both cohorts: LLS is a selected cohort of offspring from long-lived parents and their unrelated spouses, whereas NTR is an unselected population sample. Also, 59% of the LLS participants were non-fasting (but correction for fasting status was applied in the analyses), whereas 93–96% (dependent on subsample) of all NTR participants in this study were fasting. However, separate analysis of the data from fasting and non-fasting LLS participants did not substantially alter the results (data not shown).

Hexose showed only moderate MZ correlation, which is consistent with the notion that the majority of serum samples from NTR included in the current study were obtained from participants after an overnight fast. Dur-

ing fasting, serum concentrations of glucose (~90–95% of hexose; Goek et al., 2012) are expected to be low and little inter-individual variation is to be expected (Krug et al., 2012). The relatively high MZ correlation observed for free carnitine (C0), obtained using samples from fasting participants, suggests considerable genetic variation in the enzymes and transporters that are involved in the palmitoyl-CoA carnitine transferase II shuttle which is active during fasting (Krug et al., 2012). The observed considerable heritability of various amino acids (notably glutamine, serine, and threonine, which displayed both considerable MZ correlations and considerable midparent–offspring regression coefficients) may be due to these biomolecules being precursors of glucose during gluconeogenesis, a process also occurring during fasting (Berg et al., 2012). Interestingly, there appears to be an increasing trend in MZ twin correlations for phosphatidylcholines with increasing numbers of carbon atoms in their acyl chains. A bifurcation seems to occur in this respect between the diacyl (aa) and acyl-alkyl (ae) phosphatidylcholines (see Figure 1), where diacyl- or acyl-alkyl phosphatidylcholines with approximately the same number of carbon atoms appear to display similar heritability. This observation is in line with what we observed previously for triglycerides in a pilot study of heritability in lipidomics data, where triglycerides with increasing numbers of carbon atoms in their fatty acid side chains and/or increasing numbers of double bonds displayed increasing heritability (Draisma, 2011). The bifurcation in the effects observed for the diacyl- and acyl-alkyl (ether) phospholipids might be due to the fact that the biosynthetic pathways for these two classes of lipids are spatially separated and in some aspects different. We speculate that the patterns observed in heritability for the phosphatidylcholines are due to different numbers of metabolic conversion rounds during either fatty acid synthesis or beta-oxidation for phosphatidylcholines with different numbers of carbon atoms in their acyl chains.

In conclusion, we have demonstrated several aspects of familial resemblance for metabolite concentrations in serum. Our results suggest a substantial heritable component of variation for most metabolites, as has also been suggested by various genome-wide association studies based on metabolomics-based metabolic variables (e.g., Gieger et al., 2008; Illig et al., 2010; Kettunen et al., 2012). Our findings have implications for instance for biomarker research, suggesting that the metabolites showing high heritability may have better value as biomarkers for diseases that are mediated primarily by genetic factors compared with diseases that are driven primarily by environmental influences.

## Acknowledgments

We thank all participants in the studies described in this article. Professor Jerzy Adamski, Dr C. Prehn and J. Scarpa from the Helmholtz Zentrum München, Institute

of Experimental Genetics, performed the metabolomics measurements at the Genome Analysis Centre (GAC) of the HMGU. The European Network of Genomic and Genetic Epidemiology (ENGAGE) contributed to funding to perform the metabolomics measurements in LLS and in the NTR parent–offspring sample. We acknowledge the support from Pfizer Inc. and the late Professor David Cox in collection and metabolomics analysis of the NTR MZ twins included in the current study, as well as funding from the European Union's Seventh Framework Programme (FP7/2007–2011) under grant agreement no. 259679; CMSB: Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192); the Neuroscience Campus Amsterdam (NCA); the European Science Foundation (ESF): Genomewide analyses of European twin and population cohorts (EU/QLRT-2001-01254); Genotype/phenotype database for behavior genetic and genetic epidemiological studies (NWO 40-0056-98-9032), the European Science Council (ERC) Genetics of Mental Illness (230374), the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, BBMRI-NL (Biobanking and Biomolecular Resources Research Infrastructure Netherlands), and Unilever Colworth. HHMD is supported by an EMGO+ Fellowship (Mental Health research program of the EMGO+ Institute for Health and Care Research).

## Supplementary Material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/thg.2013.59>.

**Supplementary Table S1.** Full Names and Summary Statistics for Metabolites Measured in Each Cohort (LLS and NTR)

**Supplementary Table S2.** Monozygotic Twin Correlations for the Levels of All Measured Metabolites That Passed Quality Control

**Supplementary Table S3.** Midparent–Offspring Regression Coefficients for the Levels of All Measured Metabolites That Passed Quality Control

**Supplementary Table S4.** NTR: Spouse Correlations for the Levels of All Measured Metabolites That Passed Quality Control

**Supplementary Table S5.** NTR: Regression of Within-Spouse Pair Absolute Differences on Self-Reported Relationship Duration

**Supplementary Table S6.** NTR: Regression of Within-MZ Twin Pair Absolute Differences on Age

**Supplementary Table S7.** LLS: Spouse Correlations for the Levels of All Measured Metabolites That Passed Quality Control

**Supplementary Table S8.** Comparison of MZ Correlations With Estimates of Familiarity in Previous Studies

## References

- Alul, F. Y., Cook, D. E., Shchelochkov, O. A., Fleener, L. G., Berberich, S. L., Murray, J. C., & Ryckman, K. K. (2013). The heritability of metabolic profiles in newborn twins. *Heredity*, *110*, 253–258.
- Berg, J. M., Tymoczko, J. L., Stryer, L., & Gatto, G. J. Jr. (2012). *Biochemistry*. New York: Palgrave Macmillan.
- Di Castelnuovo, A., Quacquarello, G., Donati, M. B., de Gaetano, G., & Iacoviello, L. (2009). Spousal concordance for major coronary risk factors: A systematic review and meta-analysis. *American Journal of Epidemiology*, *169*, 1–8.
- Draisma, H. H. M. (2011). *Analysis of metabolomics data from twin families*. Leiden: Leiden University.
- Dunn, W. B., Broadhurst, D. I., Atherton, H. J., Goodacre, R., & Griffin, J. L. (2011). Systems level studies of mammalian metabolomes: The roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews*, *40*, 387–426.
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed., pp. 160–183). Harlow, England: Pearson Education.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.
- Floegel, A., Stefan, N., Yu, Z., Mühlenbruch, K., Drogan, D., Joost, H.-G., . . . Pischon, T. (2013). Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes*, *62*, 639–648.
- Gieger, C., Geistlinger, L., Altmaier, E., Hrabě de Angelis, M., Kronenberg, F., Meitinger, T., . . . Suhre, K. (2008). Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum. *PLoS Genetics*, *4*, e1000282.
- Goek, O.-N., Döring, A., Gieger, C., Heier, M., Koenig, W., Prehn, C., . . . Meisinger, C. (2012). Serum metabolite concentrations and decreased GFR in the general population. *American Journal of Kidney Diseases*, *60*, 197–206.
- Griffiths, W. J., Koal, T., Wang, Y., Kohl, M., Enot, D. P., & Deigner, H. P. (2010). Targeted metabolomics for biomarker discovery. *Angewandte Chemie International Edition*, *49*, 5426–5445.
- He, Y., Yu, Z., Giegling, I., Xie, L., Hartmann, A. M., Prehn, C., . . . Rujescu, D. (2012). Schizophrenia shows a unique metabolomics signature in plasma. *Translational Psychiatry*, *2*, e149.
- Illig, T., Gieger, C., Zhai, G., Römisch-Margl, W., Wang-Sattler, R., Prehn, C., . . . Suhre, K. (2010). A genome-wide perspective of genetic variation in human metabolism. *Nature Genetics*, *42*, 137–141.
- Jourdan, C., Petersen, A.-K., Gieger, C., Döring, A., Illig, T., Wang-Sattler, R., . . . Linseisen, J. (2012). Body fat free mass is associated with the serum metabolite profile in a population-based study. *PLoS ONE*, *7*, e40009.
- Kettunen, J., Tukiainen, T., Sarin, A.-P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L.-P., . . . Ripatti, S. (2012). Genome-wide association study identifies multiple loci

- influencing human serum metabolite levels. *Nature Genetics*, 44, 269–276.
- Knoblauch, H., Busjahn, A., Münter, S., Nagy, Z., Faulhaber, H. D., Schuster, H., & Luft, F. C. (1997). Heritability analysis of lipids and three gene loci in twins link the macrophage scavenger receptor to HDL cholesterol concentrations. *Arteriosclerosis, Thrombosis and Vascular Biology*, 17, 2054–2060.
- Krug, S., Kastenmüller, G., Stücker, F., Rist, M. J., Skurk, T., Sailer, M., . . . Daniel, H. (2012). The dynamic range of the human metabolome revealed by challenges. *FASEB Journal*, 26, 2607–2619.
- Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*, 95, 221–227.
- Menni, C., Zhai, G., Macgregor, A., Prehn, C., Römisch-Margl, W., Suhre, K., . . . Valdes, A. M. (2013). Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics*, 9, 506–514.
- Mittelstrass, K., Ried, J. S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., . . . Illig, T. (2011). Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genetics*, 7, e1002215.
- Nicholson, G., Rantalainen, M., Maher, A. D., Li, J. V., Malmodin, D., Ahmadi, K. R., . . . Holmes, C. C. (2011). Human metabolic profiles are stably controlled by genetic and environmental variation. *Molecular Systems Biology*, 7, 525.
- Posthuma, D., & Boomsma, D. I. (2000). A note on the statistical power in extended twin designs. *Behavior Genetics*, 30, 147–158.
- R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.
- Robinson, H. F., Comstock, R. E., & Harvey, P. H. (1949). Genotypic and phenotypic correlation in corn and their implications in selection. *Agronomy Journal*, 43, 282–287.
- Römisch-Margl, W., Prehn, C., Bogumil, R., Röhring, C., Suhre, K., & Adamski, J. (2012). Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics*, 8, 133–142.
- Schoenmaker, M., de Craen, A. J. M., de Meijer, P. H. E. M., Beekman, M., Blauw, G. J., Slagboom, P. E., & Westendorp, R. G. J. (2006). Evidence of genetic enrichment for exceptional survival using a family approach: The Leiden Longevity Study. *European Journal of Human Genetics*, 14, 79–84.
- Shah, S. H., Hauser, E. R., Bain, J. R., Muehlbauer, M. J., Haynes, C., Stevens, R. D., . . . Kraus, W. E. (2009). High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. *Molecular Systems Biology*, 5, 258.
- Snieder, H., van Doornen, L. J., & Boomsma, D. I. (1999). Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: Lessons from twin studies. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 19, 2826–2834.
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45, 1–67.
- Van Dongen, J., Slagboom, P., Draisma, H. H. M., Martin, N., & Boomsma, D. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics*, 13, 640–653.
- Van Grootheest, D. S., van den Berg, S. M., Cath, D. C., Willemsen, G., & Boomsma, D. I. (2008). Marital resemblance for obsessive–compulsive, anxious and depressive symptoms in a population-based sample. *Psychological Medicine*, 38, 1731–1740.
- Vance, D. E., & Vance, J. E. (2008). *Biochemistry of lipids, lipoproteins and membranes* (5th ed.). Amsterdam: Elsevier.
- Wang-Sattler, R., Yu, Z., Herder, C., Messias, A. C., Floegel, A., He, Y., . . . Illig, T. (2012). Novel biomarkers for pre-diabetes identified by metabolomics. *Molecular Systems Biology*, 8, 615.
- Westendorp, R. G. J., van Heemst, D., Rozing, M. P., Frölich, M., Mooijaart, S. P., Blauw, G.-J., . . . Leiden Longevity Study Group. (2009). Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *Journal of the American Geriatrics Society*, 57, 1634–1637.
- Willemsen, G., de Geus, E. J., Bartels, M., van Beijsterveldt, C. E., Brooks, A. I., Estourgie-Van Burk, G. F., . . . Boomsma, D. I. (2010). The Netherlands Twin Register biobank: A resource for genetic epidemiological studies. *Twin Research and Human Genetics*, 13, 231–245.
- Willemsen, G., Vink, J. M., Abdellaoui, A., den Braber, A., van Beek, J. H. D. A., Draisma, H. H. M., . . . Boomsma, D. I. (2013). The Adult Netherlands Twin Register: Twenty-five years of survey and biological data collection. *Twin Research and Human Genetics*, 16, 271–281.
- Wu, R., Huang, Y. H., Elinder, L. S., & Frostegård, J. (1998). Lysophosphatidylcholine is involved in the antigenicity of oxidized LDL. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 18, 626–630.
- Xu, T., Holzapfel, C., Dong, X., Bader, E., Yu, Z., Prehn, C., . . . Wang-Sattler, R. (2013). Effects of smoking and smoking cessation on human serum metabolite profile: Results from the KORA cohort study. *BMC Medicine*, 11, 60.
- Yu, Z., Zhai, G., Singmann, P., He, Y., Xu, T., Prehn, C., . . . Wang-Sattler, R. (2012). Human serum metabolic profiles are age dependent. *Aging Cell*, 11, 960–967.
- Zhai, G., Wang-Sattler, R., Hart, D. J., Arden, N. K., Hakim, A. J., Illig, T., & Spector, T. D. (2010). Serum branched-chain amino acid to histidine ratio: A novel metabolomic biomarker of knee osteoarthritis. *Annals of the Rheumatic Diseases*, 69, 1227–1231.

**Appendix: Overview of all detected metabolites and their monozygotic twin correlations**

Short and full names for metabolites measured in each cohort (LLS and NTR), and metabolite class. rMZ is the correlation in monozygotic twins for the levels of all measured metabolites that passed quality control and rMZ\_lbound and rMZ\_ubound are the lower and upper bounds of 95% confidence interval for rMZ.

Metabolite	Biochemical name	Group	rMZ	rMZ_lbound	rMZ_ubound
C0	Carnitine	Acylcarnitines	0.66	0.57	0.74
C12:1	Dodecenoylcarnitine	Acylcarnitines	0.38	0.25	0.5
C14:1	Tetradecenoylcarnitine	Acylcarnitines	0.38	0.25	0.5
C14:2	Tetradecadienylcarnitine	Acylcarnitines	0.37	0.24	0.49
C16	Hexadecanoylcarnitine	Acylcarnitines	0.35	0.21	0.47
C18	Octadecanoylcarnitine	Acylcarnitines			
C18:1	Octadecenoylcarnitine	Acylcarnitines	0.43	0.3	0.54
C18:2	Octadecadienylcarnitine	Acylcarnitines	0.29	0.16	0.42
C2	Acetylcarnitine	Acylcarnitines	0.42	0.29	0.53
C3	Propionylcarnitine	Acylcarnitines	0.5	0.38	0.6
C4	Butyrylcarnitine	Acylcarnitines	0.62	0.52	0.7
C5	Valerylcarnitine	Acylcarnitines	0.41	0.28	0.52
C8:1	Octenoylcarnitine	Acylcarnitines	0.52	0.41	0.62
Arg	Arginine	Amino acids	0.42	0.29	0.53
Gln	Glutamine	Amino acids	0.56	0.45	0.65
Gly	Glycine	Amino acids	0.6	0.5	0.68
His	Histidine	Amino acids	0.21	0.07	0.34
Met	Methionine	Amino acids	0.27	0.13	0.4
Orn	Ornithine	Amino acids	0.41	0.28	0.52
Phe	Phenylalanine	Amino acids	0.32	0.18	0.44
Pro	Proline	Amino acids	0.59	0.49	0.68
Ser	Serine	Amino acids	0.77	0.7	0.82
Thr	Threonine	Amino acids	0.64	0.54	0.72
Trp	Tryptophan	Amino acids	0.26	0.12	0.39
Tyr	Tyrosine	Amino acids	0.33	0.19	0.45
Val	Valine	Amino acids	0.4	0.27	0.51
xLeu	Leucine/isoleucine	Amino acids	0.35	0.21	0.47
lysoPC a C14:0	lysoPhosphatidylcholine acyl C14:0	Glycerophospholipids	0.35	0.21	0.47
lysoPC a C16:0	lysoPhosphatidylcholine acyl C16:0	Glycerophospholipids	0.36	0.22	0.48
lysoPC a C16:1	lysoPhosphatidylcholine acyl C16:1	Glycerophospholipids	0.48	0.36	0.59
lysoPC a C17:0	lysoPhosphatidylcholine acyl C17:0	Glycerophospholipids	0.52	0.4	0.62
lysoPC a C18:0	lysoPhosphatidylcholine acyl C18:0	Glycerophospholipids	0.5	0.39	0.61
lysoPC a C18:1	lysoPhosphatidylcholine acyl C18:1	Glycerophospholipids	0.5	0.39	0.61
lysoPC a C18:2	lysoPhosphatidylcholine acyl C18:2	Glycerophospholipids	0.48	0.36	0.58
lysoPC a C20:3	lysoPhosphatidylcholine acyl C20:3	Glycerophospholipids	0.41	0.29	0.53
lysoPC a C20:4	lysoPhosphatidylcholine acyl C20:4	Glycerophospholipids	0.56	0.45	0.65
lysoPC a C24:0	lysoPhosphatidylcholine acyl C24:0	Glycerophospholipids	0.47	0.34	0.57
lysoPC a C26:0	lysoPhosphatidylcholine acyl C26:0	Glycerophospholipids	0.49	0.38	0.6
lysoPC a C26:1	lysoPhosphatidylcholine acyl C26:1	Glycerophospholipids	0.57	0.46	0.66
lysoPC a C28:0	lysoPhosphatidylcholine acyl C28:0	Glycerophospholipids	0.55	0.44	0.65
lysoPC a C28:1	lysoPhosphatidylcholine acyl C28:1	Glycerophospholipids	0.53	0.41	0.62
PC aa C26:0	Phosphatidylcholine diacyl C26:0	Glycerophospholipids	0.48	0.35	0.58
PC aa C28:1	Phosphatidylcholine diacyl C28:1	Glycerophospholipids	0.58	0.47	0.67
PC aa C30:0	Phosphatidylcholine diacyl C30:0	Glycerophospholipids	0.41	0.28	0.52
PC aa C32:0	Phosphatidylcholine diacyl C32:0	Glycerophospholipids	0.51	0.39	0.61
PC aa C32:1	Phosphatidylcholine diacyl C32:1	Glycerophospholipids	0.55	0.44	0.64
PC aa C32:2	Phosphatidylcholine diacyl C32:2	Glycerophospholipids	0.43	0.3	0.54
PC aa C32:3	Phosphatidylcholine diacyl C32:3	Glycerophospholipids	0.45	0.33	0.56
PC aa C34:1	Phosphatidylcholine diacyl C34:1	Glycerophospholipids	0.55	0.44	0.65
PC aa C34:2	Phosphatidylcholine diacyl C34:2	Glycerophospholipids	0.59	0.49	0.68
PC aa C34:3	Phosphatidylcholine diacyl C34:3	Glycerophospholipids	0.55	0.44	0.64
PC aa C34:4	Phosphatidylcholine diacyl C34:4	Glycerophospholipids	0.48	0.36	0.58
PC aa C36:0	Phosphatidylcholine diacyl C36:0	Glycerophospholipids	0.57	0.47	0.66
PC aa C36:1	Phosphatidylcholine diacyl C36:1	Glycerophospholipids	0.53	0.41	0.62
PC aa C36:2	Phosphatidylcholine diacyl C36:2	Glycerophospholipids	0.59	0.49	0.68
PC aa C36:3	Phosphatidylcholine diacyl C36:3	Glycerophospholipids	0.54	0.43	0.63
PC aa C36:4	Phosphatidylcholine diacyl C36:4	Glycerophospholipids	0.65	0.55	0.72
PC aa C36:5	Phosphatidylcholine diacyl C36:5	Glycerophospholipids	0.61	0.51	0.7
PC aa C36:6	Phosphatidylcholine diacyl C36:6	Glycerophospholipids	0.58	0.47	0.67
PC aa C38:0	Phosphatidylcholine diacyl C38:0	Glycerophospholipids	0.59	0.48	0.68
PC aa C38:1	Phosphatidylcholine diacyl C38:1	Glycerophospholipids	0.58	0.48	0.67
PC aa C38:3	Phosphatidylcholine diacyl C38:3	Glycerophospholipids	0.52	0.41	0.62
PC aa C38:4	Phosphatidylcholine diacyl C38:4	Glycerophospholipids	0.64	0.55	0.72
PC aa C38:5	Phosphatidylcholine diacyl C38:5	Glycerophospholipids	0.66	0.57	0.73
PC aa C38:6	Phosphatidylcholine diacyl C38:6	Glycerophospholipids	0.57	0.46	0.66
PC aa C40:1	Phosphatidylcholine diacyl C40:1	Glycerophospholipids	0.59	0.49	0.68
PC aa C40:2	Phosphatidylcholine diacyl C40:2	Glycerophospholipids	0.72	0.64	0.78
PC aa C40:3	Phosphatidylcholine diacyl C40:3	Glycerophospholipids	0.69	0.6	0.76
PC aa C40:4	Phosphatidylcholine diacyl C40:4	Glycerophospholipids	0.6	0.5	0.69

**Appendix**  
**Continued**

Metabolite	Biochemical name	Group	rMZ	rMZ_lbound	rMZ_ubound
PC aa C40:5	Phosphatidylcholine diacyl C40:5	Glycerophospholipids	0.67	0.58	0.74
PC aa C40:6	Phosphatidylcholine diacyl C40:6	Glycerophospholipids	0.61	0.5	0.69
PC aa C42:0	Phosphatidylcholine diacyl C42:0	Glycerophospholipids	0.58	0.47	0.67
PC aa C42:1	Phosphatidylcholine diacyl C42:1	Glycerophospholipids	0.55	0.44	0.65
PC aa C42:2	Phosphatidylcholine diacyl C42:2	Glycerophospholipids	0.68	0.6	0.75
PC aa C42:4	Phosphatidylcholine diacyl C42:4	Glycerophospholipids	0.67	0.58	0.74
PC aa C42:5	Phosphatidylcholine diacyl C42:5	Glycerophospholipids	0.57	0.46	0.66
PC aa C42:6	Phosphatidylcholine diacyl C42:6	Glycerophospholipids	0.48	0.36	0.59
PC ae C30:0	Phosphatidylcholine acyl-alkyl C30:0	Glycerophospholipids	0.52	0.41	0.62
PC ae C30:2	Phosphatidylcholine acyl-alkyl C30:2	Glycerophospholipids	0.62	0.52	0.7
PC ae C32:1	Phosphatidylcholine acyl-alkyl C32:1	Glycerophospholipids	0.53	0.42	0.63
PC ae C32:2	Phosphatidylcholine acyl-alkyl C32:2	Glycerophospholipids	0.52	0.4	0.62
PC ae C34:0	Phosphatidylcholine acyl-alkyl C34:0	Glycerophospholipids	0.45	0.33	0.56
PC ae C34:1	Phosphatidylcholine acyl-alkyl C34:1	Glycerophospholipids	0.52	0.4	0.62
PC ae C34:2	Phosphatidylcholine acyl-alkyl C34:2	Glycerophospholipids			
PC ae C34:3	Phosphatidylcholine acyl-alkyl C34:3	Glycerophospholipids	0.57	0.47	0.66
PC ae C36:0	Phosphatidylcholine acyl-alkyl C36:0	Glycerophospholipids	0.57	0.46	0.66
PC ae C36:1	Phosphatidylcholine acyl-alkyl C36:1	Glycerophospholipids	0.6	0.5	0.69
PC ae C36:2	Phosphatidylcholine acyl-alkyl C36:2	Glycerophospholipids	0.56	0.46	0.66
PC ae C36:3	Phosphatidylcholine acyl-alkyl C36:3	Glycerophospholipids	0.59	0.49	0.68
PC ae C36:4	Phosphatidylcholine acyl-alkyl C36:4	Glycerophospholipids	0.56	0.45	0.65
PC ae C36:5	Phosphatidylcholine acyl-alkyl C36:5	Glycerophospholipids	0.61	0.51	0.69
PC ae C38:0	Phosphatidylcholine acyl-alkyl C38:0	Glycerophospholipids	0.66	0.57	0.73
PC ae C38:1	Phosphatidylcholine acyl-alkyl C38:1	Glycerophospholipids	0.74	0.67	0.8
PC ae C38:2	Phosphatidylcholine acyl-alkyl C38:2	Glycerophospholipids	0.67	0.58	0.74
PC ae C38:3	Phosphatidylcholine acyl-alkyl C38:3	Glycerophospholipids	0.66	0.57	0.74
PC ae C38:4	Phosphatidylcholine acyl-alkyl C38:4	Glycerophospholipids	0.56	0.45	0.65
PC ae C38:5	Phosphatidylcholine acyl-alkyl C38:5	Glycerophospholipids	0.62	0.52	0.7
PC ae C38:6	Phosphatidylcholine acyl-alkyl C38:6	Glycerophospholipids	0.59	0.49	0.68
PC ae C40:0	Phosphatidylcholine acyl-alkyl C40:0	Glycerophospholipids	0.59	0.48	0.68
PC ae C40:1	Phosphatidylcholine acyl-alkyl C40:1	Glycerophospholipids	0.66	0.57	0.73
PC ae C40:2	Phosphatidylcholine acyl-alkyl C40:2	Glycerophospholipids			
PC ae C40:3	Phosphatidylcholine acyl-alkyl C40:3	Glycerophospholipids	0.77	0.7	0.82
PC ae C40:4	Phosphatidylcholine acyl-alkyl C40:4	Glycerophospholipids	0.73	0.65	0.79
PC ae C40:5	Phosphatidylcholine acyl-alkyl C40:5	Glycerophospholipids	0.68	0.59	0.75
PC ae C40:6	Phosphatidylcholine acyl-alkyl C40:6	Glycerophospholipids	0.59	0.48	0.67
PC ae C42:0	Phosphatidylcholine acyl-alkyl C42:0	Glycerophospholipids	0.6	0.49	0.68
PC ae C42:1	Phosphatidylcholine acyl-alkyl C42:1	Glycerophospholipids	0.67	0.59	0.75
PC ae C42:2	Phosphatidylcholine acyl-alkyl C42:2	Glycerophospholipids	0.62	0.52	0.7
PC ae C42:3	Phosphatidylcholine acyl-alkyl C42:3	Glycerophospholipids	0.67	0.58	0.74
PC ae C42:4	Phosphatidylcholine acyl-alkyl C42:4	Glycerophospholipids			
PC ae C42:5	Phosphatidylcholine acyl-alkyl C42:5	Glycerophospholipids	0.67	0.58	0.74
PC ae C44:3	Phosphatidylcholine acyl-alkyl C44:3	Glycerophospholipids			
PC ae C44:4	Phosphatidylcholine acyl-alkyl C44:4	Glycerophospholipids			
PC ae C44:5	Phosphatidylcholine acyl-alkyl C44:5	Glycerophospholipids	0.68	0.59	0.75
PC ae C44:6	Phosphatidylcholine acyl-alkyl C44:6	Glycerophospholipids	0.61	0.51	0.69
SM (OH) C14:1	Hydroxysphingomyeline C14:1	Sphingolipids	0.63	0.54	0.71
SM (OH) C16:1	Hydroxysphingomyeline C16:1	Sphingolipids	0.54	0.43	0.64
SM (OH) C22:1	Hydroxysphingomyeline C22:1	Sphingolipids	0.55	0.44	0.65
SM (OH) C22:2	Hydroxysphingomyeline C22:2	Sphingolipids	0.47	0.35	0.58
SM (OH) C24:1	Hydroxysphingomyeline C24:1	Sphingolipids	0.46	0.34	0.57
SM C16:0	Sphingomyeline C16:0	Sphingolipids	0.5	0.38	0.6
SM C16:1	Sphingomyeline C16:1	Sphingolipids	0.53	0.42	0.63
SM C18:0	Sphingomyeline C18:0	Sphingolipids	0.46	0.33	0.57
SM C18:1	Sphingomyeline C18:1	Sphingolipids	0.45	0.33	0.56
SM C20:2	Sphingomyeline C20:2	Sphingolipids	0.25	0.1	0.38
SM C24:0	Sphingomyeline C24:0	Sphingolipids	0.58	0.47	0.67
SM C24:1	Sphingomyeline C24:1	Sphingolipids	0.47	0.35	0.58
SM C26:0	Sphingomyeline C26:0	Sphingolipids			
SM C26:1	Sphingomyeline C26:1	Sphingolipids	0.21	0.07	0.35
H1	Hexose	Sugars	0.32	0.19	0.45