CHAPTER 5

Contribution of Genetic and Environmental Factors to Variation in the Human Blood Plasma Metabolome: a Multivariate Study in Twins and Siblings

Harmen H.M. Draisma,[1] Theo H. Reijmers,[1] Jacqueline J. Meulman,[2] Dorret I. Boomsma,[3] Jan van der Greef,[1] and Thomas Hankemeier[1]

Adapted from Draisma et al., *submitted for publication*

[1]Leiden University, LACDR, Leiden, The Netherlands.
[2]Leiden University, Mathematical Institute, Leiden, The Netherlands.
[3]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands.

5.1 Abstract

Phenotypic data obtained in a genetically informative population sample of individuals can be used for quantitative genetic analyses to elucidate the relative contributions of genetic and environmental variance components to the observed phenotypic variation. Metabolomics aims at the comprehensive measurement in a given sample of all small molecules that are intermediate or end-products of cellular metabolism. Therefore, data as provided by metabolomics experiments represent a snapshot of the physiological state of an organism and are particularly informative of actual phenotypic traits such as disease. By structural equation modeling, we analyzed data obtained with two metabolomics methods in blood plasma samples from in total 163 participants (healthy mono- and dizygotic twins and their sex-matched nontwin siblings). Relative concentrations were obtained of 59 individual lipid metabolites by 'targeted' liquid chromatography–mass spectrometry (LC–MS); a 'global' overview of the relative concentrations of metabolites from different metabolite classes was provided by proton nuclear magnetic resonance (^{1}H NMR) spectroscopy. Univariate quantitative genetic analyses of the LC–MS data revealed potentially biologically relevant differences in heritability for different lipids. In multivariate analysis, we observed that in particular lipids of the same lipid class shared genetic causes of phenotypic variance. In contrast, the heterogeneity of genetic causes of phenotypic variation among different metabolites was relatively large in the ^{1}H NMR data. In conclusion, in this study we have shown the potential of uni- and multivariate quantitative genetic analyses to generate biological insight into the importance of genetic variation for variation observed in human metabolomics data.

5.2 Introduction

Recently, the results of the first genome-wide association (GWA) studies have been reported linking genomic variation and variation in human metabolomics data.[12,51,52] Metabolomics is the comprehensive study of the reagents, intermediate products, or end products of cellular metabolism.[2] Being intermediate phenotypes, with respect to studies on the level of actual phenotypes metabolites provide more insight into biological pathways underlying phenotypic variation.[10–12] The study of (endo)phenotypic variation in quantitative traits, such as metabolite levels measured in body fluids might be of relevance for our understanding of the causes of common diseases.[47,66] Among the various endophenotypes that are measured at "omics" scale (*e.g.*, proteins, gene transcripts), metabolites have the most direct link to cellular physiology and functioning.[8,9,160] The measurement at "omics" scale using the currently available analytical techniques provide an unprecedented scale of resolution, which can be even more directly linked to cellular physiology than is possible on the basis of measurements of 'conventional metabolites'.[161] Metabolomics studies

aim to obtain a comprehensive view of all metabolites from particular metabolite classes (in a so-called "targeted" approach), or of the metabolites from all classes (in a "global" approach). Both approaches allow for the discovery of previously unknown biological pathways on the basis of patterns of relationships among different metabolites, which would be much harder to achieve in a classical reductionist approach that focuses only at select compounds.[162,163]

Here we report the results of quantitative genetic analyses of metabolomics data obtained in a genetically informative sample of individuals, *i.e.* in mono- and dizygotic twin pairs and their nontwin siblings. Instead of elucidating the measureable or 'manifest' genotypic variables (*i.e.*, single-nucleotide polymorphisms indicating quantitative trait loci), as is done in *e.g.* GWA studies,[49,50] in such analyses the causes of phenotypic variation are often modeled as latent variables in a structural equation model.[1,38] Analysis of the covariance structure in phenotypic data by structural equation modeling (SEM)[32] allows for the decomposition of phenotypic (co)variance into variance components attributable to genetic variation and to environmental variation. Using select study designs it is also possible to elucidate the relative contribution of gene-environment interaction to phenotypic (co)variation of traits.[29,38,68]

However, this is not possible on the basis of the classical twin design, which is based upon the comparison of the phenotypic covariances of mono- and dizygotic twins raised together. Monozygotic (MZ) twins, who are fertilized from the same egg, share 100% of their additive genetic variance.[29] Dizygotic (DZ) twins, who are fertilized from two separate egg cells, share only on average 50% of their segregating genes; this percentage is the same for biological nontwin siblings. Therefore, any excess phenotypic correlation between MZ co-twins over that between DZ co-twins is an indication that genetic effects contribute to the variance of a trait.[1]

In SEM, such reasoning is formalized in the structural model and its consistency with the observed data is statistically tested.[32] Analysis of the phenotypic covariances for a single trait of mono- and dizygotic twin pairs raised together allows for the estimation of the heritability of this trait, *i.e.* of the proportion of phenotypic variation attributable to genetic variation among individuals.[37] Next to such a univariate analysis, multivariate analysis is used to elucidate the contribution of genetic and environmental effects to the phenotypic covariance among multiple traits, and it increases statistical power to detect genetic effects.[164,165]

In addition to MZ and DZ twins, sex-matched nontwin siblings of these twins were included in this study because this is known to enhance the power to detect genetic as well as shared environmental effects.[166] In this study we performed both uni- and multivariate quantitative genetic analyses using two types of metabolomics data obtained in blood plasma samples from the same participants. That is, we analyzed data from liquid chromatography–mass spectrometry (LC–MS) of plasma lipids, and from one-dimensional proton nuclear magnetic resonance (^{1}H NMR) spectroscopy. The LC–MS data provide a 'targeted' view of the lipid metabolites present in the samples; lipids are

involved in a number of important (patho-)physiological processes.[153] Proton NMR spectroscopy, on the other hand, aims at a more 'global' view of metabolites from different classes, for example amino acids, lipoproteins and carbohydrates.[2] However, with this latter method one can not discriminate among for example the individual lipid metabolites that are detected by the targeted LC–MS method used in this study. Also, with NMR spectroscopy typically only metabolites present in higher concentrations in a sample can be detected.[2]

In our analyses of the LC–MS data, we observed marked heritability for a number of lipids, but also different degrees of heritability among lipids belonging to different classes. In particular, we found a potentially biologically relevant pattern of heritabilities among the lipids of the triglyceride class. In multivariate analysis, in general lipids of the same class tended to cluster together. This suggests that positive phenotypic correlation among blood plasma lipids from the same lipid class is caused by pleiotropic genes.

Probably due to the "global" nature of the used ^{1}H NMR method, the results of the multivariate analyses of the ^{1}H NMR data suggested a much larger diversity in genetic causes of variance for different metabolites than in case of the LC–MS data.

5.3 Materials and methods

5.3.1 Participants

Twins and biological nontwin siblings were recruited from the Netherlands Twin Register.[154] Collection of fasting blood samples from all participants, and sample preparation were performed as described previously.[155–157] Zygosity was determined for all twin pairs by DNA genotyping.

5.3.2 Measures

Semiquantitative metabolomics analyses of the samples obtained from all study participants were performed in two "blocks", where in the first block samples from different participants were analyzed than in the second block (see Chapter 3). Blood plasma was analyzed both with an LC–MS method targeted at the analysis of lipids, and with ^{1}H NMR spectroscopy, as described in Chapter 3 as well. In a metabolomics context, the term "semiquantitative" indicates that no absolute concentrations were measured for the individual metabolites. Rather, we measured either the concentrations of lipids with respect to those of a limited number of so-called "internal standards" (in case of the LC–MS analyses), or the relative concentrations of metabolites with respect to each other (in case of the ^{1}H NMR analyses).

The measurements of the second 'block' were performed almost one year after those of the first 'block'; samples from members of the same family were always measured in the same block. For the data obtained with both methods, the nonbiological systematic differences between the normalized data from the

two measurement blocks were removed by quantile equating (see Chapter 3). This allowed the combination of data from the same variables measured in both blocks into one common data set that can be analyzed with methods like those used in this chapter. After equating, replicate measurement data were averaged per study sample before entering them into SEM as described below.

In this chapter, individual lipid compounds (*e.g.*, C16:1_LPC) as measured with LC–MS are denoted as follows: the number of carbon atoms (*e.g.*, C16) as well as the number of double bonds (*e.g.*, 1) in the lipid, separated by a colon are followed by the class abbreviation (*e.g.*, "LPC" for lysophosphatidylcholines).[127] Proton NMR variables are denoted by the chemical shift values that correspond to the detected features (see Chapter 3).

5.3.3 Genetic analysis

With respect to the quantitative genetic analyses, in this study we followed a similar strategy as was pursued by Schmitt *et al.* in the analysis of voxel-based magnetic resonance imaging data.[167] That is, first we performed univariate genetic analyses to estimate the proportions of phenotypic variance of each variable separately attributable to genetic and specific environmental variance.

Then, we performed all possible bivariate analyses to estimate the genetic and non-genetic components of covariance between all pairs of variables within each data set. The results of these multiple bivariate analyses populated for each data set a genetic correlation matrix, which was subsequently subjected to hierarchical clustering analysis. Schmitt *et al.*, in their 2008-paper, refer to this methodology as "multistep multivariate analysis". A "multistep multivariate" analysis strategy is actually a workaround that provides "semimultivariate" results in cases where existing covariance-based multivariate data analysis methods can not directly be applied to analyze the data for all variables within a data set simultaneously.[168] Typically, as is the case in maximum likelihood-based SEM, data that consist of a relatively small number of objects (participants) and a very large number of (correlated) measured variables prohibit the straightforward use of such existing methods because variance-covariance matrices computed on the basis of such data are non-positive definite.

Variance components were estimated by SEM approach using full information maximum likelihood (FIML) under normal theory using the raw data as input. FIML allows structural equation models to be fitted in the presence of missing values in the data (*e.g.* on twin pairs without nontwin sibling). For SEM we used the novel package OpenMx (version 0.4.1-1320),[169] which is implemented in the statistical computing environment `R`[158] (version 2.10.1).

Univariate analyses

Before fitting variance component models to the data, we established the likelihood resulting from fitting saturated models, where as many characteristics of the observed data (means, variances, covariances) as possible are freely esti-

mated. Then, we equated means and variances within families, and compared the resulting likelihood with that of fitting the saturated model to the data using a likelihood ratio chi-square test. The significance of variance components was tested in a similar way, *i.e.* by comparing the likelihood of the more complex model with that of a more parsimonious model.

We based our choice for a particular genetic variance components model on the customary rules of fit and parsimony, *i.e.*, overall for most variables the fit of the genetic model had to be non-significantly different from that of a saturated model, and overall for most variables the fit of a more parsimonious model (*e.g.*, "E") had to be significantly different from that of the more complex model (*e.g.*, "AE"). Here, the capitals "A" and "E" denote the latent additive genetic and non-shared environmental sources of phenotypic variance, respectively.[38] p-values lower than 0.05 were considered statistically significant. We chose one variance components model (*i.e.*, the "AE" model) to be used for the analysis of all variables, as the sample size was relatively small in the current study. One consideration for doing so was that the estimated values of variance components are always (slightly) dependent on the particular model used, and therefore to be able to compare variance component estimates among different variables it is important that they all have been estimated under the same model.

The homogeneity of means, variances, and variance components across sexes was assessed by comparing the fit of variance components models fitted to the data for males and females separately, with the fit of models where these parameters had been equated across the sexes. For the analysis of all data sets we used data for nontwin siblings only if they were of the same sex as their twin siblings, because the statistical significance of the estimated variance components was higher than when we included opposite-sex nontwin siblings as well (not shown).

Standardized variance components estimates were obtained by dividing the squared values by total variance.[170] Confidence intervals (CIs) for the standardized genetic variance components were likelihood-based.[171]

Bivariate analyses

The components of covariance for each pair of variables within each data set were estimated by fitting a bivariate model based upon a so-called Cholesky composition of the expected covariance matrix (see Fig. 1.4). For initial analysis a multivariate model based upon Cholesky composition is attractive because it is relatively hypothesis-free.[172] For the bivariate analyses, the relative contributions of the same latent sources of phenotypic variance (*i.e.*, "A" and "E") were estimated as in the univariate analyses.[173]

Genetic correlations were computed from the results of the bivariate analyses as follows:[38]

$$r_{\mathrm{x,y}} = \frac{varA_{\mathrm{xy}}}{\sqrt{(varA_{\mathrm{x}} \times varA_{\mathrm{y}})}} \tag{5.1}$$

where $r_{x,y}$ is the genetic correlation between a pair of variables, $varA_{xy}$ is the unstandardized genetic component of the covariance between the two variables, and $varA_x$ and $varA_y$ are the unstandardized genetic components of variance for the respective variables. For each data set, the genetic correlations for each pair of variables were aggregated into a square genetic 'correlation matrix', of which the dimensions equal the number of variables in the data set.[167]

5.3.4 Hierarchical clustering analysis

We used hierarchical clustering analysis to discover patterns of relationships among different variables in the genetic correlation matrices.[174,175] The aim of hierarchical clustering analysis is to group (cluster) variables on the basis of their relative similarities and differences, such that variables that are relatively similar will be grouped together, and variables that are relatively dissimilar will be in different clusters. For hierarchical clustering, we computed the dissimilarities among variables as (1−correlation).[41,176] Then, we subjected the resulting 'dissimilarity matrix' to hierarchical clustering, using the average linkage clustering algorithm. It has been noted[41] that average linkage in practice often performs satisfactorily. The results of hierarchical clustering were visualized using the "heatmap.2" function from the "gplots" package in R.

Of note, as an alternative to hierarchical clustering analysis, eigenvalue decomposition (spectral decomposition, EVD) of the genetic correlation matrix could be used to visualize the patterns of genetic relationships among different metabolites. The eigenvectors as resulting from EVD of a correlation matrix are equivalent to the "loadings" that would result from a principal component analysis on the autoscaled original two-mode (*i.e.*, objects × variables) data matrix on which the correlation matrix was based.[23] By EVD of the genetic correlation matrix, the genetic covariance among variables (metabolites) can be summarized by projecting the original variables onto new orthogonal variables, the so-called principal components (PCs), on the basis of the dominant direction of the genetic covariance among all metabolites. However, we do not show the results of EVD of the genetic correlation matrix here, because hierarchical clustering analysis was used in the remainder of this thesis to summarize the relationships among either objects or variables.

5.4 Results and discussion

5.4.1 Participants

The combined data sets, based on the measurements obtained in the two measurement blocks, comprised data for in total 130 twins and 33 sex-matched nontwin siblings for both LC–MS and ^{1}H NMR. The LC–MS data set contained data on 59 lipids detected in the sample from each participant. Lipids from the following five classes were detected: lysophosphatidylcholines (LPCs); phosphatidylcholines (PCs); sphingomyelins (SPMs); cholesterol esters (ChEs);

Table 5.1: Basic description of participants [a]

	MZM	MZF	DZM	DZF	Nontwin siblings	Total
Number of participants	34	40	20	36	33	163
Average age in years (standard deviation)	18.1 (0.2)	18.1 (0.2)	18.2 (0.2)	18.2 (0.2)	19.0 (4.7)	18.3 (2.1)

[a]MZM, monozygotic male; MZF, monozygotic female; DZM, dizygotic male; DZF, dizygotic female.

and triglycerides (TGs). The ^{1}H NMR data set contained data on 74 features (peaks) detected in each spectrum.

In total 67 participants were male and 96 were female; participants originated from in total 65 families (see Table 5.1). All DZ twin pairs included in the study were same-sex pairs.

5.4.2 Univariate variance components analyses

Genetic models that incorporated heterogeneity of means, variances and covariances across sexes did not fit differently to the data than models where the values for these parameters had been equated across males and females. Therefore, we estimated covariance components for males and females together.

For all data sets, the likelihood-based CIs were rather large, due to sampling error because of the relatively small number of participants in this study. The univariate results specific for each data set are given below.

LC–MS lipids

The heritability estimates per lipid, as well as the 95%-CI, are shown in Figure 5.1. For all measured lysophosphatidylcholines (LPCs), the estimates for the standardized genetic variance components were rather high (range, [0.64–0.75]). The phosphatidylcholines (PCs), on the other hand, displayed relatively much heterogeneity with respect to their heritability: whereas for some lipids (notably C36:2_PC) the estimated heritability was very low, for others (*e.g.*, C36:4_PC) it was rather high. The total range of the estimated heritabilities for the lipids in this class was [0.25–0.77]. The heritability estimates for the sphingomyelins (SPMs) displayed a similar pattern as those for the LPCs: the estimated values for all lipids in this class were rather high (range, [0.47–0.71]). The cholesterol esters (ChEs) displayed a remarkable heterogeneity in their estimated heritabilities when considering the number of C-atoms in the fatty acid: whereas for the measured ChEs with 16 or 18 C-atoms in the fatty acid the estimates were moderate and in the range [0.42–0.48], for the lipids in this class with 20 or 22 C-atoms in the fatty acid the estimates were notably higher and in the range [0.71–0.74].

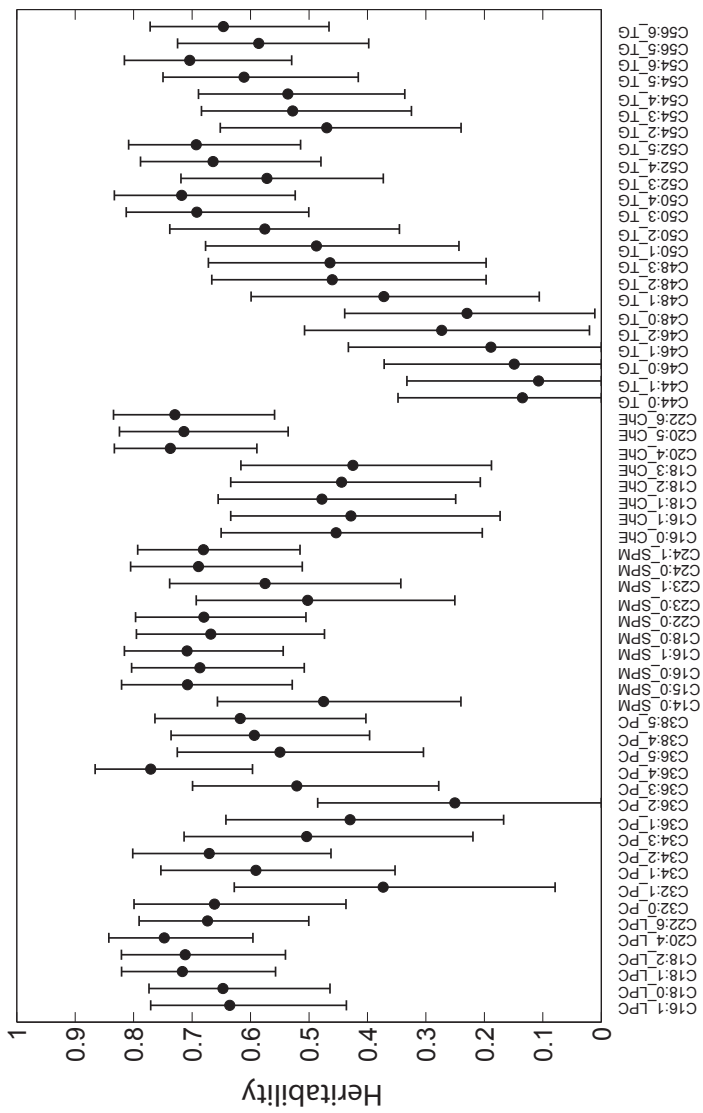

Figure 5.1: LC–MS lipid data: heritability estimates for all lipids under "AE" model. Dots indicate the original estimate for the standardized "$varA$" variance component under a univariate "AE" model; the whiskers denote the maximum likelihood-based 95%-confidence interval for this estimate. Because under the "AE" model, the values of the standardized "$varA$" and "$varE$" variance components add up to a value of one, the values of the "$varE$" variance component estimates can be inferred from this figure as well. For denotation of lipids, see Section 5.3.2.

In the triglycerides (TGs), the pattern was even more striking when considering the combination of the number of C-atoms as well as the number of double bonds in the fatty acids. For 44 up to 50 carbon atoms in the triglyceride, on average the heritabilities increased with additional carbon atoms in the fatty acid. From 50 up to 56 carbon atoms, on average the heritabilities did not change much. However, for each group of TGs with the same number of carbon atoms, with exception of C44, we observed a consistent upward trend in the heritability with increasing numbers of double bonds in the fatty acids. For example, for the TGs with 54 carbon atoms, the estimate for the heritability was always larger for lipids with larger numbers of double bonds in the fatty acids. These remarkable differences in heritability among TGs with different numbers of carbon atoms as well as different numbers of double bounds in the fatty acids might be due to different numbers of conversions by enzymes involved in both catabolism and anabolism of fatty acids. The apparently lower average heritabilities of TGs with numbers of carbon atoms decreasing from 50 up to 44, are perhaps due to increasing numbers of C2-fragment cleavages from the fatty acid backbone (during anabolism) by β-ketoacyl-CoA thiolase, and/or smaller numbers of C2-fragment attachments to the fatty acid backbone (during catabolism) by fatty acid synthase.[177] Similarly, the increases in heritability of TGs with the same number of carbon atoms but increasing numbers of double bonds in the fatty acid backbones, are perhaps due to increasing numbers of actions by enoyl-CoA isomerase and/or 2,4-dienyol-CoA reductase and 3,2-enoyl-CoA isomerase (during fatty acid catabolism), and/or smaller numbers of conversions by fatty acyl-CoA desaturases during fatty acid anabolism. Overall, for the TGs the heritability estimates were in the range [0.11–0.72].

Plasma ^{1}H NMR

The heritability estimates per variable, as well as the 95%-CI, are shown in Figure 5.2. Within the plasma ^{1}H NMR data there was much heterogeneity in the estimated heritabilities among different variables; this is as expected because in contrast to for instance the targeted LC–MS method used to generate the lipid data described in this chapter, NMR is considered a 'global' metabolomics method that should be able to detect metabolites of a much larger number of different classes (*e.g.*, amino acids, carbohydrates). It is conceivable that different classes of metabolites are subject to different (genetic and/or environmental) mechanisms that influence their phenotypic variance. Therefore, in data from a global method like NMR, it is expected that (widely) different relative contributions of genetic and environmental causes of variance are estimated for different metabolites.

Assignment of compound names on basis of the estimated heritabilities of the features detected in the ^{1}H NMR spectra alone is difficult, amongst others because the same compound may have a signal at multiple positions in the spectrum. Also, it is often difficult to elucidate which metabolites corre-

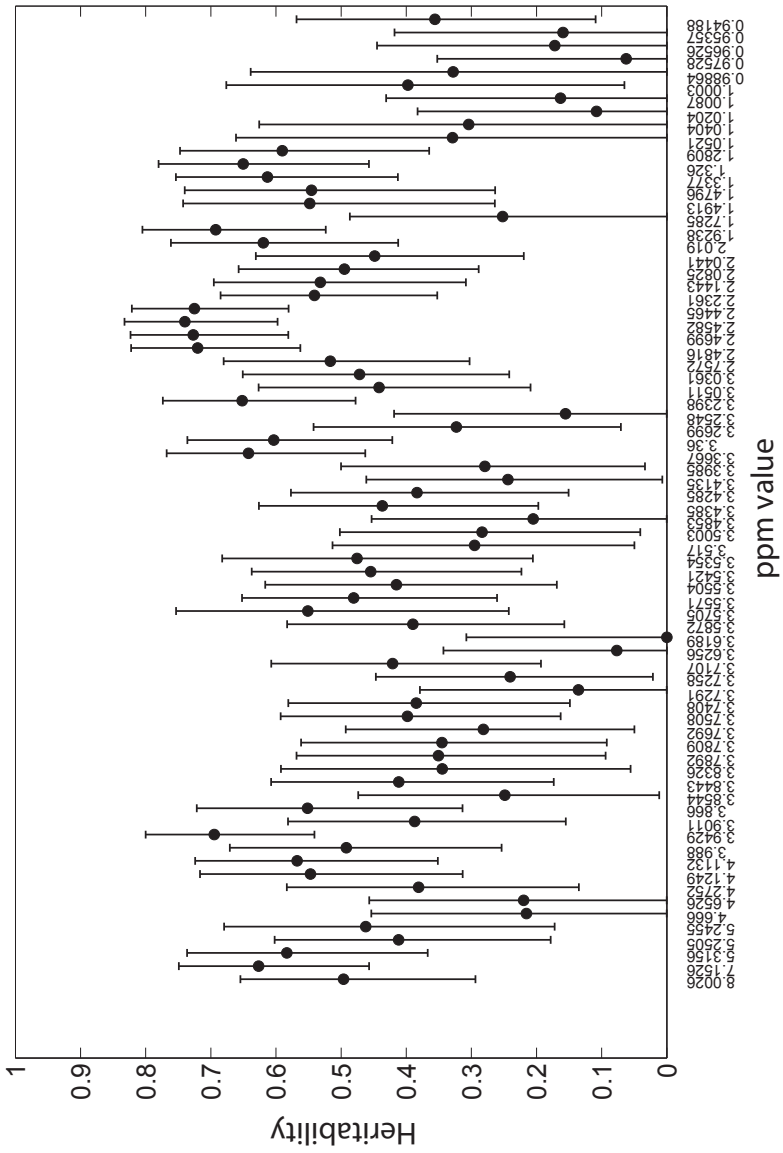

Figure 5.2: ^{1}H NMR data: heritability estimates for all features under "AE" model. For legend, see Figure 5.1. Features (variables) are indicated by their corresponding chemical shift (ppm) value; the variables are sorted from left to right along the horizontal axis in this figure in keeping with the order of the features as these occur in the original NMR spectrum, *i.e.* from high to low.

spond to the measured ppm values because peaks of multiple metabolites may overlap.[178] In our case, most compounds that we putatively linked to a particular combination of features (ppm values) on basis of an in-house reference database, did not display a consistent pattern of heritability for all features within such a combination.

5.4.3 Multivariate analyses

The heritabilities as computed on basis of the bivariate analyses resembled those as resulting from the univariate analyses; this is in line with previous findings.[173] The results of the multivariate analyses specific for each of the two data sets are given below.

LC–MS lipids

Figure 5.3 displays a heatmap of the dissimilarities that result from rescaling the genetic correlations, as well as the associated dendrogram resulting from hierarchical clustering based on these dissimilarities. For most pairs of lipids, the genetic correlations were larger than zero: the median correlation was 0.49 (range, [-0.41; 1]). This suggests that most of the lipids detected in this study have at least some common genetic causes of phenotypic variance. All LPCs clustered together perfectly; the TGs also clustered together very well although one PC (C36:2_PC) clustered together with the TGs because of a very high genetic correlation. For the clustering among the TGs, the number of double bonds in the fatty acid appears to be important: in Figure 5.3, two main clusters of TGs can be observed where one cluster consists of TGs with up to two double bonds in the fatty acid, whereas the TGs in the other cluster have two or more double bonds in their fatty acid chains. This may indicate the action of different enzymes in the metabolism of the TGs in the two different clusters. The SPMs also clustered together rather well, although the results suggest that they share genetic causes of variance with three ChEs (*i.e.*, C16:0_ChE, C18:1_ChE, and C18:2_ChE) as well. The PCs also have a tendency to cluster, although the clustering pattern suggests that also the lipids in this class share some genetic causes of variance with notably the ChEs.

Plasma ^{1}H NMR

Figure 5.4 shows the heatmap of the dissimilarities that result from rescaling the genetic correlations, as well as the associated dendrogram resulting from hierarchical clustering based on these dissimilarities. The median genetic correlation among the ^{1}H NMR variables was 0.08; range [-1; 1]. Note that this is in contrast with the situation for the LC–MS data, where almost all genetic correlations were larger than zero. This contrast might indeed be due to the fact that the LC–MS method used to generate the data analyzed in this study is a 'targeted' method that detects metabolites of the same class (in this case lipids) that indeed may share an important part of their biological pathways

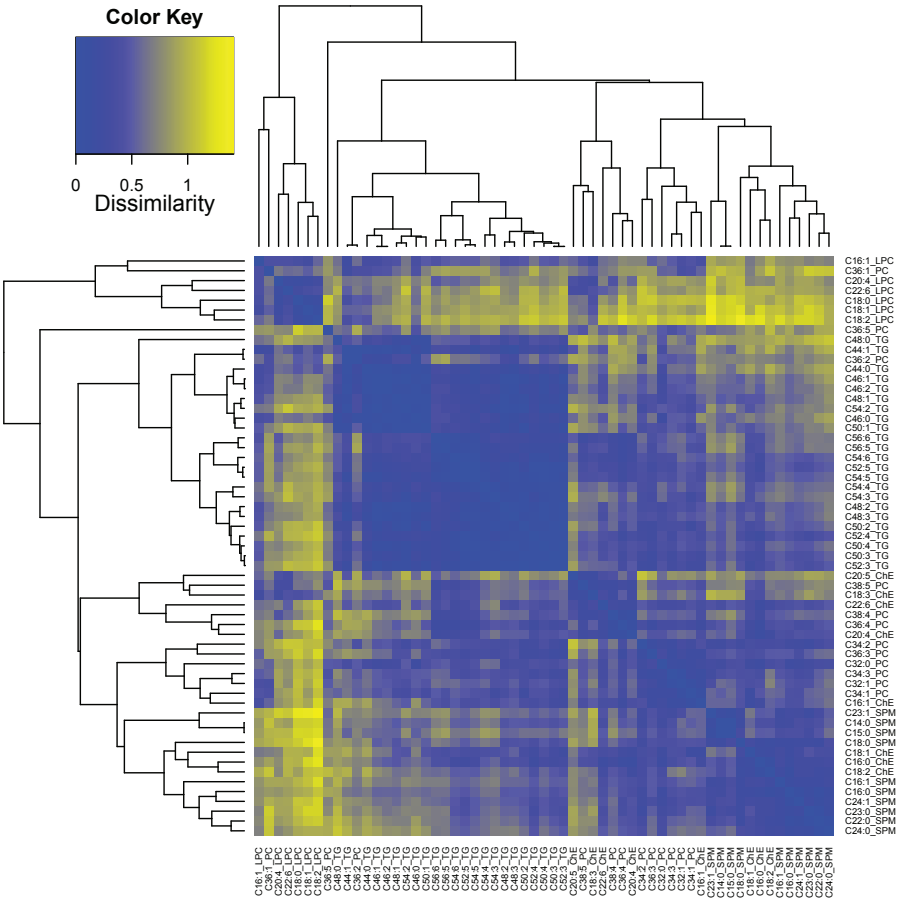

Figure 5.3: LC–MS lipid data: analysis of genetic correlation matrix through dissimilarities. The heatmap indicates with a color code for each pair of lipids the dissimilarity that results from rescaling of the genetic correlation as explained in Section 5.3. For example, a dissimilarity equal to zero as displayed in this figure corresponds to a genetic correlation of 1; a dissimilarity of 1 corresponds to a genetic correlation equal to zero. The average linkage algorithm was used for hierarchical clustering based on these dissimilarities; the resulting dendrogram is shown both along the horizontal and the vertical axes of the ordered heatmap. The Pearson correlation between the cophenetic distance matrix estimated from the dendrogram, and the original dissimilarity matrix based on genetic correlations, was equal to 0.77. The dissimilarity matrix was treated as being symmetric for producing this figure. Therefore, this figure is symmetric with the diagonal of the heatmap as the axis of symmetry; the dendrograms along the horizontal and vertical axes of the heatmap are mirrors of each other. For explanation of lipid labeling, see Section 5.3.2.

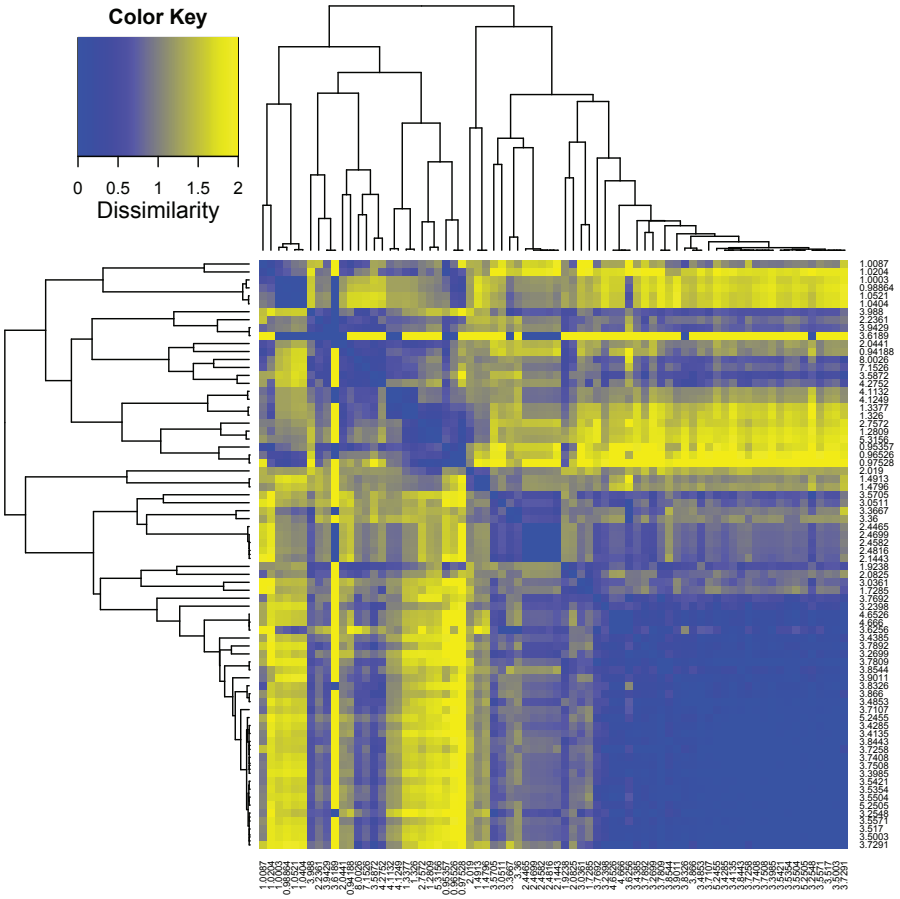

Figure 5.4: ^{1}H NMR data: analysis of genetic correlation matrix through dissimilarities. For explanation, see the legend to Figure 5.3. The Pearson correlation between the cophenetic distance matrix estimated from the dendrogram, and the original dissimilarity matrix based on genetic correlations, was equal to 0.79. Variables are denoted by the chemical shift values that correspond to the detected features.

to phenotypic variation. The ^{1}H NMR data, however, were generated using a 'global' method where indeed metabolites of a much larger number of classes may be detected that will share less biological pathways leading to phenotypic variation.

As already noted in the discussion of the results of our univariate analyses, interpretation of the results for the ^{1}H NMR data was often difficult due to the inherent properties of this metabolomics method. Nevertheless, we suspect that in future studies, hierarchical clustering on the basis of the genetic correlations among different peaks might be useful to reveal genetic relationships among different metabolites.

5.4.4 Quantitative genetic analyses of metabolomics data as reported in the literature

Other authors have given heritability estimates for metabolites as well. However, with the exception of a study by Shah and colleagues,[161] in all publications that we are aware of, the number of metabolites studied was too small and/or the resolution was too low (*e.g.*, all triglycerides lumped into one summary measure denoted "total triglyceride concentration") to denote the phenotypic data as "metabolomics data".[179–184] Furthermore, of note, a graphical example of variance components analysis of twin metabolomics data is given in Rahmioğlu *et al.* (their Figure 4).[185]

The reported study that is probably the closest to our current study is the one by Shah and co-workers.[161] In that study, quantitative measurements of 66 metabolites were performed belonging to acylcarnitine species, amino acids and free fatty acids, in blood plasma from eight (nontwin) families (total 117 individuals) heavily burdened with premature coronary artery disease. Univariate heritabilities were computed with methods that are equivalent to those employed in this chapter; however, the authors did not perform multivariate quantitative genetic analyses. Consistent with our findings, the authors report heritabilities within a large range over all investigated metabolites: for the metabolites for which the heritability estimate was statistically significant, they found heritabilities within the range [0.23–0.82].

In another interesting study, Pilia *et al.*[176] analyzed data on 98 quantitative traits relevant for cardiovascular function and personality, measured in a large cohort of Sardinians. These traits included levels of 'conventional metabolites' on the basis of clinical chemical measurements. In accordance with the current study, the authors used quantitative genetic methods in a "multistep multivariate" fashion to assess the genetic and environmental components of the phenotypic variances and covariances.

5.5 Conclusions

We have presented the results of a pilot investigation into the relative contribution of genetic variation to the variation observed in human blood plasma metabolite levels. Our analyses were based on the data obtained with two frequently used metabolomics platforms, *i.e.* LC–MS and ^{1}H NMR spectroscopy.

Notably, univariate quantitative genetic analysis of the lipid LC–MS data revealed a remarkable pattern in the heritabilities of TGs with different numbers of C-atoms and/or different numbers of double bonds in the fatty acids that may warrant further biochemical investigation. In multivariate analysis we found genetic covariance among lipids from the same lipid class (LC–MS). Therefore, we envision that the methods employed in this study can be used to discover novel biological pathways on the basis of "omics" type data obtained in families.

Due to the inherent properties of ^{1}H NMR, interpretation of the results based on these data was difficult. However, in general we found higher genetic covariance observed among variables observed with the 'targeted' lipid LC–MS platform with respect to those observed with the 'global' ^{1}H NMR metabolomics platform.

In conclusion, our study has demonstrated the use of uni- and multivariate quantitative genetic analysis to elucidate the importance of genetic variation to quantitative variation observed in human blood plasma metabolites. The statistical significance of our findings should be enhanced by replication in a larger cohort of families.

5.6 Acknowledgments

We thank all the twins and siblings who participated in this study. We gratefully acknowledge dr. MC Neale (Virginia Commonwealth University, VA, USA) and dr. MHM de Moor (VU University Amsterdam, Amsterdam, The Netherlands) for assistance with the OpenMx software. Furthermore we would like to acknowledge support from the Netherlands Bioinformatics Centre (NBIC) through its research programme BioRange (project number: SP 3.3.1); the Netherlands Metabolomics Centre; Spinozapremie NWO/SPI 56-464-14192; the Center for Medical Systems Biology (CMSB); Twin-family database for behavior genetics and genomics studies (NWO-MaGW 480-04-004) and NWO-MaGW Vervangingsstudie (NWO no. 400-05-717).