Article

Genetics and Not Shared Environment Explains Familial Resemblance in Adult Metabolomics Data

René Pool[1,2]* , Fiona A. Hagenbeek[1,2]* , Anne M. Hendriks[1,2] , Jenny van Dongen[1,2] , Gonneke Willemsen[1] , Eco de Geus[1,2] BBMRI Metabolomics Consortium[3], Ko Willems van Dijk[4,5,6] , Aswin Verhoeven[7] , H. Eka Suchiman[8] , Marian Beekman[8] , P. Eline Slagboom[8] , Amy C. Harms[9,10] , Thomas Hankemeier[9,10] and Dorret I. Boomsma[1,2]

[1]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands, [2]Amsterdam Public Health Research Institute, Amsterdam, the Netherlands, [3]Members of the BBMRI Metabolomics Consortium are listed after the abstract, [4]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands, [5]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands, [6]Department of Internal Medicine Division Endocrinology, Leiden University Medical Center, Leiden, the Netherlands, [7]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, the Netherlands, [8]Department of Biomedical Data Sciences, Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands, [9]Division of Analytical Biosciences, Leiden Academic Center for Drug Research, Leiden University, Leiden, the Netherlands and [10]The Netherlands Metabolomics Centre, Leiden, the Netherlands

Abstract

Metabolites are small molecules involved in cellular metabolism where they act as reaction substrates or products. The term 'metabolomics' refers to the comprehensive study of these molecules. The concentrations of metabolites in biological tissues are under genetic control, but this is limited by environmental factors such as diet. In adult mono- and dizygotic twin pairs, we estimated the contribution of genetic and shared environmental influences on metabolite levels by structural equation modeling and tested whether the familial resemblance for metabolite levels is mainly explained by genetic or by environmental factors that are shared by family members. Metabolites were measured across three platforms: two based on proton nuclear magnetic resonance techniques and one employing mass spectrometry. These three platforms comprised 237 single metabolic traits of several chemical classes. For the three platforms, metabolites were assessed in 1407, 1037 and 1116 twin pairs, respectively. We carried out power calculations to establish what percentage of shared environmental variance could be detected given these sample sizes. Our study did not find evidence for a systematic contribution of shared environment, defined as the influence of growing up together in the same household, on metabolites assessed in adulthood. Significant heritability was observed for nearly all 237 metabolites; significant contribution of the shared environment was limited to 6 metabolites. The top quartile of the heritability distribution was populated by 5 of the 11 investigated chemical classes. In this quartile, metabolites of the class lipoprotein were significantly overrepresented, whereas metabolites of classes glycerophospholipids and glycerolipids were significantly underrepresented.

Keywords: Classical twin design; enrichment analysis; heritability; metabolite classes; shared environment

(Received 9 March 2020; accepted 5 May 2020)

BBMRI Metabolomics Consortium

Cohort Collection and Sample Management Group: M. Beekman[1], H. E. D. Suchiman[1], N. Amin[2], J. W. Beulens[3,4], J. A. van der Bom[5–8], N. Bomer[9], A. Demirkan[2], J. A. van Hilten[10], J. M. T. A. Meessen[11], R. Pool[12], M. H. Moed[1], J. Fu[13,14], G. L. J. Onderwater[15], F. Rutters[3], C. So-Osman[10], W. M. van der Flier[3,16], A. A. W. A. van der Heijden[17], A. van der Spek[2], F. W. Asselbergs[18], E. Boersma[19],

P. M. Elders[20,21], J. M. Geleijnse[22], M. A. Ikram[2,23,24], M. Kloppenburg[8,25], I. Meulenbelt[1], S. P. Mooijaart[26], R. G. H. H. Nelissen[27], M. G. Netea[28,29], B. W. J. H. Penninx[21,30], C.D.A. Stehouwer[31,32], C.E. Teunissen[33], G. M. Terwindt[15], L. M. 't Hart[1,3,21,34,35], A. M. J. M. van den Maagdenberg[36], P. van der Harst[8], I. C. C. van der Horst[37], C. J. H. van der Kallen[31,32], M. M. J. van Greevenbroek[31,32], W. E. van Spil[38], C. Wijmenga[13], A. H. Zwinderman[39], A. Zhernikova[13], J. W. Jukema[40]

Database & Catalogue: J. J. H. Barkey Wolf[1], M. Beekman[1], D. Cats[1], H. Mei[1,41], M. Slofstra[13], M. Swertz[13]

Quality Control: E. B. van den Akker[1,42,43], J. J. H. Barkey Wolf[1], J. Deelen[1,44], M. J. T. Reinders[42,43]

Steering Committee: D. I. Boomsma[21,45], C. M. van Duijn[2], P.E. Slagboom[1]

[1]Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.

[2]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands.

[3]Department of Epidemiology and Biostatistics, Amsterdam University Medical Center, Amsterdam, the Netherlands.

[4]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands.

[5]Centre for Clinical Transfusion Research, Sanquin Research, Leiden, the Netherlands.

[6]Jon J van Rood Centre for Clinical Transfusion Research, Leiden University Medical Centre, Leiden, the Netherlands.

[7]TIAS, Tilburg University, Tilburg, the Netherlands.

[8]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, the Netherlands.

[9]Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

[10]Center for Clinical Transfusion Research, Sanquin Research, Leiden, the Netherlands.

[11]Department of Orthopedics, Leiden University Medical Centre, Leiden, The Netherlands.

[12]Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands.

[13]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

[14]Department of Pediatrics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

[15]Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands.

[16]Department of Neurology and Alzheimer Center, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, the Netherlands.

[17]Department of General Practice, The EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands.

[18]Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands.

[19]Thorax Centre, Erasmus Medical Centre, Rotterdam, the Netherlands.

[20]Department of General Practice and Elderly Care Medicine, VU University Medical Center, Amsterdam, the Netherlands.

[21]Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, the Netherlands.

[22]Division of Human Nutrition and Health, Wageningen University, Wageningen, the Netherlands.

[23]Department of Radiology, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands.

[24]Department of Neurology, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands.

[25]Department of Rheumatology, Leiden University Medical Center, the Netherlands.

[26]Department of Internal Medicine, Division of Gerontology and Geriatrics, Leiden University Medical Centre, Leiden, the Netherlands.

[27]Department of Orthopaedics, Leiden University Medical Center, Leiden, the Netherlands.

[28]Department of Internal Medicine, Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands.

[29]Department for Genomics & Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany.

[30]Department of Psychiatry, VU University Medical Center, Amsterdam, the Netherlands.

[31]Department of Internal Medicine, Maastricht University Medical Center (MUMC+), Maastricht, the Netherlands.

[32]School for Cardiovascular Diseases (CARIM), Maastricht University, Maastricht, the Netherlands.

[33]Neurochemistry Laboratory, Clinical Chemistry Department, Amsterdam University Medical Center, Amsterdam Neuroscience, the Netherlands.

[34]Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands.

[35]Department of General practice, Amsterdam University Medical Center, Amsterdam, the Netherlands.

[36]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands.

[37]Department of Critical Care, University Medical Center Groningen, Groningen, the Netherlands.

[38]UMC Utrecht, Department of Rheumatology & Clinical Immunology, Utrecht, the Netherlands.

[39]Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands.

[40]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands.

[41]Sequencing Analysis Support Core, Leiden University Medical Center, Leiden, the Netherlands.

[42]Leiden Computational Biology Center, Leiden University Medical Center, Leiden, the Netherlands.

[43]Department of Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, the Netherlands.

[44]Max Planck Institute for Biology of Ageing, Cologne, Germany.

[45]Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands.

Metabolites (small molecules involved in biological processes) are important intermediates in understanding how a person's genotype translates to health or disease. Many different types of metabolites can be distinguished, including amino acids, lipids, and sugars (Adamski & Suhre, 2013). Due to their diversity, metabolites have various functions in the human body, including energy storage, signaling, and forming structures, such as proteins or cell walls (Dunn et al., 2011). Thus, metabolites can be considered the building blocks of the body (Hasirci & Hasirci, 2018). The complete set of metabolites found within a specific biological sample is referred to as the metabolome (Wishart, 2007), the study thereof is termed metabolomics (Fiehn, 2002). The metabolome is downstream of gene transcription, protein translation and protein function; therefore, metabolites are close to observable phenotypes in health and disease (Draisma et al., 2013; Goodacre et al., 2004). Metabolomics has been successful in identifying disease biomarkers, unraveling biological mechanisms, and for drug discovery and development (Pang et al., 2019).

The metabolome differs between people, as metabolite levels are influenced by many exogenous (originating from outside an organism) and endogenous factors (originating from inside an organism). Exogenous factors influencing the human metabolome include lifestyle, diet or medication use. For example, metabolite levels differ between current-, former- and never-smokers, between individuals on a low fat, low glycemic, or very low

Table 1. Participant characteristics by the Nightingale, lipidomics and NMR-LUMC platforms

Platform	Age (SD)	BMI (SD)	F_{Fml}	F_{Smkng}	F_{LLMd}	N_{MZPrs}	N_{DZPrs}
Nightingale	35.23 (10.31)	23.92 (3.79)	0.69	0.21	0.05	886	601
Lipidomics	35.71 (10.32)	23.97 (3.91)	0.69	0.21	0.05	643	524
NMR-LUMC	35.10 (10.78)	24.02 (3.84)	0.68	0.21	0.04	663	407

Note: Age (SD) denotes the mean age and standard deviation, BMI (SD) the mean BMI and standard deviation, F_{Fml} the fraction of female subjects, F_{Smkng} the fraction of current smokers, F_{LLMd} the fraction of subjects using lipid lowering medication, N_{MZPrs} the number of complete monozygotic twin pairs and N_{DZPrs} the number of complete dizygotic twin pairs.

carbohydrate diet, and among users of various medication classes (Esko et al., 2017; J. Liu et al., 2020; Xu et al., 2013). Endogenous factors influencing the human metabolome include sex, age or body mass index (BMI). For example, metabolite levels differ between males and females, younger and older individuals and obese and nonobese individuals (Chaleckis et al., 2016; Fan et al., 2018; Rangel-Huerta et al., 2019). Endogenous factors also include genetic influences, either directly on metabolite levels or indirectly through the effect on behavior or lifestyle (e.g., smoking; M. Liu et al., 2019). The metabolome can differ between cases and controls — for example, in major depressive disorder (Bot et al., 2019). Here it is observed that associations between metabolites and the case/control status are attenuated by antidepressant use, while the causalities of the associations remain unknown as of yet. Thus, metabolite levels reflect individual differences in genetics, physiology, lifestyle and behavior or responses to environmental changes (Fiehn, 2002).

Genetic factors account for approximately 50% of the individual differences in metabolite levels (Shin et al., 2014; Yet et al., 2016). The average proportion of phenotypic variance in metabolite levels ascribed to genetic factors (i.e., heritability; h^2) differs per type of metabolite. The median heritability for lipids is approximately 37%, with heritability estimates for sphingolipids (e.g., sphingomyelins) and glycerolipids (e.g., triglycerides) often, but not consistently, higher than for phospholipids (e.g., phosphatidylcholines; Bellis et al., 2014; Darst et al., 2019; Frahnow et al., 2017). Similarly, while the median heritability for amino acids is approximately 40% (Darst et al., 2019), amino acids that the body is able (nonessential) or unable (essential) to synthesize *de novo* differ in mean heritability. Specifically, levels of essential amino acids are less heritable than levels of nonessential amino acids (Rhee et al., 2013). These differences in heritability among metabolite classes also occur for single-nucleotide polymorphism (SNP)-based heritability (i.e., heritability estimates derived from genomewide SNPs; Rhee et al., 2016; Tabassum et al., 2019). In fact, the observed differences in additive heritability estimates among metabolite classes are rarely significant, while differences in heritability estimates based on known genetic variants are frequently significant (Hagenbeek et al., 2020).

Whereas the contribution of genetic variants to metabolite levels is fairly well established, also through genome-wide association and (exome-) sequencing studies (Hagenbeek et al., 2020; Kastenmüller et al., 2015; Yazdani et al., 2019), the contribution of the shared or common environment (c^2) to metabolite levels is not as well characterized. Not all studies investigating the genetic contribution to metabolite levels were based on the classic twin design (e.g., Draisma et al., 2013; Tremblay et al., 2019). Instead, these types of methods estimate the familial resemblance (or generalized heritability) of metabolite levels, which comprises both additive genetic effects and common environment effects shared by family members (Rice, 2008). Studies using the classic

twin design to investigate the contributions to metabolite levels vary widely with respect to how metabolites are influenced by common environment as well as in the estimate of the effects of the common environment. Overall, it would appear that studies in smaller samples more frequently report larger contributions of the common environment to metabolite levels (e.g., Frahnow et al., 2017; Kettunen et al., 2012). Large-scale twin studies that estimate the contribution of shared environment tend to be scarce.

In the current study, we aim to expand our understanding of the contribution of common environment shared by family members to variation in fasting blood metabolites and analyzed data from multiple metabolomics platforms from a large cohort of twins (between 1037 and 1407 twin pairs per platform), representing a general population. First, a series of power analyses were performed, estimating the statistical power to detect shared environment in the classical twin design, given the number of monozygotic (MZ) and dizygotic (DZ) twin pairs available in our study. The power to detect shared environment in quantitative genetic studies, employing the classical twin design, is influenced by effect size, the heritability of the trait, the sample size, the probability level that is chosen, and the homogeneity of means and variances in the MZ and DZ groups of the sample (Martin et al., 1978; Posthuma & Boomsma, 2000). Sample size and probability were given, and we investigated different values for the proportion of variation explained by shared environment against a background of different heritability values that were chosen based on what is typically reported in the literature for metabolomics. Next, we determined the heritability of all metabolic traits by structural equation modeling where contributions to additive genetic effects (A), shared environmental effects between siblings (C) and unique environmental effects (E) were estimated. By computing the significance of the C variance component when comparing ACE model outcomes to AE model outcomes, we were able to assess whether C contributes to the total variance observed in all metabolic traits. Finally, to obtain insight into the distribution of chemical classes our metabolites belong to over the range of calculated heritabilities and contributions of shared environment, we performed enrichment analyses. The metabolic traits were grouped in heritability and shared environment estimate quartiles. By counting the number of metabolites of a given class per quartile and comparing these to the counts of the entire range of heritabilities or common environment contributions, we determined class enrichment factors per quartile and assessed their statistical significance by Fisher's exact tests.

Methods

Participants

At the Netherlands Twin Register (NTR; Ligthart et al., 2019) metabolomics data for twins were available for 886 complete MZ pairs and 601 complete DZ pairs (c.f. Table 1). All

Table 2. Chemical class counts for each platform used

	Platform			
Chemical class	Nightingale	Lipidomics	NMR-LUMC	ALL
Lipoprotein	64			64
Glycerophospholipids	2	61		63
Glycerolipids		37		37
Carboxylic acids and derivatives	8		22	30
Sphingolipids		20		20
Organooxygen compounds			8	8
Hydroxy acids and derivatives	2		4	6
Steroids and steroid derivatives	1	2		3
Keto acids and derivatives			2	2
Diradylglycerols		2		2
Organonitrogen compounds			2	2

Note: The numbers in the platform columns represent the number of compounds of a chemical class, where '—' means that there are none.

measurements were performed in blood samples that were collected from participants of the NTR biobank projects (Willemsen et al., 2010; Willemsen et al., 2013). Blood samples were collected after overnight fasting. Fertile women were bled in their pill-free week or on day 2−4 of their menstrual cycle. For the Nightingale Health metabolomics platform (see below), data were acquired in several shipments (subsets). After completing the preprocessing of the metabolomics data, each platform subset (if applicable) was merged into a single per platform dataset, randomly retaining a single observation per platform whenever multiple observations were available. Characteristics for the sample of individuals included in the analyses can be found in Table 1. Informed consent was obtained from all participants. Projects were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance-FWA00017598; IRB/institute codes and NTR 03-180).

Metabolite Profiling

Below, we briefly describe the methods used for metabolite profiling. For more detailed information, see Hagenbeek et al. (2020).

Nightingale Health ^{1}H-NMR platform. Metabolic biomarkers were quantified from plasma samples using high-throughput proton nuclear magnetic resonance spectroscopy (^{1}H-NMR) metabolomics (Nightingale Health Ltd, Helsinki, Finland). This method provides simultaneous quantification of routine lipids, lipoprotein subclass profiling with lipid concentrations within 14 subclasses, fatty acid composition, and various low-molecular weight metabolites, including amino acids, ketone bodies and glycolysis-related metabolites in molar concentration units. Details of the experimentation and epidemiological applications

of the NMR metabolomics platform have been reviewed previously (Soininen et al., 2015; Würtz et al., 2017).

UPLC-MS lipidomics platform. Plasma lipid profiling was performed at the division of Analytical Biosciences at the Leiden Academic Center for Drug Research at Leiden University/ Netherlands Metabolomics Centre. The lipids were analyzed with an Ultra-High-Performance Liquid Chromatograph directly coupled to an Electrospray Ionization Quadruple Time-of-Flight high-resolution mass spectrometer (UPLC-ESI-Q-TOF; Agilent 6530, San Jose, CA, USA) that uses reference mass correction. For liquid chromatographic separation, a ACQUITY UPLC HSS T3 column (1.8 μm, 2.1 × 100 mm) was used with a flow of 0.4 ml/min over a 16-min gradient. Lipid detection was performed using a full scan in the positive ion mode. The raw MS data were pre-processed using Agilent MassHunter Quantitative Analysis software (Agilent, Version B.04.00). Detailed descriptions of lipid profiling and quantification have been described previously (Dane et al., 2014; Gonzalez-Covarrubias et al., 2013).

Leiden ^{1}H-NMR platform (for small metabolites). The Leiden ^{1}H-NMR spectroscopy experiment of ethylenediaminetetraacetic acid plasma samples used a 600-MHz Bruker Advance II spectrometer (Bruker BioSpin, Karlsruhe, Germany). The peak deconvolution method used for this platform has been previously described (Demirkan et al., 2015; Verhoeven et al., 2017).

Metabolomics Data Preprocessing

To ensure our data were consistent with Hagenbeek, Pool, van Dongen, Draisma, Boomsma et al. (2020), we excluded participants if they were on lipid-lowering medication at the time of blood draw or if they had not adhered to the fasting protocol (~4 % of the sample of each platform). Preprocessing of the metabolomics data was executed for each of the platforms and measurement/shipment batches per platform separately. Metabolites were excluded from analysis when the mean coefficient of variation exceeded 25% or the missing rate exceeded 5%. Metabolite measurements were set to missing if they were below the lower limit of detection or quantification or could be classified as an outlier (five standard deviations greater or smaller than the mean). Metabolite measurements that were set to missing because they fell below the limit of detection/quantification were imputed with half of the value of this limit, or when this limit was unknown with half of the lowest observed level for this metabolite. All remaining missing values were imputed using multivariate imputation by chained equations ('mice'; van Buuren & Groothuis-Oudshoorn, 2011). On average, nine values had to be imputed for each metabolite ($SD = 12$; range: 1−151). Data for each metabolite on the lipidomics platform and both ^{1}H-NMR platforms were normalized by inverse normal rank transformation (Demirkan et al., 2015; Kettunen et al., 2016).

We computed heritability for 237 single metabolic traits (i.e., no ratios or composite variables, see Supplementary Table S1). These traits are members of 11 different chemical classes, as listed by the human metabolome database (Hagenbeek et al., 2020; Wishart et al., 2018). As is shown in Table 2, most metabolites are lipid species.

To account for confounding by age and sex, we used the residuals of the linear fit of model $M_i \sim$ Age + Sex for each metabolite i (M_i) as input for the statistical analyses. The above data processing steps were performed in the Jupyterlab

Table 3. Statistical power estimates as a function of values for A, C, r_{MZ} and r_{DZ} for the metabolomics platforms in this work

				Nightingale		Lipidomics		NMR-LUMC	
A	C	r_{MZ}	r_{DZ}	Power	χ^2	Power	χ^2	Power	χ^2
0.2	0.1	0.3	0.2	0.212	1.343	0.192	1.177	0.163	0.943
0.2	0.2	0.4	0.3	0.685	5.964	0.630	5.255	0.533	4.182
0.2	0.3	0.5	0.4	0.975	15.383	0.958	13.625	0.906	10.766
0.3	0.1	0.4	0.25	0.225	1.449	0.204	1.275	0.172	1.017
0.3	0.2	0.5	0.35	0.728	6.585	0.675	5.826	0.574	4.610
0.3	0.3	0.6	0.45	0.986	17.339	0.975	15.419	0.935	12.117
0.4	0.1	0.5	0.3	0.243	1.591	0.220	1.407	0.184	1.115
0.4	0.2	0.6	0.4	0.775	7.377	0.726	6.555	0.622	5.157
0.4	0.3	0.7	0.5	0.993	19.799	0.987	17.666	0.960	13.820
0.5	0.1	0.6	0.35	0.265	1.773	0.240	1.574	0.199	1.240
0.5	0.2	0.7	0.45	0.824	8.368	0.779	7.462	0.676	5.842
0.5	0.3	0.8	0.55	0.997	22.849	0.994	20.432	0.978	15.937
0.6	0.1	0.7	0.4	0.292	1.999	0.266	1.782	0.218	1.396
0.6	0.2	0.8	0.5	0.871	9.578	0.833	8.562	0.734	6.682
0.6	0.3	0.9	0.6	0.999	25.562	0.998	23.760	0.990	18.528
0.7	0.1	0.8	0.45	0.325	2.272	0.296	2.030	0.242	1.585
0.7	0.2	0.9	0.55	0.912	11.017	0.880	9.851	0.791	7.687

Note: A, additive genetic effects; C, shared environmental effects between siblings; E, unique environmental effects; MZ, monozygotic; DZ, dizygotic. For determining the estimates, we applied N(MZ pairs) = 848 and N(DZ pairs) = 559 for the Nightingale platform, N(MZ pairs) = 612 and N(DZ pairs) = 504 for the lipidomics platform, and N(MZ pairs) = 648 and N(DZ pairs) = 389 for the NMR-LUMC platform.

environment (v0.35.4), running an IPython kernel (v5.1.1: python v3.7.3) and utilizing modules pandas (v0.24.2), scipy (v1.2.1) and statsmodels (v0.9.0). The complete lists with full names of all detected metabolites that survived QC and preprocessing for all platforms can be found in Supplementary Table S1.

Statistical Analyses

Power analyses. We calculated covariance matrices for multiple combinations of heritability and the proportion of variation explained by common environment (C). Background heritability differed between 0.2 and 0.7, which are typical values estimated for metabolites (Hagenbeek et al., 2020). Within heritability class, the effect of C was increased from 0.1 to 0.3. Power analyses were carried out in the statistical software package Mx (Maes et al., 2009; Neale, 1997), with estimation of parameters by normal theory maximum likelihood. Goodness-of-fit testing was based on likelihood ratio tests. First, an ACE model was considered, and next the influence of C was constrained at 0 and power was obtained for a 1 degree of freedom test with $p = .05$. Note that we do not take into account the number of tests performed, which would lower the probability level (and hence lower the power to detect C).

Genetic analyses. For each metabolite M_i, we estimated the additive genetic contribution (A) and the contributions of common and unique environment (C and E). We applied genetic structural equation modeling using maximum likelihood estimation on both ACE and AE models. By comparing the outcomes of both models, for each metabolite, we applied a threshold of 3.84 on the χ^2 statistic, above which we considered the contribution of C significant. The analyses were performed in the R software package

(v3.5.2) using the OpenMx (v2.13.2.161) library (applying the NPSOL optimizer) for running the ACE and AE models (Boker et al., 2011).

Analyses of chemical class enrichment over heritability percentile groups. We subdivided the outcomes of the additive genetic components of the heritability or common environment variance component into three groups: (1) the 0−25 percentile group, (2) the 25−75 percentile group and (3) the 75−100 percentile group. In each group, we counted the number of class memberships of the metabolites. Within each group, we performed Fisher's exact tests (Fisher, 1922) of the group chemical class counts with respect to their counts in the entire sample.

Results

Power Analyses

The outcomes of the power analyses are summarized in Table 3. As can be observed from these results, for all three platforms, when A is in the range of 0.2–0.6, a larger sample size is needed to enable detection of C ≤ 0.2. This implies that when we observe a statistically significant C variance component, its minimal value needs to be approximately 0.2.

Genetic Analyses

The results of the ACE models for each platform and across the platforms are listed in Table 4. The results of the ACE and AE models are summarized graphically in Figure 1. On average, the analyses included 694 (range 612–848) and 503 (range 389–559) MZ and DZ pairs, respectively. For 6 of the 237 metabolites, we observe a significant contribution of c^2 ranging

Table 4. Summary of the AE (a) and ACE models (b and c) per platform and combined across platforms. Table 4b lists the ACE component means and variances only for the traits that exhibited a significant contribution of C. Table 4c lists the ACE component means and variances irrespective of the significance of the contribution of C

Dataset	a^2	e^2	N_M	$N_{MZPairs}$	$N_{DZPairs}$	$N_{TwinPairs}$
(a)						
Nightingale	0.56 [0.25, 0.65]	0.44 [0.35, 0.75]	76	848	559	1407
Lipidomics	0.42 [0.17, 0.57]	0.59 [0.43, 0.86]	122	612	504	1116
NMR-LUMC	0.41 [0.15, 0.74]	0.60 [0.27, 0.85]	33	648	389	1037
ALL	0.46 [0.14, 0.74]	0.54 [0.27, 0.86]	231	695	506	1200

Dataset	a^2	c^2	e^2	N_M	$N_{MZPairs}$	$N_{DZPairs}$	$N_{TwinPairs}$
(b)							
Nightingale	0.37 [0.37, 0.37]	0.18 [0.18, 0.18]	0.45 [0.45, 0.45]	1	848	559	1407
Lipidomics	NA	NA	NA	0	NA	NA	NA
NMR-LUMC	0.20 [0.16, 0.25]	0.27 [0.20, 0.34]	0.53 [0.46, 0.64]	5	648	389	1037
ALL	0.23 [0.16, 0.37]	0.25 [0.18, 0.34]	0.52 [0.45, 0.64]	6	681	417	1099

Dataset	a^2	c^2	e^2	N_M	$N_{MZPairs}$	$N_{DZPairs}$	$N_{TwinPairs}$
(c)							
Nightingale	0.50 [0.18, 0.63]	0.06 [0.00, 0.18]	0.44 [0.35, 0.75]	77	848	559	1407
Lipidomics	0.40 [0.14, 0.57]	0.01 [0.00, 0.16]	0.59 [0.43, 0.86]	122	612	504	1116
NMR-LUMC	0.35 [0.08, 0.74]	0.07 [0.00, 0.34]	0.58 [0.27, 0.85]	38	648	389	1037
ALL	0.43 [0.08, 0.74]	0.04 [0.00, 0.34]	0.54 [0.27, 0.86]	237	695	506	1200

Note: A, additive genetic effects; C, shared environmental effects between siblings; E, unique environmental effects; MZ, monozygotic; DZ, dizygotic. For each metabolomics platform, the additive genetic variance component a^2 and the variance component of the common environment c^2 are listed. These outcomes are based on N_M metabolites of which N_M(ACE) have a significant C component. The analyses are based on $N_{MZPairs}$ and $N_{DZPairs}$ of MZ and DZ twin pairs. The bottom row lists the statistics over all platforms. Values for the individual metabolites are listed in Supplementary Table S1.

from 0.175 to 0.423 (mean = 0.254). Over 6 ACE models and 231 AE models, the mean additive genetic variance component a^2 is 0.456 (range 0.136–0.735) and the mean of the residual error and unique environment variance component e^2 is 0.527 (range 0.265–0.864). Supplementary Figures S1 and S2 depict the model outcomes for the AE and ACE models, respectively. Supplementary Figure S2 shows the MZ and DZ twin correlations for all metabolites.

Analyses of Chemical Class Enrichment Over Percentile Groups of Heritabilities and Shared Environment Variance Components

Only six metabolic traits exhibit a significant shared environment variance component, rendering it futile to perform an enrichment analysis over C percentile groups. As can be observed in Figure 1, it seems that the lipoprotein chemical class is overrepresented in the top heritability percentile group and underrepresented in percentiles 25–50 and 0–25. This is confirmed by enrichment analyses of the chemical groups across the different heritability percentile groups (Table 5). Moreover, chemical classes glycerophospholipids and glycerolipids are significantly underrepresented in the top percentile group 75–100. The parallel coordinates plots in Figure 2 summarize these findings graphically.

Discussion

Metabolites have an important role in the relationship between the genotype and health and disease. Therefore, characterizing the factors influencing metabolite levels is a vital first step toward elucidating the mechanisms underlying health and disease. It is well established that metabolite levels are influenced by a complex interplay of genetic and environmental influences; however, the role of the common, or shared, environment in metabolite levels remains unclear. Here, we investigated the contribution of the common environment to variation in fasting blood metabolite levels in a large twin cohort. Although our study had sufficient power to detect a contribution of the common environment to the metabolite levels of 20% or higher, we found little evidence for a contribution of this size. In contrast, but in line with expectations, we found that additive genetics contributed significantly to metabolite levels. We found that the top 25% most heritable metabolites included mostly lipoproteins, while lipoproteins were underrepresented in the other.

Our conclusions need to be placed in the context of the design, platforms and biofluid that were used. First, while we established that with a fixed sample size, a probability level of 5% and depending on the background heritability, a contribution of common environment of 20% can be detected with reasonable power. A contribution of the common environment of 30% had very high statistical power for all platforms. However, when the contribution of the common environment is small, even for the largest sample sizes in our study (Nightingale platform with 848 MZ and 559 DZ twin pairs), the power to detect common environment shared by twins is low. Second, our study was cross-sectional, not longitudinal. While the fasting metabolite levels of individuals are generally stable over time (Lacruz et al., 2018), age is a
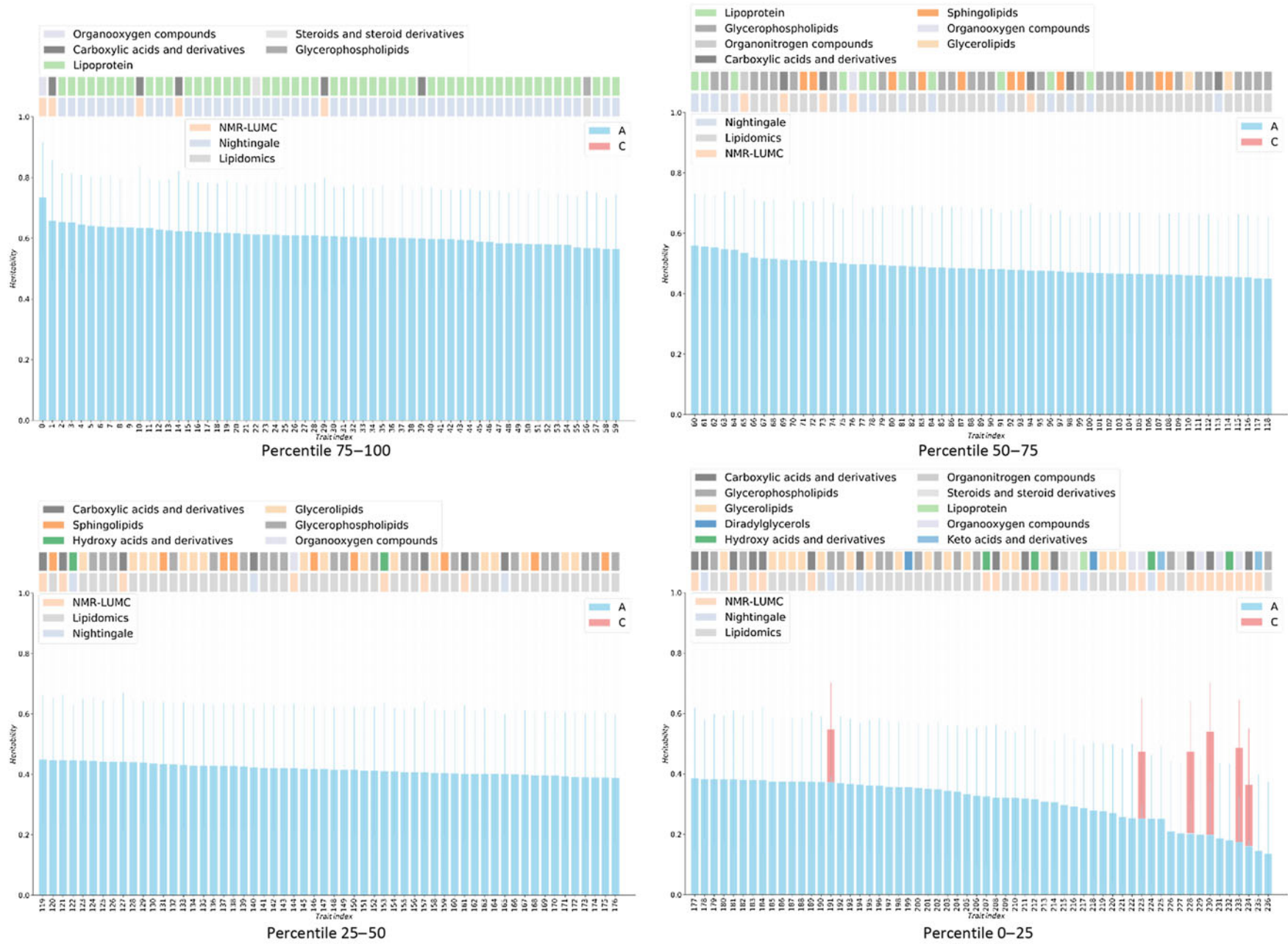
Fig. 1. Heritabilities for twin AE or ACE models. For the six metabolites showing a red bar, a significant contribution of C was observed ($\chi^2 > 3.84$). For all other metabolites, only estimates for the additive genetic component are shown, as determined by AE models. On the top of bar plot, two color bars are depicted that indicate the chemical class of the metabolite (top bars) and the metabolomics platform of the metabolite was reported from (bottom bars). The x-axes denote the indices of the metabolites, listed in Supplementary Table S1.

Note: A, additive genetic effects; C, shared environmental effects between siblings; E, unique environmental effects.
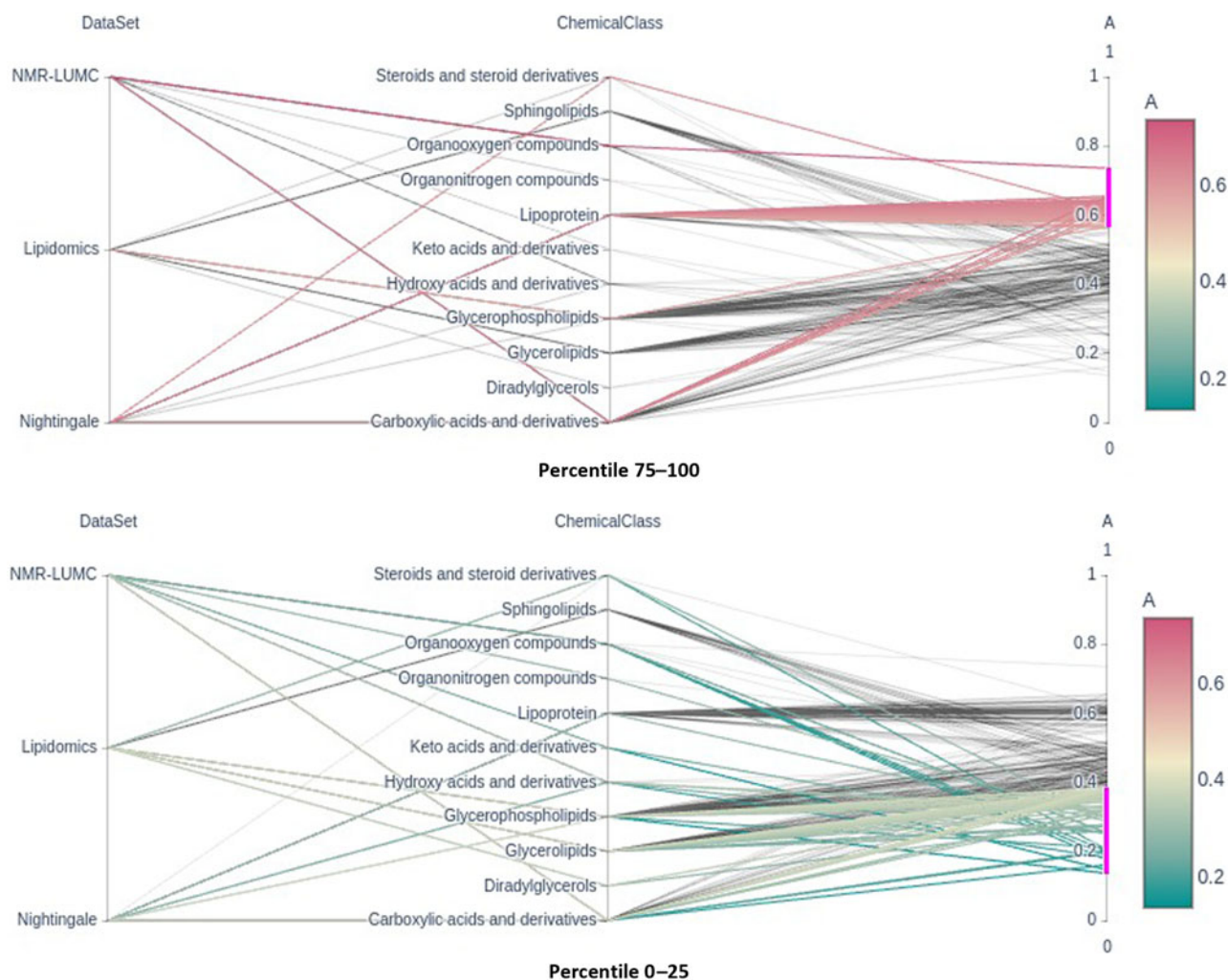
Fig. 2. Parallel coordinates plots of the top and bottom heritability quartiles showing the enrichment of the lipoprotein chemical class in the top percentile and its underrepresentation in the bottom percentile groups. Note that classes glycerophospholipids and glycerolipids are underrepresented in the top percentile group.

well-established factor influencing metabolite levels, with higher levels observed in adults as compared to children (Ellul et al., 2019). Third, this work involves metabolic traits measured in blood. Results reported here do not necessarily represent those expected for other tissue types. Fourth, we investigated 'fasting state' metabolite levels. As such, this can be considered a relatively stable homeostatic state, in contrast to states of high energy consumption or after food and/or drink intake. Finally, this study, as most others, includes platforms favoring lipids, limiting the conclusions we may draw with regard to nonlipid species (Hagenbeek et al., 2020).

We observed a limited contribution of common environment to the total variation in metabolite levels: we estimated a significant contribution of the common environment only for 6 out of 237 metabolites. For these 6 traits, the average contribution to the common environment was 0.25 (range: 0.17−0.34). Our findings are in line with previous studies reporting an influence of the common environment for a minority of the investigated metabolites, with average contributions of the common environment frequently lower even than the 0.25 we observed (Kettunen et al., 2012; Menni et al., 2013; Yet et al., 2016). The limited contribution of the common environment on metabolite levels is also consistent with observations for other molecular traits, such

as expression (Ouwens et al., 2020; Wright et al., 2014) or methylation levels (van Dongen et al., 2016), which were measured in the same group of twin pairs as in this work. This is in stark contrast to a recent family-based study that reported substantial familial resemblance in metabolite levels to which common environment had the strongest contribution (Tremblay et al., 2019). A strong contribution of the common environment has also been observed for specific lipid species, while other lipid species had no or only small contributions of the common environment (Frahnow et al., 2017). It must be noted that those studies reporting substantial influence of the common environment frequently have small sample sizes and are generally younger than our own samples.

The observation that an effect of the common environment is more frequently reported in younger participants is in line with observations for other complex traits. For several complex traits with a strong influence of the common environment during childhood, this decreases greatly with age, often completely disappearing in adulthood (Haworth et al., 2010; Lamb et al., 2010). Currently, longitudinal genetic investigations of metabolite levels starting in childhood are lacking. Therefore, we may not definitely conclude that a similar pattern holds for metabolite levels. Moreover, in general for studies including adult samples, the question arises what the common environment comprises

Table 5. Representation of chemical classes across the percentiles of the additive genetic component (A) values as determined by the AE models

Chemical class	Enrichment percentile 0–25				Enrichment percentile 25–50				Enrichment percentile 50–75				Enrichment percentile 75–100				Total
	N	OR	p	p (fdr)	N	OR	p	p (fdr)	N	OR	p	p (fdr)	N	OR	p	p (fdr)	N
Lipoprotein	1	3.1E−02	8.2E−09	9.0E−08	0	0.0E+00	4.0E−10	4.3E−09	12	6.2E−01	2.4E−01	6.5E−01	51	8.1E+01	6.5E−31	7.2E−30	64
Glycerolipids	18	3.6E+00	8.2E−04	4.5E−03	17	3.2E+00	3.1E−03	1.7E−02	2	1.4E−01	1.7E−03	9.1E−03	0	0.0E+00	1.4E−05	5.2E−05	37
Sphingolipids	0	0.0E+00	2.8E−03	1.0E−02	9	2.7E+00	5.4E−02	1.5E−01	11	4.3E+00	2.5E−03	9.2E−03	0	0.0E+00	5.0E−03	1.4E−02	20
Carboxylic acids and derivatives	13	2.6E+00	2.3E−02	5.8E−02	7	9.1E−01	1.0E+00	1.0E+00	5	5.7E−01	3.7E−01	6.7E−01	5	5.7E−01	3.7E−01	6.7E−01	30
Organooxygen compounds	5	5.3E+00	2.7E−02	5.8E−02	1	4.2E−01	6.8E−01	1.0E+00	1	4.2E−01	6.8E−01	8.3E−01	1	4.2E−01	6.8E−01	1.0E+00	8
Hydroxy acids and derivatives	4	6.3E+00	3.7E−02	6.8E−02	2	1.5E+00	6.4E−01	1.0E+00	0	0.0E+00	3.4E−01	6.7E−01	0	0.0E+00	3.4E−01	6.7E−01	6
Keto acids and derivatives	2	Inf	6.3E−02	8.7E−02	0	0.0E+00	1.0E+00	1.0E+00	0	0.0E+00	1.0E+00	1.0E+00	0	0.0E+00	1.0E+00	1.0E+00	2
Diradylglycerols	2	Inf	6.3E−02	8.7E−02	0	0.0E+00	1.0E+00	1.0E+00	0	0.0E+00	1.0E+00	1.0E+00	0	0.0E+00	1.0E+00	1.0E+00	2
Steroids and steroid derivatives	2	6.1E+00	1.6E−01	1.9E−01	0	0.0E+00	5.8E−01	1.0E+00	0	0.0E+00	5.8E−01	7.9E−01	1	1.5E+00	1.0E+00	1.0E+00	3
Glycerophospholipids	12	6.2E−01	2.4E−01	2.6E−01	23	2.2E+00	1.7E−02	6.3E−02	27	3.3E+00	2.8E−04	3.1E−03	1	3.2E−02	1.8E−08	9.9E−08	63
Organonitrogen compounds	1	3.0E+00	4.4E−01	4.4E−01	0	0.0E+00	1.0E+00	1.0E+00	1	3.1E+00	4.4E−01	6.9E−01	0	0.0E+00	1.0E+00	1.0E+00	2
Total	60				59				59				59				237

Note: The Count columns list the metabolite counts per chemical class. The enrichment statistics are coming from Fisher's exact test performed over all metabolite class counts per subset with respect to the entire sample. OR columns list the odds ratios per chemical class, p columns the associated p values and p (fdrb) the Benjamini–Hochberg false discovery rate corrected ones.

(Finnicum et al., 2019; Zaitlen et al., 2013). In adult twin pairs living apart the common environment may represent lasting influences of their time spend living together. Alternatively, the common environment could reflect similar living environments, including the same (or similar) neighborhood, exposure to similar pollutants or shared lifestyles. Thus, while many environmental factors influencing metabolite levels have been identified, the distinction between common and unique environmental factors remains unclear.

We found an overrepresentation of lipoproteins with heritability estimates in the highest quartile. Interestingly, we observed no differences in total heritability estimates between lipoproteins and other lipid classes in the same sample using genome-wide SNP data; however, the portion of heritability due to known metabolite loci did differ between lipid classes (Hagenbeek et al., 2020). This could indicate that the genetic relationship estimates from genome-wide SNP data do not comprise enough information to assess this enrichment. Combining the enrichment results with our previous findings that class-specific metabolite loci have a larger contribution to metabolite heritability than nonclass loci suggests the merit of investigating the genetics of biochemical pathways rather than individual metabolites. The strength of leveraging knowledge of biochemical pathways in genetic investigations has been previously established. Genetic investigations of metabolite ratios reflecting enzymatic conversions in biochemical pathways have led to stronger associations than studying single metabolites (Suhre et al., 2016). Similarly, multivariate genome-wide association studies of correlated metabolic networks have proven effective (Inouye et al., 2012).

Conclusion

Based on structural equation modeling in twins, a significant heritability was found for the blood concentrations of 231 metabolites, explaining 46% of their variance on average. We find that in the top 25% most heritable metabolites, those of class lipoprotein are overrepresented, while metabolites of classes glycerophospholipids and glycerolipids are underrepresented. In contrast, we find little evidence of shared environment influences on individual differences in metabolite levels. These findings show that familial resemblance in the concentrations of metabolites is due to genetic factors with minimal contribution of the shared environment.

References

Adamski, J., & Suhre, K. (2013). Metabolomics platforms for genome wide association studies-linking the genome to the metabolome. *Current Opinion in Biotechnology*, 24, 39–47.

Bellis, C., Kulkarni, H., Mamtani, M., Kent, J. W., Wong, G., Weir, J. M., ... **Curran, J. E.** (2014). Human plasma lipidome is pleiotropically associated with cardiovascular risk factors and death. *Circulation: Cardiovascular Genetics*, 7, 854–863.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., ... **Fox, J.** (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, 76, 306–317.

Bot, M., Milaneschi, Y., Al-Shehri, T., Amin, N., Garmaeva, S., Onderwater, G. L. J., ... **Penninx, B. W. J. H.** (2019). Metabolomics profile in depression: A pooled analysis of 230 metabolic markers in 5283 cases with depression and 10,145 controls. *Biological Psychiatry*, 87, 409–418.

Chaleckis, R., Murakami, I., Takada, J., Kondoh, H., & Yanagida, M. (2016). Individual variability in human blood metabolites identifies age-related differences. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 4252–4259.

Dane, A. D., Hendriks, M. M. W. B., Reijmers, T. H., Harms, A. C., Troost, J., Vreeken, R. J., ... **Hankemeier, T.** (2014). Integrating metabolomics profiling measurements across multiple biobanks. *Analytical Chemistry*, 86, 4110–4114.

Darst, B. F., Koscik, R. L., Hogan, K. J., Johnson, S. C., & Engelman, C. D. (2019). Longitudinal plasma metabolomics of aging and sex. *Aging*, 11, 1262–1282.

Demirkan, A., Henneman, P., Verhoeven, A., Dharuri, H., Amin, N., van Klinken, J. B., ... **van Dijk, K. W.** (2015). Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses. *PLoS Genetics*, 11, e1004835.

Draisma, H. H. M., Beekman, M., Pool, R., van Ommen, G.-J. B., Vaarhorst, A. A. M., de Craen, A. J. M., ... **Boomsma, D. I.** (2013). Familial resemblance for serum metabolite concentrations. *Twin Research and Human Genetics*, 16, 948–961.

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.

Ellul, S., Wake, M., Clifford, S. A., Lange, K., Würtz, P., Juonala, M., ... **Saffery, R.** (2019). Metabolomics: Population epidemiology and concordance in Australian children aged 11–12 years and their parents. *BMJ Open*, 9, 106–117.

Esko, T., Hirschhorn, J. N., Feldman, H. A., Hsu, Y. H. H., Deik, A. A., Clish, C. B., ... **Ludwig, D. S.** (2017). Metabolomic profiles as reliable biomarkers of dietary composition. *American Journal of Clinical Nutrition*, 105, 547–554.

Fan, S., Yeon, A., Shahid, M., Anger, J. T., Eilber, K. S., Fiehn, O., & Kim, J. (2018). Sex-associated differences in baseline urinary metabolites of healthy adults. *Scientific Reports*, 8, 11883.

Fiehn, O. (2002). Metabolomics–the link between genotypes and phenotypes. *Plant Molecular Biology*, 48, 155–171.

Finnicum, C. T., Beck, J. J., Dolan, C. V., Davis, C., Willemsen, G., Ehli, E. A., ... **De Geus, E. J. C.** (2019). Cohabitation is associated with a greater resemblance in gut microbiota which can impact cardiometabolic and inflammatory risk. *BMC Microbiology*, 19, Article no. 230.

Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, 85, 87–94.

Frahnow, T., Osterhoff, M. A., Hornemann, S., Kruse, M., Surma, M. A., Klose, C., ... **Pfeiffer, A. F. H.** (2017). Heritability and responses to high fat diet of plasma lipidomics in a twin study. *Scientific Reports*, 7, 1–11.

Gonzalez-Covarrubias, V., Beekman, M., Uh, H. W., Dane, A., Troost, J., Paliukhovich, I., ... **Slagboom, E. P.** (2013). Lipidomics of familial longevity. *Aging Cell*, 12, 426–434.

Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G., & Kell, D. B. (2004). Metabolomics by numbers: Acquiring and understanding global metabolite data. *Trends in Biotechnology*, 22, 245–252.

Hagenbeek, F. A., Pool, R., van Dongen, J., Draisma, H. H. M., Jan Hottenga, J., Willemsen, G., ... **Boomsma, D. I.** (2020). Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nature Communications*, 11, 39.

Hasirci, V., & Hasirci, N. (Eds.). (2018). *Fundamentals of biomaterials*. New York, NY: Springer.

Haworth, C. M. A., Wright, M. J., Luciano, M., Martin, N. G., De Geus, E. J. C., Van Beijsterveldt, C. E. M., ... **Plomin, R.** (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 15, 1112–1120.

Inouye, M., Ripatti, S., Kettunen, J., Lyytikäinen, L.-P., Oksala, N., Laurila, P.-P., ... **de Bakker, P. I. W.** (2012). Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genetics*, 8, e1002907.

Kastenmüller, G., Raffler, J., Gieger, C., & Suhre, K. (2015). Genetics of human metabolism: An update. *Human Molecular Genetics*, 24, R93–R101.

Kettunen, J., Demirkan, A., Wurtz, P., Draisma, H. H. M., Haller, T., Rawal, R., ... **Ala-Korpela, M.** (2016). Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nature Communications*, 7, Article no. 11122.

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.

Lacruz, M. E., Kluttig, A., Tiller, D., Medenwald, D., Giegling, I., Rujescu, D., ... **Kastenmüller, G.** (2018). Instability of personal human metabotype is linked to all-cause mortality. *Scientific Reports*, 8, Article no. 9810.

Lamb, D. J., Middeldorp, C. M., van Beijsterveldt, C. E. M., Bartels, M., van der Aa, N., Polderman, T. J. C., & Boomsma, D. I. (2010). Heritability of anxious-depressive and withdrawn behavior: Age-related changes during adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49, 248–255.

Ligthart, L., van Beijsterveldt, C. E. M., Kevenaar, S. T., de Zeeuw, E., van Bergen, E., Bruins, S., ... **Boomsma, D. I.** (2019). The Netherlands Twin Register: Longitudinal research based on twin and twin-family designs. *Twin Research and Human Genetics*, 22, 623–626.

Liu, J., Lahousse, L., Nivard, M. G., Bot, M., Chen, L., van Klinken, J. B., ... **van Duijn, C. M.** (2020). Integration of epidemiologic, pharmacologic, genetic and gut microbiome data in a drug–metabolite atlas. *Nature Medicine*, 26, 110–117.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., ... **Vrieze, S.** (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *In Nature Genetics*, 51, 237–244.

Maes, H. H., Neale, M. C., Medland, S. E., Keller, M. C., Martin, N. G., Heath, A. C., & Eaves, L. J. (2009). Flexible Mx specification of various extended twin kinship designs. *Twin Research and Human Genetics*, 12, 26–34.

Martin, N. G., Eaves, L. J., Kearsey, M. J., & Davies, P. (1978). The power of the classical twin study. *Heredity*, 40, 97–116.

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.

Neale, M. C. (1997). *Mx: Statistical modeling*. Richmond, VA: Virginia Commonwealth University,

Ouwens, K. G., Jansen, R., Nivard, M. G., van Dongen, J., Frieser, M. J., Hottenga, J.-J Boomsma, D. I. (2020). A characterization of cis- and trans-heritability of RNA-Seq-based gene expression. *European Journal of Human Genetics*, 28, 253–263.

Pang, H., Jia, W., & Hu, Z. (2019). Emerging applications of metabolomics in clinical pharmacology. *Clinical Pharmacology and Therapeutics*, 106, 544–556.

Posthuma, D., & Boomsma, D. I. (2000). A note on the statistical power in extended twin designs. *Behavior Genetics*, 30, 147–158.

Rangel-Huerta, O. D., Pastor-Villaescusa, B., & Gil, A. (2019). Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. *Metabolomics*, 15, 93.

Rhee, E. P., Ho, J. E., Chen, M.-H., Shen, D., Cheng, S., Larson, M. G., . . . Gerszten, R. E. (2013). A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metabolism*, *18*, 130–143.

Rhee, E. P., Yang, Q., Yu, B., Liu, X., Cheng, S., Deik, A., . . . Gerszten, R. E. (2016). An exome array study of the plasma metabolome. *Nature Communications*, *7*, 12360.

Rice, T. K. (2008). Familial resemblance and heritability. *Advances in Genetic*, *60*, 35–49.

Shin, S.-Y., Fauman, E. B., Petersen, A.-K., Krumsiek, J., Santos, R., Huang, J., . . . Soranzo, N. (2014). An atlas of genetic influences on human blood metabolites. *Nature Genetics*, *46*, 543–550.

Soininen, P., Kangas, A. J., Würtz, P., Suna, T., & Ala-Korpela, M. (2015). Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circulation: Cardiovascular Genetics*, *8*, 192–206.

Suhre, K., Raffler, J., & Kastenmüller, G. (2016). Biochemical insights from population studies with genetics and metabolomics. *Archives of Biochemistry and Biophysics*, *589*, 168–176.

Tabassum, R., Rämö, J. T., Ripatti, P., Koskela, J. T., Kurki, M., Karjalainen, J., . . . Ripatti, S. (2019). Genetic architecture of human plasma lipidome and its link to cardiovascular disease. *Nature Communications*, *10*, 4329.

Tremblay, B. L., Guénard, F., Lamarche, B., Pérusse, L., & Vohl, M. C. (2019). Familial resemblances in human plasma metabolites are attributable to both genetic and common environmental effects. *Nutrition Research*, *61*, 22–30.

Van Buuren, S. ,& Groothuis-Oudshoorn, K. (2011). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van Dongen, J., Nivard, M. G., Willemsen, G., Hottenga, J.-J., Helmer, Q., Dolan, C. V., . . . Boomsma, D. I. (2016). Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nature Communications*, *7*, 11115.

Verhoeven, A., Slagboom, E., Wuhrer, M., Giera, M., & Mayboroda, O. A. (2017). Automated quantification of metabolites in blood-derived samples by NMR. *Analytica Chimica Acta*, *976*, 52–62.

Willemsen, G., de Geus, E. J. C., Bartels, M., van Beijsterveldt, C. E. M. T., 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.

Wishart, D. S. (2007). Current progress in computational metabolomics. *Briefings in Bioinformatics*, *8*, 279–293.

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., . . . Scalbert, A. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research*, *46*, D608–D617.

Wright, F. A., Sullivan, P. F., Brooks, A. I., Zou, F., Sun, W., Xia, K., . . . Boomsma, D. I. (2014). Heritability and genomics of gene expression in peripheral blood. *Nature Genetics*, *46*, 430–437.

Würtz, P., Kangas, A. J., Soininen, P., Lawlor, D. A., Davey Smith, G., & Ala-Korpela, M. (2017). Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: A primer on -omic technology. *American Journal of Epidemiology*, *186*, 1–13.

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*, Article no. 60.

Yazdani, A., Yazdani, A., Elsea, S. H., Schaid, D. J., Kosorok, M. R., Dangol, G., & Samiei, A. (2019). Genome analysis and pleiotropy assessment using causal networks with loss of function mutation and metabolomics. *BMC Genomics*, *20*, 395.

Yet, I., Menni, C., Shin, S. Y., Mangino, M., Soranzo, N., Adamski, J., . . . Bell, J. T. (2016). Genetic influences on metabolite levels: A comparison across metabolomic platforms. *PLoS ONE*, *11*, e0153672.

Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., & Price, A. L. (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genetics*, *9*, e1003520.