Integrated analysis of genomics, longitudinal metabolomics, and Alzheimer's risk factors among 1,111 cohort participants

Burcu F. Darst[1,2,#1,*], **Qiongshi Lu**[1,3], **Sterling C. Johnson**[1,4,5,6], **Corinne D. Engelman**[1,2,4,6]

[1]University of Wisconsin, Madison, WI, 53706, USA

[2]Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI, 53726, USA

[3]Department of Biostatistics & Medical Informatics, Madison, WI, 53792, USA

[4]Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI, 53719, USA

[5]Geriatric Research Education and Clinical Center, Wm. S. Middleton Memorial VA Hospital, Madison, WI, 53705, USA

[6]Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI, 53792, USA

Abstract

Although Alzheimer's disease (AD) is highly heritable, genetic variants known to be associated with AD only explain a small proportion of its heritability. Genetic factors may only convey disease risk in individuals with certain environmental exposures, suggesting that a multi-omics approach could reveal underlying mechanisms contributing to complex traits, such as AD. We developed an integrated network to investigate relationships between metabolomics, genomics, and AD risk factors using Wisconsin Registry for Alzheimer's Prevention participants. Analyses included 1,111 non-Hispanic Caucasian participants with whole blood expression for 11,376 genes (imputed from dense genome-wide genotyping), 1,097 fasting plasma metabolites, and 17 AD risk factors. A subset of 155 individuals also had 364 fasting cerebral spinal fluid (CSF) metabolites. After adjusting each of these 12,854 variables for potential confounders, we developed an undirected graphical network, representing all significant pairwise correlations upon adjusting for multiple testing. There were many instances of genes being indirectly linked to AD risk factors through metabolites, suggesting that genes may influence AD risk through particular metabolites. Follow-up analyses suggested that glycine mediates the relationship between *CPS1* and measures of cardiovascular and diabetes risk, including body mass index, waist-hip ratio, inflammation, and

*Correspondence: Burcu F. Darst, 323-442-0078, bdarst@usc.edu.
#1Present address: Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, 90033, USA

insulin resistance. Further, 38 CSF metabolites explained more than 60% of the variance of CSF levels of tau, a detrimental protein that accumulates in the brain of AD patients and is necessary for its diagnosis. These results further our understanding of underlying mechanisms contributing to AD risk while demonstrating the utility of generating and integrating multiple omics data types.

Keywords

Genomics; metabolomics; multi-omics; integrated analysis; Alzheimer's disease

Introduction

Genome-wide association studies (GWAS) have identified tens of thousands of single nucleotide polymorphism (SNP)-trait associations(MacArthur et al., 2017). However, these variants tend to have very small effect sizes and typically explain a small portion of trait heritability. Late onset Alzheimer's disease (AD) is an example of such a trait: 53% of its phenotypic variance can be explained by genomic variants, collectively (*i.e.*, SNP heritability); yet, the 21 GWAS variants identified in a meta-analysis to be associated with AD only account for 31% of its genetic variance, leaving 69% unaccounted for(Ridge et al., 2016). In order to more comprehensively understand the disease risk conveyed by genetic factors, it is crucial to consider genomics in combination with other omics data types and to use integrative multi-omics approaches that can capture intricate relationships.

Although there has been great interest recently in the integration of multi-omics datasets, progress in this field is still fairly limited and it faces many challenges(Bersanelli et al., 2016; Buescher & Driggers, 2016; Gligorijevic & Przulj, 2015; S. Huang, Chaudhary, & Garmire, 2017; Lopez de Maturana, Pineda, Brand, Van Steen, & Malats, 2016; Ritchie, Holzinger, Li, Pendergrass, & Kim, 2015). However, studies have been able to show that the use of multiple omics data types is more predictive than single data types(Buescher & Driggers, 2016; Mankoo, Shen, Schultz, Levine, & Sander, 2011). A recent study with dense longitudinal omics data displayed the utility of integrating such data with regards to personalized medicine(Price et al., 2017). Although limited by its sample size of 108 participants, this investigation identified meaningful systems biology relationships that were able to improve the health of its participants. As it is becoming more feasible and common to acquire multiple omics data types, it is essential that we move towards systems biology approaches of understanding complex diseases, rather than focusing on single data types that are unable to capture the intricacies imposed by biology.

Recent technological advances have made metabolomics studies increasingly favorable among investigations of AD(Enche Ady et al., 2017), obesity(Zhang, Sun, & Wang, 2017), and cardiovascular disease(Ussher, Elmariah, Gerszten, & Dyck, 2016), to name a few. An appeal of the metabolome is that of the biological systems, metabolomics could offer an effective way to accurately capture individual-level environmental exposures; it is the most proximal to the development of the phenotype(Horgan & Kenny, 2011) and many metabolites have a low heritability(Long et al., 2017; Shin et al., 2014), implying that such metabolites are more strongly influenced by the environment than genomics. Metabolomic

variations that precede disease onset could prove to be highly informative for predictive models as well as preventative and therapeutic medicine. Pathological changes that cause AD are known to begin decades before the diagnosis of AD(Jack et al., 2009). As such, an integrated approach of studying the genomics and metabolomics of risk factors that precede an AD diagnosis could provide a better understanding of the underlying biological and environmental mechanisms that lead to the onset of AD.

We developed an integrative network to investigate relationships between plasma metabolomics, cerebral spinal fluid (CSF) metabolomics, genomics, and AD risk factors using 1,111 participants with deep longitudinal phenotypes from the Wisconsin Registry for Alzheimer's Prevention (WRAP). AD risk factors included neuropsychological measures of cognitive function, CSF levels of the two proteins required for an AD diagnosis that are known to accumulate in the brains of AD patients, amyloid-beta (Aβ) and tau, and measures of cardiovascular disease and diabetes risk, two diseases that are known to increase AD risk. Further, in order to understand whether plasma metabolite levels are representative of metabolites in CSF, which may be a more relevant tissue for neurological diseases, we also assessed the correlation of plasma and CSF metabolite levels.

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(Johnson et al., 2018; Sager, Hermann, & La Rue, 2005). Participants included in this analysis had genetic ancestry that was primarily of European descent, had both genomic and metabolomic data available, and up to seventeen AD risk factors (Table 1; of note, cholesterol is not included in this table because it was measured on the metabolite panel). 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 and CSF collection and sample handling

Fasting blood samples for this study were drawn the morning of each study visit. Plasma samples were stored in ethylenediaminetetraacetic acid (EDTA) tubes at −80°C. 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.

As previously described(Darst et al., 2017), CSF was collected via lumbar puncture (LP) in the morning after a 12-hour fast, not necessarily on the same day as a study visit (LPs were drawn within a median of 120 days of the study visit, ranging from 0–661 days). LPs were performed using a Sprotte 25- or 24-gauge spinal needle at the L3/4 or L4/5 interspace using

gentle extraction into polypropylene syringes. CSF (22 mL) was then gently mixed and centrifuged at 2000g for 10 minutes. Supernatants were frozen in 0.5 mL aliquots in polypropylene tubes and stored at −80°C.

Plasma and CSF samples were never thawed before being shipped overnight on dry ice to Metabolon (Durham, NC), where they were again stored in −80°C freezers and thawed once before testing.

CSF biomarker quantification

CSF $A\beta_{42}$, total tau (T-tau), and phosphorylated tau (P-tau) were quantified with sandwich ELISAs (INNOTEST β-amyloid1–42, hTAU-Ag, and Phospho-Tau[181P], respectively; Fujirebio Europe, Ghent, Belgium). CSF levels of $A\beta_{42}$ and $A\beta_{40}$ (a less amyloidogenic Aβ fragment as compared to $A\beta_{42}$) were used to calculate the ratio of $A\beta_{42}/A\beta_{40}$ were quantified by electrochemiluminescence (ECL) using an Aβ triplex assay (MSD Human Aβ peptide Ultra-Sensitive Kit, Meso Scale Discovery, Gaithersburg, MD). A total of 223 samples with CSF biomarkers among 141 individuals were available for this analysis.

Plasma and CSF metabolomic profiling and quality control

Untargeted plasma and CSF metabolomic analyses and quantification were performed by Metabolon (Durham, NC) using Ultrahigh Performance Liquid Chromatography-Tandom Mass Spectrometry (UPLC-MS/MS)(Evans et al., 2014); details are outlined in the Supplemental Note. Metabolites within eight super pathways were identified: amino acids, carbohydrates, cofactors and vitamins, energy, lipids, nucleotides, peptides, and xenobiotics.

Up to three longitudinal plasma samples were available for each participant. Plasma metabolites with an interquartile range of zero (*i.e.*, those with very low or no variability) were excluded from analyses (178 metabolites). After removing these metabolites, samples were missing a median of 11.7% plasma metabolites, while plasma metabolites were missing in a median of 1.2% of samples.

Up to four longitudinal CSF samples were available for each participant. Similarly, CSF metabolites with an interquartile range of zero were excluded from analyses (48 CSF metabolites). After removing these metabolites, samples were missing a median of 6.9% CSF metabolites, while CSF metabolites were missing in a median of 0.3% of samples.

Missing plasma and CSF metabolite values were imputed to the lowest level of detection for each metabolite(Menni et al., 2013). Metabolite values were median-scaled and log-transformed to normalize metabolite distributions(van den Berg, Hoefsloot, Westerhuis, Smilde, & van der Werf, 2006). If a participant reported that they did not fast or withhold medications and caffeine for at least eight hours prior to the blood draw, the plasma sample was excluded from analyses (159 plasma samples), leaving 1,097 plasma metabolites among 2,189 plasma samples (1,111 individuals) for analyses. Similarly, if a participant reported that they did not fast for at least eight hours prior to the LP, the CSF sample was excluded from analyses (4 CSF samples), leaving 364 CSF metabolites among 346 CSF samples (155 individuals) for analyses.

CSF and plasma metabolite correlations

A total of 326 metabolites were captured in both CSF and plasma. The correlations of these metabolites between tissue types were calculated using the Pearson correlation coefficient. In order to reduce variability due to the time interval between plasma and CSF sample collection, correlations were based on 141 pairs of plasma and CSF samples that were collected within a timespan of four months of each other. After removing these samples, plasma and CSF samples were collected a median of 27 days apart.

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, Hampton, NH) analyzed on the Synergy 2 Multi-Detection Microplate Reader (Biotek Instruments, Winooski, VT). Samples were normalized 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 S1). 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 the sibling relationships present 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(Conomos, Miller, & Thornton, 2015). 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(Genomes Project et al., 2015). First, kinship coefficients (KCs) were calculated between all pairs of individuals using genomic data with the Kinship-based Inference for Gwas (KING)-robust method(Manichaikul et al., 2010). 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(Conomos, Reiner, Weir, & Thornton, 2016). 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(Das et al., 2016), using the Haplotype Reference Consortium (HRC) v. r1.1 2016(McCarthy et al., 2016) as the reference panel and Eagle2 v2.3(Loh et al., 2016) for phasing. Prior to imputation, the HRC Imputation Checking Tool(Rayner, Robertson, Mahajan, & McCarthy, 2016) 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(Chang et al., 2015) and VCFtools v0.1.14(Danecek et al., 2011). Coordinates are based on GRCh37 assembly hg19.

Whole blood gene expression imputation

The resulting 10,499,994 imputed and genotyped variants were used to impute whole blood gene expression using PrediXcan(Gamazon et al., 2015) with the Depression Genes and Networks reference dataset(Battle et al., 2014), PrediXcan's largest reference sample consisting of 922 individuals with RNA sequencing on whole blood and GWAS data. PrediXcan filters results to only include genes that are imputed with reasonable accuracy, using a false discovery rate of 0.05. After removing genes with zero variability between individuals (162 genes), whole blood gene expression data for 11,376 genes were available for analyses.

Integrative network analysis

The analytic approach we used for our network analysis was similar to that of Price et al., 2017(Price et al., 2017). A total of 12,856 variables, including 11,376 expressed genes, 1,097 plasma metabolites, 364 CSF metabolites, and 17 AD risk factors, were available for the network analysis. Linear mixed models, as implemented by the lme4 package in R(Bates, Machler, Bolker, & Walker, 2015), were used to adjust each variable for age and sex and included a random intercept for individual to account for repeated measures and family to account for sibling relationships. Further adjustments were made specific to the variable being assessed: imputed gene expression was also adjusted for the first four principal components to account for ancestry; CSF and plasma metabolites were adjusted for cholesterol lowering medication use and sample storage time; the executive function and delayed recall composite scores were adjusted for practice effects; and systolic and diastolic blood pressure were adjusted for ace inhibitor and beta blocker medication use. For longitudinal traits (such as metabolites), random intercepts were used as the new outcomes for each individual, whereas for constant traits (such as imputed gene expression values), residuals were used as the new outcomes for each individual. These adjusted outcomes were used to assess all 82,606,231 pairwise correlations between traits using Spearman rank, and significance was determined using a Bonferroni-adjusted P-value (0.05/82,606,231=6.1e–10). To identify relationships between omics data, significant inter-omic associations and significant associations with an AD risk factor were used to develop an integrative network,

which was created using the igraph R package(Csardi & Nepusz, 2006). Dense subgraphs were identified using a community detection algorithm that maximizes the modularity of the network, such that there is high connectivity within communities (or groups of distinct variables), but low connectivity between communities(Clauset, Newman, & Moore, 2004).

Targeted mediation and interaction analyses

Results from the integrated network analysis were used to identify potential mediation and interactions between imputed gene expression and metabolite levels that could impact AD risk factors, as a proof of concept. Although our network analysis suggested many potentially meaningful mediation or interaction relationships, we only investigated gene-metabolite correlations with the most consistent biological support from the GWAS catalog(MacArthur et al., 2017) (www.ebi.ac.uk/gwas, date accessed: May 9, 2018), to illustrate the utility of the network analysis results. Such relationships were investigated using the longitudinal data (2,198 observations among 1,111 individuals) with linear mixed models, again as implemented by the lme4 package in R(Bates et al., 2015), including random intercepts for within-individual repeated measures and within-family relationships. To assess whether a metabolite mediated the relationship between imputed gene expression and an AD risk factor, models were run to assess whether: 1) the gene predicted the AD risk factor, 2) the gene predicted metabolite levels, 3) the metabolite predicted the AD risk factor, and 4) the gene predicted the AD risk factor while adjusting for the metabolite. The causal mediation effect, or the indirect effect of a gene on an AD risk factor through a metabolite, was calculated as the difference between the effect of the gene in model 1 and model 4, as implemented in the R mediation package(Imai, Keele, & Tingley, 2010). To determine whether this difference was significant, standard errors and *P*-values were estimated using the quasi-Bayesian Monte Carlo method with 1,000 simulations. Because the mediation package can only handle mixed models with one random effect, the mediation analysis was run with models 1 and 4 excluding the random effect for family. As a sensitivity analysis, the mediation analysis was rerun limiting models 1 and 4 to unrelated individuals (n=898 with 1,774 observations). A fifth linear mixed model was used to assess interactions by adding a gene*metabolite interaction term to model 4. Model 5 did not use the mediation package and was thus able to include random intercepts for both within-individual repeated measures and within-family relationships. All models including a gene had covariates for age, sex, and the first four PCs, while models including a metabolite had covariates for age, sex, cholesterol lowering medication use, and sample storage time.

Secondary Analyses

Secondary analyses were performed to further evaluate network analysis results and assess the biological validity of our findings. To better understand resulting communities of particular interest, we performed a joint pathway analysis integrating the genes and metabolites of a community in a joint pathway analysis conducted in MetaboAnalyst 4.0(Chong et al., 2018). This joint pathway analysis takes as input gene names and metabolite IDs from the Human Metabolome Database (HMDB)(Wishart et al., 2007) and identifies pathways that are over-represented.

Results that led to the identification of numerous metabolites were expanded upon to investigate the combined influence of metabolites. Multiple metabolites were summarized into a single measurement using a principal component analysis (PCA), with input being limited to the metabolites under study. The resulting first principal component (PC1) was then used to determine whether the same results could be captured, repeating analyses using PC1 in place of the metabolites of interest.

Results

Participants

A total of 1,111 WRAP participants had both genomic and plasma metabolomic data. At baseline, 68.9% of participants were female and participants were 61.0 years old with a bachelor's degree, on average (Table 2). Participants each had 1,097 plasma metabolites available for analyses, 347 (31.6%) of which were of unknown chemical structure, whole blood gene expression for 11,376 genes, and up to 17 AD risk factors. A subset of 155 individuals also had 364 CSF metabolites available for analyses, 56 (15.4%) of which were of unknown chemical structure. Participants with CSF metabolomic data had similar characteristics as the full sample (Table 2). Properties of each plasma and CSF metabolite, such as biochemical name, super pathway, and sub pathway are described in Tables S1 and S2, and numbers of metabolites within each super pathway are summarized in Table S3.

Correlation between plasma and CSF metabolomics

The median correlation between the 326 metabolites common to both plasma and CSF was r=0.26, with some variability existing between different metabolite pathways (Figure 1). Xenobiotics had the highest median correlation (r=0.53), while lipids had the lowest (r=0.11). Overall, metabolite correlations ranged from |r|=0.0002 (inosine, a nucleotide) to |r|=0.88 (quinate, a xenobiotic). Interestingly, one of the highest correlations was caffeine (r=0.81). Correlations between each of the 326 CSF and plasma metabolites are described in Table S4.

Integrated network

After applying a Bonferroni correction for multiple testing, a total of 90,308 significant correlations (edges) among 10,869 variables (nodes) were used to develop an overall 'hairball' network (Figure S2). Notably, although there were far fewer metabolites than genes in the network (1,387 metabolites versus 9,481 genes), there were more edges between metabolites than genes (49,499 versus 37,473 edges, respectively).

The inter-omic network is shown in Figure 2 (a labeled version is shown in Figure S3), and its corresponding community partitions are shown in Figure S4. This network had 1,224 edges and 635 nodes, including 171 metabolite-gene and 833 metabolite-AD risk factor edges. Of these, there were only four CSF metabolite-gene edges and 73 CSF metabolite-AD risk factor edges, likely due to the much smaller number of CSF metabolomic samples. No genes were directly linked to AD risk factors; however, many genes were indirectly linked to AD risk factors through metabolites, as described below. Each of the 1,224 correlations is described in Table S5.

The largest community contained 680 edges among 289 nodes, which included 264 plasma metabolites, ten CSF metabolites, eight genes, and seven AD risk factors related to cardiovascular disease and diabetes: body mass index (BMI), waist-hip ratio (WHR), homeostatic model assessment of insulin resistance (HOMA-IR), interleukin 6 (IL-6), metabolic equivalents (METs), diastolic blood pressure (DBP), and systolic blood pressure (SBP) (Figure S5A). Expression levels of these eight genes were all indirectly linked to AD risk factors within this community through plasma metabolites. *CPS1* expression levels were negatively correlated with plasma gamma-glutamylglycine, proprionylglycine, and glycine levels, all of which were negatively correlated with BMI, WHR, IL-6, and/or HOMA-IR (Figure 3). *TMEM229B* and *PLEKHH1* were both negatively correlated with two glycerophosphatidylcholines (1-(1-enyl-palmitoyl)-2-palmitoleoyl-GPC (P-16:0/16:1) and 1-(1-enyl-palmitoyl)-2-palmitoyl-GPC (P-16:0/16:0)), which were also negatively correlated with BMI, WHR, and/or HOMA-IR. *NAALAD2* was negatively correlated with an amino acid beta-citrylglutamate, which was positively correlated with BMI, WHR, IL-6, and HOMA-IR. *ZNF655* and *ZKSCAN1* were both positively correlated with X-12063, which was also positively correlated with BMI, WHR, and HOMA-IR. *CHRNA5* was positively correlated with 5-hydroxylysine, which was positively correlated with BMI, WHR, IL-6, and HOMA-IR, and negatively correlated with METs. *ARVCF* was negatively correlated with X-11593, which was positively correlated with BMI, IL-6, and HOMA-IR.

We identified several key pathways involved in the cardiovascular and diabetes community using a joint genetic and metabolomic pathway analysis (Figure S5B). The analysis included all eight genes and 153 of the 264 plasma metabolites in this community, due to the remaining metabolites not having an identifying HMDB ID needed for analyses. The strongest pathway was valine, leucine, and isoleucine biosynthesis (P=6.3e-6), followed by aminoacyl-tRNA biosynthesis (P=7.3e-4) and alanine, aspartate, and glutamate metabolism (P=1.9e-3). Of these top three pathways, the alanine, aspartate, and glutamate metabolism pathway was the only one that included one of the eight genes in the cardiovascular community, which was *CPS1*.

Several genes outside of the cardiovascular and diabetes community were indirectly linked to AD risk factors within this community. Gene expression of *FOSL2*, *KRTCAP3*, and *ZNF513* were positively correlated, while *IFT172*, *NRBP1*, *PPM1G*, and *ZNF512* were negatively correlated, with levels of plasma mannose, a carbohydrate that was positively correlated with BMI, WHR, IL-6, and HOMA-IR (Figure S6A). *CABP1*, *SPPL3*, and *UNC119B* expression levels were negatively correlated with plasma butyrylcarnitine (C4), which was positively correlated with BMI, WHR, IL-6, and HOMA-IR (Figure S6B). *SLC27A4*, *PHYHD1*, *ENDOG*, and *SH3GLB2* expression levels were negatively correlated with plasma 2'-O-methyluridine and 2'-O-methylcytidine levels, both nucleotides involved in pyrimidine metabolism, and the latter nucleotide is also negatively correlated with BMI and WHR (Figure S6C). *PHYHD1* was also negatively correlated with CSF levels of 2'-O-methylcytidine.

The only correlations identified among the CSF biomarkers (*i.e.*, amyloid and tau) are shown in Figure 4. Higher CSF T-tau and P-tau levels were correlated with higher levels of 38 CSF metabolites, collectively. These metabolites included 13 lipids (six

phosphatidylcholines, two lysophosphatidylcholines, five sphingolipids, and cholesterol), seven amino acids, five carbohydrates, one nucleotide, one energy metabolite, one cofactor and vitamin metabolite, one xenobiotic, and nine unknown metabolites. However, none of the CSF amyloid biomarkers were correlated with CSF metabolites. We investigated how much of the variance of T-tau and P-tau could be explained by these metabolites with linear mixed models that included random intercepts for within-subject repeated measures and within-family relationships, using the R^2 statistic for mixed models as defined by Edwards et al., 2008(Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008) and implemented in the r2glmm R package. After removing the variation explained by age and sex, the 37 metabolites correlated with T-tau explained 60.7% of the variation of T-tau, while the 35 metabolites correlated with P-tau explained 64.0% of the variation of P-tau. Because a small number of identified metabolites had a HMDB ID, we were unable to perform pathway analyses to further understand this community.

Targeted mediation and interaction analyses

Targeted mediation and interaction analyses were focused on a particular pathway identified within the large cardiovascular and diabetes community involving *CPS1*, glycine plasma metabolites (glycine, proprionylglycine, and gamma-glutamylglycine), BMI, WHR, IL-6, and HOMA-IR. Associations between *CPS1* variants and glycine have been reported in at least nine studies(Demirkan et al., 2015; Draisma et al., 2015; Kettunen et al., 2016; Long et al., 2017; Raffler et al., 2015; Shin et al., 2014; Suhre et al., 2011; Xie et al., 2013; Yu et al., 2014), more than any of the other gene-metabolite associations identified in our network analysis, and these studies were based not only on Caucasian populations, but also on Japanese and African American populations. Many previous studies have also reported associations between glycine and cardiovascular risk factors, including BMI, waist circumference, inflammation, and HOMA-IR(Cheng et al., 2012; Demirkan et al., 2015; Ding et al., 2015; El Hafidi, Perez, & Banos, 2006; Felig, Marliss, & Cahill, 1969; Geidenstam et al., 2017; Kraus et al., 2016; Takashina et al., 2016; Thalacker-Mercer et al., 2014). This evidence made this pathway a strong candidate for mediation and interaction analyses.

Figure 5 shows results from the mediation analyses using glycine as the mediator, including the total effect (*i.e.*, the effect of *CPS1* in the model unadjusted for glycine), the direct effect (*i.e.*, the effect of *CPS1* in the model adjusted for glycine), and the indirect effect (*i.e.*, the effect of *CPS1* due to the effect of *CPS1* on glycine) for BMI (Figure 5A and Figure 5B), WHR (Figure 5C and Figure 5D), IL-6 (Figure 5E and Figure 5F), and HOMA-IR (Figure 5G and Figure 5H). The total effect of *CPS1* was null for each of these four outcomes, likely due to the negative association between *CPS1* and glycine coupled with the negative association between glycine and the AD risk factor, resulting in direct and indirect effects that had opposing directions(Richiardi, Bellocco, & Zugna, 2013). Our results show that lower levels of *CPS1* expression lead to increased glycine levels, and higher glycine levels lead to decreased BMI, WHR, IL-6, and HOMA-IR. Thus, with glycine as a mediator, lower levels of *CPS1* lead to decreased BMI, WHR, IL-6, and HOMA-IR. Mediation analyses using propionylglycine and gamma-glutamylglycine as the mediator showed similar results and can be found in Figure S7 and Figure S8. We did not identify any interactions between

CPS1 and the three glycine metabolites that were associated with BMI, WHR, IL-6, or HOMA-IR (all *P*-values>0.07). Similar results were observed in sensitivity analyses limited to unrelated individuals (due to the inability to include a random effect for family members in mediation analyses).

To expand on these mediation results, we conducted a PCA on the three glycine plasma metabolites and repeated mediation analyses. PC1 explained 71.8% of the variance of the three metabolites and was used to further test the mediation effect identified. Results similarly showed that while *CPS1* expression did not have a direct effect on BMI, WHR, IL-6, or HOMA-IR, it had a very strong indirect effect, when including PC1 as a mediator, on all four outcomes (*P*<2.0e-16), further suggesting that lower *CPS1* levels are protective against high levels of these cardiovascular risk factors (Figure S9).

Discussion

We developed an integrative network to investigate relationships between genomics, plasma metabolomics, CSF metabolomics, and AD risk factors. Although no gene expression levels were directly correlated with AD risk factors, there were many instances of genes being indirectly correlated with AD risk factors though metabolites. Building on one such instance, we found that glycine mediated the pathway between *CPS1* expression and cardiovascular and diabetes risk factors. This suggests that our results may have generated many valid hypotheses that warrant further investigation. We also found that correlations between plasma and CSF metabolites ranged widely but typically had low correlations. This could suggest that most plasma metabolites are not representative of certain metabolic changes occurring in the brain, although we cannot rule out the possibility that the low average correlation is, at least partially, due to the time difference between the plasma and CSF sample collection.

The low correlation we observed between plasma and CSF metabolite levels could be related to ~98% of small molecules not being able to pass the blood-brain barrier (BBB)(Pardridge, 2005). Cholesterol is an example of a lipid metabolite that typically cannot pass the BBB(Bjorkhem & Meaney, 2004), and was not correlated between tissues (r=−0.07). On the other hand, caffeine (a xenobiotic) readily crosses the BBB(McCall, Millington, & Wurtman, 1982) and it was highly correlated between tissues (r=0.81), as was 5-acetylamino-6-amino-3-methyluracil (r=0.82), which is a caffeine metabolite, and theophylline (r=0.82), which is structurally and pharmacologically similar to caffeine. This could contribute to lipids having the weakest average correlation and xenobiotics having the strongest average correlation between plasma and CSF tissues. However, it is important to note that metabolites within a given pathway can vary widely from each other and should be considered on an individual basis, accordingly, as the averages presented here may not reflect a particular metabolite's unique properties. The hypothesis about plasma and CSF differing due to the BBB is also supported by the only correlations in the network analysis involving CSF biomarkers (*i.e.*, tau) being with CSF metabolites, although we cannot rule out the possibility that this correlation is related to CSF biomarkers and CSF metabolomics being analyzed from the same sample and thus, not having time-related variation.

Our network analysis revealed that 38 CSF metabolites were highly predictive of CSF T-tau and P-tau, collectively explaining 60.7% and 64.0% of the variance of T-tau and P-tau, respectively. Further investigations of these CSF metabolites could lead to a better understanding of mechanisms and pathways that influence the development of tau tangles. In contrast, no CSF metabolites were correlated with CSF amyloid biomarkers, which could have implications about the biological function of amyloid versus tau. It is possible that we did not capture the small molecules that amyloid may be associated with, or that amyloid is generally not associated with small molecules. Although our CSF findings were limited by their small sample size, they offer potentially novel information regarding the interface between CSF biomarkers and CSF metabolites, as we have not identified previous studies investigating these relationships.

One advantage of using imputed gene expression data is that it only represents the genetically regulated component of gene expression, reducing the risk of confounding due to environmental factors and reverse causality in mediation analyses. We found that glycine mediated the relationship between *CPS1* and BMI, WHR, IL-6, and HOMA-IR, such that lower *CPS1* expression was associated with higher levels of glycine, which were associated with lower BMI, WHR, IL-6, and HOMA-IR. Relationships between *CPS1*, glycine, and cardiovascular risk factors have been hypothesized recently, but not clearly defined(Matone et al., 2016; Xie et al., 2013). The *CPS1* (Carbamoyl-Phosphate Synthase 1) gene encodes for a mitochondrial enzyme that catalyzes the first step of the hepatic urea cycle by synthesizing carbamoyl phosphate from ammonia, bicarbonate, and two molecules of ATP, and is important for removal of urea from cells(Haberle et al., 2011). Notably, all genes encoding enzymes involved in the urea cycle are expressed in the brain, including *CPS1*(Hansmannel et al., 2010), and levels of enzymes and metabolic intermediates involved in the urea cycle are altered in AD patients(Griffin & Bradshaw, 2017). *CPS1* variants have been linked to CPS1 deficiency(Haberle et al., 2011), neonatal pulmonary hypertension(Pearson et al., 2001), vascular function(Summar et al., 2004), traits related to blood clotting, such as fibrinogen levels and platelet count(Astle et al., 2016; Danik et al., 2009; de Vries et al., 2016; Sabater-Lleal et al., 2013), homocysteine levels(Lange et al., 2010; Pare et al., 2009; van Meurs et al., 2013; Williams et al., 2014), HDL cholesterol(Willer et al., 2013), kidney function and disease(Gorski et al., 2017; Kottgen et al., 2010; Mahajan et al., 2016; Pattaro et al., 2016), AD(Jun et al., 2016), and BMI(Locke et al., 2015; Melen et al., 2013). Higher adipose tissue expression of *CPS1* has been associated with detrimental traits, including weight gain(Matone et al., 2016). At least nine studies have reported associations between *CPS1* variants and glycine(Demirkan et al., 2015; Draisma et al., 2015; Kettunen et al., 2016; Long et al., 2017; Raffler et al., 2015; Shin et al., 2014; Suhre et al., 2011; Xie et al., 2013; Yu et al., 2014) and others have reported associations with betaine, a derivative of glycine(Hartiala et al., 2016; Long et al., 2017; Shin et al., 2014). Notably, in previous work we found that glycine has a high heritability of 71.7% (Darst, Koscik, Hogan, Johnson, & Engelman, 2019), supporting the potential of glycine to mediate a genetic factor. Glycine is a common amino acid involved in the production of DNA, phospholipids, and collagen, and in the release of energy. Previous studies have identified negative correlations between glycine and cardiovascular and diabetes risk factors such as BMI, waist circumference, HOMA-IR, obesity and visceral obesity, subcutaneous

and visceral fat area, hypertension, and acute myocardial infarction(Cheng et al., 2012; Demirkan et al., 2015; Ding et al., 2015; El Hafidi et al., 2006; Felig et al., 1969; Geidenstam et al., 2017; Kraus et al., 2016; Takashina et al., 2016; Thalacker-Mercer et al., 2014). These previous findings are in the same direction as our findings and are highly supportive of the biological relevance of our results, which lead us to hypothesize that the *CPS1*-cardiovascular risk pathway is linked through the mediation of glycine.

One particular *CPS1* variant, rs715, has been linked to urine and blood glycine levels(Demirkan et al., 2015; Long et al., 2017; Raffler et al., 2015; Shin et al., 2014; Xie et al., 2013), blood levels of betaine(Hartiala et al., 2016; Long et al., 2017; Shin et al., 2014), blood levels of fibrinogen(de Vries et al., 2016; Sabater-Lleal et al., 2013), and BMI(Locke et al., 2015). This is a common variant, with a MAF=0.27 based on 62,784 whole genome sequences from Trans-Omics for Precision Medicine (TOPMed)(TNT, 2017). The minor C allele of rs715 decreases *CPS1* expression(Hartiala et al., 2016). To further test our findings, we conducted additional mediation analyses using this variant and found highly consistent results, suggesting that having one or two minor alleles of rs715 (which decreases *CPS1* expression; MAF=0.33 in our data) increases levels of the three glycine plasma metabolites, which decreases BMI, WHR, IL-6, and HOMA-IR (P<2.0e-16 among 14 of the 16 tests) (Figures S10–S12). Thus, the minor C allele of rs715 may have a protective role in cardiovascular risk.

To further understand the cardiovascular and diabetes community, which was the largest community in our model and contained the glycine mediator, we performed a joint genetic and metabolomic pathway analysis, which led to strong evidence for the involvement of the valine, leucine, and isoleucine biosynthesis pathway in this community. Collectively, these three essential amino acids are referred to as branched-chain amino acids (BCAAs) and are acquired solely from external food(Y. Huang, Zhou, Sun, & Wang, 2011). Higher BCAA levels have been linked to the onset of type 2 diabetes, cardiovascular disease, and insulin resistance(Asghari et al., 2018; Bhattacharya et al., 2014; Wang et al., 2011). This further conveys the biological relevance of the factors identified in this community to cardiovascular disease and diabetes risk.

A notable challenge of this investigation is identifying an appropriate cohort to replicate novel findings. Although previous literature supports the biological plausibility of *CPS1* mediating the relationship between glycine and cardiovascular risk factors, it will be crucial to replicate this finding, along with the high variance of T-tau and P-tau explained by the 38 CSF metabolites we identified. However, the robustness of our mediation results is strengthened by our consistent findings across several plasma glycine metabolites (glycine, proprionylglycine, and gamma-glutamylglycine, with r^2 ranging from 0.17 to 0.76 between these three metabolites), four independent cardiovascular risk outcomes (BMI, WHR, IL-6, HOMA-IR, with r^2 ranging from 0.02 to 0.14 between these four outcomes), and imputed *CPS1* expression as well as rs715 within *CPS1* (across these 24 tests, 20 had a mediation effect P<2.0e-16). It was further strengthened by similar results from analyses using PC1 from a PCA of the three glycine metabolites as a mediator of *CPS1* and cardiovascular risk. While we were encouraged by this internal validation, finding that many of our results replicated previous investigations, and biologically relevant pathways identified in the

cardiovascular and diabetes community pathway analysis, this does not preclude the need for external replication of novel findings.

Due to computational burdens, our network analysis did not fully utilize the longitudinal aspect of our data. Further, our sample sizes for CSF biomarkers and metabolites were limited, which is likely why we had few CSF findings in our network analysis. Plasma and CSF samples typically were not collected on the same day, which could influence our correlation results between these two tissues. However, this may not have influenced our network analysis to a large extent because we used individual level random effects for longitudinal traits. We were unable to include smoking behavior in our network analysis due to the prohibitive number of smokers in our cohort (n=48).

One of the primary strengths of this analysis is that it shows the feasibility of performing integrated omics analyses and the potential utility of such approaches. It is becoming more common for cohorts to collect such datasets; for example, the National Institutes of Health is sponsoring the new TOPMed nation-wide consortium that aims to deeply phenotype its participants utilizing omics technologies (www.nhlbiwgs.org). It is anticipated that initiatives such as TOPMed will greatly advance our knowledge of many complex diseases and traits. However, to fully utilize these rich data, it will be crucial to identify effective means of integrating them and maximize their potential to provide a more holistic understanding of the disease process. While there is still a great need for such methods, our inter-omic network analysis and subsequent targeted follow-up analyses outlines one approach to effectively integrate omics data.

Opponents of the "big data" era have criticized omics approaches because they are not hypothesis driven and do not follow the standard scientific method(Duncan, 2007; Stieb, Boot, & Turner, 2017). However, we know biology to be complex far beyond our current understanding. To believe that we currently have the ability to generate valid biological hypotheses to understand complex conditions without data would be a fallacy. This was a lesson learned in the years preceding the completion of the human genome sequence in 2001 when research efforts were heavily invested into targeted genetic loci and genome-wide linkage screens of ~500 loci(Fallin, Duggal, & Beaty, 2016). This approach was successful for genes that follow Mendelian patterns, such as highly penetrant variants in the *BRCA1* and *BRCA2* genes that are responsible for inherited forms of breast cancer and the *APP*, *PSEN1*, and *PSEN2* genes that cause the inherited early onset form of AD. However, it had limited success for traits that follow complex inheritance patterns(Fallin et al., 2016). The utility of omics data, and particularly integrated omics approaches, is the ability to generate data driven hypotheses. Our knowledge of biology has been evolving for centuries; however, with the data we are able to generate due to recent biotechnological advances, we now have the opportunity to advance our knowledge of biology at an unprecedented rate. Such data could lead to dramatic improvements in the state of preventative and therapeutic medicine, particularly for complex diseases such as AD, for which few such preventative or therapeutic methods exist and little is known about the underlying biological mechanisms.

By integrating genomics, metabolomics, and clinical risk factors for AD, we were able to identify complex relationships that offer insight into the onset of AD and risk factors

associated with its onset. Our research has generated many promising hypotheses that could drive subsequent experimental investigations and potentially offer clinicians and researchers new insights regarding the development of tau tangles. As the generation of omics data accelerates across investigations of a variety of research fields, continued efforts to navigate statistical and computational issues will be critical. The work presented here represents early efforts to integrate omics data, but much more research is needed to identify the most effective means of doing so and thereby maximize the utility of such rich sources of data. The success of precision medicine is heavily reliant on the advancement of computational biology and the ability to translate millions of biological data points into individual clinical implications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Asghari G, Farhadnejad H, Teymoori F, Mirmiran P, Tohidi M, & Azizi F (2018). High dietary intake of branched-chain amino acids is associated with an increased risk of insulin resistance in adults. J Diabetes, 10(5), 357–364. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29281182. doi: 10.1111/1753-0407.12639 [PubMed: 29281182]

Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, … Soranzo N. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell, 167(5), 1415–1429 e1419. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27863252. doi: 10.1016/j.cell.2016.10.042 [PubMed: 27863252]

Bates D, Machler M, Bolker BM, & Walker SC (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–48. Retrieved from <Go to ISI>://WOS: 000365981400001.

Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, … Koller D. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res, 24(1), 14–24. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24092820. doi:10.1101/gr.155192.113 [PubMed: 24092820]

Bersanelli M, Mosca E, Remondini D, Giampieri E, Sala C, Castellani G, & Milanesi L (2016). Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics, 17 Suppl 2, 15 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26821531. doi:10.1186/ s12859-015-0857-9 [PubMed: 26821531]

Bhattacharya S, Granger CB, Craig D, Haynes C, Bain J, Stevens RD, … Shah SH. (2014). Validation of the association between a branched chain amino acid metabolite profile and extremes of coronary

artery disease in patients referred for cardiac catheterization. Atherosclerosis, 232(1), 191–196. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24401236. doi:10.1016/j.atherosclerosis. 2013.10.036 [PubMed: 24401236]

Bjorkhem I, & Meaney S (2004). Brain cholesterol: long secret life behind a barrier. Arterioscler Thromb Vasc Biol, 24(5), 806–815. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14764421. doi:10.1161/01.ATV.0000120374.59826.1b [PubMed: 14764421]

Buescher JM, & Driggers EM (2016). Integration of omics: more than the sum of its parts. Cancer Metab, 4, 4 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26900468. doi:10.1186/s40170-016-0143-y [PubMed: 26900468]

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, & Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4, 7 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25722852. doi:10.1186/s13742-015-0047-8 [PubMed: 25722852]

Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, … Wang TJ (2012). Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation, 125(18), 2222–2231. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22496159. doi:10.1161/CIRCULATIONAHA.111.067827 [PubMed: 22496159]

Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, … Xia J. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res, 46(W1), W486–W494. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/29762782. doi:10.1093/nar/gky310 [PubMed: 29762782]

Clauset A, Newman MEJ, & Moore C (2004). Finding community structure in very large networks. Physical Review E, 70(6). Retrieved from <Go to ISI>://WOS:000226299200018. doi:Artn066111 10.1103/Physreve.70.066111

Conomos MP, Miller MB, & Thornton TA (2015). Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol, 39(4), 276–293. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25810074. doi:10.1002/gepi.21896 [PubMed: 25810074]

Conomos MP, Reiner AP, Weir BS, & Thornton TA (2016). Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet, 98(1), 127–148. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26748516. doi:10.1016/j.ajhg.2015.11.022 [PubMed: 26748516]

Csardi G, & Nepusz T (2006). The igraph software package for complex network research. InterJournal, Complex Systems, 1695 Retrieved from http://igraph.org.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, … Genomes Project Analysis, G. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21653522. doi:10.1093/bioinformatics/btr330 [PubMed: 21653522]

Danik JS, Pare G, Chasman DI, Zee RY, Kwiatkowski DJ, Parker A, … Ridker PM (2009). Novel loci, including those related to Crohn disease, psoriasis, and inflammation, identified in a genome-wide association study of fibrinogen in 17 686 women: the Women's Genome Health Study. Circ Cardiovasc Genet, 2(2), 134–141. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20031577. doi:10.1161/CIRCGENETICS.108.825273 [PubMed: 20031577]

Darst BF, Koscik RL, Hogan KJ, Johnson SC, & Engelman CD (2019). Longitudinal plasma metabolomics of aging and sex. Aging (Albany NY), 11(4), 1262–1282. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/30799310. doi:10.18632/aging.10183 [PubMed: 30799310]

Darst BF, Koscik RL, Racine AM, Oh JM, Krause RA, Carlsson CM, … Engelman CD (2017). Pathway-Specific Polygenic Risk Scores as Predictors of Amyloid-beta Deposition and Cognitive Function in a Sample at Increased Risk for Alzheimer's Disease. J Alzheimers Dis, 55(2), 473–484. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27662287. doi:10.3233/JAD-160195 [PubMed: 27662287]

Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, … Fuchsberger C. (2016). Next-generation genotype imputation service and methods. Nat Genet, 48(10), 1284–1287. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27571263. doi:10.1038/ng.3656 [PubMed: 27571263]

de Vries PS, Chasman DI, Sabater-Lleal M, Chen MH, Huffman JE, Steri M, … Dehghan A. (2016). A meta-analysis of 120 246 individuals identifies 18 new loci for fibrinogen concentration. Hum Mol

Genet, 25(2), 358–370. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26561523. doi: 10.1093/hmg/ddv454 [PubMed: 26561523]

Demirkan A, Henneman P, Verhoeven A, Dharuri H, Amin N, van Klinken JB, … van Dijk KW (2015). Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses. PLoS Genet, 11(1), e1004835 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25569235. doi:10.1371/journal.pgen.1004835 [PubMed: 25569235]

Ding Y, Svingen GF, Pedersen ER, Gregory JF, Ueland PM, Tell GS, & Nygard OK (2015). Plasma Glycine and Risk of Acute Myocardial Infarction in Patients With Suspected Stable Angina Pectoris. J Am Heart Assoc, 5(1). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26722126. doi:10.1161/JAHA.115.002621

Draisma HHM, Pool R, Kobl M, Jansen R, Petersen AK, Vaarhorst AAM, … Boomsma DI (2015). Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun, 6, 7208 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26068415. doi:10.1038/ncomms8208 [PubMed: 26068415]

Duncan MW (2007). Omics and its 15 minutes. Exp Biol Med (Maywood), 232(4), 471–472. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17392481. [PubMed: 17392481]

Edwards LJ, Muller KE, Wolfinger RD, Qaqish BF, & Schabenberger O (2008). An R2 statistic for fixed effects in the linear mixed model. Stat Med, 27(29), 6137–6157. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18816511. doi:10.1002/sim.3429 [PubMed: 18816511]

El Hafidi M, Perez I, & Banos G (2006). Is glycine effective against elevated blood pressure? Curr Opin Clin Nutr Metab Care, 9(1), 26–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16444815. [PubMed: 16444815]

Enche Ady CNA, Lim SM, Teh LK, Salleh MZ, Chin AV, Tan MP, … Ramasamy K. (2017). Metabolomic-guided discovery of Alzheimer's disease biomarkers from body fluid. J Neurosci Res, 95(10), 2005–2024. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28301062. doi: 10.1002/jnr.24048 [PubMed: 28301062]

Evans AM, Bridgewater BR, Liu Q, Mitchell MW, Robinson RJ, Dai H, … Miller LAD (2014). High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. Metabolomics, 4(1), 1–7. doi:10.4172/2153-0769.1000132

Fallin MD, Duggal P, & Beaty TH (2016). Genetic Epidemiology and Public Health: The Evolution From Theory to Technology. Am J Epidemiol, 183(5), 387–393. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26905340. doi:10.1093/aje/kww001 [PubMed: 26905340]

Felig P, Marliss E, & Cahill GF Jr. (1969). Plasma amino acid levels and insulin secretion in obesity. N Engl J Med, 281(15), 811–816. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5809519. doi:10.1056/NEJM196910092811503 [PubMed: 5809519]

Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, … Im HK (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat Genet, 47(9), 1091–1098. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26258848. doi:10.1038/ng.3367 [PubMed: 26258848]

Geidenstam N, Magnusson M, Danielsson APH, Gerszten RE, Wang TJ, Reinius LE, … Ridderstrale M (2017). Amino Acid Signatures to Evaluate the Beneficial Effects of Weight Loss. Int J Endocrinol, 2017, 6490473 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28484491. doi: 10.1155/2017/6490473 [PubMed: 28484491]

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, … Abecasis GR (2015). A global reference for human genetic variation. Nature, 526(7571), 68–74. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/26432245. doi:10.1038/nature15393 [PubMed: 26432245]

Gligorijevic V, & Przulj N (2015). Methods for biological data integration: perspectives and challenges. J R Soc Interface, 12(112). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26490630. doi:10.1098/rsif.2015.0571

Gorski M, van der Most PJ, Teumer A, Chu AY, Li M, Mijatovic V, … Fuchsberger C (2017). 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. Sci Rep, 7, 45040 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28452372. doi:10.1038/srep45040 [PubMed: 28452372]

Griffin JW, & Bradshaw PC (2017). Amino Acid Catabolism in Alzheimer's Disease Brain: Friend or Foe? Oxid Med Cell Longev, 2017, 5472792 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28261376. doi:10.1155/2017/5472792 [PubMed: 28261376]

Haberle J, Shchelochkov OA, Wang J, Katsonis P, Hall L, Reiss S, … Summar M (2011). Molecular defects in human carbamoy phosphate synthetase I: mutational spectrum, diagnostic and protein structure considerations. Hum Mutat, 32(6), 579–589. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21120950. doi:10.1002/humu.21406 [PubMed: 21120950]

Hansmannel F, Sillaire A, Kamboh MI, Lendon C, Pasquier F, Hannequin D, … Lambert JC (2010). Is the urea cycle involved in Alzheimer's disease? J Alzheimers Dis, 21(3), 1013–1021. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20693631. doi:10.3233/JAD-2010-100630 [PubMed: 20693631]

Hartiala JA, Tang WH, Wang Z, Crow AL, Stewart AF, Roberts R, … Allayee H (2016). Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. Nat Commun, 7, 10558 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26822151. doi:10.1038/ncomms10558 [PubMed: 26822151]

Horgan RP, & Kenny LC (2011). 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. The Obstetrician & Gynaecologist, 13, 189–195. doi:10.1576/toag.13.3.189.27672

Huang S, Chaudhary K, & Garmire LX (2017). More Is Better: Recent Progress in Multi-Omics Data Integration Methods. Front Genet, 8, 84 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28670325. doi:10.3389/fgene.2017.00084 [PubMed: 28670325]

Huang Y, Zhou M, Sun H, & Wang Y (2011). Branched-chain amino acid metabolism in heart disease: an epiphenomenon or a real culprit? Cardiovasc Res, 90(2), 220–223. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21502372. doi:10.1093/cvr/cvr070 [PubMed: 21502372]

Imai K, Keele L, & Tingley D (2010). A general approach to causal mediation analysis. Psychol Methods, 15(4), 309–334. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/20954780. doi:10.1037/a0020761 [PubMed: 20954780]

Jack CR Jr., Lowe VJ, Weigand SD, Wiste HJ, Senjem ML, Knopman DS, … Alzheimer's Disease Neuroimaging, I. (2009). Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. Brain, 132(Pt 5), 1355–1365. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19339253. doi:10.1093/brain/awp062 [PubMed: 19339253]

Johnson SC, Koscik RL, Jonaitis EM, Clark LR, Mueller KD, Berman SE, … Sager MA (2018). The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current directions. Alzheimers Dement (Amst), 10, 130–142. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/29322089. doi:10.1016/j.dadm.2017.11.007 [PubMed: 29322089]

Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert JC, Chung J, Naj AC, … Farrer LA (2016). A novel Alzheimer disease locus located near the gene encoding tau protein. Mol Psychiatry, 21(1), 108–117. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25778476. doi:10.1038/mp.2015.23 [PubMed: 25778476]

Kettunen J, Demirkan A, Wurtz P, Draisma HH, Haller T, Rawal R, … Ala-Korpela M (2016). Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun, 7, 11122 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27005778. doi:10.1038/ncomms11122 [PubMed: 27005778]

Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, … Fox CS (2010). New loci associated with kidney function and chronic kidney disease. Nat Genet, 42(5), 376–384. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20383146. doi:10.1038/ng.568 [PubMed: 20383146]

Kraus WE, Pieper CF, Huffman KM, Thompson DK, Kraus VB, Morey MC, … Newgard CB (2016). Association of Plasma Small-Molecule Intermediate Metabolites With Age and Body Mass Index Across Six Diverse Study Populations. J Gerontol A Biol Sci Med Sci, 71(11), 1507–1513. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26984390. doi:10.1093/gerona/glw031 [PubMed: 26984390]

Lange LA, Croteau-Chonka DC, Marvelle AF, Qin L, Gaulton KJ, Kuzawa CW, … Mohlke KL (2010). Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. Hum Mol Genet, 19(10), 2050–

2058. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20154341. doi:10.1093/hmg/ddq062 [PubMed: 20154341]

Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, … Speliotes EK (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature, 518(7538), 197–206. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25673413. doi:10.1038/nature14177 [PubMed: 25673413]

Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, … A LP (2016). Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet, 48(11), 1443–1448. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27694958. doi:10.1038/ng.3679 [PubMed: 27694958]

Long T, Hicks M, Yu HC, Biggs WH, Kirkness EF, Menni C, … Telenti A (2017). Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet, 49(4), 568–578. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28263315. doi:10.1038/ng.3809 [PubMed: 28263315]

Lopez de Maturana E, Pineda S, Brand A, Van Steen K, & Malats N (2016). Toward the integration of Omics data in epidemiological studies: still a "long and winding road". Genet Epidemiol, 40(7), 558–569. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27432111. doi:10.1002/gepi.21992 [PubMed: 27432111]

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, … Parkinson H (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res, 45(D1), D896–D901. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27899670. doi:10.1093/nar/gkw1133 [PubMed: 27899670]

Mahajan A, Rodan AR, Le TH, Gaulton KJ, Haessler J, Stilp AM, … Franceschini N (2016). Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity. Am J Hum Genet, 99(3), 636–646. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27588450. doi:10.1016/j.ajhg.2016.07.012 [PubMed: 27588450]

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, & Chen WM (2010). Robust relationship inference in genome-wide association studies. Bioinformatics, 26(22), 2867–2873. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20926424. doi:10.1093/bioinformatics/btq559 [PubMed: 20926424]

Mankoo PK, Shen R, Schultz N, Levine DA, & Sander C (2011). Time to recurrence and survival in serous ovarian tumors predicted from integrated genomic profiles. PLoS One, 6(11), e24709 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22073136. doi:10.1371/journal.pone.0024709 [PubMed: 22073136]

Matone A, Scott-Boyer MP, Carayol J, Fazelzadeh P, Lefebvre G, Valsesia A, … Hager J (2016). Network Analysis of Metabolite GWAS Hits: Implication of CPS1 and the Urea Cycle in Weight Maintenance. PLoS One, 11(3), e0150495 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26938218. doi:10.1371/journal.pone.0150495 [PubMed: 26938218]

McCall AL, Millington WR, & Wurtman RJ (1982). Blood-brain barrier transport of caffeine: dose-related restriction of adenine transport. Life Sci, 31(24), 2709–2715. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7154859. [PubMed: 7154859]

McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, … Haplotype Reference, C. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet, 48(10), 1279–1283. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27548312. doi:10.1038/ng.3643 [PubMed: 27548312]

Melen E, Granell R, Kogevinas M, Strachan D, Gonzalez JR, Wjst M, … Lasky-Su J (2013). Genome-wide association study of body mass index in 23 000 individuals with and without asthma. Clin Exp Allergy, 43(4), 463–474. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23517042. doi:10.1111/cea.12054 [PubMed: 23517042]

Menni C, Kastenmuller G, Petersen AK, Bell JT, Psatha M, Tsai PC, … Valdes AM (2013). Metabolomic markers reveal novel pathways of ageing and early development in human populations. Int J Epidemiol, 42(4), 1111–1119. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23838602. doi:10.1093/ije/dyt094 [PubMed: 23838602]

Pardridge WM (2005). The blood-brain barrier: bottleneck in brain drug development. NeuroRx, 2(1), 3–14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15717053. doi:10.1602/neurorx.2.1.3 [PubMed: 15717053]

Pare G, Chasman DI, Parker AN, Zee RR, Malarstig A, Seedorf U, … Ridker PM (2009). Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. Circ Cardiovasc Genet, 2(2), 142–150. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20031578. doi:10.1161/CIRCGENETICS.108.829804 [PubMed: 20031578]

Pattaro C, Teumer A, Gorski M, Chu AY, Li M, Mijatovic V, … Fox CS (2016). Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. Nat Commun, 7, 10023 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26831199. doi:10.1038/ncomms10023 [PubMed: 26831199]

Pearson DL, Dawling S, Walsh WF, Haines JL, Christman BW, Bazyk A, … Summar ML (2001). Neonatal pulmonary hypertension--urea-cycle intermediates, nitric oxide production, and carbamoyl-phosphate synthetase function. N Engl J Med, 344(24), 1832–1838. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11407344. doi:10.1056/NEJM200106143442404 [PubMed: 11407344]

Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, … Hood L (2017). A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nature Biotechnology, 35(8), 747–756. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28714965. doi:10.1038/nbt.3870

Raffler J, Friedrich N, Arnold M, Kacprowski T, Rueedi R, Altmaier E, … Suhre K (2015). Genome-Wide Association Study with Targeted and Non-targeted NMR Metabolomics Identifies 15 Novel Loci of Urinary Human Metabolic Individuality. PLoS Genet, 11(9), e1005487 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26352407. doi:10.1371/journal.pgen.1005487 [PubMed: 26352407]

Rayner NW, Robertson N, Mahajan A, & McCarthy MI (2016). A suite of programs for pre- and post-imputation data checking. Paper presented at the The American Society of Human Genetics, Vancouver, Canada http://www.ashg.org/2016meeting/

Richiardi L, Bellocco R, & Zugna D (2013). Mediation analysis in epidemiology: methods, interpretation and bias. Int J Epidemiol, 42(5), 1511–1519. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24019424. doi:10.1093/ije/dyt127 [PubMed: 24019424]

Ridge PG, Hoyt KB, Boehme K, Mukherjee S, Crane PK, Haines JL, … Alzheimer's Disease Genetics, C. (2016). Assessment of the genetic variance of late-onset Alzheimer's disease. Neurobiol Aging, 41, 200 e213–200 e220. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/27036079. doi:10.1016/j.neurobiolaging.2016.02.024

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, & Kim D (2015). Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet, 16(2), 85–97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25582081. doi:10.1038/nrg3868 [PubMed: 25582081]

Sabater-Lleal M, Huang J, Chasman D, Naitza S, Dehghan A, Johnson AD, … O'Donnell CJ (2013). Multiethnic meta-analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci but no strong evidence of a causal association between circulating fibrinogen and cardiovascular disease. Circulation, 128(12), 1310–1324. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23969696. doi:10.1161/CIRCULATIONAHA.113.002251 [PubMed: 23969696]

Sager MA, Hermann B, & La Rue A (2005). 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, 18(4), 245–249. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16306248. doi:10.1177/0891988705281882 [PubMed: 16306248]

Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, … Soranzo N (2014). An atlas of genetic influences on human blood metabolites. Nat Genet, 46(6), 543–550. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24816252. doi:10.1038/ng.2982 [PubMed: 24816252]

Stieb DM, Boot CR, & Turner MC (2017). Promise and pitfalls in the application of big data to occupational and environmental health. BMC Public Health, 17(1), 372 Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28482822. doi:10.1186/s12889-017-4286-8 [PubMed: 28482822]

Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, … Gieger C (2011). Human metabolic individuality in biomedical and pharmaceutical research. Nature, 477(7362), 54–60. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21886157. doi:10.1038/nature10354 [PubMed: 21886157]

Summar ML, Gainer JV, Pretorius M, Malave H, Harris S, Hall LD, … Brown NJ (2004). Relationship between carbamoyl-phosphate synthetase genotype and systemic vascular function. Hypertension, 43(2), 186–191. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14718356. doi: 10.1161/01.HYP.0000112424.06921.52 [PubMed: 14718356]

Takashina C, Tsujino I, Watanabe T, Sakaue S, Ikeda D, Yamada A, … Nishimura M (2016). Associations among the plasma amino acid profile, obesity, and glucose metabolism in Japanese adults with normal glucose tolerance. Nutr Metab (Lond), 13, 5 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26788116. doi:10.1186/s12986-015-0059-5 [PubMed: 26788116]

Thalacker-Mercer AE, Ingram KH, Guo F, Ilkayeva O, Newgard CB, & Garvey WT (2014). BMI, RQ, diabetes, and sex affect the relationships between amino acids and clamp measures of insulin action in humans. Diabetes, 63(2), 791–800. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24130332. doi:10.2337/db13-0396 [PubMed: 24130332]

TNT C (2017). Trans-Omics for Precision Medicine (TOPMed). Retrieved April 14, 2018 https://bravo.sph.umich.edu

Ussher JR, Elmariah S, Gerszten RE, & Dyck JR (2016). The Emerging Role of Metabolomics in the Diagnosis and Prognosis of Cardiovascular Disease. J Am Coll Cardiol, 68(25), 2850–2870. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28007146. doi:10.1016/j.jacc.2016.09.972 [PubMed: 28007146]

van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, & van der Werf MJ (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics, 7, 142 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16762068. doi: 10.1186/1471-2164-7-142 [PubMed: 16762068]

van Meurs JB, Pare G, Schwartz SM, Hazra A, Tanaka T, Vermeulen SH, … Ahmadi KR (2013). Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. Am J Clin Nutr, 98(3), 668–676. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23824729. doi:10.3945/ajcn.112.044545 [PubMed: 23824729]

Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, … Gerszten RE (2011). Metabolite profiles and the risk of developing diabetes. Nat Med, 17(4), 448–453. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21423183. doi:10.1038/nm.2307 [PubMed: 21423183]

Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, … Global Lipids Genetics, C. (2013). Discovery and refinement of loci associated with lipid levels. Nat Genet, 45(11), 1274–1283. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24097068. doi:10.1038/ng.2797 [PubMed: 24097068]

Williams SR, Yang Q, Chen F, Liu X, Keene KL, Jacques P, … Framingham Heart, S. (2014). Genome-wide meta-analysis of homocysteine and methionine metabolism identifies five one carbon metabolism loci and a novel association of ALDH1L1 with ischemic stroke. PLoS Genet, 10(3), e1004214 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24651765. doi:10.1371/journal.pgen.1004214 [PubMed: 24651765]

Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, … Querengesser L (2007). HMDB: the Human Metabolome Database. Nucleic Acids Res, 35(Database issue), D521–526. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17202168. doi:10.1093/nar/gkl923 [PubMed: 17202168]

Xie W, Wood AR, Lyssenko V, Weedon MN, Knowles JW, Alkayyali S, … Walker M (2013). Genetic variants associated with glycine metabolism and their role in insulin sensitivity and type 2 diabetes. Diabetes, 62(6), 2141–2150. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23378610. doi:10.2337/db12-0876 [PubMed: 23378610]

Yu B, Zheng Y, Alexander D, Morrison AC, Coresh J, & Boerwinkle E (2014). Genetic determinants influencing human serum metabolome among African Americans. PLoS Genet, 10(3), e1004212 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24625756. doi:10.1371/journal.pgen.1004212 [PubMed: 24625756]

Zhang AH, Sun H, & Wang XJ (2017). Emerging role and recent applications of metabolomics biomarkers in obesity disease research. Rsc Advances, 7(25), 14966–14973. Retrieved from <Go to ISI>://WOS:000396290400002. doi:10.1039/c6ra28715h

[dataset] Johnson SC, Koscik RL, Jonaitis EM, Clark LR, Mueller KD, Berman SE, Bendline BB, Engelman CD, Okonkwo OC, Hogan KJ, Asthana S, Carlsson CM, Hermann BP, Sager MA (2018). The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current directions. Alzheimers Dement (Amst), 10, 130–142. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/29322089. doi:10.1016/j.dadm.2017.11.007 [PubMed: 29322089]
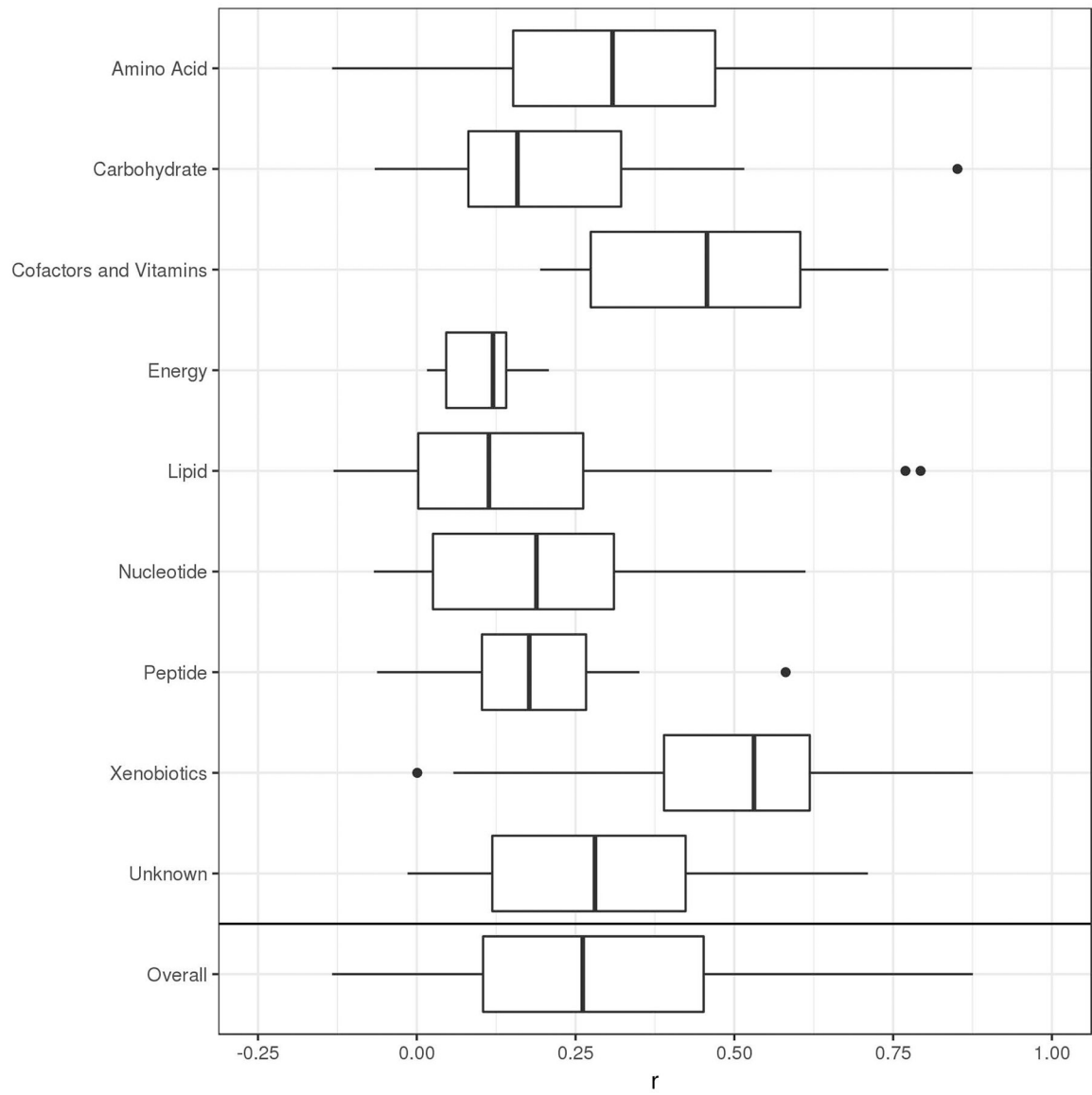
Figure 1.
Correlations between plasma and CSF metabolites by super pathway. Vertical bars represent median correlations; box width represents the first and third quartiles; horizontal bars (whiskers) represent the range of correlations that are within 1.5 times the interquartile range; and dots represent outlier correlations that exceed 1.5 times the interquartile range.
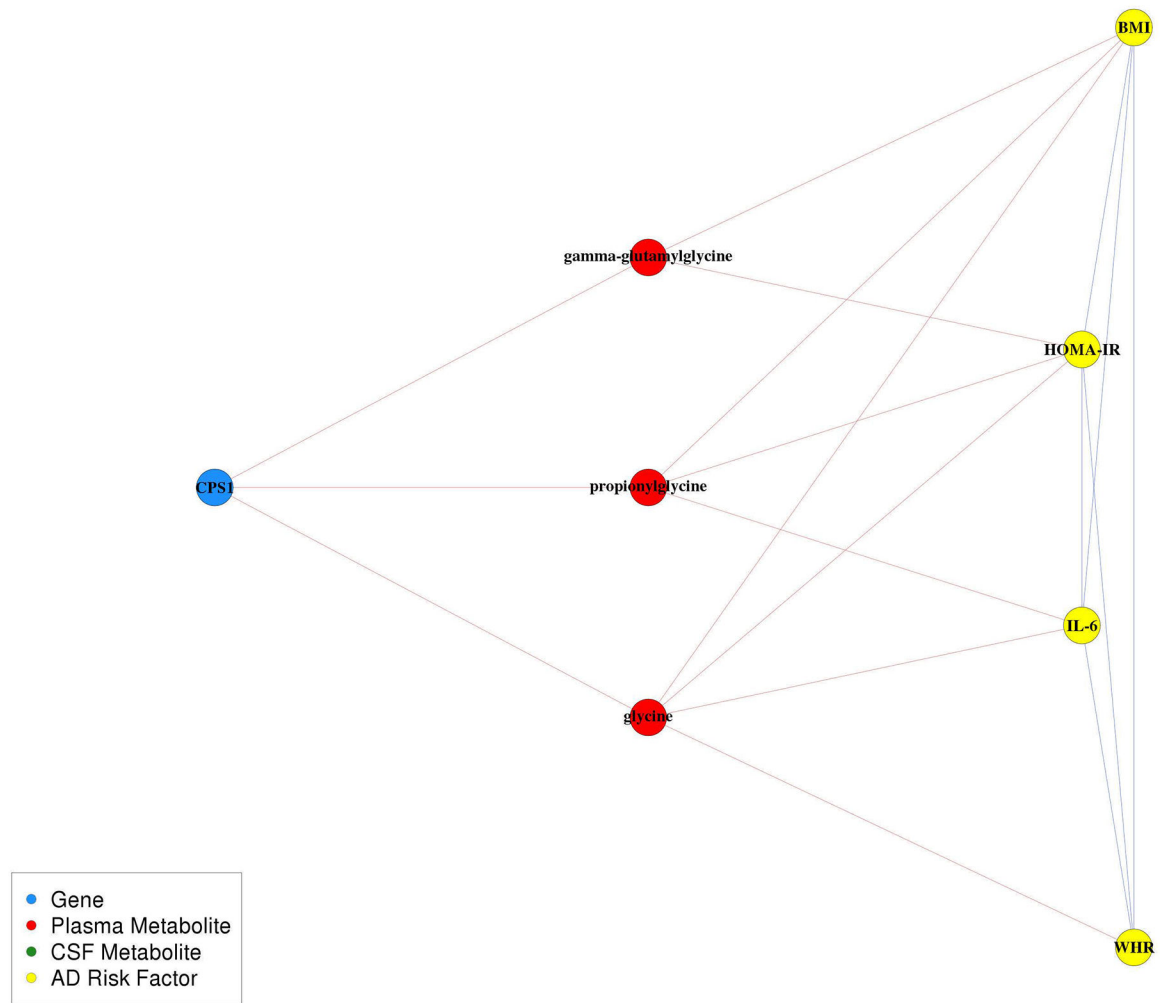
Figure 2.
Inter-omic network. This network has 1,224 edges and 635 nodes, which included 171 metabolite-gene edges, 833 metabolite-AD risk factor edges. Of these, 73 were CSF metabolite-AD risk factor edges (CSF T-tau and P-tau, exclusively) and 4 were CSF metabolite-gene edges. Red edges indicate negative correlations and blue edges indicate positive correlations.

Figure 3.
CPS1, glycine, and cardiovascular and diabetes sub-network. Relationships within this pathway are highly cited; however, the pathway as a whole is not understood as well. Red edges indicate negative correlations and blue edges indicate positive correlations.
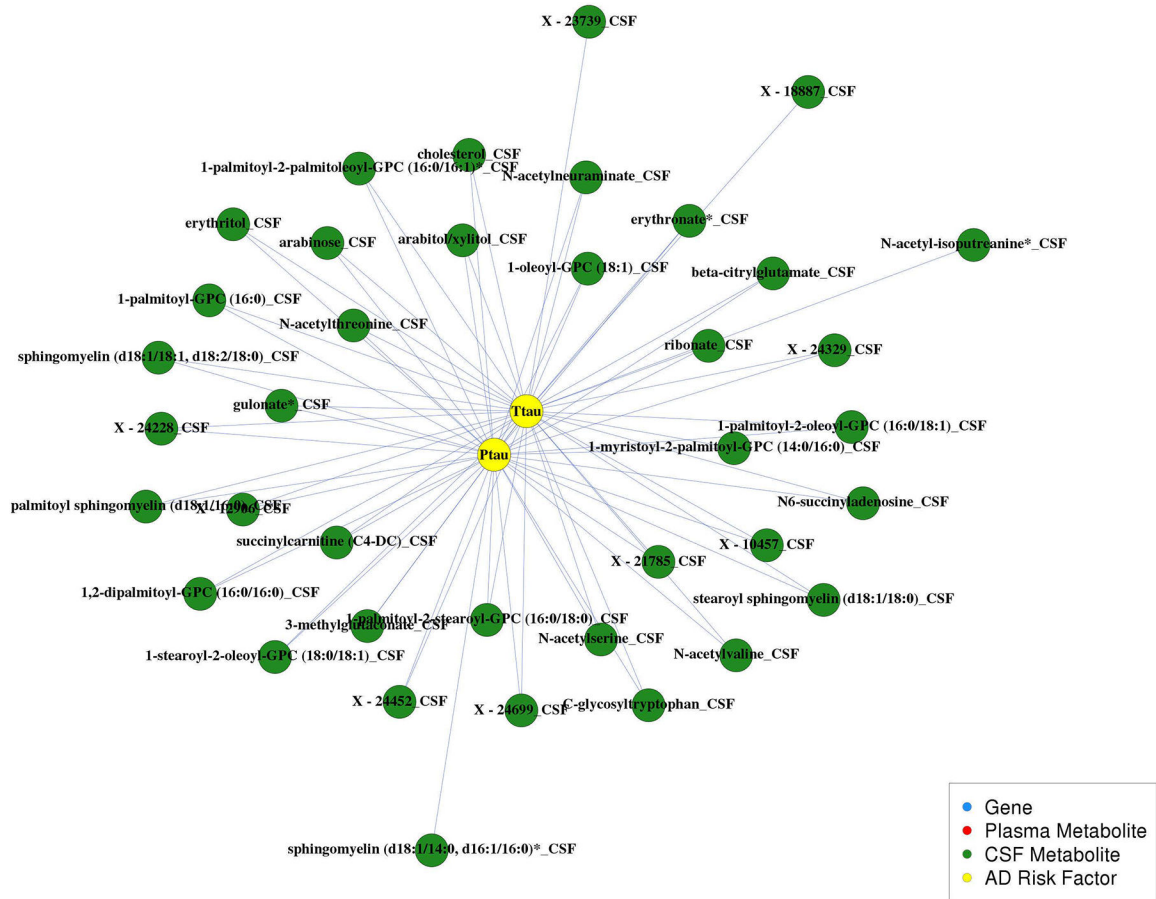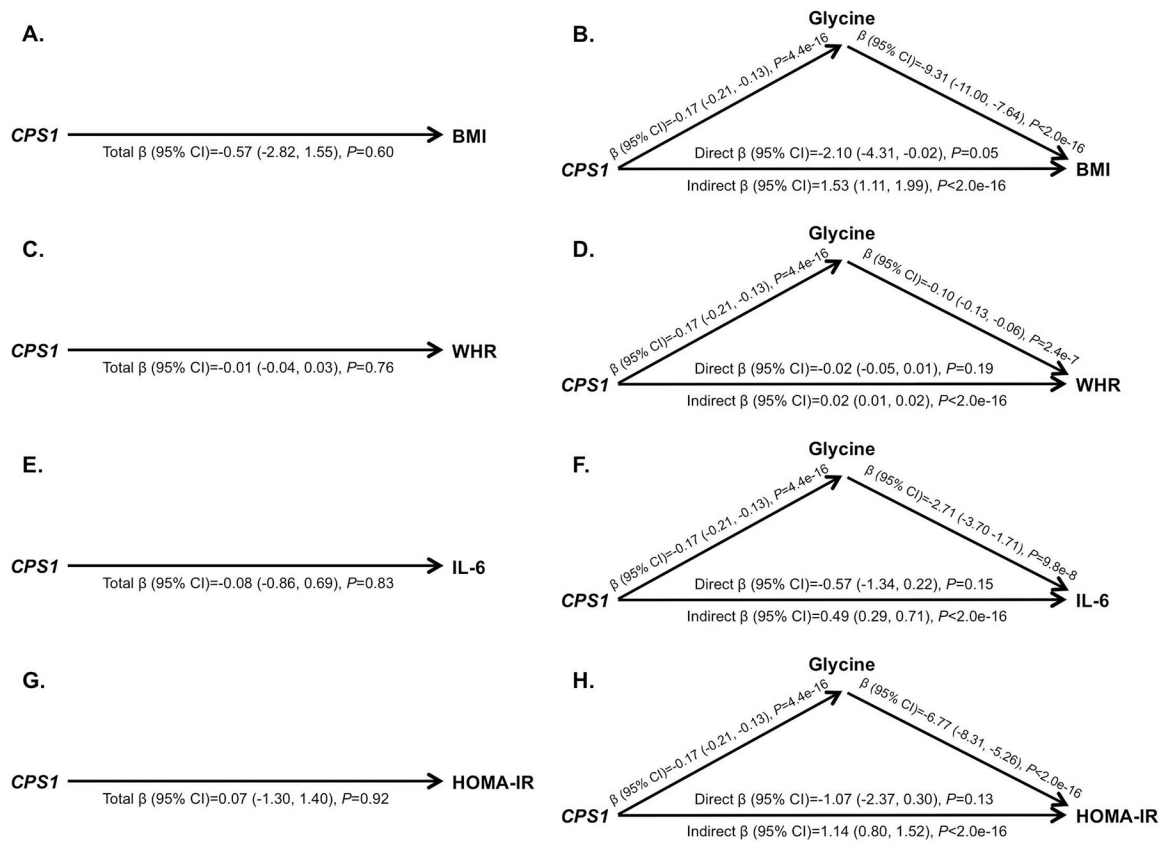
Figure 4.
CSF biomarker community. This network has 73 edges among 38 CSF metabolites and CSF biomarkers T-tau and P-tau. Red edges indicate negative correlations and blue edges indicate positive correlations.

Figure 5.
Mediation analyses to assess whether plasma glycine mediates the relationships between imputed *CPS1* expression, BMI, WHR, IL-6, and HOMA-IR. A. Total effect of *CPS1* on BMI. B. Direct and indirect effects of *CPS1* on BMI. C. Total effect of *CPS1* on WHR. D. Direct and indirect effects of *CPS1* on WHR. E. Total effect of *CPS1* on IL-6. F. Direct and indirect effects of *CPS1* on IL-6. F. Total effect of *CPS1* on HOMA-IR. G. Direct and indirect effects of *CPS1* on HOMA-IR. All models adjusted for age and sex; models including *CPS1* additionally adjusted for the first four PCs; models that included glycine additionally adjusted for cholesterol lowering medication use and sample storage time.

Table 1.

Seventeen AD Risk Factors Included in Network Analysis

Category of Risk Factor	Risk Factor	N
Cognitive	Executive function Composite Score	1,096
	Delayed Recall Composite Score	1,107
	Education	1,111
	Mom's age at memory loss	608
	Dad's age at memory loss	340
Cerebral Spinal Fluid	$A\beta_{42}$	141
	T-tau	141
	P-tau	141
	$A\beta_{42}/A\beta_{40}$	141
Cardiovascular/Diabetic	BMI	1,111
	WHR	1,111
	METs	1,108
	Alcohol use	1,104
	IL-6	1,088
Cardiovascular	SBP	1,111
	DBP	1,111
Diabetic	HOMA-IR	1,107

AD: Alzheimer's disease, $A\beta_{42}$: β-Amyloid42, T-tau: Total-tau, P-tau: Phosphorylated-tau, $A\beta_{40}$: β-Amyloid40, BMI: Body-mass index, WHR: Waist-hip ratio, METs: Metabolic equivalents, IL-6: Interleukin 6, SBP: Systolic blood pressure, DBP: Diastolic blood pressure, HOMA-IR: Homeostatic model assessment of insulin resistance

Alcohol use=(#drinks/day)*(#days/week)

Table 2.

WRAP Participant Characteristics at Baseline Sample. Mean (SD) or N (%).

Characteristic	Overall (N= 1,111, obs=2,191)	CSF Metabolomics (N=155, obs=346)
Age (years)	61.0 (6.7)	61.2 (6.6)
Female	766 (68.9)	103 (66.5)
Years of education	16.4 (2.8)	16.7 (2.9)
Parental history of AD	803 (72.3)	112 (72.3)
Use of cholesterol-lowering medication	354 (31.9)	45 (29.0)
Number of Visits	2.0 (0.6)	2.2 (1.0)

obs: observations, CSF: cerebral spinal fluid, AD: Alzheimer's disease