Longitudinal plasma metabolomics of aging and sex

Burcu F. Darst[1], Rebecca L. Koscik[2], Kirk J. Hogan[2,3], Sterling C. Johnson[2,4,5], Corinne D. Engelman [1,2,5]

[1]Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA
[2]Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA
[3]Department of Anesthesiology, University of Wisconsin School of Medicine and Public Health, Madison, WI 53792, USA
[4]Geriatric Research Education and Clinical Center, Wm. S. Middleton Memorial VA Hospital, Madison, WI 53705, USA
[5]Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI 53792, USA

Correspondence to: Burcu F. Darst; **email:** bdarst@wisc.edu
Keywords: metabolomics, lipidomics, genomics, aging, sex, heritability
Received: September 20, 2018 **Accepted:** February 17, 2019 **Published:** February 24, 2019

ABSTRACT

Understanding how metabolites are longitudinally influenced by age and sex could facilitate the identification of metabolomic profiles and trajectories that indicate disease risk. We investigated the metabolomics of age and sex using longitudinal plasma samples from the Wisconsin Registry for Alzheimer's Prevention (WRAP), a cohort of participants who were dementia free at enrollment. Metabolomic profiles were quantified for 2,344 fasting plasma samples among 1,212 participants, each with up to three study visits. Of 1,097 metabolites tested, 623 (56.8%) were associated with age and 695 (63.4%) with sex after correcting for multiple testing. Approximately twice as many metabolites were associated with age in stratified analyses of women versus men, and 68 metabolite trajectories significantly differed by sex, most notably including sphingolipids, which tended to increase in women and decrease in men with age. Using genome-wide genotyping, we also report the heritabilities of metabolites investigated, which ranged dramatically (0.2–99.2%); however, the median heritability of 36.2% suggests that many metabolites are highly influenced by a complex combination of genomic and environmental influences. These findings offer a more profound description of the aging process and may inform many new hypotheses regarding the role metabolites play in healthy and accelerated aging.

INTRODUCTION

The metabolome represents the functional endpoints of a complex network of biological events, including genomic, epigenomic, transcriptomic, proteomic, and environmental factors [1]. Being the final downstream product, the metabolome is the closest to the phenotype among the biological systems [2], making it particularly relevant to investigate. Age is known to be the single largest risk factor of most prevalent diseases in developed countries [3]. A better understanding of how the metabolome changes with age could further reveal the mechanisms by which age influences disease risk and could facilitate the identification of high-risk metabolomic profiles that are suggestive of the early stages of particular diseases.

Previous studies have provided important evidence that age and sex influence the metabolome [4-10]. While informative, these studies are limited by their cross-sectional designs and the relatively small number of metabolites assessed by most. According to the Human Metabolite Database (HMDB) v4.0, there are an estimated 25,424 blood metabolites [11]. However, due to current technical limitations in identifying and quantifying metabolites, most recent studies have only been able to confidently capture ~100-600 of these. A larger panel of metabolites will provide a more comprehensive understanding of the metabolomics of age and sex. Further, in order to assess the metabolomics of aging, it is crucial to use a longitudinal study design that can capture age-related phenomena, particularly due to the high variability of metabolites [12]. Longitudinal assessments also facilitate the examination of metabolite trajectories, which can address important biological questions.

Using longitudinal plasma samples from the Wisconsin Registry for Alzheimer's Prevention (WRAP), we investigated how a large panel of metabolites is influenced by age and sex, and whether metabolite trajectories vary by sex. To facilitate the interpretation of our results and determine whether identified metabolites are more strongly influenced by genetic or environmental factors, we used genome-wide genotyping data to assess the heritability (h^2) of metabolites.

RESULTS

Participants

A total of 1,212 WRAP participants with 2,344 longitudinal fasting plasma samples were available for analyses. At the baseline visit for the current study, participants were 61 years old on average, 69% were female, and 94% were Caucasian (Table 1). Most individuals were unrelated (n=825), but 147 families had >1 individual (family sizes ranged from 1-9 members, with an average of 1.2 individuals per family). Analyses stratified by sex included 838 women and 374 men, who had similar characteristics with the exception of more men taking cholesterol lowering medications than women. Participants each had 1,097 plasma metabolites available for analyses, 347 (31.6%) of which were of unknown chemical structure. Correlations between metabolites were assessed using Pearson r and the first available sample for each individual (*i.e.*, using a cross-sectional approach). Metabolites were largely uncorrelated with each other (Figure S1). Properties of each metabolite, such as biochemical name, super pathway, and sub pathway, are described in Table S1.

Metabolome-wide association study

Associations were tested using linear mixed effects regression models implemented in the SAS MIXED procedure. Primary predictors included age and sex, which were assessed within the same models. To examine effect modification of the metabolomics trajectories by sex, analyses were repeated stratifying the sample by sex. To assess the statistical significance of the effect modification, separate models were run that included an interaction term for age-by-sex using the full sample (men and women combined). All models included random intercepts for within-subject correlations (due to repeated measures) and within-family correlations (due to siblings). Models included fixed effects for age, sex, self-reported race, body-mass in-

Table 1. WRAP Participant Characteristics at Baseline for the Current Study. Mean (SD) or N (%).

Characteristic	Overall (N=1,212, obs=2,344)	Male (n=374, obs=731)	Female (n=838, obs=1,613)
Age (years)	60.8 (6.7)	61.2 (6.9)	60.7 (6.6)
Caucasian	1,135 (93.7)	351 (93.9)	784 (93.6)
Cholesterol lowering medication	387 (31.9)	146 (39.0)*	241 (28.8)*
Sample storage (days)	1,510.5 (415.7)	1,511.2 (424.3)	1,510.2 (412.0)

obs=number of longitudinal observations
*Differs between men and women with *P*=3.9e-4

dex, sample storage time, and cholesterol lowering medication use, which was the most commonly used class of medications in our sample. Since our cohort has increased risk for Alzheimer's disease, we performed sensitivity analyses including additional fixed effects for parental history of Alzheimer's and participant cognitive impairment status, and results were largely unchanged. Each set of analyses was corrected for multiple testing using the Benjamini-Hochberg [13] adjustment with an alpha of 0.05.

Aging metabolomics

All metabolome-wide association results are summarized in Table 2 and detailed in Table S1. After adjusting for multiple testing, the levels of 623 metabolites (56.8% of metabolites assessed) significantly changed with age, 523 of which increased with age (Figure S2A and Figure 1). Of the total 34 steroid lipids

tested, 29 significantly decreased with age (including 19/22 androgenic, 5/5 progestin, 4/4 pregnenolone, and 1/3 corticosteroids), while two, 11-ketoetiocholanolone glucuronide, an androgenic steroid, and cortisol, significantly increased with age. Higher levels of most fatty acid lipids were associated with increased age (including 13/14 long chain fatty acids, 28/34 acyl-carnitines, and 42/78 other fatty acids). Higher levels of sphingolipids tended to be associated with increased age (24/25 associated sphingolipids).

The majority of amino acids associated with age increased with age (87.6% or 92/105 associated amino acids), including glutamine and tyrosine of the 20 common amino acids that are encoded directly by the genetic code. Five other common amino acids decreased with age (histidine, threonine, tryptophan, leucine, and serine), while the 13 others were not associated with age.
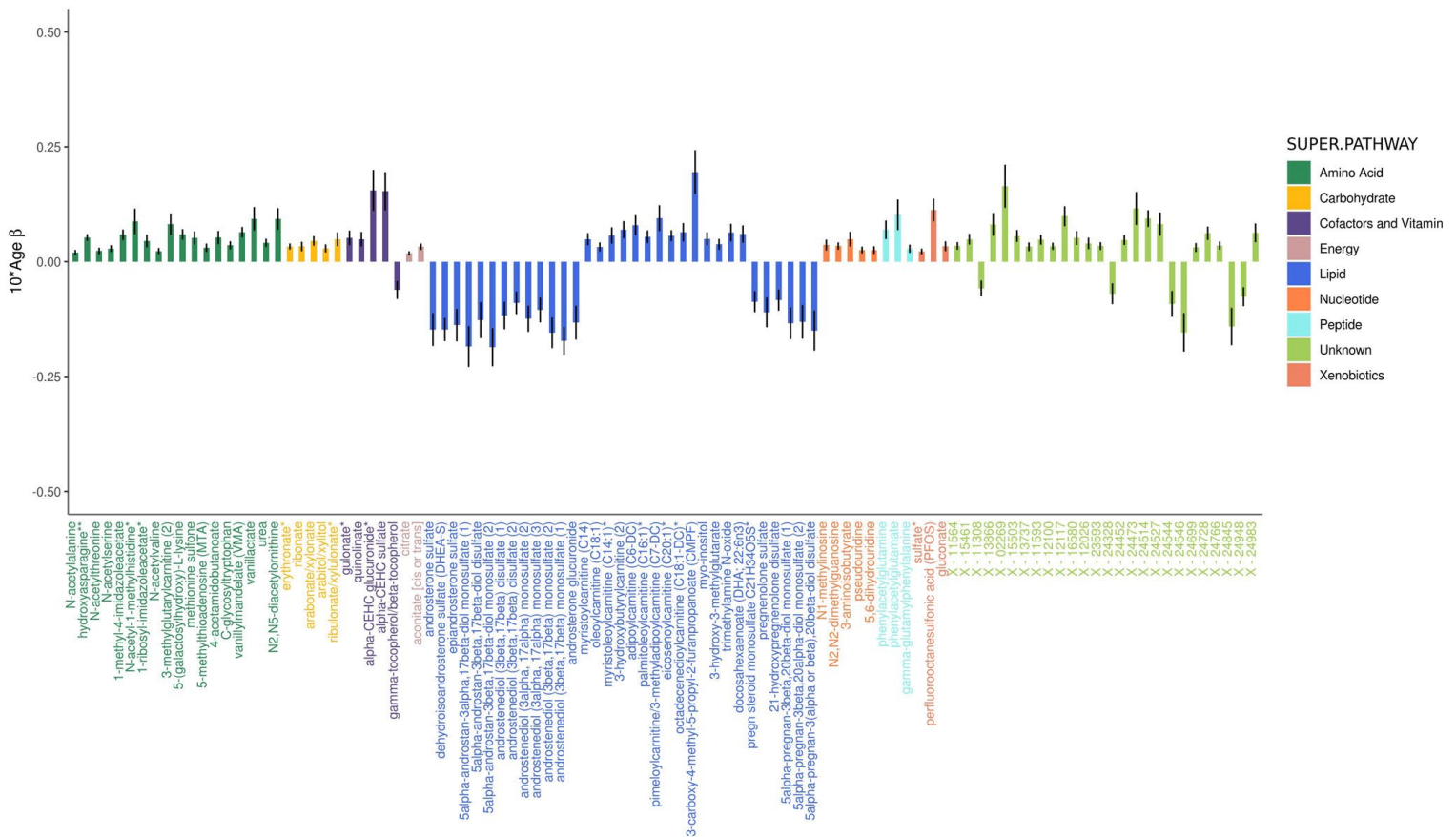

Figure 1. Adjusted effects of a 10-year increase in age on the top 100 metabolites most strongly influenced by age. Positive values indicate the amount a metabolite increased over 10 years, whereas negative values indicate the amount a metabolite decreased over 10 years. Black vertical lines indicate standard errors.

Table 2. Metabolome-wide association results summary. Number of metabolites associated with each trait by pathways and recurrent sub pathways after correcting for multiple comparisons.

Metabolite Super/Sub Pathways (#Metabolites)	Age			Sex			Age in Women			Age in Men			Age*Sex (female/male)				
	Tot	+β	−β	Tot	+β	−β	Tot	+β	−β	Tot	+β	−β	Tot	+/+	−/−	+/−	−/+
Amino Acids (175)	105	92	13	112	27	85	98	90	8	42	36	6	7	2	0	4	1
Common Amino Acids (20)	7	2	5	15	2	13	6	3	3	0	0	0	0	0	0	0	0
Carbohydrates (23)	16	16	0	13	6	7	17	17	0	5	5	0	1	1	0	0	0
Cofactors and Vitamins (28)	20	18	2	15	7	8	20	17	2	9	7	2	2	1	0	1	0
Energy (8)	6	6	0	5	3	2	4	4	0	2	2	0	0	0	0	0	0
Lipids (353)	194	152	42	252	177	75	187	155	32	72	38	34	46	12	12	20	2
Fatty Acids (126)	88	83	4	90	60	30	88	87	1	24	22	2	6	4	0	1	1
Acylcarnitines (34)	28	28	0	26	9	17	31	31	0	7	7	0	4	3	0	1	0
Phospholipids (65)	20	12	8	52	48	4	15	13	2	15	6	9	8	0	3	5	0
Lysophospholipids (24)	2	1	1	17	15	2	2	2	0	1	0	1	4	0	0	4	0
Phosphatidylcholines (19)	8	5	3	17	16	1	7	5	2	10	4	6	4	0	3	1	0
Phosphatidylethanolamine (9)	4	3	1	7	7	0	3	3	0	1	0	1	0	0	0	0	0
Sphingolipids (40)	25	24	1	35	35	0	26	25	1	4	0	4	15	4	2	9	0
Steroids (34)	31	2	29	29	2	27	30	2	28	19	1	18	8	0	7	0	1
Androgenic (22)	20	1	19	20	0	20	20	1	19	15	0	15	4	0	2	0	1
Progestin (5)	5	0	5	2	0	2	5	0	5	0	0	0	5	0	5	0	0
Pregnenolone (4)	4	0	4	4	0	4	4	0	4	3	0	3	0	0	0	0	0
Corticosteroids (3)	2	1	1	3	2	1	1	1	0	1	1	0	0	0	0	0	0
Nucleotides (35)	20	19	1	24	1	23	17	17	0	11	10	1	1	1	0	0	0
Partially Characterized Molecules (5)	4	4	0	2	0	2	5	5	0	0	0	0	0	0	0	0	0
Peptides (22)	18	18	0	16	3	13	17	17	0	4	4	0	0	0	0	0	0
Xenobiotics (101)	45	40	5	52	18	34	35	31	4	16	15	1	2	0	0	2	0
Unknown (347)	196	158	37	204	67	137	165	138	27	94	72	22	9	3	1	2	3
Total (1,097)	623	523	100	695	309	386	565	492	73	255	189	66	68	20	13	29	6

Shaded rows represent super pathways, which sum to the "Total" row. Sub pathways are indented. In the Sex columns, + means the metabolite was higher in women, whereas − means the metabolite was higher in men. For all other columns, + means the metabolite increased with age, whereas − means it decreased with age. In the Age*Sex columns, +/+ means the metabolite increased with age in both women and men, −/− means it decreased with age in both women and men, +/− means it increased with age in women and decreased with age in men, and −/+ means it decreased with age in women and increased with age in men. Results from the Age and Sex columns were assessed within the same model; results from the Age in Women and Age in Men columns were assessed within separate models stratifying the sample by sex; and results from the Age*Sex column were assessed within a separate model including an age-by-sex interaction term.
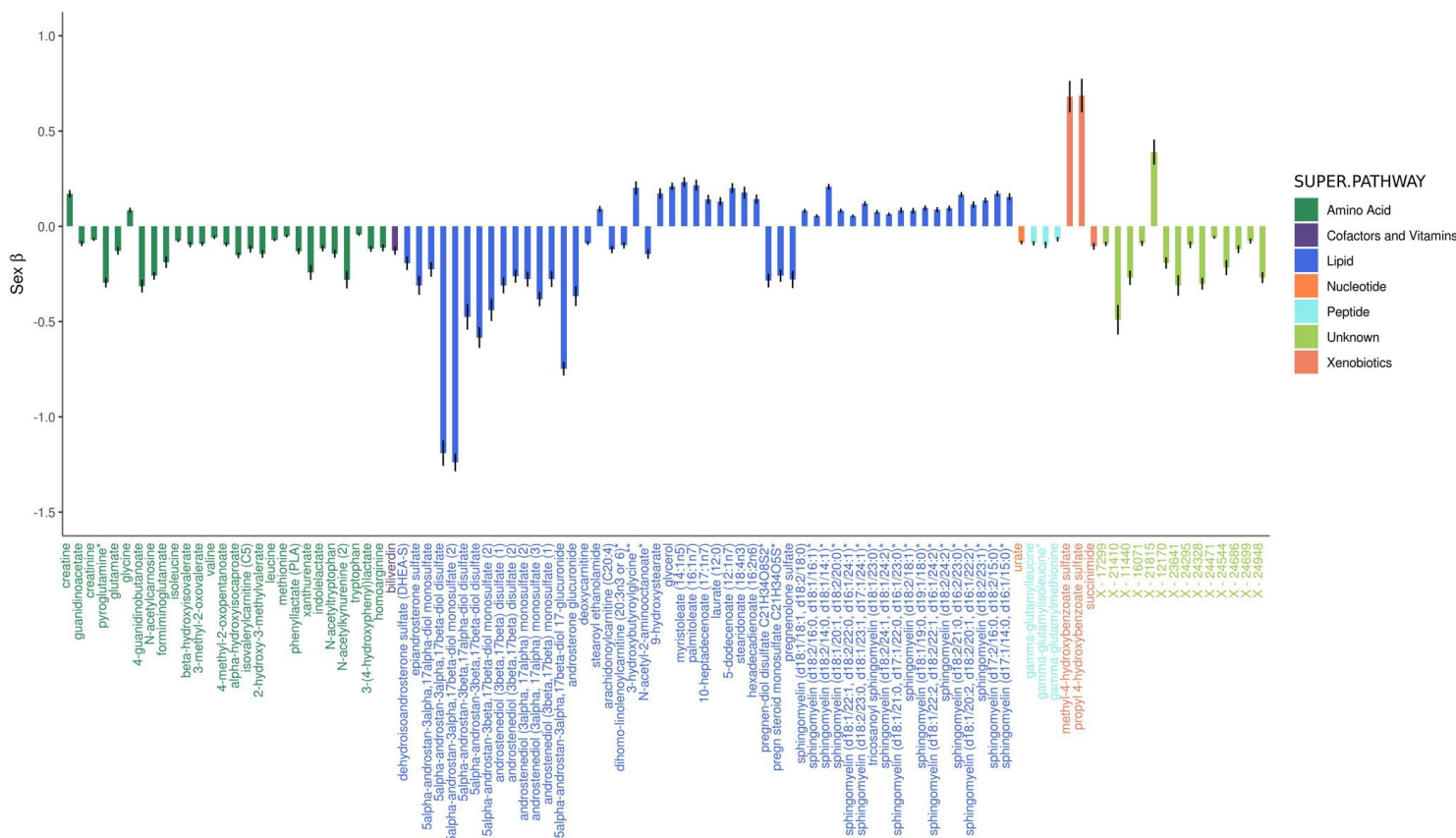
Figure 2. Adjusted effects of the top 100 metabolites most strongly influenced by sex. Positive values indicate that the metabolite was higher in women, whereas negative values indicate that the metabolite was higher in men. Black vertical lines indicate standard errors.

Sex metabolomics

Six hundred and ninety-five metabolites (63.4% of metabolites assessed) significantly differed by sex, with the slight majority (386 metabolites or 55.5%) found in lower levels in women (Figure 2B and Figure 2). Of the metabolites associated with sex, 405 were also associated with age. Twenty-nine steroid lipids were associated with sex, all of which were found in significantly lower levels in women, with the exception of two corticosteroids (cortisol and corticosterone), which were found in higher levels in women. Androgenic steroids constituted the three metabolites most strongly associated with sex (5alpha-androstan-3alpha, 17beta-diol monosulfate, P=1.4e-311, 5alpha-androstan-3alpha, 17beta-diol 17-glucuronide, P=3.1e-228, and 5alpha-androstan-3alpha, 17beta-diol disulfate, P=5.7e-185).

Ninety fatty acids were associated with sex, 60 of which were found in higher levels in women. Acylcarnitine fatty acids were an exception, as 17/26 significantly associated acylcarnitines were found in lower levels in women. Among all tested phospholipids, 73.8% (48/65) were higher in women, as were 87.5% (35/40) of all tested sphingolipids.

The majority of amino acids associated with sex were found in lower levels in women (75.9% or 85/112), including 13 of the 20 common amino acids (alanine, tyrosine, methionine, arginine, proline, aspartate, asparagine, tryptophan, glutamate, phenylalanine, and the three branched-chain amino acids (BCAAs): leucine, isoleucine, and valine), while two were found in higher levels in women (glycine and serine). The remaining five did not significantly differ by sex.

Effect modification of metabolomics trajectories by sex

Analyses stratified by sex identified 565 metabolites (51.5% of metabolites assessed) that were significantly associated with age among women (Figure S3A and Figure S4) and 255 metabolites (23.2% of metabolites assessed) among men (Figure S3B and Figure S5), with 188 being common to both groups.

The trajectories of 68 metabolites (6.2%) significantly differed over time by sex (Figures S3C and S6). The three most significant metabolites were sphingolipids, which were also the largest group of metabolites whose trajectories differed by sex (22.1% or 15/68). Nine of these sphingolipids increased with age among women and decreased with age among men. Several other groups of metabolites had trajectories that also differed by sex, including six fatty acids, five of which showed larger increases with age among women than men; eight steroid lipids, seven of which showed larger decreases with age among women than men; eight phospholipids, five of which increased in women and decreased in men with age; and cholesterol, which increased in women and decreased in men with age.

Metabolite heritability estimates

The h^2 of each metabolite was estimated using a variance components method that jointly models narrow-sense h^2 and the h^2 explained by genotyped variants [14], which allows for the inclusion of both closely and distantly related individuals, as implemented in GCTA [15]. A genetic relationship matrix was created from 272,839 weakly linked ($R^2 < 0.50$) and common (MAF>0.05) directly genotyped variants. Analyses of h^2 were cross-sectional, using the first available metabolomics sample for 1,111 Caucasians that had both metabolomic and genomic data, and adjusted for sex and age. To assess whether metabolite h^2 could influence the effect of age or sex on metabolite levels, Pearson r was used to calculate correlations between h^2 estimates and the strength of associations (*i.e.*, *P*-values) for age and sex.

Metabolite h^2 estimates ranged widely (0.2–99.1%) and had a median h^2 of 36.3%, with a first quartile (Q_1) of 25.5% and a third quartile (Q_3) of 49.7% (Figure 3 and Table S1). The metabolites with the lowest h^2 were three
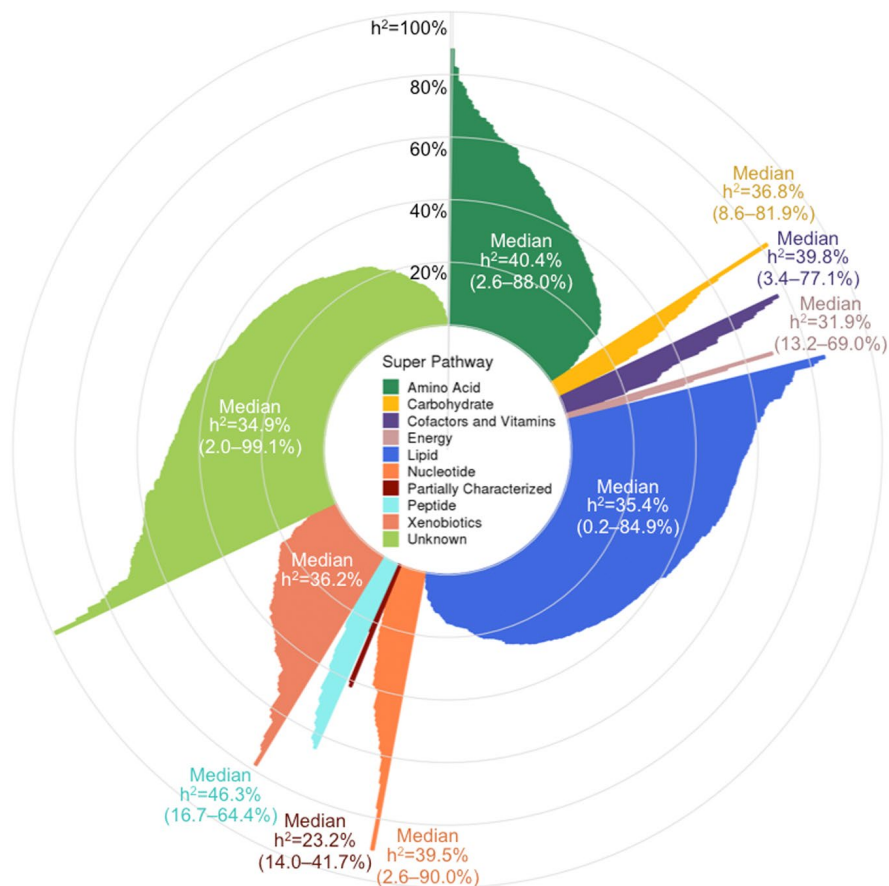

Figure 3. Pinwheel plot of metabolite heritabilities. Each bar indicates the heritability of the corresponding metabolite. Heritability estimates can also be found in Table S1.

lipids: adipoylcarnitine (C6-DC), an acylcarnitine (h^2 = 0.2%), 15-methylpalmitate (i17:0), a branched fatty acid (h^2=0.2%), and glycosyl-N-stearoyl-sphingosine (d18:1/18:0), a ceramide lipid (h^2=0.6%). The metabolites with the highest h^2 estimates were two unknown metabolites (X-12093 and X-24328, h^2=99.1% and 91.1%, respectively) and a nucleotide involved in purine metabolism (N2, N2-dimethylguanosine, h^2=90.0%). Metabolon recently identified X-12093 as N2-acetyl, N6 methyllysine, an amino acid in the lysine catabolic sub pathway.

Super pathway median h^2 estimates ranged from 23.2–46.3%, with peptides having the highest median, followed by amino acids (40.4%), and partially characterized molecules having the lowest median, although the latter pathway only contained five metabolites. Among the metabolite subgroups that were recurrent themes in our association results (*i.e.*, sub pathways highlighted in Table 2), the 20 common amino acids had a median h^2 of 49.3% (Q_1–Q_3: 36.9–65.1%); fatty acids overall had a median h^2 of 30.3% (Q_1–Q_3: 16.9–42.4%), while acylcarnitines had a slightly higher median h^2 of 41.3% (Q_1–Q_3: 26.6–56.0%); phospholipids overall had a median h^2 of 35.9 (Q_1–Q_3: 24.7–53.3%), while phosphatidylcholines had a slightly lower median h^2 of 30.6% (Q_1–Q_3: 22.6–39.6%); sphingolipids had a median h^2 of 42.0% (Q_1–Q_3: 31.8–51.7%); and steroid lipids overall had a median h^2 of 39.6% (Q_1–Q_3: 35.0–50.7%), while androgenic steroids had a median h^2 of 42.5% (Q_1–Q_3: 37.6–50.7%).

Metabolites associated with age and sex had h^2 estimates that were representative of overall metabolite h^2 estimates. Among the 623 metabolites associated with age, the median h^2=36.1% and Q_1–Q_3: 26.2–50.0%. Similarly, among the 695 metabolites associated with sex, the median h^2=37.2% and Q_1-Q_3: 25.6–50.7%. Overall, metabolite h^2 estimates were not correlated with the strength of associations for age or sex (Pearson r=-0.01 and -0.02, respectively).

DISCUSSION

To our knowledge, this is the first longitudinal metabolomics assessment of aging and sex and uses one of the largest panels of metabolites reported to date. Our results provide strong evidence that most plasma metabolite levels are highly influenced by aging and that aging has a broader effect on metabolites in women than men. Metabolites are also highly influenced by sex, with men and women having substantially different metabolomic profiles. We report h^2 estimates on more metabolites than previously reported and find that most are influenced by a complex combination of genetic and environmental factors, consistent with previous studies [16, 17]. How heritable a metabolite was did not appear to influence the effect of age or sex on metabolite levels.

Differences in levels of plasma lipid steroids, including androgens, progestins, and pregnenolones, were among the most significant findings for both age and sex. Steroid differences by sex serve as a proof of concept, as it is well established that androgens are present in lower levels in women than men [18]. Androgens are also known to decrease with age among men [19, 20] and women [21, 22].

Plasma metabolites we identified to be associated with sex and age are consistent with findings from previous cross-sectional studies. The UK Adult Twin Registry (TwinsUK) study reported 165 out of 280 (58.9%) tested serum and plasma metabolites to be associated with age in cross-sectional analyses [5]. Our data had 114 of these 165 metabolites, of which 72 were significantly associated with age, and 66 had effects that were in the same direction as those reported in the TwinsUK study (Table S2). The metabolites that had the opposite direction of effect between studies were four amino acids (dimethylarginine, leucine, serine, and tryptophan), one nucleotide (uridine), and one xenobiotic (theophylline), all of which we reported decreased with age, with the exception of dimethyl-arginine, which increased with age, contradictory to findings from the TwinsUK study. However, other studies have reported that serum tryptophan levels decrease with age [4, 9]. Among the 66 metabolites with the same effect, 27 were lipids, all of which increased with age (the majority were fatty acids, including 10 long chain fatty acids, six polyunsaturated fatty acids, and four other fatty acids), and 14 were amino acids (including glutamine and tyrosine, which both increased with age, and histidine, which decreased with age).

The Cooperative Health Research in the Region of Augsburg (KORA F4) study, which was also cross-sectional, reported 180 out of 507 (35.5%) tested serum metabolites to be associated with sex [7]. Our data had 98 of these 180 metabolites, of which 84 were significantly associated with sex, and all had effects that were in the same direction as those reported in KORA F4 (Table S3). Among these were 33 amino acids (including 11 common amino acids, all of which were lower in women except glycine and serine, which is also consistent with Mittelstrass et al. [6]); 18 lipids (including five long chain fatty acids and three medium chain acids, all of which were higher in women, and three androgenic steroids, all of which were lower in women); and 18 unknown metabolites (all but one were

lower in women). The single most significant finding in the KORA F4 study was the third most significant in our study (5alpha-androstan-3beta,17beta-diol disulfate, an androgenic steroid; the two other andro-genic steroids that were our first and second most significant sex findings were not tested in the KORA F4 study). Also consistent with our findings, other studies have reported serum and plasma phosphatidylcholines and sphingo-lipids levels to be higher in women than men [6, 8, 23], and serum acylcarnitines to be lower in women [6].

Consistent with results from our sex-stratified analyses, a previous KORA F4 publication also reported serum sphingolipids to increase in concentrations with age among women and acylcarnitines to increase with age among both women and men [4]. It has been shown that higher levels of acylcarnitines are associated with higher risk for type 2 diabetes and obesity, which are increasingly common conditions in the US, and correlate with poor glycemic control [24]. Follow up research is needed to investigate whether acylcarnitines are causally associated with obesity and could serve as a target for obesity intervention. The KORA F4 study, which had a sample of 1,038 women and 1,124 men, also similarly found twice as many metabolites associated with age among women than men. This suggests that our similar observation may not be driven solely by the differences in sample sizes between women and men in our study and that it may have biological implications; i.e., aging may influence a wider breadth of metabolites in women than men. A probable cause for such a difference may be that during menopause, women experience very abrupt and dramatic hormone changes and loss of ovarian function, whereas during "andropause", men experience a gradual loss of hormones and decline in fertility[25]. These hormonal changes could be associated with other metabolic changes as well. Post-menopausal women have higher levels of sphingomyelins, fatty acids, acylcarnitines, lysophosphatidylcholines, and several amino acids than pre-menopausal women [26, 27], and a recent study found that plasma and urine metabo-lomics can be used to predict menopause status with 90% accuracy. Moreover, androgenic steroids have been linked to lipid levels in postmenopausal women [28]. Given that the baseline average ages of women and men in our sample are each ~61 years old, it is likely that our results are indicative of hormonal changes that occur in later ages and that most of our female participants have undergone menopause. It will be crucial to replicate these findings with a metabo-lomics panel that captures a larger proportion of the ~25,000 known blood metabolites in order to determine the validity of this hypothesis.

Among the 68 metabolites with different trajectories between women and men were sphingolipids, phos-phatidylcholines, and cholesterol. Metabolites from the latter two subgroups have been previously reported to have similar trajectories as what we identified, i.e., increasing with age in women and decreasing in men [26]. To our knowledge, a decrease of sphingolipid levels in men as our results suggest has not been previously reported. However, it has been reported that women have greater sphingomyelin increases with age than men [29] and that women with high sphingomyelin levels have reduced risk of AD, while men with high levels of sphingomyelins have increased risk of AD [30]. This could suggest that among men, declining levels of sphingolipids are a typical trait of healthy aging. While impaired sphingolipid metabolism is thought to be involved in AD [31], follow-up inves-tigations are needed to verify whether declining sphingolipids indicate healthy aging in men but increase AD risk in women.

Understanding how metabolites differ by sex and change with age could have implications for cancer. A recent study found that men with higher levels of serum androgenic steroids, which decrease with age in healthy men, measured up to 25 years prior to a diagnosis of prostate cancer were prospectively associated with increased risk of prostate cancer death [32]. Establishing "healthy" metabolite trajectories could help identify these high-risk individuals at different stages of life and be used to better understand changes occurring in the tumor microenvironment.

We compared our metabolite h^2 estimates to those recently estimated from a twin study of 1,930 individuals in the TwinsUK cohort [17]. Among the 466 metabolites overlapping with our study, h^2 estimates were only moderately correlated (Pearson r=0.36) and our estimates were 9.6 percentage points lower on average. However, our metabolite h^2 estimates were 8.9 percentage points higher on average (and had a lower correlation of r=0.25) when comparing 191 overlapping metabolite h^2 estimates from an earlier twin study based on 7,824 individuals from both the KORA F4 and TwinsUK cohorts [16]. Interestingly, despite having some overlapping participants, h^2 estimates between these two previous studies were only moderately correlated: among 163 overlapping metabolite h^2 estimates, Pearson r was 0.38, with estimates based on the TwinsUK cohort being 18.8 percentage points higher on average than the combined KORA and TwinsUK study. Differences in h^2 estimates may be driven by differences in population composition and size, phenotypic variation, and analytic approaches.

Although the strength of aging and sex metabolite associations were not associated with metabolite heritability, several of our aging and sex metabolites were identified to be associated with genetic variants (mQTLs, or metabolomic quantitative trait loci) in a previous study [16]. These aging and sex mQTLs are summarized in Tables S4 and S5, respectively. One of the androgenic steroids linked strongly to both sex and age in our analyses (5alpha-androstan-3beta, 17beta-diol disulfate, h^2=41.1%) is associated with a variant in the *CYP3A5* gene (*P*=1.17e-29). CYP3A enzymes play a critical role in the metabolism of ~30% of clinically used drugs, and the capacity to metabolize drugs declines with age [33]. Expression of cytochrome P450 (CYP) enzymes typically increases with age and has been shown to be influenced by interactions between age and sex [34]. It is likely that some of our observed metabolomic changes with age and differences by sex are linked to these CYP changes. An in-depth pharmacogenomics investigation into relationships between CYP enzymes, androgenic steroids, age, and sex could further elucidate factors of aging that influence drug metabolism.

This study was not without limitations. Although our analyses adjusted for cholesterol lowering medication use, there could be residual confounding due to differences in duration or type of cholesterol lowering medication, which could be influencing the apparent lower lipid levels in men. An in-depth investigation into medication use could be informative. Our findings are likely driven by our panel of metabolites, and it is possible that a different panel of metabolites could produce different results. Many of our findings are in accordance with previous publications, thereby strengthening confidence in our results that have not been previously investigated with regards to age and sex. Accordingly, it will be crucial to replicate novel findings with an external cohort. However, we also identified several inconsistencies between our study and others regarding h^2 estimates and a few of our association results, which could have been due to differences in study designs and sample populations. This challenge is common [35], as the field of metabolomics is rapidly developing and widely accepted standards for quality control techniques are forthcoming. Differences in platforms, quantification techniques, statistical analysis methods, laboratory techniques for sample handling (*i.e.*, anti-coagulation method, preservation, storage duration), and fasting status at the time of the sample draw may result in large variations from one study to another [36]. The metabolomics quality control process we have outlined here as well as that described in Voyle et al. [37] could serve as guidelines for future studies. Many of our findings included metabolites that had unknown chemical structures, which is a current limitation of the field of metabolomics, as it can be difficult and costly to accurately identify metabolites. Further, we only investigated linear effects of age, but non-linear age effects may exist and should be investigated in future investigations.

Using a large panel of longitudinal metabolomics data, we conducted a comprehensive investigation of the influence of aging and sex on metabolomics. Our findings suggest that levels of most metabolites are highly influenced by sex and age, and that sex differentially influences levels and trajectories of many metabolites. These findings underscore the importance of incorporating age and sex in the design and analysis of metabolomics investigations and offer a deeper understanding of the aging process that could inform many novel hypotheses regarding the role of metabolites in healthy and accelerated aging.

MATERIALS AND METHODS

Participants

Study participants were from WRAP, a longitudinal study of initially dementia free middle-aged adults that allows for the enrollment of siblings and is enriched for a parental history of Alzheimer's disease. Further details of the study design and methods used have been previously described [38, 39]. For the current analyses, follow-up occurred every two years. This study was conducted with the approval of the University of Wisconsin Institutional Review Board and all subjects provided signed informed consent before participation.

Plasma collection and sample handling

Fasting blood samples for this study were drawn the morning of each study visit. Blood was collected in 10 mL ethylenediaminetetraacetic acid (EDTA) vacutainer tubes. They were immediately placed on ice, and then centrifuged at 3000 revolutions per minute for 15 minutes at room temperature. Plasma was pipetted off within one hour of collection. Plasma samples were aliquoted into 1.0 mL polypropylene cryovials and placed in -80°C freezers within 30 minutes of separation. Samples were never thawed before being shipped overnight on dry ice to Metabolon, Inc. (Durham, NC), where they were again stored in -80°C freezers and thawed once before testing.

Metabolomic profiling and quality control

An untargeted plasma metabolomics analysis was performed by Metabolon, Inc. using Ultrahigh Performance Liquid Chromatography-Tandom Mass Spectro-

metry (UPLC-MS/MS). Quantification was performed as previously described [40]; details are outlined in the Supplemental Note. Metabolites within nine super pathways were identified: amino acids, carbohydrates, cofactors and vitamins, energy, lipids, nucleotides, partially characterized molecules, peptides, and xenobiotics.

Up to three longitudinal plasma samples were available for each participant. Metabolites with an interquartile range of zero (*i.e.*, those with very low or no variability) were excluded from analyses (n=178 metabolites). After removing these metabolites, samples were missing a median of 11.7% metabolites, while metabolites were missing in a median of 1.2% of samples. Missing metabolite values were imputed to the lowest level of detection for each metabolite. Metabolite values were median-scaled and log-transformed to normalize metabolite distributions [41]. If a participant reported that they did not fast or withhold medications and caffeine for at least eight hours, the sample was excluded from analyses (n=159 samples). A total of 1,097 metabolites among 2,344 samples remained for analyses.

DNA collection and genomics quality control

DNA was extracted from whole blood samples using the PUREGENE® DNA Isolation Kit (Gentra Systems, Inc., Minneapolis, MN). DNA concentrations were quantified using the Invitrogen™ Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Fisher Scientific, Inc., Hampton, NH) analyzed on the Synergy 2 Multi-Detection Microplate Reader (Biotek Instruments, Inc., Winooski, VT). Samples were diluted to 50 ng/ul following quantification.

A total of 1,340 samples were genotyped using the Illumina Multi-Ethnic Genotyping Array at the University of Wisconsin Biotechnology Center (Figure S7). Thirty-six blinded duplicate samples were used to calculate a concordance rate of 99.99%, and discordant genotypes were set to missing. Sixteen samples missing >5% of variants were excluded, while 35,105 variants missing in >5% of individuals were excluded. No samples were removed due to outlying heterozygosity. Six samples were excluded due to inconsistencies between self-reported and genetic sex.

Due to sibling relationships in the WRAP cohort, genetic ancestry was assessed using Principal Components Analysis in Related Samples (PC-AiR), a method that makes robust inferences about population structure in the presence of relatedness [42]. This approach included several iterative steps and was based on 63,503 linkage disequilibrium (LD) pruned (r^2<0.10) and common (MAF>0.05) variants, using the 1000 Genomes data as reference populations [43]. First, kin-

ship coefficients (KCs) were calculated between all pairs of individuals using genomic data with the Kinship-based Inference for Gwas (KING)-robust method [44]. PC-AiR was used to perform principal components analysis (PCA) on the reference populations along with a subset of unrelated individuals identified by the KCs. Resulting principal components (PCs) were used to project PC values onto the remaining related individuals. All PCs were then used to recalculate the KCs taking ancestry into account using the PC-Relate method, which estimates KCs robust to population structure [45]. PCA was performed again using the updated KCs, and KCs were also estimated again using updated PCs. The resulting PCs identified 1,198 WRAP participants whose genetic ancestry was primarily of European descent. This procedure was repeated within this subset of participants (excluding 1000 Genomes individuals) to obtain PC estimates used to adjust for population stratification in subsequent genomic analyses. Among European descendants, 160 variants were not in Hardy-Weinberg equilibrium (HWE) and 327,064 were monomorphic and thus, removed.

A total of 1,294,660 bi-allelic autosomal variants among 1,198 European descendants remained for imputation, which was performed with the Michigan Imputation Server v1.0.3 [46], using the Haplotype Reference Consortium (HRC) v. r1.1 2016 [47] as the reference panel and Eagle2 v2.3 [48] for phasing. Prior to imputation, the HRC Imputation Checking Tool [49] was used to identify variants that did not match those in HRC, were palindromic, differed in MAF>0.20, or that had non-matching alleles when compared to the same variant in HRC, leaving 898,220 for imputation. A total of 39,131,578 variants were imputed. Variants with a quality score R^2<0.80, MAF<0.001, or that were out of HWE were excluded, leaving 10,400,394 imputed variants. These were combined with the genotyped variants, leading to 10,499,994 imputed and genotyped variants for analyses. Data cleaning and file preparation were completed using PLINK v1.9 [50] and VCFtools v0.1.14 [51]. Coordinates are based on GRCh37 assembly hg19.

AUTHOR CONTRIBUTIONS

B.F.D. conceptualization of the study, performed analyses, and wrote the manuscript; R.L.K., K.J.H, and S.C.J. provided critical manuscript revisions; and C.D.E. provided guidance for all aspects of the study.

ACKNOWLEDGEMENTS

providing Illumina Infinium genotyping services. We especially thank the WRAP participants.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

REFERENCES

1. Deidda M, Piras C, Bassareo PP, Dessalvi CC, Mercuro G. Metabolomics, a promising approach to translational research in cardiology. IJC Metab Endocr. 2015; 9:31–38. https://doi.org/10.1016/j.ijcme.2015.10.001

2. Horgan RP, Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. Obstet Gynaecol. 2011; 13:189–95. https://doi.org/10.1576/toag.13.3.189.27672

3. Niccoli T, Partridge L. Ageing as a risk factor for disease. Curr Biol. 2012; 22:R741–52. https://doi.org/10.1016/j.cub.2012.07.024

4. Yu Z, Zhai G, Singmann P, He Y, Xu T, Prehn C, Römisch-Margl W, Lattka E, Gieger C, Soranzo N, Heinrich J, Standl M, Thiering E, et al. Human serum metabolic profiles are age dependent. Aging Cell. 2012; 11:960–67. https://doi.org/10.1111/j.1474-9726.2012.00865.x

5. Menni C, Kastenmüller G, Petersen AK, Bell JT, Psatha M, Tsai PC, Gieger C, Schulz H, Erte I, John S, Brosnan MJ, Wilson SG, Tsaprouni L, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol. 2013; 42:1111–19. https://doi.org/10.1093/ije/dyt094

6. Mittelstrass K, Ried JS, Yu Z, Krumsiek J, Gieger C, Prehn C, Roemisch-Margl W, Polonikov A, Peters A, Theis FJ, Meitinger T, Kronenberg F, Weidinger S, et al. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PLoS Genet. 2011; 7:e1002215. https://doi.org/10.1371/journal.pgen.1002215

7. Krumsiek J, Mittelstrass K, Do KT, Stückler F, Ried J, Adamski J, Peters A, Illig T, Kronenberg F, Friedrich N, Nauck M, Pietzner M, Mook-Kanamori DO, et al. Gender-specific pathway differences in the human serum metabolome. Metabolomics. 2015; 11:1815–33. https://doi.org/10.1007/s11306-015-0829-0

8. Rist MJ, Roth A, Frommherz L, Weinert CH, Krüger R, Merz B, Bunzel D, Mack C, Egert B, Bub A, Görling B, Tzvetkova P, Luy B, et al. Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. PLoS One. 2017; 12:e0183228. https://doi.org/10.1371/journal.pone.0183228

9. Dunn WB, Lin W, Broadhurst D, Begley P, Brown M, Zelena E, Vaughan AA, Halsall A, Harding N, Knowles JD, Francis-McIntyre S, Tseng A, Ellis DI, et al. Molecular phenotyping of a UK population: defining the human serum metabolome. Metabolomics. 2015; 11:9–26. https://doi.org/10.1007/s11306-014-0707-1

10. Chaleckis R, Murakami I, Takada J, Kondoh H, Yanagida M. Individual variability in human blood metabolites identifies age-related differences. Proc Natl Acad Sci USA. 2016; 113:4252–59. https://doi.org/10.1073/pnas.1603023113

11. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, Sajed T, Johnson D, Li C, Karu N, Sayeeda Z, Lo E, Assempour N, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 2018; 46:D608–17. https://doi.org/10.1093/nar/gkx1089

12. Mäkinen VP, Ala-Korpela M. Metabolomics of aging requires large-scale longitudinal studies with replication. Proc Natl Acad Sci USA. 2016; 113:E3470. https://doi.org/10.1073/pnas.1607062113

13. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J R Stat Soc B. 1995; 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

14. Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. PLoS Genet. 2013; 9:e1003520. https://doi.org/10.1371/journal.pgen.1003520

15. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

16. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, Arnold M, Erte I, Forgetta V, Yang TP, Walter K, Menni C, Chen L, et al, and Multiple Tissue Human Expression Resource (MuTHER) Consortium. An atlas of genetic influences on human blood

metabolites. Nat Genet. 2014; 46:543–50.
https://doi.org/10.1038/ng.2982

17. Long T, Hicks M, Yu HC, Biggs WH, Kirkness EF, Menni C, Zierer J, Small KS, Mangino M, Messier H, Brewerton S, Turpaz Y, Perkins BA, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet. 2017; 49:568–78.
https://doi.org/10.1038/ng.3809

18. Goodman-Gruen D, Barrett-Connor E. Sex differences in the association of endogenous sex hormone levels and glucose tolerance status in older men and women. Diabetes Care. 2000; 23:912–18.
https://doi.org/10.2337/diacare.23.7.912

19. Ferrini RL, Barrett-Connor E. Sex hormones and age: a cross-sectional study of testosterone and estradiol and their bioavailable fractions in community-dwelling men. Am J Epidemiol. 1998; 147:750–54.
https://doi.org/10.1093/oxfordjournals.aje.a009519

20. Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR, and Baltimore Longitudinal Study of Aging. Longitudinal effects of aging on serum total and free testosterone levels in healthy men. J Clin Endocrinol Metab. 2001; 86:724–31.
https://doi.org/10.1210/jcem.86.2.7219

21. Spencer JB, Klein M, Kumar A, Azziz R. The age-associated decline of androgens in reproductive age and menopausal Black and White women. J Clin Endocrinol Metab. 2007; 92:4730–33.
https://doi.org/10.1210/jc.2006-2365

22. Davison SL, Bell R, Donath S, Montalto JG, Davis SR. Androgen levels in adult females: changes with age, menopause, and oophorectomy. J Clin Endocrinol Metab. 2005; 90:3847–53.
https://doi.org/10.1210/jc.2005-0212

23. Gonzalez-Covarrubias V, Beekman M, Uh HW, Dane A, Troost J, Paliukhovich I, van der Kloet FM, Houwing-Duistermaat J, Vreeken RJ, Hankemeier T, Slagboom EP. Lipidomics of familial longevity. Aging Cell. 2013; 12:426–34.
https://doi.org/10.1111/acel.12064

24. Mihalik SJ, Goodpaster BH, Kelley DE, Chace DH, Vockley J, Toledo FG, DeLany JP. Increased levels of plasma acylcarnitines in obesity and type 2 diabetes and identification of a marker of glucolipotoxicity. Obesity (Silver Spring). 2010; 18:1695–700.
https://doi.org/10.1038/oby.2009.510

25. Andropause VA. Maturitas. 2000; 34:5–15.
https://doi.org/10.1016/S0378-5122(99)00075-4

26. Auro K, Joensuu A, Fischer K, Kettunen J, Salo P, Mattsson H, Niironen M, Kaprio J, Eriksson JG, Lehtimäki T, Raitakari O, Jula A, Tiitinen A, et al. A metabolic view on menopause and ageing. Nat Commun. 2014; 5:4708.
https://doi.org/10.1038/ncomms5708

27. Ke C, Hou Y, Zhang H, Yang K, Wang J, Guo B, Zhang F, Li H, Zhou X, Li Y, Li K. Plasma Metabolic Profiles in Women are Menopause Dependent. PLoS One. 2015; 10:e0141743.
https://doi.org/10.1371/journal.pone.0141743

28. Noyan V, Yucel A, Sagsoz N. The association of androgenic sex steroids with serum lipid levels in postmenopausal women. Acta Obstet Gynecol Scand. 2004; 83:487–90. https://doi.org/10.1111/j.0001-6349.2004.00417.x

29. Mielke MM, Bandaru VV, Han D, An Y, Resnick SM, Ferrucci L, Haughey NJ. Factors affecting longitudinal trajectories of plasma sphingomyelins: the Baltimore Longitudinal Study of Aging. Aging Cell. 2015; 14:112–21. https://doi.org/10.1111/acel.12275

30. Mielke MM, Haughey NJ, Han D, An Y, Bandaru VV, Lyketsos CG, Ferrucci L, Resnick SM. The Association Between Plasma Ceramides and Sphingomyelins and Risk of Alzheimer's Disease Differs by Sex and APOE in the Baltimore Longitudinal Study of Aging. J Alzheimers Dis. 2017; 60:819–28.
https://doi.org/10.3233/JAD-160925

31. van Echten-Deckert G, Walter J. Sphingolipids: critical players in Alzheimer's disease. Prog Lipid Res. 2012; 51:378–93.
https://doi.org/10.1016/j.plipres.2012.07.001

32. Huang J, Weinstein SJ, Moore SC, Derkach A, Hua X, Mondul AM, Sampson JN, Albanes D. Pre-diagnostic Serum Metabolomic Profiling of Prostate Cancer Survival. J Gerontol A Biol Sci Med Sci. 2018.
https://doi.org/10.1093/gerona/gly128

33. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther. 2013; 138:103–41.
https://doi.org/10.1016/j.pharmthera.2012.12.007

34. Yang X, Zhang B, Molony C, Chudin E, Hao K, Zhu J, Gaedigk A, Suver C, Zhong H, Leeder JS, Guengerich FP, Strom SC, Schuetz E, et al. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. Genome Res. 2010; 20:1020–36. https://doi.org/10.1101/gr.103341.109

35. Enche Ady CN, Lim SM, Teh LK, Salleh MZ, Chin AV, Tan MP, Poi PJ, Kamaruzzaman SB, Abdul Majeed AB, Ramasamy K. Metabolomic-guided discovery of Alzheimer's disease biomarkers from body fluid. J Neurosci Res. 2017; 95:2005–24.
https://doi.org/10.1002/jnr.24048

36. González-Domínguez R, Sayago A, Fernández-Recamales Á. Metabolomics in Alzheimer's disease: the need of complementary analytical platforms for the identification of biomarkers to unravel the underlying pathology. J Chromatogr B Analyt Technol Biomed Life Sci. 2017; 1071:75–92. https://doi.org/10.1016/j.jchromb.2017.02.008

37. Voyle N, Kim M, Proitsi P, Ashton NJ, Baird AL, Bazenet C, Hye A, Westwood S, Chung R, Ward M, Rabinovici GD, Lovestone S, Breen G, et al, and Alzheimer's Disease Neuroimaging Initiative. Blood metabolite markers of neocortical amyloid-β burden: discovery and enrichment using candidate proteins. Transl Psychiatry. 2016; 6:e719. https://doi.org/10.1038/tp.2015.205

38. ager MA, Hermann B, La Rue A. Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. J Geriatr Psychiatry Neurol. 2005; 18:245–49. https://doi.org/10.1177/0891988705281882

39. Johnson SC, Koscik RL, Jonaitis EM, Clark LR, Mueller KD, Berman SE, Bendlin BB, Engelman CD, Okonkwo OC, Hogan KJ, Asthana S, Carlsson CM, Hermann BP, Sager MA. The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current directions. Alzheimers Dement (Amst). 2017; 10:130–42. https://doi.org/10.1016/j.dadm.2017.11.007

40. Evans AM, Bridgewater BR, Liu Q, Mitchell MW, Robinson RJ, Dai H, Stewart SJ, DeHaven CD, Miller LA. High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. Metabolomics. 2014; 4:132. https://doi.org/10.4172/2153-0769.1000132

41. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006; 7:142. https://doi.org/10.1186/1471-2164-7-142

42. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol. 2015; 39:276–93. https://doi.org/10.1002/gepi.21896

43. 1000 Genomes Project Consortium. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. Nature. 2015; 526:68–74. https://doi.org/10.1038/nature15393

44. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010; 26:2867–73. https://doi.org/10.1093/bioinformatics/btq559

45. Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet. 2016; 98:127–48. https://doi.org/10.1016/j.ajhg.2015.11.022

46. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016; 48:1284–87. https://doi.org/10.1038/ng.3656

47. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, et al, and Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016; 48:1279–83. https://doi.org/10.1038/ng.3643

48. Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, Schoenherr S, Forer L, McCarthy S, Abecasis GR, Durbin R, L Price A. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016; 48:1443–48. https://doi.org/10.1038/ng.3679

49. Rayner NW, Robertson N, Mahajan A, McCarthy MI. A suite of programs for pre- and post-imputation data checking. The American Society of Human Genetics. 2016. (Vancouver, Canada).

50. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4:7. https://doi.org/10.1186/s13742-015-0047-8

51. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, and 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156–58. https://doi.org/10.1093/bioinformatics/btr330

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY NOTE

Metabolite profiling

Plasma metabolites were profiled by Metabolon (Durham, NC) using Ultrahigh Performance Liquid Chromatography-Tandom Mass Spectrometry (UPLC-MS/MS). Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18-2.1x100 mm, 1.7 μm) using water and methanol, containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the same afore mentioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, however with 6.5mM Ammonium Bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate, pH 10.8. The MS analysis alternated between MS and data-dependent MSn scans using dynamic exclusion. The scan range varied slighted between methods but covered 70-1000 m/z. Raw data files are archived and extracted as described below.

Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. These systems are built on a web-service platform utilizing Microsoft's .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library

spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds have been acquired and registered into LIMS for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

A variety of curation procedures were carried out to ensure that a high quality data set was made available for statistical analysis and data interpretation. The QC and curation processes were designed to ensure accurate and consistent identification of true chemical entities, and to remove those representing system artifacts, misassignments, and background noise. Metabolon data analysts use proprietary visualization and interpretation software to confirm the consistency of peak identification among the various samples. Library matches for each compound were checked for each sample and corrected if necessary.

Peaks were quantified using area-under-the-curve. A data normalization setp was performed to correct variation resulting from instrument inter-day tuning differences. Essentially, each compound was corrected in run-day blocks by registering the medians to equal one and normaling each data point proportionately.

SUPPLEMENTARY FIGURES
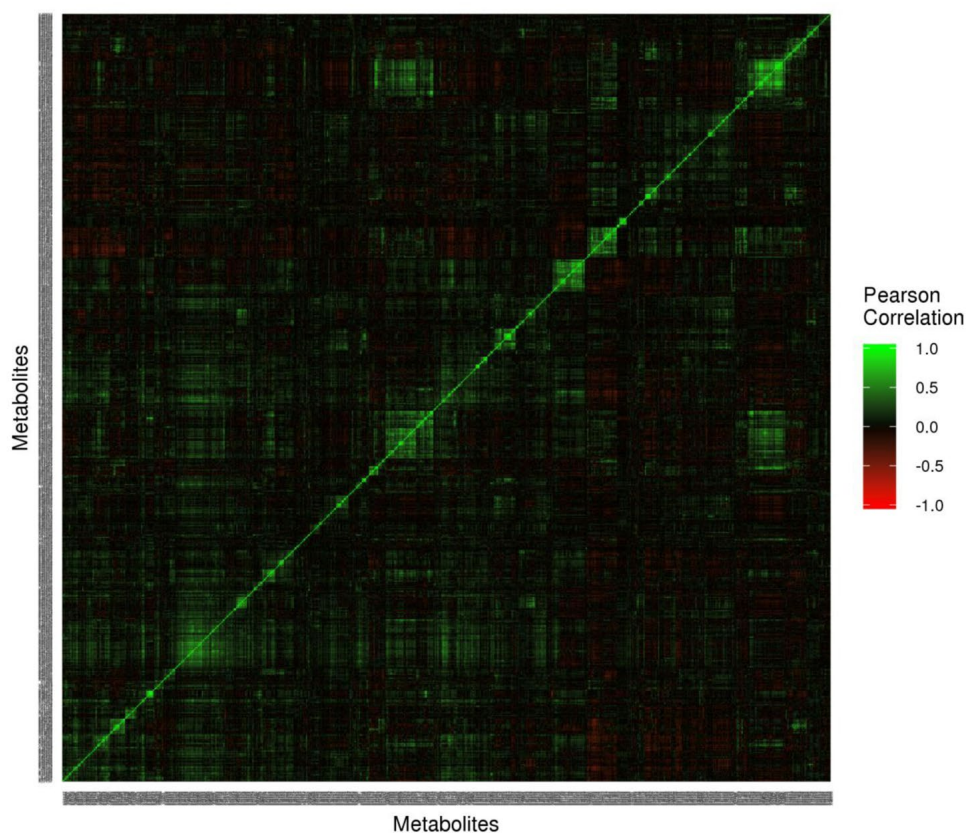
Figure S1. Heatmap of metabolite correlations. Pearson correlations (r) are displayed and hierarchical clustering was used to sort metabolites. Clusters of strong positive correlations are largely between amino acid–amino acid and between lipid–lipid (i.e., correlations between super pathway had weaker correlations). Negative correlations are less common and all have an r≥-0.62.

Figure S2. Manhattan plots of metabolome-wide associations results. (**A**) Age significantly influenced 623 metabolites. (**B**) Sex significantly influenced 695 metabolites. Both sets of results use a Benjamini-Hochberg adjusted p-value threshold (red horizontal line).
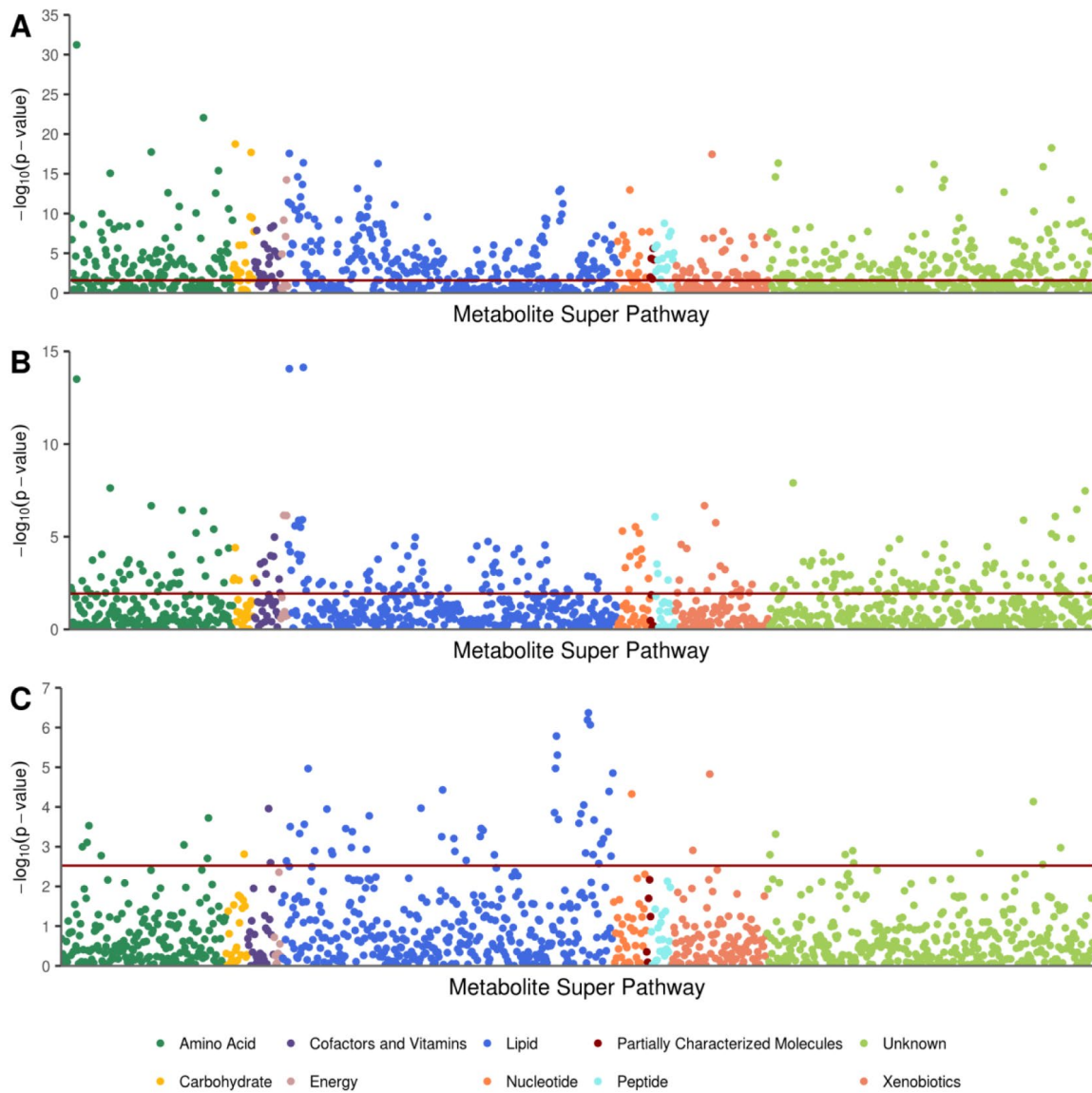
Figure S3. Manhattan plots of metabolome-wide associations results. (A) Age significantly influenced 565 metabolites in women. (()) Age signi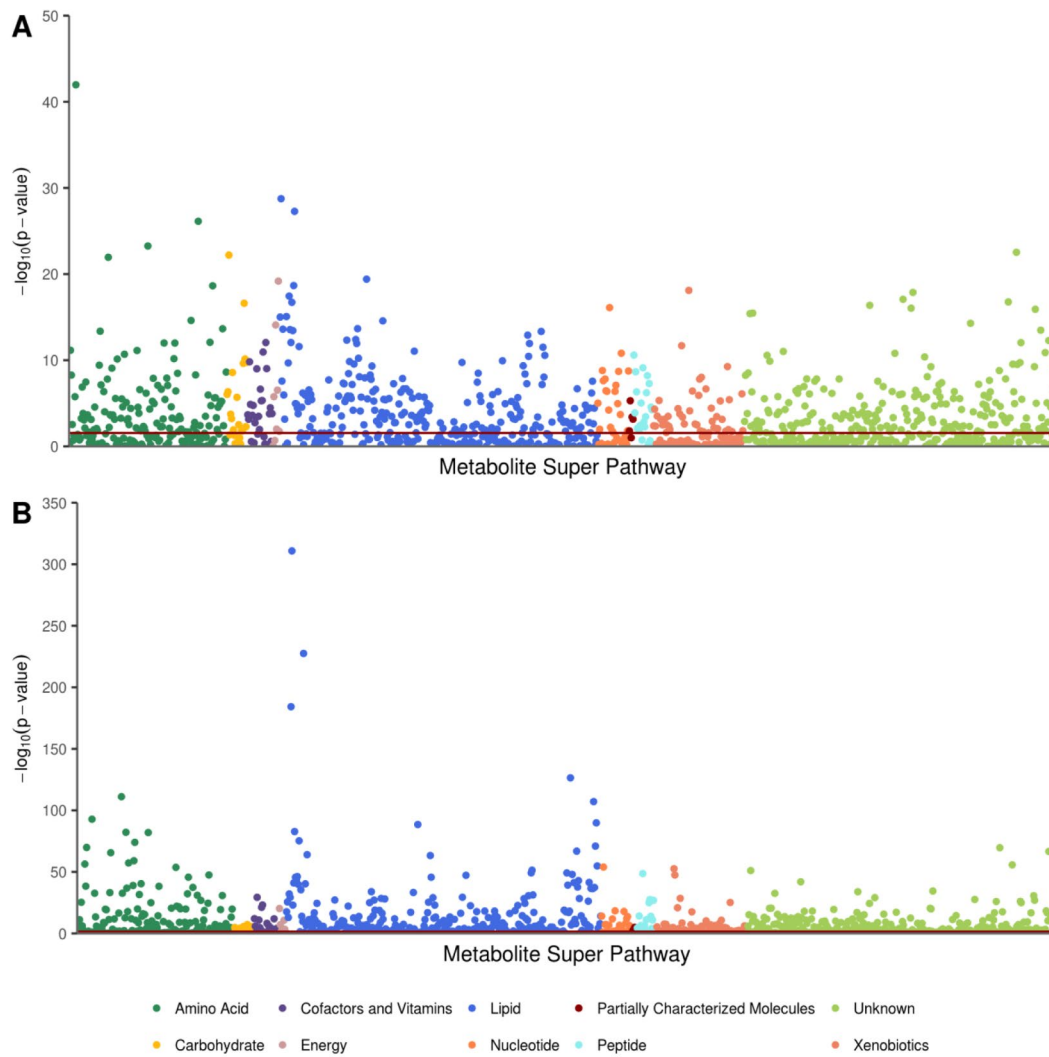ficantly influenced 255 metabolites in men. (()) Trajectories of 68 metabolites significantly differ by sex. Each set of results uses a Benjamini-Hochberg adjusted p-value threshold (red horizontal line).
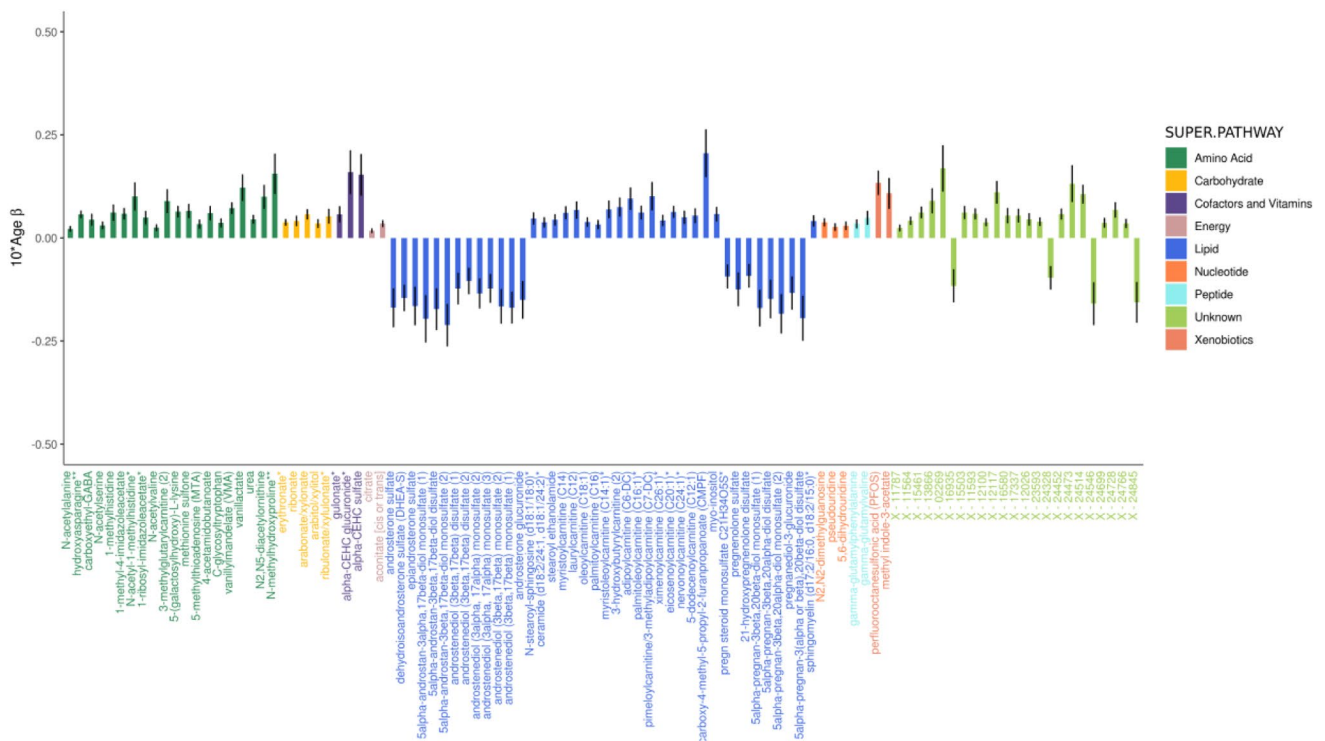
Figure S4. Age stratified by sex: Adjusted effects of a 10-year increase in age on the top 100 metabolites most strongly influenced by age in women. Positive values indicate the amount a metabolite increased over 10 years, whereas negative values indicate the amount a metabolite decreased over 10 years. Black vertical lines indicate standard errors.
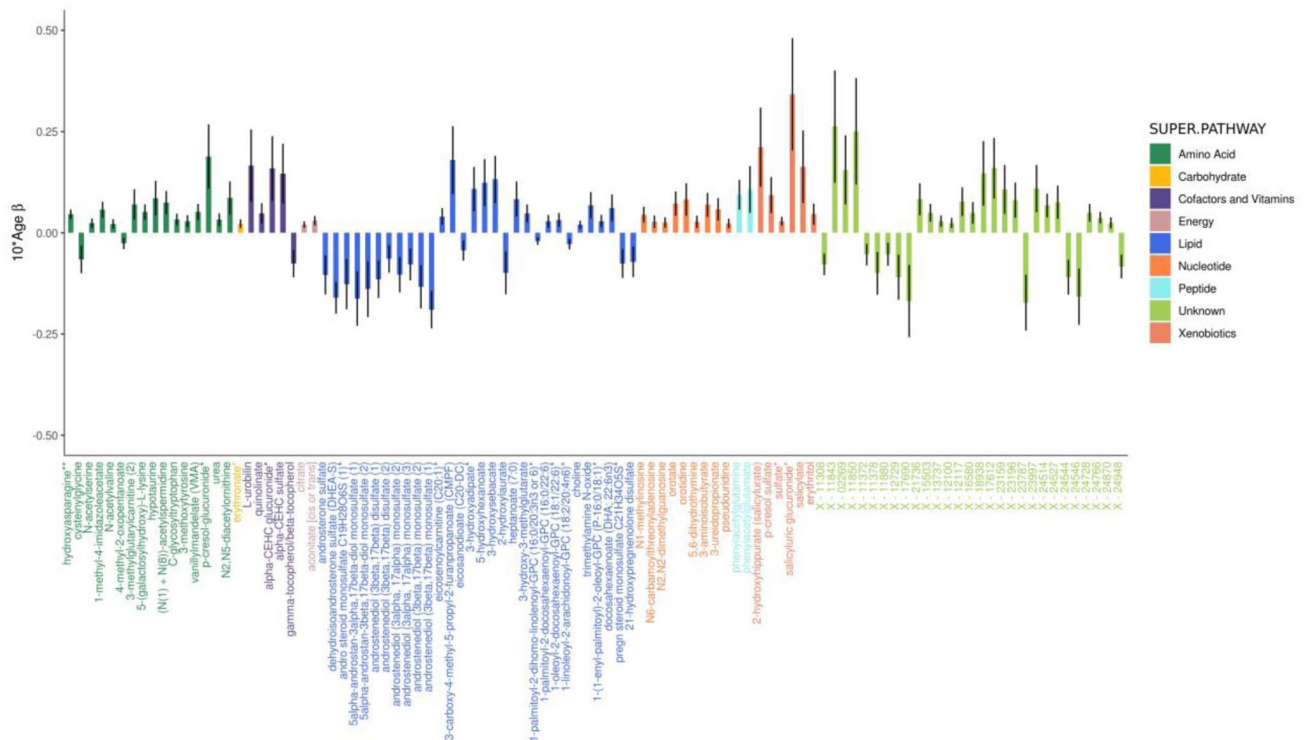
Figure S5. Age stratified by sex: Adjusted effects of a 10-year increase in age on the top 100 metabolites most strongly influenced by age in men. Positive values indicate that the metabolite increased with 10 years of age, whereas negative values indicate that the metabolite decreased with 10 years of age. Black vertical lines indicate standard errors.

Figure S6. Adjusted effects of a 10-year increase in age on the 68 metabolites with trajectories that significantly differ by sex. For each metabolite, the bar on the left represents the change in metabolite level over 10 years in women, whereas the bar on the right represents the change in metabolite level over 10 years in men. Black vertical lines indicate standard errors.

WRAP GWAS QC Flowchart

1,340 samples
1,779,819 variants
-Sample missingness, remove chr sex & 0

Samples missing >5%: 16 samples removed
Chr 0: 11,468 variants removed
Sex chr: 61,006 variants removed

1,324 samples
1,707,345 variants
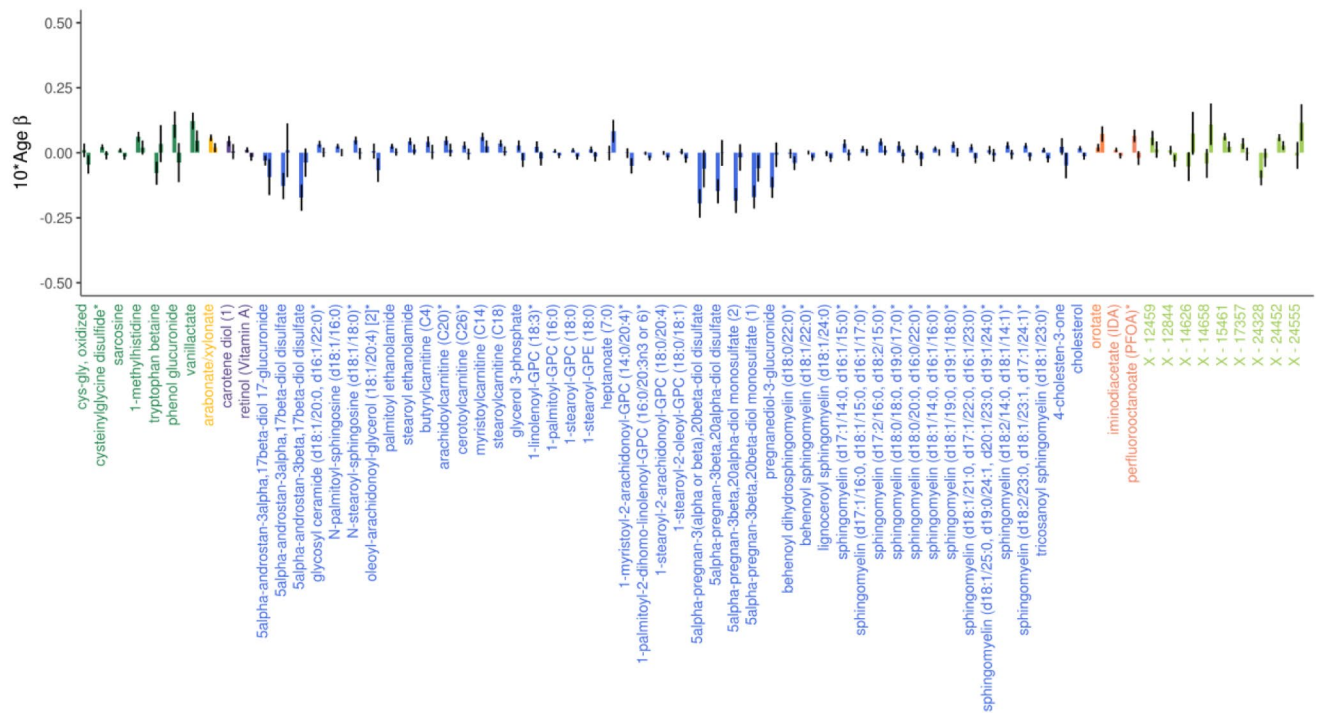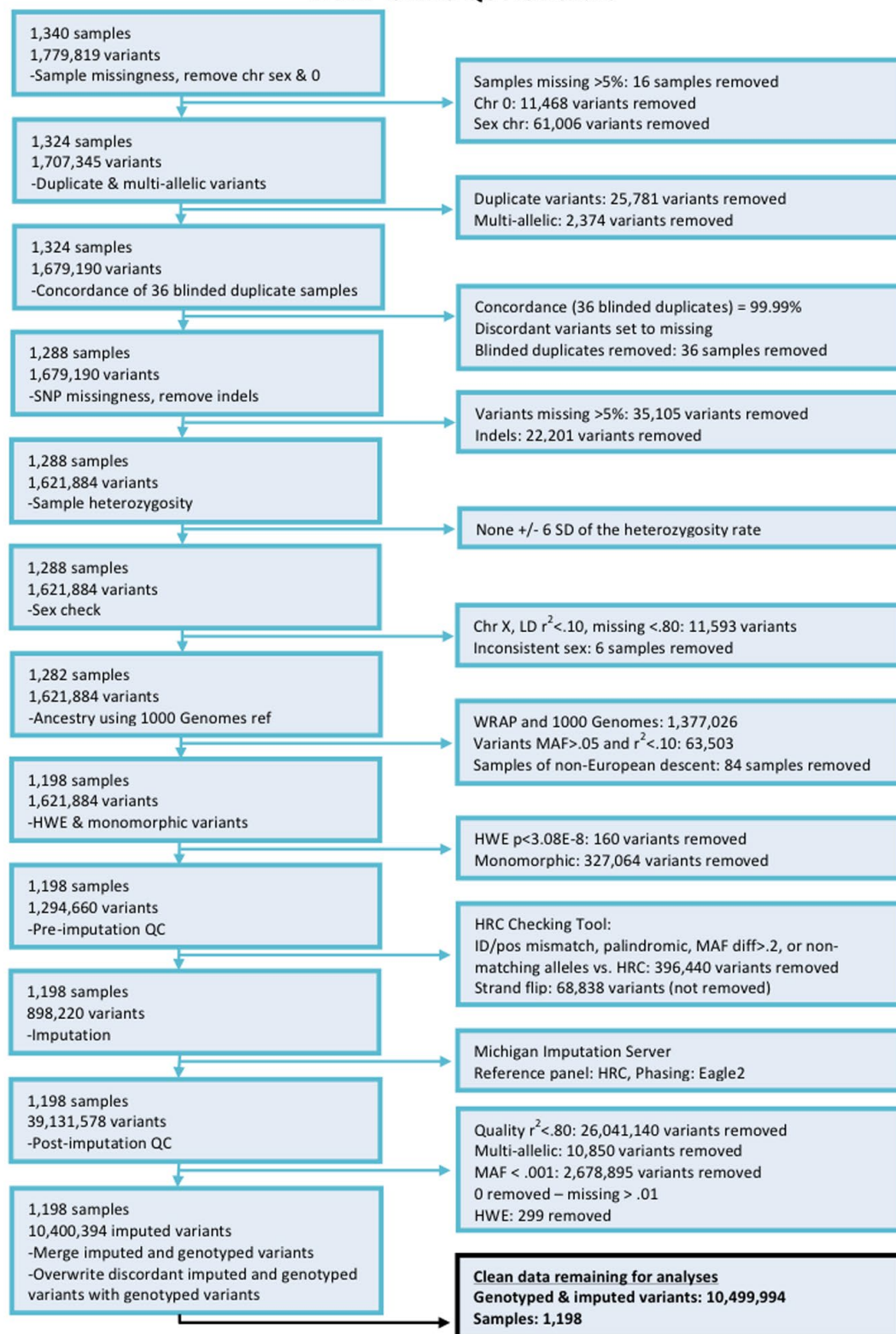-Duplicate & multi-allelic variants

Duplicate variants: 25,781 variants removed
Multi-allelic: 2,374 variants removed

1,324 samples
1,679,190 variants
-Concordance of 36 blinded duplicate samples

Concordance (36 blinded duplicates) = 99.99%
Discordant variants set to missing
Blinded duplicates removed: 36 samples removed

1,288 samples
1,679,190 variants
-SNP missingness, remove indels

Variants missing >5%: 35,105 variants removed
Indels: 22,201 variants removed

1,288 samples
1,621,884 variants
-Sample heterozygosity

None +/- 6 SD of the heterozygosity rate

1,288 samples
1,621,884 variants
-Sex check

Chr X, LD r^2<.10, missing <.80: 11,593 variants
Inconsistent sex: 6 samples removed

1,282 samples
1,621,884 variants
-Ancestry using 1000 Genomes ref

WRAP and 1000 Genomes: 1,377,026
Variants MAF>.05 and r^2<.10: 63,503
Samples of non-European descent: 84 samples removed

1,198 samples
1,621,884 variants
-HWE & monomorphic variants

HWE $p<3.08E-8$: 160 variants removed
Monomorphic: 327,064 variants removed

1,198 samples
1,294,660 variants
-Pre-imputation QC

HRC Checking Tool:
ID/pos mismatch, palindromic, MAF diff>.2, or non-matching alleles vs. HRC: 396,440 variants removed
Strand flip: 68,838 variants (not removed)

1,198 samples
898,220 variants
-Imputation

Michigan Imputation Server
Reference panel: HRC, Phasing: Eagle2

1,198 samples
39,131,578 variants
-Post-imputation QC

Quality r^2<.80: 26,041,140 variants removed
Multi-allelic: 10,850 variants removed
MAF < .001: 2,678,895 variants removed
0 removed – missing > .01
HWE: 299 removed

1,198 samples
10,400,394 imputed variants
-Merge imputed and genotyped variants
-Overwrite discordant imputed and genotyped variants with genotyped variants

Clean data remaining for analyses
Genotyped & imputed variants: 10,499,994
Samples: 1,198

Figure S7. GWAS QC Flowchart.

SUPPLEMENTARY TABLES

Please browse the links in Full Text version to see the data of Supplementary Tables:

1. **Table S1.** Summary statistics of metabolomic associations with age and sex.
2. **Table S2.** Comparison of metabolites associated with age in our results to those reported in Menni et al., 2013.
3. **Table S3.** Comparison of metabolites associated with sex in our results to those reported in Krumsiek et al., 2015.
4. **Table S4.** Genetic variants associated with significant aging metabolites, as reported in Shin et al., 2014.
5. **Table S5.** Genetic variants associated with significant sex metabolites, as reported in Shin et al., 2014.