The genomic architecture of blood metabolites based on a decade of genome-wide analyses

Fiona A. Hagenbeek[1,2*], René Pool[1,2], Jenny van Dongen[1,2], Harmen H.M. Draisma[1], Jouke Jan Hottenga[1], Gonneke Willemsen[1], Abdel Abdellaoui[1], Iryna O. Fedko[1], Anouk den Braber[1,3,4], Pieter Jelle Visser[3,5], Eco J.C.N. de Geus[1,2,4], Ko Willems van Dijk[6], Aswin Verhoeven[7], H. Eka Suchiman[8], Marian Beekman[8], P. Eline Slagboom[8], Cornelia M. van Duijn[9], BBMRI-NL Consortium[10], Amy C. Harms[11], Thomas Hankemeier[11], Meike Bartels[1,2,4], Michel G. Nivard[1,2,4*¥] and Dorret I. Boomsma[1,2,4*¥]

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

[2]Amsterdam Public Health research institute, Amsterdam, the Netherlands.

[3]Alzheimer Center Amsterdam, Department of Neurology, VU Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands.

[4]Amsterdam Neuroscience, Amsterdam, the Netherlands.

[5]Department of Psychiatry and Neuropsychology, School of Mental Health and Neuroscience, Alzheimer Center Limburg, Maastricht University, Maastricht, The Netherlands.

[6]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, The Netherlands; Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 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, section of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[9]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.

[10] Members of the BBMRI Metabolomics Consortium are listed before the references.

[11]Division of Analytical Biosciences, Leiden Academic Center for Drug Research, Leiden University, Leiden, the Netherlands; The Netherlands Metabolomics Centre, Leiden, The Netherlands.

[¥] These authors contributed equally.

[*]Correspondence to: Fiona A. Hagenbeek, Dorret I. Boomsma or Michel G. Nivard, Department of Biological Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 7-10, 1081 BT Amsterdam, The Netherlands. E-mail: f.a.hagenbeek@vu.nl; di.boomsma@vu.nl; m.g.nivard@vu.nl

Word count:

Abstract: 138; Main text: 2,864; Methods: 2,784;

References incl. methods: 71; Tables: 4; Figures: 3

Supplementary Material:

Supplementary Notes: 3; Supplementary Figures: 4; Supplementary Tables: 15; Supplementary Data: 1

Abstract

Metabolomics examines the small molecules involved in cellular metabolism. Approximately 50% of total phenotypic differences in metabolite levels is due to genetic variance, but heritability estimates differ across metabolite classes and lipid species. From the literature we aggregate > 800 class-specific metabolite loci that influence metabolite levels. In a twin-family cohort (*N* = 5,117) these metabolite loci were leveraged to simultaneously estimate total heritability (h^2_{total}), SNP-based heritability (h^2_{SNP}) and the proportion of heritability captured by known metabolite loci ($h^2_{GW\text{-}loci}$) for 309 lipids and 52 organic acids. Our study revealed significant differences in h^2_{SNP} and $h^2_{GW\text{-}loci}$ among different classes of lipids and organic acids. Furthermore, phosphatidylcholines with a higher degree of unsaturation had higher $h^2_{GW\text{-}loci}$ estimates. This study highlights the importance of common genetic variants for metabolite levels and elucidates the genetic architecture of metabolite classes and lipid species.

Metabolites are the small molecules involved in cellular metabolism, while the metabolome is typically defined as the collection of metabolites produced by cells[1]. Metabolomics aims at providing a holistic overview of the metabolome[1], and allows for the elucidation of underlying biological mechanisms and metabolic disturbances in diseases. At the same time metabolomics may offer potential new therapeutic targets or new biomarkers for disease diagnosis[2]. Variation in metabolite levels can arise due to gender[3], and age[4], as well as physiologic effects, behavior, and lifestyle, such as diet[5]. Genetic differences may be a source of direct variation in metabolomics profiles or may exert their effects on metabolite profiles through the genetic influences on behavior or physiology.

Systematic investigations of common genetic variants in human metabolism by genome- and metabolome-wide analysis successfully identified genetically influenced metabotypes (GIMs)[6]. The first genome-wide association study (GWAS) in 2008 (N = 284 participants) identified four genetic variants associated with metabolite levels[7]. Thereafter, GWAS with increasing sample sizes, and in diverse populations, have resulted in the identification of hundreds of Single Nucleotide Polymorphism (SNP) associations with metabolites from a wide range of metabolite classes[6]. Additional metabolite loci have been identified by leveraging low-frequency and rare-variant analyses using (exome-) sequencing. We conducted a comprehensive review of all quantitative trait locus (QTL) discovery for metabolites and supply the complete reference list in **Supplementary Note 1**.

Twin and family studies estimated the heritability (h^2; proportion of phenotypic variance due to genetic variance) for metabolite levels at around 50%, ranging from a heritability of 0% to 80% [5,8–15]. Several studies reported differences in heritability estimates among different classes of lipid species[12,14] or lipoprotein subclasses[13]. For example, Rhee et al. (2013) reported higher heritability estimates for amino acids than for lipids[11]. Essential amino acids, which cannot be synthesized by an organism *de novo*[16], had lower heritability than non-essential amino acids[11] that are synthesized within the body[16].

Intriguingly, phosphatidylcholines[10] and triglycerides (TGs)[15] show increasing heritability as the number of carbon atoms and/or double bonds in their fatty acyl side chains increases. Draisma et. al speculated this might be attributed to differences in the number of metabolic conversion rounds for phosphatidylcholines or TGs with a variable number of carbon atoms[10].

An improved understanding of the genetic architecture of intermediate phenotypes such as metabolites may benefit insight into the aetiology of diseases and traits, such as cardiometabolic diseases[17], migraine[18], psychiatric disorders[19], and cognition[20]. We aim to expand our understanding of the contribution of genetic factors to variation in fasting blood metabolic measures (referred to as metabolites in the remainder of the text for brevity) and analyzed data from multiple metabolomics platforms from a large cohort of twins and family members ($N = 5{,}117$). Combining SNP and family data allows for the simultaneous estimation of SNP heritability (h^2_{SNP}) and total heritability (h^2_{total})[21]. We further extended this approach to estimate the proportion of variance explained by metabolite loci identified by GWAS or rare-variant analysis ($h^2_{GW\text{-}loci}$; **Supplementary Data 1**). The $h^2_{GW\text{-}loci}$ consisted of two sub-fractions, a fraction composed of all metabolite loci associated with metabolites of a specific superclass ($h^2_{GW\text{-}Class}$) and a fraction composed of all other metabolite loci ($h^2_{GW\text{-}Notclass}$).

After characterizing all published metabolite-SNP associations by metabolite classification, we present the h^2_{total}, h^2_{SNP} and $h^2_{GW\text{-}loci}$ results for 361 metabolites (**Figure 1**). Next, we further expand on the current knowledge of the genetic aetiology of metabolite classes by employing mixed-effect meta-regression models to test for differences in heritability estimates among metabolite classes and among lipid species. To distinguish between the effects of the number of carbon atoms or number of double bonds in the fatty acyl side chains of phosphatidylcholines and TGs additional univariate follow-up analyses were conducted.

Results

Metabolite classification

In the period of November 2008 to October 2018, 40 GWA and (exome-) sequencing studies have identified 242,580 metabolite-SNP or metabolite ratio-SNP associations (see **Supplementary Note 1**). These associations included 1,804 unique metabolites or ratios and 49,231 unique SNPs (43,830 after converting all SNPs to build 37; **Supplementary Data 1**). For all metabolites their Human Metabolome Database (HMDB)[22–24] identifiers were retrieved in order to extract information with regards to their hydrophobicity and chemical classification (see **Methods**). Excluding the ratios and unidentified metabolites, 953 metabolites could be classified into 12 'super classes' (**Table 1**), 43 'classes', or 77 'subclasses' based on the HMDB classification (**Supplementary Data 1**). The majority of the metabolites were classified as 'lipids' and 'organic acids'. The 'lipids' could be subdivided into 8 classes, with 1 to 95,795 metabolite-SNP associations per class (mean = 17,589; SD = 32,553), and in 32 subclasses, with 1-40,440 metabolites-SNP associations of per subclass (mean = 4,673; SD = 9,124). The 'organic acids and derivatives' could be divided in 9 classes, with 1 to 26,832 metabolite-SNP associations per class (mean = 3,374; SD = 8,832), and 17 'organic acid' subclasses, including 1 to 26,448 metabolite-SNP associations per subclass (mean = 1,786; SD = 6,371; **Supplementary Data 1**).

For 5,117 individuals, data were available from four different metabolomics platforms: the Nightingale Health ^{1}H-NMR platform, a UPLC-MS Lipidomics platform, the Leiden ^{1}H-NMR platform and the Biocrates Absolute-IDQ^{TM} p150 platform. All participants were registered with the Netherlands Twin Register (NTR)[25] and came from 2,445 nuclear families. Metabolomics and SNP data were available for all participants. Background and demographic characteristics for the sample can be found in **Table 2**. Across all four platforms 427 metabolites were assessed. After excluding the ratios (17) and the metabolites of super classes not included in the curated metabolite-SNP association list (8), data were available for 402 metabolites. The 402 metabolites could be classified as 336 'lipids', 53 'organic acids', 9 'organic oxygen compounds', 3 'proteins' and one 'organic nitrogen compound'. In the remainder of this paper we solely focus on the 369 metabolites classified as 'lipids' or 'organic acids and derivatives'. The full list of metabolites, with their classifications and the quartile values of the untransformed levels, are included in **Supplementary Table 1**.

Characterization of the heritable influences on lipid and organic acid levels

For the 369 metabolites that passed QC, we estimated total heritability (h^2_{total}), the proportion of phenotypic variance explained by measured SNPs (h^2_{SNP}), the proportion attributable to metabolite superclass-specific loci ($h^2_{GW\text{-}Class}$) and the proportion of variance attributable to non-superclass metabolite loci ($h^2_{GW\text{-}Notclass}$) in twin and family members. The four-variance component analyses were performed in the genome-wide complex trait analysis (GCTA) software[26]. The analyses were performed separately for 'lipids' and 'organic acids', using unique superclass-specific and non-superclass genetic relationship matrices (GRMs; created in LDAK[27,28]) in both sets of analyses (**Figure 1**). The 'lipid' analyses employed a superclass-specific GRM of 479 'lipid' loci and a non-superclass GRM including 596 SNPs (**Figure 1**). The 'organic acid' analyses included a superclass-specific GRM with 397 loci and a non-superclass GRM with 683 SNPs (**Figure 1**). Before analyses, the metabolite data were normalized (log-normal or inverse rank; **see Methods**). All models included age at blood draw, sex, the first 10 principal components (PCs) from SNP genotype data, genotyping chip and metabolomics measurement batch as covariates.

Supplementary Table 2 includes the estimates for h^2_{total} , h^2_{SNP}, and $h^2_{GW\text{-}loci}$ from the four-variance genetic component model for all 369 metabolites. The genomic relatedness matrix residual maximum likelihood (GREML) algorithm converged successfully for 361 (97.8%) of the 53 'organic acids' and 316 'lipids'. Poor convergence of the GREML algorithm was observed for 6 metabolites (1.6%). The analyses for 2 metabolites (0.5%) were not completed due to non-invertible variance-covariance matrices. The estimates for h^2_{total} of the 309 'lipids' ranged from 0.11 to 0.66 (mean = 0.47; mean s.e. = 0.04). The h^2_{SNP} estimates ranged from -0.54 to 0.71 (mean = 0.05; mean s.e. = 0.24). The estimates for $h^2_{GW\text{-}loci}$ ranged from -0.05 to 0.16 (mean = 0.06; mean s.e. = 0.03; **Table 3**). The 52 'organic acids' had h^2_{total} estimates ranging from 0.14 to 0.72 (mean = 0.41; mean s.e. = 0.04). The estimates for h^2_{SNP} ranged from -0.42 to 0.46 (mean = 0.05; mean s.e. = 0.24) and for $h^2_{GW\text{-}loci}$ ranged from -0.08 to 0.11 (mean = 0.01; mean s.e. = 0.02; **Table 3**). On average, for both 'lipids' and 'organic acids' the h^2_{class} was higher than the $h^2_{Notclass}$, with $h^2_{GW\text{-}Class}$ ranging from -0.02 to 0.16 (0.06; mean s.e. = 0.02) for 'lipids' and from -0.04 to 0.14 for 'organic acids' (mean = 0.01; mean s.e. = 0.02). For both 'lipids' and 'organic acids' $h^2_{GW\text{-}Notclass}$ was zero (mean s.e. = 0.02), ranging from -0.06 to 0.12 for 'lipids' and from -0.06 to 0.05 for 'organic acids' (**Table 3**).

Including multiple metabolomics platforms allowed for a comparison of metabolites as measured on multiple platforms. An earlier study showed 29 out of 43 overlapping metabolites across two platforms to exhibit moderate heritability on both platforms[29]. In the current study, 61 metabolites were measured on multiple platforms, with moderate h^2_{total} on each of the platforms and on average a medium positive correlation between the h^2_{total} of the same metabolite assessed on different platforms (mean r h^2_{total} = 0.36; **Supplementary Table 3**).

Differential heritability among metabolite classes and lipid-species

Figure 2 shows variation in median heritability among the different classes of 'organic acids': 'keto acids', 'hydroxy acids' and 'carboxylic acids' (see **Supplementary Table 1** for metabolites per class). 'Keto acids', followed by 'carboxylic acids', had the highest median h^2_{total}, h^2_{SNP} and $h^2_{GW\text{-}Class}$ estimates (**Figure 2**). While 'hydroxy acids' had the highest median $h^2_{GW\text{-}Notclass}$ and $h^2_{GW\text{-}loci}$ estimates, the lowest median h^2_{total}, h^2_{SNP} and $h^2_{GW\text{-}Class}$ estimates were observed for these metabolites (**Figure 2**). To investigate whether heritability differs significantly among classes of ‘organic acids’, we applied multivariate mixed-effect meta-regression, corrected for metabolite platform effects (see **Methods**). The multivariate mixed-effect meta-regression models showed that h^2_{total} and $h^2_{GW\text{-}Class}$ for the ‘organic acid’ classes did not differ significantly. Significant differences among the ‘organic acid’ classes, though, were observed for the h^2_{SNP} estimates ($F(4, 47) = 7.48$, FDR-adjusted p-value = 0.02), the $h^2_{GW\text{-}loci}$ estimates ($F(4, 47) = 3.44$, FDR-adjusted p-value = 0.03), and the $h^2_{GW\text{-}Notclass}$ estimates ($F(4,47) = 19.95$, FDR-adjusted p-value = $1.25\text{x}10^{-08}$; **Supplementary Table 4**).

The multivariate mixed-effect meta-regressions were also applied to assess the significance of heritability differences among essential and non-essential amino acids (subdivision of ‘carboxylic acids’; see **Supplementary Table 5**) and among ‘lipid’ classes (see **Supplementary Table 1** for metabolites per ‘lipid’ class). None of the observed mean differences among essential and non-essential amino acids (**Table 4**) were significant in the meta-regressions (**Supplementary Table 4**). Small but significant median heritability differences were observed among the different classes of ‘lipids’ (**Figure 3**). For ‘lipid’ classes the $h^2_{GW\text{-}loci}$ estimates differed significantly ($F(8, 300) = 8.47$; FDR-adjusted p-value = 0.004; **Supplementary Table 4**).

Finally, we explored whether heritability of phosphatidylcholines and TGs increases with a larger number of carbon atoms and/or double bonds in their fatty acyl side chains. To this end we employed both uni- and multivariate mixed-effect meta-regression models separately for the TGs, diacyl phosphatidylcholines (PCaa) and acyl-alkyl phosphatidylcholines (PCae; see **Methods**). The platform specific heritability estimates for each of these lipid species has been depicted in **Supplementary Figure 1**. Variation in the number of carbon atoms and double bonds was significantly associated with $h^2_{GW\text{-}loci}$ estimates for PCaa’s ($F(3, 52) = 7.05$; FDR-adjusted p-value = 0.009) and PCae’s ($F(3, 45) = 3.41$; FDR-adjusted p-value = 0.05; **Supplementary Table 4**). Phosphatidylcholines with a larger number of carbon atoms showed lower heritability estimates and phosphatidylcholines with a larger number of double bonds had higher heritability estimates (**Supplementary Table 4**). The differences among the phosphatidylcholines with a variable number of carbon atoms and/or double bonds could be contributed to differential h^2_{Class} estimates. Univariate models confirmed the pattern for the number of double bonds in PCaa's and PCae, though they were not significant after correction for multiple testing (**Supplementary Table 6**).

Discussion

We carried out a comprehensive assessment of GWA-metabolomics studies and created a repository of all studies reporting on associations of SNPs and blood metabolites in European ancestry samples. This led to 241,965 genome-wide associations that were curated, lifted to NCBI build 37 and for which all associated metabolites were classified. The complete, categorized, overview of all blood metabolite-SNP associations is provided in **Supplementary Data 1**, with the complete list of references in **Supplementary Note 1**. The information from the repository served to construct six GRMs which then served as predictors in the analysis of 369 metabolites. The metabolite data in our study derived from four metabolomics platforms and two metabolite super classes. By mapping all metabolites to the Human Metabolome Database (HMDB)[22–24] we were able to classify both the measured metabolites and all previously published metabolites as either 'lipids' or 'organic acids'. Because the participants in the study (N = 5,117) came from a large cohort of MZ and DZ twin-families we could evaluate the total heritability (h^2_{total}) and the contributions of genome-wide SNPs (h^2_{SNP}) on 'lipids' and 'organic acids'. A unique feature of the study was the ability to disentangle the role of superclass-specific ($h^2_{GW\text{-}Class}$) and non-superclass ($h^2_{GW\text{-}Notclass}$) metabolite loci on heritability differences among metabolite classes and lipid species.

To evaluate differences among metabolite classes and lipid species in the estimates for h^2_{total}, h^2_{SNP}, $h^2_{GW\text{-}loci}$, $h^2_{GW\text{-}Class}$, and $h^2_{GW\text{-}Notclass}$ multivariate mixed-effect meta-regression models were applied. No significant differences in h^2_{total} estimates existed among any of the metabolite classes. Congruent with a previous twin-family study[9], none of the heritability estimates differed significantly among essential and non-essential amino acids. Both h^2_{SNP} and $h^2_{GW\text{-}loci}$ showed significant differences among the different classes of 'organic acids'. 'Keto acids' had significantly higher h^2_{SNP} and significantly lower $h^2_{GW\text{-}loci}$ estimates as compared with 'carboxylic acids'. Class-specific metabolite loci heritability estimates for 'fatty acyls', 'lipoproteins' and 'steroids' were significantly higher. Similarly, significant heterogeneity in lipid class heritability, with lower h^2_{total} and h^2_{SNP} for phospholipids than for sphingolipids or glycerolipids has been described[12,14,30]. Lastly, we assessed whether heritability increases with added complexity in lipid species[10,15]. We found that this indeed held for $h^2_{GW\text{-}loci}$ estimates in more complex diacyl and acyl-alkyl phosphatidylcholines but not for more complex TGs. Previous research reported significant higher h^2_{SNP} estimates in polyunsaturated fatty acid containing lipids[14]. Furthermore, loci of traditional lipid measures explained 2% to 21% of the variance in lipid levels[14]. Together these results suggest that higher heritability in phosphatidylcholines is driven by a lower number of carbon atoms and higher number of double bonds, e.g. a larger degree of unsaturation.

Evaluating the mean heritability differences among 'lipids' and 'organic acids' it appears that 'lipids' have higher h^2_{total}, $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}loci}$ estimates than 'organic acids' (**Table 3**). However, as the GRMs used in the calculation of the heritability estimates differed among these classes, we were unable to empirically compare mean differences. Comparison of our findings with those of previous twin-family studies indicates that the heritability difference among 'lipids' and 'organic acid' is infrequently investigated[8–11]. A possible explanation for the lack of comparisons may be the shortage of balanced metabolomics platforms. The majority of metabolomics platforms have a strong focus on either 'lipids' or ‘organic acids’, which complicates such comparisons. The disproportion of metabolite classes on metabolomics platforms also affects the known metabolite loci, where ‘lipid’ studies have been overrepresented as well. As a consequence, especially the $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}loci}$ estimates of the ‘organic acids’ will be underpowered due to this imbalance. For multi-component GREML our platform-specific sample sizes were relatively small[31]. Only the Nightingale Health ^{1}H-NMR platform was sufficiently powered to obtain small s.e.’s in single-component GREML using unrelated individuals with common SNPs[32]. New[30,33–35] and future studies will increase the number of variants identified as metabolite loci. The investment in UK Biobank[36] is expected to dramatically increase sample sizes for large-scale genomic investigations of the human metabolome and subsequently the number of metabolite loci.

Applications such as two-sample Mendelian Randomization benefit greatly from the comprehensive overview of metabolite loci we identified. The identified loci are interesting to explore as instruments for metabolome-wide Mendelian Randomization studies of complex traits. Our work further offers valuable insights into the role of common genetic variants in class specific differences among metabolite classes and lipids species. Further research is required to elucidate the contribution of rare genetic variants to metabolite levels and differences among metabolite classes. A reasonable approach to tackle this issue could be to carry out a similar study in a large sample of whole-genome sequencing (WGS) data. Such an approach, using MAF- and LD-stratified GREML analysis[31], identified additional variance due to rare variants for height and BMI[37]. The extent to which our findings might generalize to populations of non-European ancestry is uncertain, with replication among different ethnicities being more likely for loci of common human metabolism pathways[38].

In conclusion, we contributed to the further elucidation of the genetic architecture of fasting blood metabolite levels and to differences in the genetic architecture among metabolite classes. Extending the GREML framework with the inclusion of known metabolite loci allowed us to simultaneously estimate h^2_{total}, h^2_{SNP}, $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}Notclass}$ for 361 metabolites. Significant differences in h^2_{SNP} or $h^2_{GW\text{-}loci}$ estimates were observed among different classes of 'lipids' and 'organic acids' and for more complex diacyl and acyl-alkyl phosphatidylcholines. Future studies need to also elucidate the proportion of metabolite variation influenced by heritable and non-heritable lifestyle factors, which may help delineate new personalized disease prevention or treatment strategies for complex disorders.

Methods

Participants

At the Netherlands Twin Register (NTR)[39] metabolomics data for twins and family members as measured in blood samples were available for 6,011 individuals of whom 5,667 were genotyped. The blood samples for the four metabolomics experiments described in this study were mainly collected in participants of the NTR biobank project[25,40]. Blood samples were collected after a minimum of two hours of fasting (1.3%), with the majority of the samples collected after overnight fasting (98.7%). Fertile women were bled in their pill-free week or on day 2-4 of their menstrual cycle. For the current paper, we excluded participants if they were not of European ancestry, were on lipid-lowering medication at the time of blood draw or if they had not adhered to the fasting protocol. The exact number of exclusions per dataset is listed in **Supplementary Table 7**. After completing the preprocessing of the metabolomics data, the separate subsets (e.g., different collection and measurement waves; see **Supplementary Table 7**) of each platform were merged into a single per platform dataset, randomly retaining a single observation per platform whenever multiple observations were available. **Supplementary Table 8** gives an overview of the overlap in participants among the different platforms, with the overlap among each metabolite that survived quality control (QC) for all four platforms available in **Supplementary Table 9**. The final number of participants included in the study was 5,117, with platform specific sample size ranging from 1,448 to 4,227 individuals from 946 to 2,179 families. Characteristics for the individuals included in the analyses can be found in **Table 2**. 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 U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance- FWA00017598; IRB/institute codes, NTR 03-180 and EMIF-AD 2014.210).

Metabolite profiling

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; formerly Brainshake Ltd.). 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[41,42].

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 (1.8µm, 2.1 $*$ 100mm) was used with a flow of 0.4 ml/min over a 16 minute gradient. Lipid detection was done 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[43,44].

Leiden ^{1}H-NMR platform (for small metabolites)

The Leiden ^{1}H-NMR spectroscopy experiment of EDTA-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[45].

Biocrates Absolute-IDQ^{TM} p150 platform

The Biocrates Absolute-IDQ^{TM} p150 (Biocrates Life Sciences AG, Innsbruck, Austria) metabolomics platform on serum samples was analysed at the Metabolomics Facility of the Genome Analysis Centre at the Helmholtz Centre in Munich, Germany. This platform utilizes flow injection analysis coupled to tandem mass spectrometry (MS/MS) and has been described in detail elsewhere[3,46,47].

Metabolomics data preprocessing

Preprocessing of the metabolomics data was done for each of the platforms and measurement batches per platform separately. Metabolites were excluded from analysis when the mean coefficient of variation exceeded 25% and 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')[48]. On average, 9 values had to be imputed for each metabolites (SD = 12; range: 1-151). Data for each metabolite on both ^{1}H-NMR platforms were normalized by inverse normal rank transformation[45,49], while the imputed values of the Biocrates metabolomics platform and the UPLC-MS lipidomics platform were normalized by natural logarithm transformation[10,50], conform previous normalization strategies applied to the data obtained using these platforms. The complete lists with full names of all detected metabolites that survived QC and preprocessing for all platforms can be found in **Supplementary Table 1**, these tables also include the quartile values of the untransformed metabolites.

Genotyping, imputation and ancestry outlier detection

Genotype information was available for 21,001 NTR participants for 6 different genotyping arrays (Affymetrix 6.0 [*N* = 8,640], Perlegen-Affymetrix [*N* = 1,238], Illumina Human Quad Bead 660 [*N* = 1,439], Affymetrix Axiom [*N* = 3,144], Illumnia GSA [*N* = 5,938] and Illumina Omni Express 1M [*N* =238]), as well as sequence data from the Netherlands reference genome project GONL (BGI full sequence at 12x (*N* = 364)[51]. For each genotyping array samples were removed if they had a genotype call rate above 90%, gender-mismatch occurred or if heterozygosity (Plink F statistic) fell outside the range of -0.10 – 0.10. SNPs removed if they were palindromic AT/GC SNPs with a minor allele frequency (MAF) range between 0.4 and 0.5, when the MAF was below 0.01, when Hardy Weinberg Equilibrium (HWE) had $p < 10^{-5}$, when the number of Mendelian errors was greater than 20 and the genotype call rate was < 0.95. After QC the six genotyping arrays were aligned to the GONL reference set (V4) and SNPs were removed if the alleles mismatched with this reference panel or the allele frequency different more than 0.10 between the genotyping array and this reference set.

The data from the six genotyping chips were subsequently merged into a single dataset (1,781,526 SNPs). Identity-by-decent (IBD) was estimated with PLINK[52] and KING[53] for all individual pairs based on the ~10.6K SNPs in common across the arrays, next IBD was compared to expected family relations and individuals were removed if this mismatched. Prior to imputation to the GONL reference data[54,55] the duplicate monozygotic pairs (*N* = 3,032) or trios (*N* = 7) and NTR GONL samples (*N* = 364) were removed and the data was cross-array phased using MACH-ADMIX[56]. Post-imputation the NTR GONL samples and the duplicated MZ pairs and trios were re-added to the data. Filtering of the imputed dataset included the removal of SNPs that were significantly associated with a single genotyping chip ($p < 10^{-5}$), had HWE $p < 10^{-5}$, the Mendelian error rate > mean + 3 SD or if the imputation quality (R^2) was below 0.90. The final cross-platform imputed dataset included 1,314,639 SNPs, including 20,792 SNPs on the X-chromosome.

The cross-platform imputed data was aligned with PERL based "HRC or 1000G Imputation preparation and checking" tool (version 4.2.5; https://www.well.ox.ac.uk/~wrayner/tools). The remaining 1,302481 SNPs were phased with EAGLE[57] for the autosomes, and SHAPEIT[58] for chromosome X and then imputed to 1000 Genomes Phase 3 (1000GP3 version 5)[59] on the Michigan Imputation server using Minimac3 following the standard imputation procedures of the server[60]. Principal Component Analysis (PCA) was used to project the first 10 PCs of the 1000 genomes references set population on the NTR cross-platform imputed data using SMARTPCA[61]. Ancestry outliers (non-Dutch ancestry; N = 1,823) were defined as individuals with PC values outside the European/British population range[62]. After ancestry outlier removal the first 10 PCs were recalculated.

Curation of metabolite loci

In October 2018 PubMed and Google Scholar were searched to identify published GWA and (exome-) sequencing studies on metabolomics or fatty acid metabolism in blood samples using ^{1}H-NMR, mass spectrometry or gas chromatography-based methods. In the period of November 2008 to October 2018 40 GWA or (exome-) sequencing studies on blood metabolomics in European samples have been published (**Supplementary Note 1**). For all studies the genome-wide significant ($p < 5\text{x}10^{-8}$) metabolite-SNP associations were extracted, including only those observations for autosomal SNPs and reporting SNP effect sizes and p-values based on the summary statistics excluding NTR samples were relevant[49,50]. Across the 40 studies, 242,580 metabolite-SNP or metabolite ratio-SNP associations were reported, these associations included 1,804 unique metabolites or ratios and 49,231 unique SNPs (**Supplementary Data 1**). For all metabolites their Human Metabolome Database (HMDB)[22–24], PubChem[63], Chemical Entities of Biological Interest (ChEBI)[64] and International Chemical Identifier (InChiKey)[65] identifiers have been retrieved. Information with regards to the ‘super class’, ‘class’ and ‘subclass’ of metabolites was extracted from HMDB, whenever no HMDB identifier was available and categorization information could not be extracted, ‘super class’, ‘class’ and ‘subclass’ were provided based on expert opinion. Excluding the ratios and unidentified metabolites, 953 metabolites could be classified into 12 ‘super classes’, 43 ‘classes’ or 77 ‘subclasses’ (**Supplementary Data 1**). Based on the metabolite identifiers we also extracted the *log(S)* value for each metabolite to assess the hydrophobicity of the metabolites. The *log(S)* value represents the log of the partition coefficient between 1-octanol and water, two fluids that hardly mix. The partition coefficient is the ratio of concentrations in water and in octanol when a substance is added to an octanol-water mixture and hence indicates the hydrophobicity of a compound. Thus, we classify a metabolite as hydrophobic if it is more hydrophobic than 1-octanol itself and hydrophilic otherwise (**Supplementary Data 1**).

The 49,231 unique SNPs reported their rsIDs or chromosome-base pair positions by different genome builds or dbSNP maps[66], therefore we lifted all SNPs to HG19 build 37[67], after which 43,830 unique SNPs remained (**Figure 1**; **Supplementary Data 1**). All bi-allelic metabolite SNPs were extracted from our 1000GP3 data, which excluded 295 tri-allelic SNPs and 4,256 SNPs could not be retrieved from 1000GP3. Next, MAF > 1% (2,067 SNPs removed), $R^2 > 0.70$ (2,002 SNPs) and HWE $P < 10^{-4}$ (72 SNPs) filtering was performed, resulting in 35,138 metabolite SNPs for NTR participants (**Figure 1**). Next, we created two ‘super class’-specific lists of metabolite loci and two ‘not-superclass’ lists of metabolite loci. To create a list of loci for the 652 unique metabolites classified as ‘lipids and lipid-like molecules’ (e.g., ‘lipids’), in 2,500 unrelated individuals we clumped (PLINK version 1.9) all 112,760 lipid-SNP associations using an LD-threshold (r^2) of 0.10 in a 500kb radius (**Figure 1**). Clumping identified 482 lead SNPs, or loci, for ‘lipids’ and an additional 12,169 SNPs were identified as LD-proxies for the lipid-loci (**Figure 1**). To obtain the ‘not-superclass’ list of lipid loci the 12,651 lipid loci and proxies were removed from the list of all metabolite-SNP associations and the resulting list was clumped to obtain the 598 ‘non-superclass’ loci

(**Figure 1**). The same clumping procedure was applied to the 26,352 organic acid-SNP associations, identifying 398 organic acids loci, 10,781 organic acid LD-proxies and 687 'non-superclass' loci (**Figure 1**).

Construction of genetic relationship matrices

In total six weighted genetic relationship matrixes (GRMs) were constructed, which were corrected for uneven and long-range LD between the SNPs (LDAK version 4.9[27,28]; **Figure 1**). In **Supplementary Note2** the use of weighted versus unweighted GRMs is compared using simulations. Two of the GRMs used the cross-platform imputed dataset as backbone and the other four GRMs were based on SNPs extracted from the 1000GP3 imputed data. For inclusion in the first GRM, after removal of ancestry outliers, the autosomal SNPs of the cross-platform imputed dataset were filtered on MAF (<1%) and all lipid and organic acid loci, their LD-proxies and 50kb surrounding both types of SNPs were removed (see **curation of metabolite loci**; **Figure 1**). The resulting LDAK GRM included 434,216 SNPs and the *V(G1)* variance component in the genomic relatedness matrix residual maximum likelihood (GREML) analyses is based on this GRM (see **heritability analyses**; **Figure 1**). The *V(G2)* variance component in the GREML analyses is based on the LDAK GRM including all autosomal SNPs with a MAF greater than 1% included on the cross-platform imputed dataset (447,794 SNPs), where ancestry outliers were removed and for all individual pairs sharing less than 0.05 of their genome their sharing was set to zero[21] (**Figure 1**). Depending on the metabolite the *V(G3)* variance component in the GREML analyses was either based on an LDAK GRM of the 1000GP3 extracted lipid loci (479 SNPs) or the organic acid loci (397 SNPs; **Figure 1**). Finally, depending on the metabolite either the 'not-lipid' LDAK GRM (596 SNPs) or the 'not-organic acid' LDAK GRM (683 SNPs) underlay the *V(G4)* variance component in the GREML analyses (**Figure 1**). **Supplementary Data 1** indicates for each listed SNP if it was included in any of the LDAK GRMs.

Statistical analyses

Heritability analyses

Mixed linear models[21], implemented in the genome-wide complex trait analysis (GCTA) software package (version 1.91.7)[26], were applied to compare three models including a variable number of covariates. **Supplementary Table 10** gives the three different models, full descriptions of the covariates and model comparison have been given in **Supplementary Note 3**. The mean and median h^2_{total} and h^2_{SNP} estimates and standard errors were highly similar across the different models, as such the most sparse model was chosen for further analyses (**Supplementary Table 11**). This final model included the first 10 genetic PCs for the Dutch population, genotyping chip, sex and age at blood draw as covariates. For metabolites of the Nightingale Health ^{1}H-NMR and Biocrates platform, measurement batch was included as covariate.

The final four-variance component model including four GRMs, allowing the estimation of the proportion of variation explained by superclass-specific significant metabolite loci ($h^2_{GW\text{-}Class}$) and non-superclass significant metabolite loci ($h^2_{GW\text{-}Notclass}$) in addition to estimating the h^2_{SNP} and total h^2 (h^2_{total}; **Figure 1**). In this extension, the total variance explained by significant metabolite loci ($h^2_{GW\text{-}loci}$) consists of the sum of $\frac{V(G3)}{Vp}$ and $\frac{V(G4)}{Vp}$, where Vp is the phenotypic variance and h^2_{SNP} is defined as the sum of $\frac{V(G1)}{Vp}$, $\frac{V(G3)}{Vp}$ and $\frac{V(G4)}{Vp}$ (**Figure 1**). To calculate the standard errors (s.e.'s) for the composite variance estimates, we have randomly sampled 10,000 instances from the parameter variance-covariance matrices for each metabolite. The s.e.'s of the specific ratio of interest were then based on the standard deviation of the ratio of interest across 10.000 samples. The four-variance component models obtained the unconstrained variance components which allowed for negative h^2_{SNP} and $h^2_{GW\text{-}loci}$ estimates. All four-variance component models applied the --reml-bendV flag where necessary to invert the variance-covariance matrix V if V was not positive definite, which may occur when variance components are negative[68]. Finally, we calculated the log likelihood of a reduced model with either *V(G3)*, *V(G4)* or both dropped from the full model and calculated the LRT and p-value (**Supplementary Table 2**).

Mixed-effect meta-regression analyses

To investigate differences in heritability estimates among metabolites of different classes we applied mixed-effect meta-regression models as implemented in the 'metafor' package (version 2.0-0) in R (version 3.5.1)[69]. Here we tested for the moderation of heritability estimates by metabolite class and metabolomics platform on all 361 successfully analyzed metabolites while including a matrix combining the phenotypic correlations (**Supplementary Table 12**) and the sample overlap (**Supplementary Table 9**) between the metabolites as random factor to correct for dependence among the metabolites and participants. This matrix includes the sample size of the metabolite on the diagonal, with the off-diagonal computed by $\frac{N_{1,2}}{\sqrt{n_1 * n_2}} * r$ (**Supplementary Table 13**), where $N_{1,2}$ is the sample overlap between the metabolites, n_1 is the sample size of metabolite one, n_2 is the sample size of metabolite two and r is the phenotypic correlation between the metabolites as calculated with Spearman's Rho. For all mixed-effect meta-regression models we obtained the robust estimates based on a sandwich-type estimator, clustered by the metabolites included in the models to correct for the sample overlap among the different metabolites[70]. First, we used multivariate mixed-effect meta-regression models to simultaneously estimate the effect of metabolite class and metabolomics platform on the h^2_{total}, h^2_{SNP} and the $h^2_{GW\text{-}loci}$, as well as the $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}Notclass}$ estimates. Subsequently, to separately assess the effect of the number of carbon atoms or double bonds in the fatty acyls chains of phosphatidylcholines and triglycerides univariate models were conducted as follow-up. To account for multiple testing the p-values were adjusted with the with the False Discovery Rate (FDR)[71] using the 'p.adjust' function in R. Multiple testing correction was done separately for the univariate and the multivariate models.

Data availability

The curated list of all published metabolite-SNP associations is included in **Supplementary Data 1** and is publicly available through the BBMRI – omics atlas (http://bbmri.researchlumc.nl/atlas/#data). All information on the metabolites in this study are in **Supplementary Table 1**; with full summary statistics for the four-variance component models included in **Supplementary Table 2**. The Nightingale Health metabolomics data may be requested through BBMRI-NL (https://www.bbmri.nl/Omics-metabolomics). All (other) data may be accessed, upon approval of the data access committee, through the Netherlands Twin Register (ntr.fgb@vu.nl). A reporting summary for this Article is available as Supplementary Information file.

Funding

This work was performed within the framework of the BBMRI Metabolomics Consortium funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO, no. 184.021.007 and 184.033.111). The European Network of Genomic and Genetic Epidemiology (ENGAGE) contributed to funding to perform the Biocrates Absolute-IDQ^{TM} p150 metabolomics measurements (European Union Seventh Framework Program: FP7/2007-2013, grant number 201413). Analyses were supported by the Netherlands Organization for Scientific Research: Netherlands Twin Registry Repository: researching the interplay between genome and environment (480-15-001/674); the European Union Seventh Framework Program (FP7/2007-2013): ACTION Consortium (Aggression in Children: Unravelling gene-environment interplay to inform Treatment and InterventiON strategies; grant number 602768). Genotyping was made possible by grants from NWO/SPI 56-464-14192, Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health, Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls (USA) and the National Institutes of Health (NIH R01 HD042157-01A1, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2

MH089995) and European Research Council (ERC-230374). EMIF-AD has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking EMIF grant agreement n°115372. DIB acknowledges her KNAW Academy Professor Award (PAH/6635). M.Bartels is supported by an ERC consolidator grant (WELL-BEING 771057 PI Bartels).

Acknowledgements

We thank all twins and family members for their participation. We thank P. M. Visscher (University of Queensland) for his helpful comments. Preliminary analyses of this paper were included in a presentation at the 46th Annual Meeting of the Behavioral Genetics Association (BGA) in June 2016, the abstract of this presentation can be found in Behav. Genet. (2016) 46:785-786.

Author contributions

Nightingale Health metabolomics data: HES, MBeekman, PES and CMvD. Leiden ^{1}H-NMR metabolomics data: KWvD and AV. UPLC-MS lipidomics data: ACH and TH. EMIF-AD data: AdB and PJV. Genotype data: JJH, AA and IOF. NTR Biobank data: GW and EJCdG. Metabolomics pre-processing: RP, HHMD and FAH. Statistical analyses: FAH and MGN. Wrote the paper: FAH, JvD, MBartels, MGN and DIB. All authors critically read and commented on the manuscript.

Competing interests statement

The authors declare no competing financial interests.

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]

Affiliations:

[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, 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.

References

1. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).

2. Kuehnbaum, N. L. & Britz-McKibbin, P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem. Rev.* **113**, 2437–68 (2013).

3. Mittelstrass, K. *et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **7**, e1002215 (2011).

4. Chaleckis, R., Murakami, I., Takada, J., Kondoh, H. & Yanagida, M. Individual variability in human blood metabolites identifies age-related differences. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4252–4259 (2016).

5. Menni, C. *et al.* Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* **9**, 506–514 (2013).

6. Kastenmüller, G., Raffler, J., Gieger, C. & Suhre, K. Genetics of human metabolism: an update. *Hum. Mol. Genet.* **24**, R93–R101 (2015).

7. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **4**, e1000282 (2008).

8. Nicholson, G. *et al.* Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol. Syst. Biol.* **7**, 525 (2011).

9. Shah, S. H. *et al.* High heritability of metabolomic profiles in families burdened with premature cardiovascular disease. *Mol. Syst. Biol.* **5**, 258 (2009).

10. Draisma, H. H. M. *et al.* Familial resemblance for serum metabolite concentrations. *Twin Res. Hum. Genet.* **16**, 948–61 (2013).

11. Rhee, E. P. *et al.* A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab.* **18**, 130–143 (2013).

12. Frahnow, T. *et al.* Heritability and responses to high fat diet of plasma lipidomics in a twin study. *Sci. Rep.* **7**, 1–11 (2017).

13. Kaess, B. *et al.* The lipoprotein subfraction profile: heritability and identification of quantitative trait loci. *J. Lipid Res.* **49**, 715–723 (2008).

14. Bellis, C. *et al.* Human Plasma Lipidome Is Pleiotropically Associated With Cardiovascular Risk Factors and Death. *Circ. Cardiovasc. Genet.* **7**, 854–863 (2014).

15. Draisma, H. H. M. Analysis of Metabolomics Data from Twin Families. (Leiden, 2011).

16. Reeds, P. J. Dispensable and Indispensable Amino Acids for Humans. *J. Nutr.* **130**, 1874S–1876S (2000).

17. Newgard, C. B. Metabolomics and Metabolic Diseases: Where Do We Stand? *Cell Metab.* **25**, 43–56 (2017).

18. Onderwater, G. L. J. *et al.* Large-scale plasma metabolome analysis reveals alterations in HDL metabolism in migraine. *Neurology* **0**, 10.1212/WNL.0000000000007313 (2019).

19. Nedic Erjavec, G. *et al.* Short overview on metabolomic approach and redox changes in psychiatric disorders. *Redox Biol.* **14**, 178–186 (2018).

20. van der Lee, S. J. *et al.* Circulating metabolites and general cognitive ability and dementia:

Evidence from 11 cohort studies. *Alzheimer's Dement.* 1–16 (2018). doi:10.1016/j.jalz.2017.11.012

21. Zaitlen, N. *et al.* Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLoS Genet.* **9**, (2013).

22. Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, D603-10 (2009).

23. Wishart, D. S. *et al.* HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41**, 801–807 (2013).

24. Wishart, D. S. *et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).

25. Willemsen, G. *et al.* The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res. Hum. Genet.* **13**, 231–45 (2010).

26. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

27. Speed, D., Hemani, G., Johnson, M. R. & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **91**, 1011–1021 (2012).

28. Speed, D., Cai, N., Johnson, M. R., Nejentsev, S. & Balding, D. J. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* (2017). doi:10.1038/ng.3865

29. Yet, I. *et al.* Genetic influences on metabolite levels: A comparison across metabolomic platforms. *PLoS One* **11**, (2016).

30. Tabassum, R. *et al.* Genetics of human plasma lipidome: Understanding lipid metabolism and its link to diseases beyond traditional lipids. *bioRxiv* (2018). doi:10.1101/457960

31. Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **47**, 1114–1120 (2015).

32. Visscher, P. M. *et al.* Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLoS Genet.* **10**, (2014).

33. Gallois, A. *et al.* A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. *bioRxiv* (2018). doi:http://dx.doi.org/10.1101/461848

34. Wittemans, L. B. L. *et al.* Assessing the causal association of glycine with risk of cardio-metabolic diseases. *Nat. Commun.* **10**, 1–13 (2019).

35. Demirkan, A. *et al.* Genome-wide association study of plasma lipids. *bioRxiv* (2019). doi:http://dx.doi.org/10.1101/621334

36. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, 1–10 (2015).

37. Wainschtein, P. *et al.* Recovery of trait heritability from whole genome sequence data. *bioRxiv* (2019). doi:http://dx.doi.org/10.1101/588020

38. Yousri, N. A. *et al.* Whole-exome sequencing identifies common and rare variant metabolic QTLs in a Middle Eastern population. *Nat. Commun.* **9**, 1–13 (2018).

39. Boomsma, D. I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res. Hum. Genet.* **9**, 849–57 (2006).

40. Willemsen, G. *et al.* The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Res. Hum. Genet.* **16**, 271–81 (2013).

41. Soininen, P., Kangas, A. J., Würtz, P., Suna, T. & Ala-Korpela, M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. *Circ. Cardiovasc. Genet.* **8**, 192–206 (2015).

42. Würtz, P. *et al.* Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technology. *Am. J. Epidemiol.* **186**, 1–13 (2017).

43. Gonzalez-Covarrubias, V. *et al.* Lipidomics of familial longevity. *Aging Cell* **12**, 426–434 (2013).

44. Dane, A. D. *et al.* Integrating metabolomics profiling measurements across multiple biobanks. *Anal. Chem.* **86**, 4110–4114 (2014).

45. Demirkan, A. *et al.* Insight in Genome-Wide Association of Metabolite Quantitative Traits by Exome Sequence Analyses. *PLoS Genet.* **11**, e1004835 (2015).

46. Goek, O. N. *et al.* Serum metabolite concentrations and decreased GFR in the general population. *Am. J. Kidney Dis.* **60**, 197–206 (2012).

47. Römisch-Margl, W. *et al.* Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. *Metabolomics* **8**, 133–142 (2012).

48. Buuren, S. van & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **45**, (2011).

49. Kettunen, J. *et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 11122 (2016).

50. Draisma, H. H. M. *et al.* Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nat. Commun.* **6**, 7208 (2015).

51. Boomsma, D. I. *et al.* The Genome of the Netherlands: design, and project goals. *Eur. J. Hum. Genet.* **22**, 221–227 (2014).

52. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

53. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

54. Fedko, I. O. *et al.* Estimation of Genetic Relationships Between Individuals Across Cohorts and Platforms: Application to Childhood Height. *Behav. Genet.* **45**, 514–528 (2015).

55. Deelen, P. *et al.* Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of the Netherlands'. *Eur. J. Hum. Genet.* **22**, 1321–1326 (2014).

56. Liu, E. Y., Li, M., Wang, W. & Li, Y. MaCH-Admix: Genotype Imputation for Admixed Populations. *Genet. Epidemiol.* **37**, 25–37 (2013).

57. Loh, P. R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).

58. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–81 (2012).

59. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

60. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–

1287 (2016).

61. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

62. Abdellaoui, A. *et al.* Population structure, migration, and diversifying selection in the Netherlands. *Eur. J. Hum. Genet.* **21**, 1277–1285 (2013).

63. Kim, S. *et al.* PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Res.* **47**, D1102–D1109 (2019).

64. Hastings, J. *et al.* ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–D1219 (2016).

65. Heller, S. R., McNaught, A., Pletnev, I., Stein, S. & Tchekhovskoi, D. InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* **7**, 1–34 (2015).

66. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–11 (2001).

67. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).

68. Hayes, J. F. & Hill, W. G. Modification of Estimates of Parameters in the Construction of Genetic Selection Indices (' Bending '). *Biometrics* **37**, 483–493 (1981).

69. Viechtbauer, W. Conducting Meta-Analyses in R with the metafor Package. *J. Stat. Softw.* **36**, 1–48 (2010).

70. Hedges, L. V., Tipton, E. & Johnson, M. C. Robust variance estimation in meta-regression with

dependent effect size estimates. *Res. Synth. Methods* **1**, 39–65 (2010).

71. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**, 289–300 (1995).

Figures

Figure 1. Flowchart describing the filtering of metabolite SNPs, GRM construction and 4-variance component models.

This flowchart describes how the 242,580 metabolite-SNP associations as identified from GWA and rare-variant analyses (**Supplementary Note 1**; **Supplementary Data 1**) were converted to NCBI build 37, extracted for NTR participants from the 1000GP3 imputed data and filtered on MAF, HWE and R^2 (blue boxes at top of the figure indicated by the red curly bracket). The metabolite-SNP associations of the filtered SNPs were clumped ($r^2 = 0.10$) to obtain the metabolite loci and LD-proxies of the lipid and the organic acids, respectively (blue). To obtain the non-superclass loci, the superclass-specific loci and LD-proxies were removed from the overall list of metabolite-SNP associations and prior to clumping (blue). The lipid-loci, not-lipid loci, organic acid loci and not-organic acid loci give rise to four GRMs, respectively, as indicated by the black boxes and arrows in the flowchart. The two additional GRMs included in the 4-variance component GREML models are based on the cross-platform imputed SNPs (see **Methods**), where the lipid and organic acid loci, LD-proxies and 50 kb surrounding these SNPs have been removed from one of the cross-platform GRMs (black boxes in flowchart). The bottom part (in orange) of the flowchart describes the 4-variance component GREML model separately for the lipid and organic acid analyses (indicated by red curly brackets). To indicate which GRMs are used to calculate which variance components orange arrows have been drawn from the GRMs to the variance components. The different (combinations) of variance components give rise to the five different heritability estimates (h^2_{total}, h^2_{SNP}, $h^2_{GW\text{-}Class}$, $h^2_{GW\text{-}Notclass}$ and $h^2_{GW\text{-}loci}$), the final part of the flowcharts provides an overview of how these heritability estimates are derived (orange).

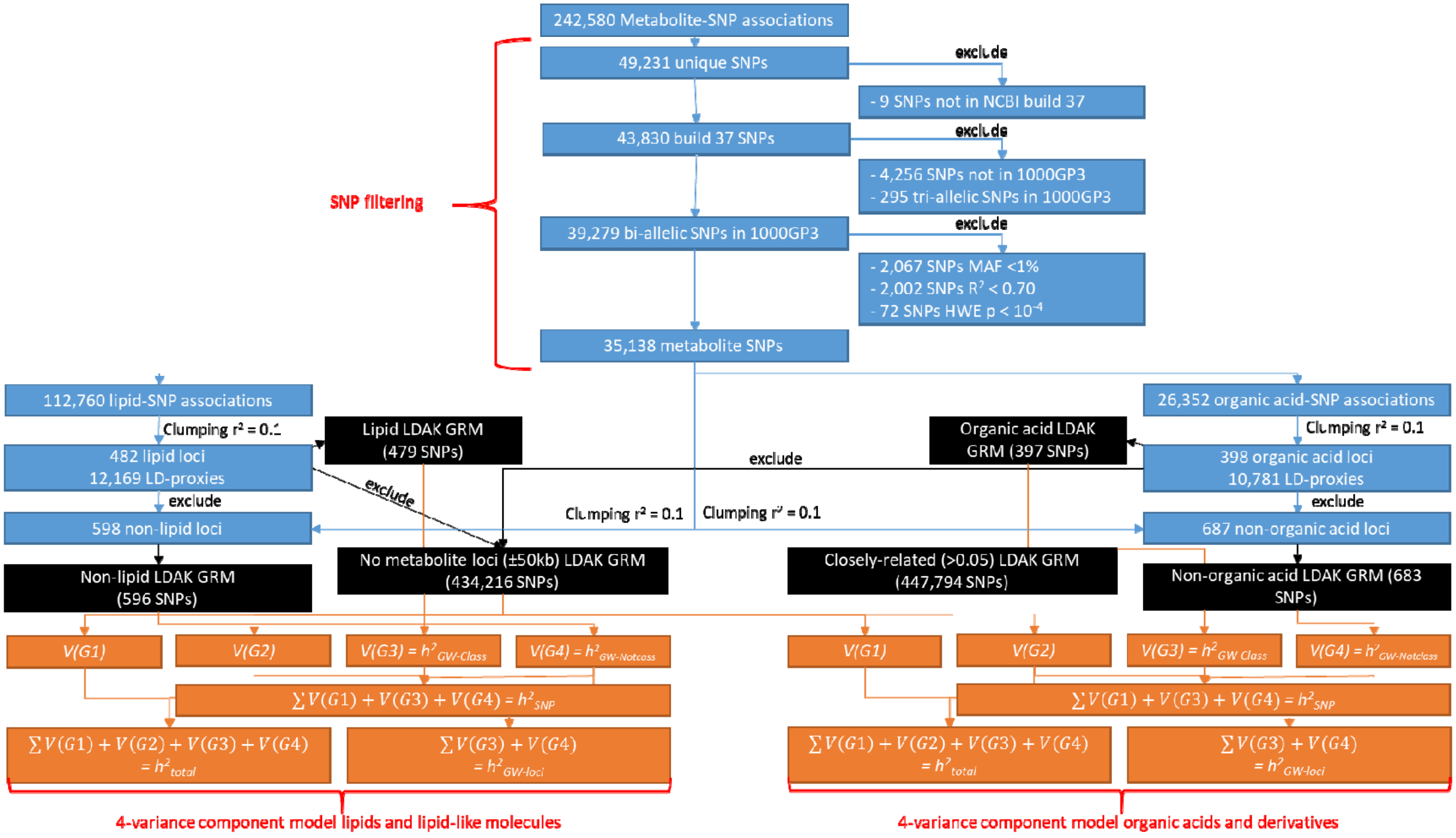

242,580 Metabolite-SNP associations
49,231 unique SNPs
exclude
- 9 SNPs not in NCBI build 37
43,830 build 37 SNPs
exclude
- 4,256 SNPs not in 1000GP3
- 295 tri-allelic SNPs in 1000GP3
39,279 bi-allelic SNPs in 1000GP3
exclude
- 2,067 SNPs MAF <1%
- 2,002 SNPs R² < 0.70
- 72 SNPs HWE p < 10⁻⁴
35,138 metabolite SNPs
SNP filtering
112,760 lipid-SNP associations
Clumping r² = 0.1
482 lipid loci
12,169 LD-proxies
exclude
598 non-lipid loci
Non-lipid LDAK GRM
(596 SNPs)
Lipid LDAK GRM
(479 SNPs)
exclude
Clumping r² = 0.1
Clumping r² = 0.1
exclude
No metabolite loci (±50kb) LDAK GRM
(434,216 SNPs)
Organic acid LDAK
GRM (397 SNPs)
Closely-related (>0.05) LDAK GRM
(447,794 SNPs)
26,352 organic acid-SNP associations
Clumping r² = 0.1
398 organic acid loci
10,781 LD-proxies
exclude
687 non-organic acid loci
Non-organic acid LDAK GRM (683
SNPs)
V(G1)
V(G2)
V(G3) = h²GW-Class
V(G4) = h²GW-Notcass
∑V(G1) + V(G3) + V(G4) = h²SNP
∑V(G1) + V(G2) + V(G3) + V(G4) = h²total
∑V(G3) + V(G4) = h²GW-loci
4-variance component model lipids and lipid-like molecules
V(G1)
V(G2)
V(G3) = h²GW Class
V(G4) = h²GW-Notclass
∑V(G1) + V(G3) + V(G4) = h²SNP
∑V(G1) + V(G2) + V(G3) + V(G4) = h²total
∑V(G3) + V(G4) = h²GW-loci
4-variance component model organic acids and derivatives

Figure 2. Heritability of all 52 ‘carboxylic acids and derivatives’ successfully analyzed across all four metabolomics platforms by class.

Box- and dotplots of the h^2_{total}, h^2_{SNP} and $h^2_{GW\text{-}loci}$ for all 52 successfully analyzed ‘carboxylic acids and derivatives’ by class. The left-hand side of the figure is a close-up of the -0.08 – 0.15 part of the heritability range, focusing on the $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}Notclass}$ estimates. The boxes denote the 25th and 75th percentile (bottom and top of box), and median value (horizontal band inside box). The whiskers indicate the values observed within up to 1.5 times the interquartile range above and below the box.

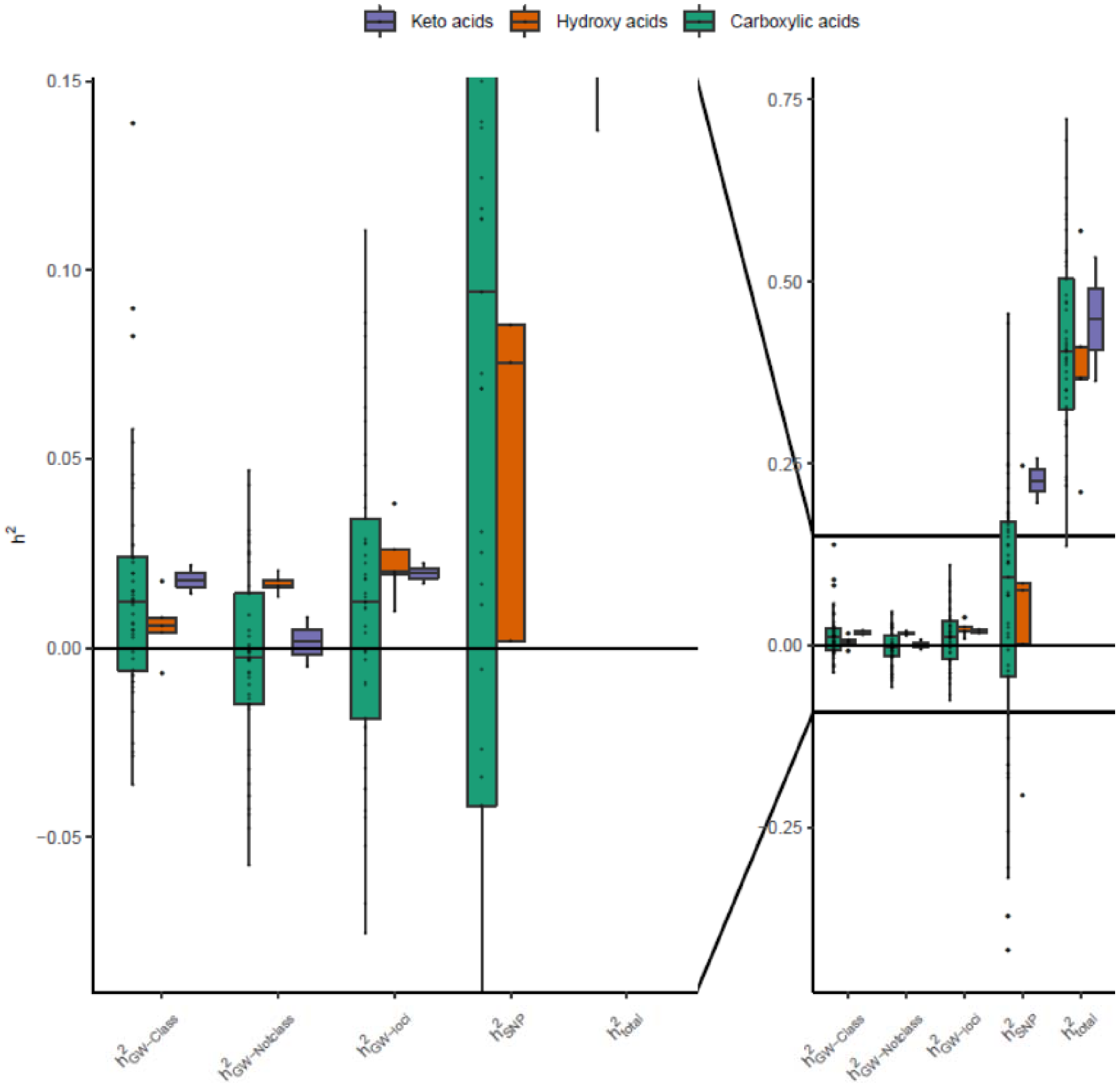
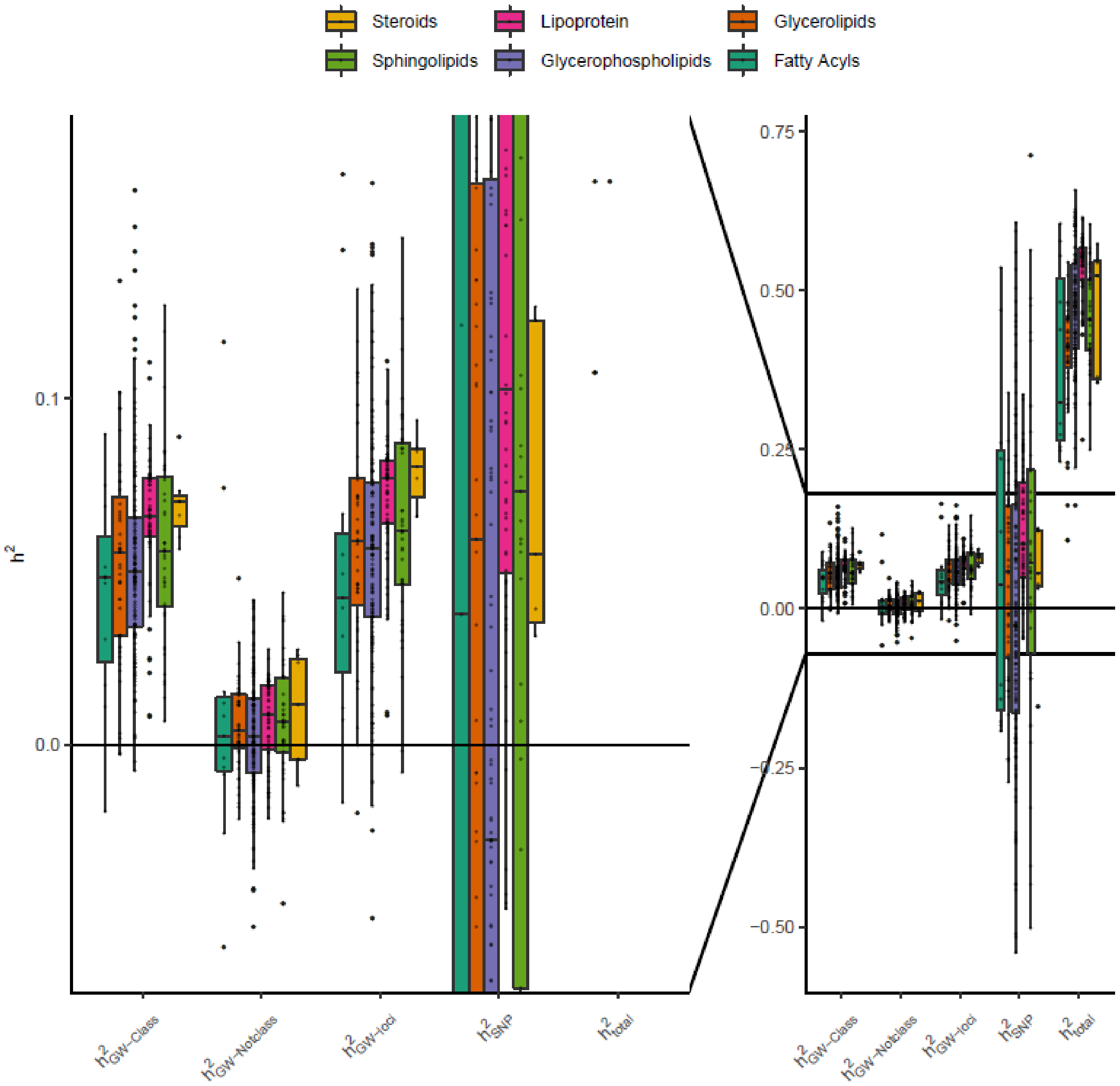

Keto acids
Hydroxy acids
Carboxylic acids
h²
0.15
0.10
0.05
0.00
−0.05
0.75
0.50
0.25
0.00
−0.25
$h^2_{GW-Class}$
$h^2_{GW-Notclass}$
$h^2_{GW-loci}$
h^2_{SNP}
h^2_{total}

Figure 3. Heritability of all 309 ed 'lipids' successfully analyzed across all four metabolomics platforms by class.

Box- and dotplots of the h^2_{total}, h^2_{SNP} and $h^2_{GW\text{-}loci}$ for all 309 successfully analyzed 'lipids' by class. The left-hand side of the figure is a close-up of the -0.06 – 0.17 part of the heritability range, focusing on the $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}Notclass}$ estimates. The boxes denote the 25th and 75th percentile (bottom and top of box), and median value (horizontal band inside box). The whiskers indicate the values observed within up to 1.5 times the interquartile range above and below the box.


Steroids
Lipoprotein
Glycerolipids
Sphingolipids
Glycerophospholipids
Fatty Acyls
h^2
0.1
0.0
$h^2_{GW-Class}$
$h^2_{GW-Notclass}$
$h^2_{GW-loci}$
h^2_{SNP}
h^2_{total}
0.75
0.50
0.25
0.00
−0.25
−0.50

Tables

Table 1. Overview of the number of unique metabolites, for which significant SNP-metabolite associations have been published, per Human Metabolome Database[22–24] ‘super class’.

See **Supplementary Data 1** for an overview of the exact metabolites classified per ‘super class’, ‘class’ and ‘subclass’, as well as the SNPs associated with each metabolite.

Super class	**Number of unique metabolites**
Lipids and lipid-like molecules (e.g., lipids)	662
Organic acids and derivatives (e.g., organic acids)	182
Organoheterocyclic compounds	45
Organic oxygen compounds	19
Nucleosides, nucleotides, and analogues	12
Benzenoids	12
Organic nitrogen compounds	11
Phenylpropanoids and polyketides	4
Proteins	3
Organic compounds	1
Trichlorophenols	1
Organooxygen compounds	1

Table 2. Participant characteristics after preprocessing per metabolomics platform.

This table gives an overview of the number of individuals (N) per platform, specifies the number of families these individuals belong to and the percentage of females and twins in each dataset. In addition, for each platform the mean and standard deviation (SD) of the age at blood draw in years, the body-mass-index (BMI), the cholesterol level in mmol/l, the low-density lipoprotein cholesterol (LDL) levels in mmol/l and the high-density lipoprotein cholesterol (HDL) levels in mmol/l are given.

Metabolomics platform	**N**	**N families**	**Age* (mean ± SD)**	**Female (%)**	**Twins (%)**	**BMI (mean ± SD)**	**Cholesterol$^{\$}$ (mean ± SD)**	**LDL$^{\$}$ (mean ± SD)**	**HDL$^{\$}$ (mean ± SD)**
All Participants	5,117	2,445	42.1 ± 14.2	62.8%	63.4%	24.8 ± 4.1	4.9 ± 1.2	3.0 ± 1.0	1.7 ± 1.0
Nightingale Health ^{1}H-NMR	4,227	2,179	40.7 ± 13.7	67.3%	69.7%	24.6 ± 4.0	4.9 ± 1.2	3.0 ± 1.0	1.7 ± 1.0
UPLC-MS Lipidomics	2,324	1,251	39.0 ± 12.9	66.6%	89.2%	24.4 ± 4.1	5.0 ± 1.0	3.0 ± 0.9	1.4 ± 0.4
Leiden ^{1}H-NMR	2,324	1,323	37.6 ± 12.5	67.0%	89.0%	24.2 ± 4.1	4.6 ± 1.3	2.7 ± 1.0	2.0 ± 1.4
Biocrates	1,448	946	45.7 ± 15.3	43.8%	39.6%	25.2 ± 3.9	4.6 ± 1.5	2.8 ± 1.1	2.3 ± 1.7

* Age at blood draw in years; $^{\$}$ levels in mmol/l.

Table 3. Summary of the heritability estimates of the four-variance component models for the 309 'lipids' and the 52 'organic acids' analyzed across all four metabolomics platforms.

The mean, median and range of the total heritability (h^2_{total}), SNP heritability (h^2_{snp}), heritability based on the 479 significant metabolite loci for the 'lipids' or the 397 significant metabolite loci for the 'organic acids' ($h^2_{GW\text{-}Class}$), the 596-683 significant metabolite loci not belonging to these classes ($h^2_{GW\text{-}Notclass}$) and the total heritability explained by metabolite loci (e.g., sum of $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}Notclass}$: $h^2_{GW\text{-}loci}$), as well as their standard errors (s.e.'s), are depicted for all 361 successfully analyzed metabolites as included on all platforms. **Supplementary Table 1** denotes which metabolites belong to each class.

		Lipids and lipid-like molecules		**Organic acids and derivatives**	
		estimate	**s.e.**	**estimate**	**s.e.**
h^2_{total}	*mean*	0.47	0.04	0.41	0.04
	median	0.47	0.03	0.40	0.03
	range	(0.11 - 0.66)	(0.02 - 0.07)	(0.14 - 0.72)	(0.02 - 0.07)
h^2_{SNP}	*mean*	0.05	0.24	0.05	0.24
	median	0.06	0.22	0.09	0.23
	range	(-0.54 - 0.71)	(0.11 - 0.35)	(-0.42 - 0.46)	(0.11 - 0.34)
$h^2_{GW\text{-}loci}$	*mean*	0.06	0.03	0.01	0.02
	median	0.06	0.03	0.02	0.02
	range	(-0.05 - 0.16)	(0.01 - 0.04)	(-0.08 - 0.11)	(0.01 - 0.04)
$h^2_{GW\text{-}Class}$	*mean*	0.06	0.02	0.01	0.02
	median	0.06	0.02	0.01	0.02
	range	(-0.02 - 0.16)	(0.01 - 0.03)	(-0.04 - 0.14)	(0.01 - 0.03)
$h^2_{GW\text{-}Notclass}$	*mean*	0.00	0.02	0.00	0.02
	median	0.01	0.02	0.00	0.02
	range	(-0.06 - 0.12)	(0.01 - 0.03)	(-0.06 - 0.05)	(0.01 - 0.03)

Table 4. Summary of the heritability estimates of the four-variance component models for the 17 essential and the 14 non-essential amino acids analyzed across all four metabolomics platforms.

The mean, median and range of the total heritability (h^2_{total}), SNP heritability (h^2_{snp}) and heritability based on the 397 significant metabolite loci for the 'organic acids' ($h^2_{GW\text{-}Class}$), the 683 significant metabolite loci not belonging to this class ($h^2_{GW\text{-}Notclass}$) and the total heritability explained by metabolite loci (e.g., sum of $h^2_{GW\text{-}Class}$ and $h^2_{GW\text{-}Notclass}$: $h^2_{GW\text{-}loci}$), as well as their standard errors (s.e.'s), are depicted for all 31 successfully analyzed essential and non-essential amino acids as included on all platforms. **Supplementary Table 1** denotes which metabolites belong to each class.

		Essential amino acids		**Non-essential amino acids**	
		estimate	**s.e.**	**estimate**	**s.e.**
h^2_{total}	*mean*	0.42	0.04	0.39	0.04
	median	0.40	0.03	0.39	0.04
	range	(0.23 - 0.64)	(0.02 - 0.07)	(0.22 - 0.69)	(0.03 - 0.07)
h^2_{SNP}	*mean*	-0.01	0.24	0.10	0.25
	median	-0.01	0.23	0.07	0.24
	range	(-0.42 - 0.46)	(0.12 - 0.34)	(-0.18 - 0.44)	(0.12 - 0.34)
$h^2_{GW\text{-}loci}$	*mean*	0.00	0.02	0.02	0.03
	median	0.00	0.02	0.01	0.03
	range	(-0.05 - 0.05)	(0.01 - 0.03)	(-0.07 - 0.11)	(0.01 - 0.04)
$h^2_{GW\text{-}Class}$	*mean*	0.01	0.02	0.03	0.02
	median	0.00	0.02	0.01	0.02
	range	(-0.03 - 0.05)	(0.01 - 0.02)	(-0.03 - 0.14)	(0.01 - 0.03)
$h^2_{GW\text{-}Notclass}$	*mean*	-0.01	0.02	0.00	0.02
	median	-0.01	0.02	0.00	0.02
	range	(-0.06 - 0.04)	(0.01 - 0.03)	(-0.04 - 0.03)	(0.01 - 0.03)