Dissecting Gene-Microbe-Exercise Interactions in Determining Host Phenotypes

Aurora C. Mattson-Hughes

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2023

Committee:

Sean Gibbons

Amy Willis

Program Authorized to Offer Degree:

Public Health Genetics

University of Washington

Abstract

Dissecting Gene-Microbe-Exercise Interactions in Determining Host Phenotypes

Aurora C. Mattson-Hughes

Chair of the Supervisory Committee:

Sean Gibbons

Department of Bioengineering

Host genetics, gut microbiome, and physical activity are all important factors in determining human health. Clinical chemistries are commonly used in research and medicine to understand, diagnose, or treat disease. Previous metabolomic research has shown that there are gene, microbe, and hybrid associations with blood metabolites. Here, I investigate the possibility that physical activity moderates the relationship between genome-driven or microbiome-driven metabolites genetics and clinical chemistries. Using the Arivale cohort, a large-scale wellness study that includes health and molecular phenotypic data, I examine monthly average distance traveled (determined by FitBit tracking) as an effect modifier in this relationship. Through ordinary least squares linear regression, I found 174 metabolites that are associated with activity, of which 35 were genetic-only, 40 were microbe-only, and 99 were under hybrid control. Of these, 7 metabolites were chosen to undergo moderation analysis which uncovered 5 metabolite-chemistry pairs where activity was a significant effect modifier (p-value < 0.05 after FDR correction). Each unique pair tells a story, which could help explain the underlying mechanisms of physical activity on health.

Background

Physical exercise and a healthful diet are key factors in preserving long-term health and preventing disease in humans. The benefits of exercise are manifold, including improved sleep, a higher quality of life, and lower risk of cancer, heart disease, and diabetes.[1,2] While the benefits of exercise are well known, the underlying mechanisms are yet to be fully understood.

The genome and gut microbiome are two distinct components of the human body that determine how we respond to environmental exposures, dietary inputs, and other lifestyle factors. It is well known that variation at a particular genetic locus, otherwise known as a single nucleotide polymorphism (SNP), can influence disease risk. Examples include SNPs in pulmonary surfactant protein genes, increasing risk of cystic fibrosis[3] and SNPs in the BRCA1 and BRCA2 genes that increase breast cancer risk.[4] In addition, multiple SNPs across many genes can cause an increase in an individual's polygenic risk score for a given disease. Polygenic risk scores have been developed for hypertension, diabetes, coronary heart disease, and stroke.[5–7] Compared to the human genome, less work has been done on quantifying the contributions of the gut microbiome to disease risk, but initial work in this area is highly promising.[8–10]

The human gut microbiome has been called a “hidden metabolic organ” because of its ability to induce changes in physiology, nutrition, metabolism, and immune function.[11] The microbiome consists of hundreds of coexisting microbial species and thousands of unique strains. When the ecological composition of the normal flora is perturbed, diseases can arise. These range from common gastroenterological ailments such as diarrhea or constipation, to more chronic conditions such as irritable bowel syndrome (IBS), obesity, or depression.[12–14]

In a recent study, Diener et al. explored how the genome and gut microbiome impact the host blood metabolome by performing variance partitioning analyses for each blood metabolite detected across a large cohort.[15] As described in their paper, the human metabolome was greatly influenced by gut microbiome composition, and to a lesser extent, by host genetics. Metabolomics can be used to define biomarkers related to disease or drug toxicity/efficacy, so the mechanisms that control them are important to biomedical research. In their study, it was found that 64% of detected metabolites were under host genetic or gut microbiome control, with 69% of these being under microbe only control, 15% under genetic only control, and 16% being under a genetic/microbiome hybrid control.

While one's genes remain – mostly – the same throughout our lifespan[7,16] the gut microbiome develops and changes throughout the course of our lives.[17] Using the Arivale cohort, which includes health and molecular phenotypic data on over 5,000 individuals, I examine how different levels of activity, recorded using FitBits, can impact the relationship between genome- and microbiome-driven blood metabolites, identified by Diener et al., and clinical health markers (i.e., a panel of common clinical chemistries).

Methods

Cohort Description and Selection

The study participants involved in this research were enrolled in the Arivale Scientific Wellness program between July 2015 and March 2018. Each participant provided consent and

authorization in order to have their information used anonymously in further research. Upon entering the program, participants underwent several measurements including fasting blood draws, stool sampling, saliva sampling, genome sequencing, and filling out an extensive lifestyle and health history questionnaire. In addition, participants received a FitBit along with personalized lifestyle coaching according to the participant's health goals. A more thorough description of this cohort can be found in Zubair et. al.[18]

Participants were between the ages of 18 and 87, predominantly female (62.7%), were slightly healthier than the general population by a number of metrics, had an average BMI of 28.4, and most resided in the pacific west of the U.S. and identified as white (80.1%). More information on the participants can be found in Table 1, and ancestry information can be found in Figure 1. Participants were selected for this research if they had paired genome and microbiome data, all matched on the earliest blood draw. This sub-cohort contained 1,738 individuals in total.

FitBit Data

For the purposes of this research, activity was measured as average monthly distance traveled (miles) as recorded by the individual's provided FitBit. The monthly distance data was matched to the earliest blood draw.

Blood Chemistries and Metabolomics

All blood draws were taken after at least 12 hours of fasting, and collected every 6 months. Only the first blood draw was used for this research. All draws were performed in CLIA-approved labs, and the labs provided reference ranges for the markers. Chemistries that were not observed

in at least 75% of samples were removed. Data points were then log-transformed and standardized. More information on the processing of clinical laboratory tests is given in Wilmanski et. al.[17]

Metabolome processing is described in full detail in Diener et. al. In brief, the blood plasma was sent to Metabolon, Inc (USA), and run through their global untargeted metabolomics pipeline (high-pressure liquid chromatography mass spectrometry). Each sample was then split into four aliquots and measured using one of four HPLC methods. Assay batch correction was performed for each batch, using a standard set of Arivale samples that were run in each batch. Metabolites not observed in at least 75% of samples were removed. Data points were then log-transformed and standardized.

Genome and Gut Microbiome Sequencing

The details on host genome and gut microbiome sequencing are also described in full in Diener et. al. To summarize, host DNA extraction was performed using Covance and whole genome sequencing was performed using an Illumina HiSeq. GWAS was performed using FastGWA, and linkage disequilibrium scores were pulled for each variant. Genome-wide significance was determined by the Bonferroni corrected p-value of 5.37×10^{-11}.[15]

For microbiome sequencing, stool samples were collected and preserved using the OMIgene Gut or DNAGenotek at-home swab kits. Sequencing was performed using MiSeq (Illumina) in one of two protocols as described. To classify taxonomy, the naive Bayes classifier in DADA2 with the SILVA database (v.128) was used.

Statistical Analysis

Statistical analysis was performed using both R and Python. To begin, metabolites were put into three bins based on previous research done by Diener et. al: genetically associated, microbially associated, and hybrid. Using Ordinary Least Square (OLS) linear regression in Python, each metabolite was determined to be associated with monthly average distance traveled, controlling for the following covariates: BMI, age, sex, and genetic kinship (the first five principal components of the genetic distance matrix). Two or more metabolites were chosen from each bin to be further analyzed. After selecting, these metabolites were run through another OLS linear regression to find which chemistries were significantly associated with the activity-associated metabolite. Each OLS regression was FDR corrected with a significance level of $p < 0.05$. Once each metabolite had their list of associated chemistries, an effect modification analysis was performed in R (also referred to as a moderation analysis).[19] The statistical model (also termed Model 3) used for this analysis is:

$$Y_i = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \delta c + \varepsilon_i$$

where Y_i denotes clinical chemistry, X denotes blood metabolite, Z denotes average monthly distance traveled, c represents the vector of covariates, and ε_i represents a random normally distributed variable with an expectation of zero.

Effect modification was deemed significant when β_3 was significantly different from zero, with a significance level of $q < 0.05$ after FDR correction. A detailed description of my models used is shown in Figure 2.

Results

To begin my analysis, an association test was performed between metabolites and activity to create a pool of metabolites to choose from for my effect modification analysis (Model 1). In total, there were 35 genetic-only associated metabolites, of which imidazole lactate and tiglylcarnitine were chosen, 40 microbe-only, of which hippurate, isoursodeoxycholate, and 1-arachidoyl-GPC were chosen, and 99 metabolites under hybrid control, of which carotene diol and indolepropionate were chosen. Each metabolite was chosen based on significance level after FDR correction and based on previous knowledge about the function of these metabolites. A full list of significant metabolites for each can be found in Table 2.

Next, another association test was done to find which clinical chemistries were associated with each of the chosen metabolites (Model 2). The list of significant chemistries for each of the chosen metabolites can be found in Table 3. Once statistically significant chemistries were identified, I performed a moderation analysis, using activity as the effect modifier (Model 3). A total of five metabolites, with at least one in each category, showed activity as a significant effect modifier with one clinical chemistry after multiple hypothesis correction (FDR p-value <0.05). The cohort was split into tertiles based on activity, and each chemistry/metabolite pair was plotted against these as seen in Figure 3. There was a negative association between the metabolites hippurate, 1-arachidoyl-GPC, and carotene diol and the clinical chemistry arachidonic acid. For each of these, increased levels of activity lead to a greater negative relationship. Between imidazole lactate and creatinine enzyme, a positive association was found, and increased activity showed a dampening effect on this relationship. Lastly, between

tiglylcarnitine and HDL cholesterol, increased exercise caused the already positive association to become steeper.

No effect modification due to activity levels was found for associations between isoursodeoxycholate or indolepropionate and clinical chemistries.

Discussion

Arachidonic acid vs. Hippurate, Arachidoyl-GPC, Carotene diol

Arachidonic acid (ARA) is an omega-6 polyunsaturated fatty acid that is essential for maintaining physiological homeostasis.[20] ARA is an important component of the cell membrane, and is kept at low levels through reacylation/deacylation cycles. This low level limits the amount of ARA that can be oxidized into other metabolites that can contribute to inflammation.[21] Here, a negative association was found in medium and high activity levels for the three metabolites associated with ARA. For each, the association was nearly absent in individuals with low activity levels.

Hippurate and carotene diol are dietary metabolites that are derived from a diet enriched in fruits and vegetables.[20,22,23] Hippurate has been associated with higher gut microbiome diversity due to this diet, which is generally thought to be a signature of health.[23,24] Carotene diol, a carotenoid, has been implicated in improving cognitive function, ocular health, and cardiovascular health.[25] As shown in the data, as hippurate and carotene diol concentrations increase, arachidonic acid levels decrease, with a greater decrease occurring in individuals who are more physically active. Together, these results indicate that both diet and exercise are important contributors to

maintaining low circulating ARA levels, which likely helps to reduce systemic inflammation in the body.[26,27]

The final metabolite associated with ARA, arachidoyl-GPC, is a product of ARA metabolism. Here, we are seeing that increased activity levels lead to higher breakdown of ARA, which is consistent with our interpretation that maintaining a healthful diet and activity level helps to reduce standing levels of ARA in the bloodstream.

Hippurate and arachidoyl-GPC were exclusively associated with the gut microbiome, while carotene diol was shown to have a hybrid association with both genetics and the microbiota. A full description of associated genera and genes are given in Table 4.

HDL Cholesterol vs. Tiglylcarnitine

HDL cholesterol (i.e., 'good' cholesterol) is a well studied biomarker that is used to predict cardiovascular disease risk.[28]

Tiglyglcarnitine is within the class of metabolites called acylcarnitines, which transport acyl-groups during beta-oxidation within mitochondria. The breakdown of these metabolites is important in regulating lipid metabolism.[28,29] Tiglylcarnitine and HDL were positively associated, with the strongest association observed in the high activity subgroup. While the mechanism remains unclear, this result indicates a possibly beneficial connection between higher circulating acylcarnitine levels, higher plasma HDL, and greater physical activity levels that merits further investigation. Prior work has indicated that higher circulating acylcarninite levels may be a

negative marker of cardiovascular health, which pushes in the opposite direction of the current analysis.[30–32] Overall, these results indicate how context-dependent and paradoxical this complex physiological interplay can be, and how much more remains to be understood.

Creatinine Enzyme Serum vs. Imidazole Lactate

Creatinine is another common clinical chemistry, which is used as a measure of renal function. Creatinine is a waste product of skeletal muscle metabolism, and clinically can be used to diagnose muscular injury and dystrophy. When creatinine levels are high, this could indicate stroke, inflammation of the heart, or impaired kidney function.[1]

Imidazole lactate is a fermentation derivative of histidine, produced by the mammalian enzyme lactate dehydrogenase, and can be found excreted in human urine.[33] In this analysis, the positive association between imidazole lactate and creatinine enzyme is dampened as activity increases. Both imidazole lactate and creatine are signatures of muscle metabolism,[34,35] and their levels in blood are correlated within individuals who show low activity levels, suggesting that activity may be crucial for clearing these potentially detrimental byproducts from circulation.

Both tiglylcarnitine and imidazole lactate were exclusively associated with host genetics and not with gut microbiome composition.

Conclusions

The addition of exercise is essential for a healthy lifestyle. Even so, it is important to understand by what mechanisms activity causes improvements in physiology. In this analysis, I looked at

whether or not activity can moderate the relationship between blood metabolites (known to be associated with genetics, the microbiome, or both) and clinical chemistries. We only explored a handful of the possible associations from this data set, which revealed several informative interactions. Moving forward, I plan to dive into a more systematic analysis of the full list of genetic, microbe, and hybrid associated metabolites, which includes 174 unique blood metabolites. In addition, I plan to explore additional activity features that can be quantified from the FitBit data, including sleep, resting heart rate, or minutes spent in HR exercise zones. I will also look into the limited dietary information available in the questionnaire data. Furthermore, the use of a mediation analysis, using the metabolites as the mediator of the impact of activity on health-related chemistries, could yield interesting results as well. In summary, the integration of activity trackers and deep-phenotyping data is an exciting avenue for better understanding physiology and health, which has potential for great impact.

Appendices

Table 1. Demographics of the selected cohort

Characteristic	N = 1738
Sex (% female)	62.7
Ancestry (% white)	80.1
Mean Age (s.d.)	48.8(11.7)
Mean BMI (s.d.)	28.4(6.67)
Mean Distance (s.d.)	3.79(1.55)
Mean Systolic Blood Pressure (s.d.)	126.0(15.9)
LDL Cholesterol (s.d.)	113.9(34.0)

Figure 1. Principal component analysis (PC1 vs. PC2) of the Arivale cohort genetic ancestry data, based on whole genome sequencing

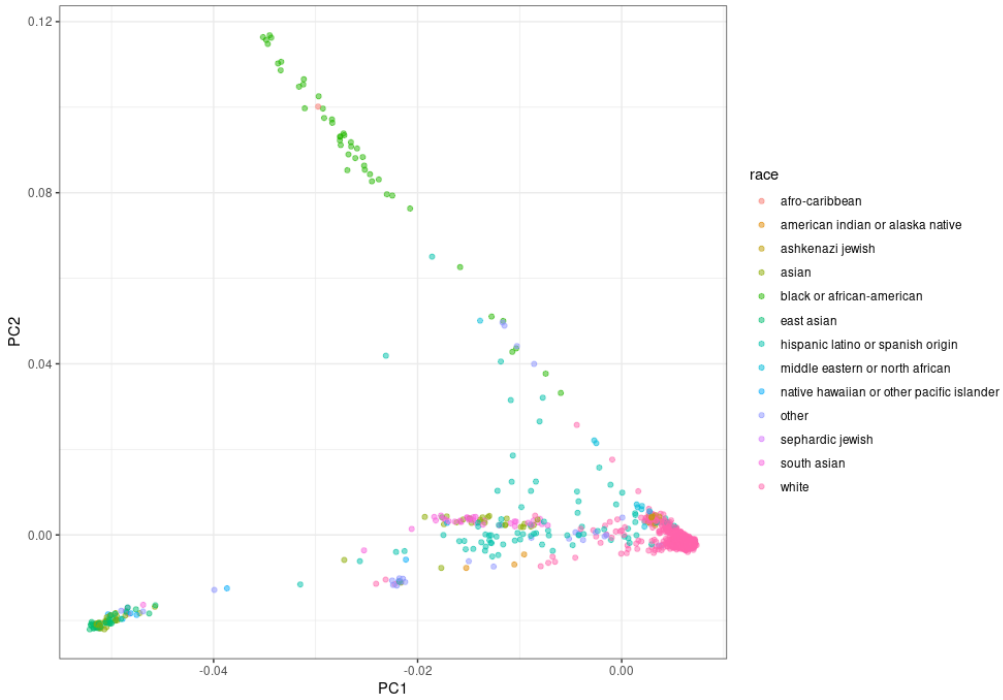

Table 2. List of all metabolites found associated with monthly average distance traveled after performing OLS linear regression and correcting for multiple testing (FDR corrected p-value <0.005)(Model 1). Metabolites are binned by association category

Category	Significant metabolites
Genetic Only	stearoyl sphingomyelin (d18:1/18:0), imidazole lactate, 10-undecenoate (11:1n1), 1-linoleoyl-GPC (18:2), 1-(1-enyl-oleoyl)-GPC (P-18:1)*, 1-(1-enyl-palmitoyl)-GPC (P-16:0)*, 1-methylnicotinamide, N-acetylglutamate, 4-vinylphenol sulfate, tiglylcarnitine (C5:1-DC), 5-hydroxylysine, bilirubin (E,E)*, 3-hydroxyhexanoate, 1-stearoyl-2-linoleoyl-GPI (18:0/18:2), N-acetylglutamine, N1-methyladenosine, N-acetylcitrulline, beta-citrylglutamate, 3-(3-amino-3-carboxypropyl)uridine*, lignoceroylcarnitine (C24)*, 1-oleoyl-GPE (18:1), N-acetyl-2-aminooctanoate*, gamma-glutamylisoleucine*, 2'-O-methyluridine, N-palmitoylglycine, sphingomyelin (d18:1/20:1, d18:2/20:0)*, choline, undecenoylcarnitine (C11:1), 1-eicosapentaenoyl-GPE (20:5)*, 5-acetylamino-6-formylamino-3-methyluracil, 1-stearoyl-2-docosahexaenoyl-GPC (18:0/22:6), 1-linoleoyl-GPI (18:2)*, orotidine, isoleucylglycine, 2-myristoyl-GPC (14:0)*
Microbe Only	1-arachidoyl-GPC (20:0), ceramide (d18:1/17:0, d17:1/18:0)*, 1-eicosenoyl-GPC (20:1)*, X - 23678, N-stearoyl-sphinganine (d18:0/18:0)*, X - 21442, X - 17351, X - 23782, X - 22789, branched chain 14:0 dicarboxylic acid**, guaiacol sulfate, trans-4-hydroxyproline, hippurate, X - 12283, sphingomyelin (d18:2/18:1)*, N-acetyl-2-aminoadipate, X - 21821, ursodeoxycholate, 3-indoxyl sulfate, glycoursodeoxycholate, X - 11847, X - 17010, X - 21607, sphingomyelin (d18:2/24:2)*, beta-cryptoxanthin, 3-hydroxyhippurate, isoursodeoxycholate, phenylacetylcarnitine, 3-hydroxypyridine sulfate, 2-oleoyl-GPE (18:1)*, 1-stearoyl-GPC (18:0), glycodeoxycholate, 1-palmitoyl-GPC (16:0), taurodeoxycholate, X - 12851, laurate (12:0), 2-hydroxydecanoate, gentisate, 12,13-DiHOME, 1-stearyl-2-arachidonoyl-GPC (O-18:0/20:4)*
Genetic and Microbe	carotene diol (2), 2-aminooctanoate, carotene diol (3), phenyllactate (PLA), carotene diol (1), N-stearoyl-sphingosine (d18:1/18:0)*, X - 21258, X - 11315, sphingomyelin (d18:1/18:1, d18:2/18:0), 2-ketocaprylate, cortolone glucuronide (1), proline, thyroxine, indolepropionate, catechol sulfate, methyl glucopyranoside (alpha + beta), sphingomyelin (d18:1/17:0, d17:1/18:0, d19:1/16:0), 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*, X - 24307, trigonelline (N'-methylnicotinate), sphingomyelin (d18:0/18:0, d19:0/17:0)*, 4-allylphenol sulfate, dopamine 3-O-sulfate, 3-carboxy-4-methyl-5-pentyl-2-furanpropionate (3-CMPFP)**, 2-hydroxyglutarate, eicosenoylcarnitine (C20:1)*, oxalate (ethanedioate), X - 23639, 1-carboxyethylleucine, 1-linoleoyl-2-linolenoyl-GPC (18:2/18:3)*, 3-(4-hydroxyphenyl)lactate, ergothioneine, 1-dihomo-linoleoyl-GPC (20:2)*, tartronate (hydroxymalonate), cinnamoylglycine, sphingomyelin (d18:2/23:1)*, tryptophan betaine, 3-phenylpropionate (hydrocinnamate), X - 13866, glycerate, 1,2-dilinoleoyl-GPC (18:2/18:2), methionine sulfone, N-delta-acetylornithine, sphingomyelin (d18:1/20:2, d18:2/20:1, d16:1/22:2)*, X - 21353, 1-carboxyethylphenylalanine, 2-linoleoyl-GPC (18:2)*, 2-methylbutyrylcarnitine (C5), X - 18899, 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*, palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*, threonate, X - 24748, sphingomyelin (d18:1/19:0, d19:1/18:0)*, 2-hydroxyoctanoate, 1-eicosapentaenoyl-GPC (20:5)*, 1-myristoyl-GPC (14:0), 4-ethylphenylsulfate, 1-oleoyl-GPI (18:1), betaine, creatine, 6-bromotryptophan, uridine, palmitoleoyl-linoleoyl-glycerol (16:1/18:2) [1]*, 1-margaroyl-2-arachidonoyl-GPC (17:0/20:4)*, 3-methyl-2-oxovalerate, 1-carboxyethylvaline, gamma-glutamylvaline, X - 22162, 1-stearoyl-2-adrenoyl-GPC (18:0/22:4)*, linoleoyl-arachidonoyl-glycerol (18:2/20:4) [1]*, 1-linoleoyl-2-docosahexaenoyl-GPC (18:2/22:6)*, pyruvate, sphingomyelin (d18:1/22:2, d18:2/22:1, d16:1/24:2)*, nervonoylcarnitine (C24:1)*, S-methylcysteine sulfoxide, citrulline, sphingomyelin (d18:1/25:0, d19:0/24:1, d20:1/23:0, d19:1/24:0)*, N-acetyltaurine, 1-palmitoyl-GPE (16:0), 1-palmityl-2-oleoyl-GPC (O-16:0/18:1)*, ceramide (d18:2/24:1, d18:1/24:2)*, ethyl beta-glucopyranoside, 3-hydroxyoctanoate, deoxycholate, 1-palmityl-2-linoleoyl-GPC (O-16:0/18:2)*, 3-hydroxydecanoate, oleoyl-linoleoyl-glycerol (18:1/18:2) [2], octadecadienedioate, diacylglycerol (16:1/18:2 [2], 16:0/18:3 [1])*, myristoyl-linoleoyl-glycerol (14:0/18:2) [1]*, andro steroid monosulfate C19H28O6S (1)*, sulfate of piperine metabolite C18H21NO3 (1)*, biliverdin, N-acetylleucine, oleoyl-linoleoyl-glycerol (18:1/18:2) [1], p-cresol glucuronide*, 3-hydroxylaurate, (R)-3-hydroxybutyrylcarnitine

Figure 2. Schematic overview of my statistical analysis, where model 1 is the initial metabolite vs. activity OLS regression, model 2 is the chemistry vs. metabolite OLS regression, and model 3 is the modification analysis using activity as an effect modifier. Color indicates the association "bin"

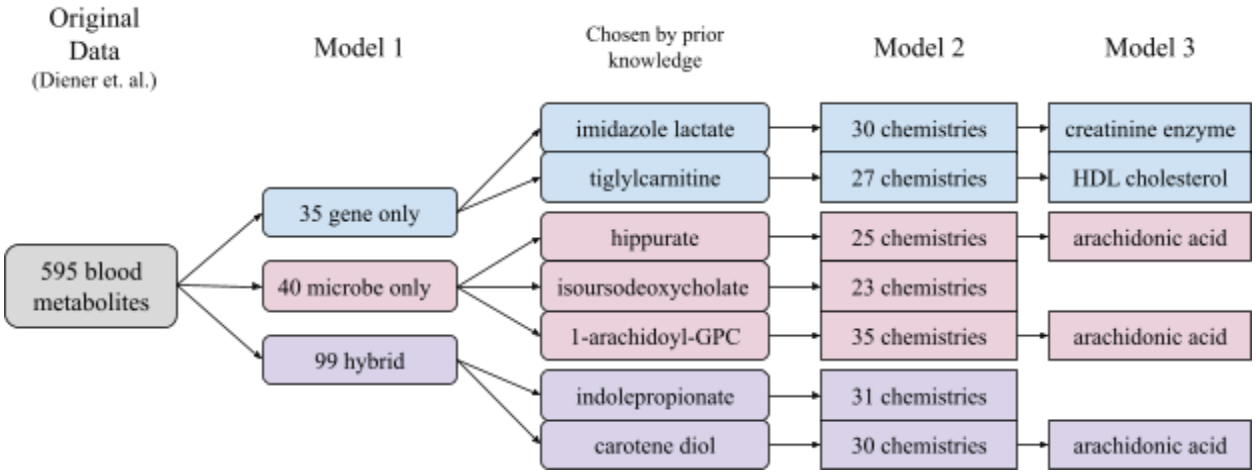

Table 3. List of all clinical chemistries and their associated microbes after performing OLS linear regression and correcting for multiple testing (FDR corrected p-value <0.005)(Model 2)

Metabolite	Significantly Associated Chemistries
Microbe and Genetic	
indolepropionate	ARACHIDONIC ACID, TRIGLYCERIDES, UREA NITROGEN, ADIPONECTIN SERUM, ALAT SGPT, BUN CREATININE RATIO, CARBON DIOXIDE, CHLORIDE, CHOLESTEROL TOTAL, CRP HIGH SENSITIVITY, FERRITIN, GGT, GLUCOSE, GLYCOHEMOGLOBIN A1C, HDL PARTICLE NUMBER, HOMA IR, HOMOCYSTEINE SERUM, LDL PARTICLE NUMBER, LDL SMALL, LDL CHOLESTEROL CALCULATION, LDL SIZE, LINOLEIC ACID, LPIR SCORE, LYMPHOCYTES, MCHC, METHYLMALONIC ACID, MONOCYTES ABSOLUTE, OMEGA 3 INDEX, OMEGA 6 TOTAL, TOTAL NEUTROPHILS ABSOLUTE, TRIGLYCERIDE HDL RATIO
carotene diol	ALKALINE PHOSPHATASE, ARACHIDONIC ACID, GGT, HOMA IR, OMEGA 6 OMEGA 3 RATIO, OMEGA 6 TOTAL, ADIPONECTIN SERUM, ALAT SGPT, CHOLESTEROL TOTAL, CRP HIGH SENSITIVITY, DHA, GLUCOSE, GLYCOHEMOGLOBIN A1C, HDL CHOL DIRECT, HDL PARTICLE NUMBER, INSULIN, LDL PARTICLE NUMBER, LDL SMALL, LDL SIZE, LINOLEIC ACID, LPIR SCORE, LYMPHOCYTES, MCHC, MCV, MONOCYTES ABSOLUTE, OMEGA 3 TOTAL, PLATELET COUNT THOUSAND, SODIUM, TOTAL NEUTROPHILS, TOTAL NEUTROPHILS AB, TRIGLYCERIDES, TRIGLYCERIDE HDL RATIO, VITAMIN D-25-OH-TOTAL, WHITE CELL COUNT
Microbe Only	
hippurate	ADIPONECTIN SERUM, ALAT SGPT, ALBUMIN, ALKALINE PHOSPHATASE, ARACHIDONIC ACID, CALCIUM, CHLORIDE, CREATININE ENZYME SERUM, CRP HIGH SENSITIVITY, FERRITIN, GFR MDRD, GGT, LDL SIZE, LDL SMALL, LINOLEIC ACID, LYMPHOCYTES, MCHC, METHYLMALONIC ACID, MONOCYTES ABSOLUTE, PROTEIN TOTAL SERUM, TOTAL NEUTROPHILS, TOTAL NEUTROPHILS AB, TRIGLYCERIDES, VITAMIN D-25-OH-TOTAL, WHITE CELL COUNT
isoursodeoxycholate	ADIPONECTIN SERUM, ALAT SGPT, ALKALINE PHOSPHATASE, CHOLESTEROL TOTAL, DHA, FERRITIN, GGT, HDL PARTICLE NUMBER, HOMA IR, INSULIN, LDL PARTICLE NUMBER, LDL SMALL, LDL SIZE, LINOLEIC ACID, LPIR SCORE, MCHC, MONOCYTES, OMEGA 6 TOTAL, TOTAL NEUTROPHILS ABSOLUTE, TRIGLYCERIDES, TRIGLYCERIDE HDL RATIO, URIC ACID, WHITE CELL COUNT
1-arachidoyl-GPC	ARACHIDONIC ACID, FERRITIN, GLUCOSE, HOMA IR, INSULIN, OMEGA 6 OMEGA 3 RATIO, ADIPONECTIN SERUM, ALAT SGPT, ALBUMIN, CALCIUM, CHLORIDE, CHOLESTEROL TOTAL, CRP HIGH SENSITIVITY, DHA, GGT, GLYCOHEMOGLOBIN A1C, HDL CHOLESTEROL DIRECT, LDL SMALL, LDL CHOL CALCULATION, LDL SIZE, LINOLEIC ACID, LPIR SCORE, LYMPHOCYTES, MCHC, MONOCYTES ABSOLUTE, OMEGA 3 INDEX, OMEGA 3 TOTAL, OMEGA 6 TOTAL, POTASSIUM, RDW, SODIUM, TOTAL NEUTROPHILS ABSOLUTE, TRIGLYCERIDE HDL RATIO, URIC ACID, WHITE CELL COUNT
Genetic Only	
imidazole lactate	ADIPONECTIN SERUM, ALBUMIN, ALKALINE PHOSPHATASE, ARACHIDONIC ACID, BILI TOTAL, CHOLESTEROL TOTAL, CREATININE ENZYME SERUM, FERRITIN, GFR MDRD, GLUCOSE, GLYCOHEMOGLOBIN A1C, HDL CHOLESTEROL DIRECT, HDL PARTICLE NUMBER, HOMA IR, HOMOCYSTEINE SERUM, INSULIN, LDL SIZE, LDL SMALL, LPIR SCORE, LYMPHOCYTES ABSOLUTE, MCHC, MCV, METHYLMALONIC ACID, MONOCYTES, MONOCYTES ABSOLUTE, PROTEIN TOTAL SERUM, RED CELL COUNT, TRIGLYCERIDE HDL RATIO, UREA NITROGEN, VITAMIN D-25-OH TOTAL
tiglylcarnitine	A G RATIO, ALAT SGPT, ALBUMIN, ALKALINE PHOSPHATASE, ASAT SGOT, BUN CREATININE RATIO, CALCIUM, CREATININE ENZYME SERUM, DHA, FERRITIN, GFR MDRD, GLUCOSE, HDL CHOLESTEROL DIRECT, HDL PARTICLE NUMBER, HEMATOCRIT, LDL SMALL, MCHC, MCV, OMEGA 3 TOTAL, OMEGA 6 OMEGA 3 RATIO, PROTEIN TOTAL SERUM, SODIUM, TRIGLYCERIDE HDL RATIO, TRIGLYCERIDES, UREA NITROGEN, URIC ACID, VITAMIN-D 25-OH TOTAL

Figure 3. OLS linear regression of clinical chemistries vs. metabolites using activity as an effect modifier. Cohort was split into tertiles ranked by activity, colors here represent those groups with blue as the lowest activity and red as the highest. Straight line indicates the regression line

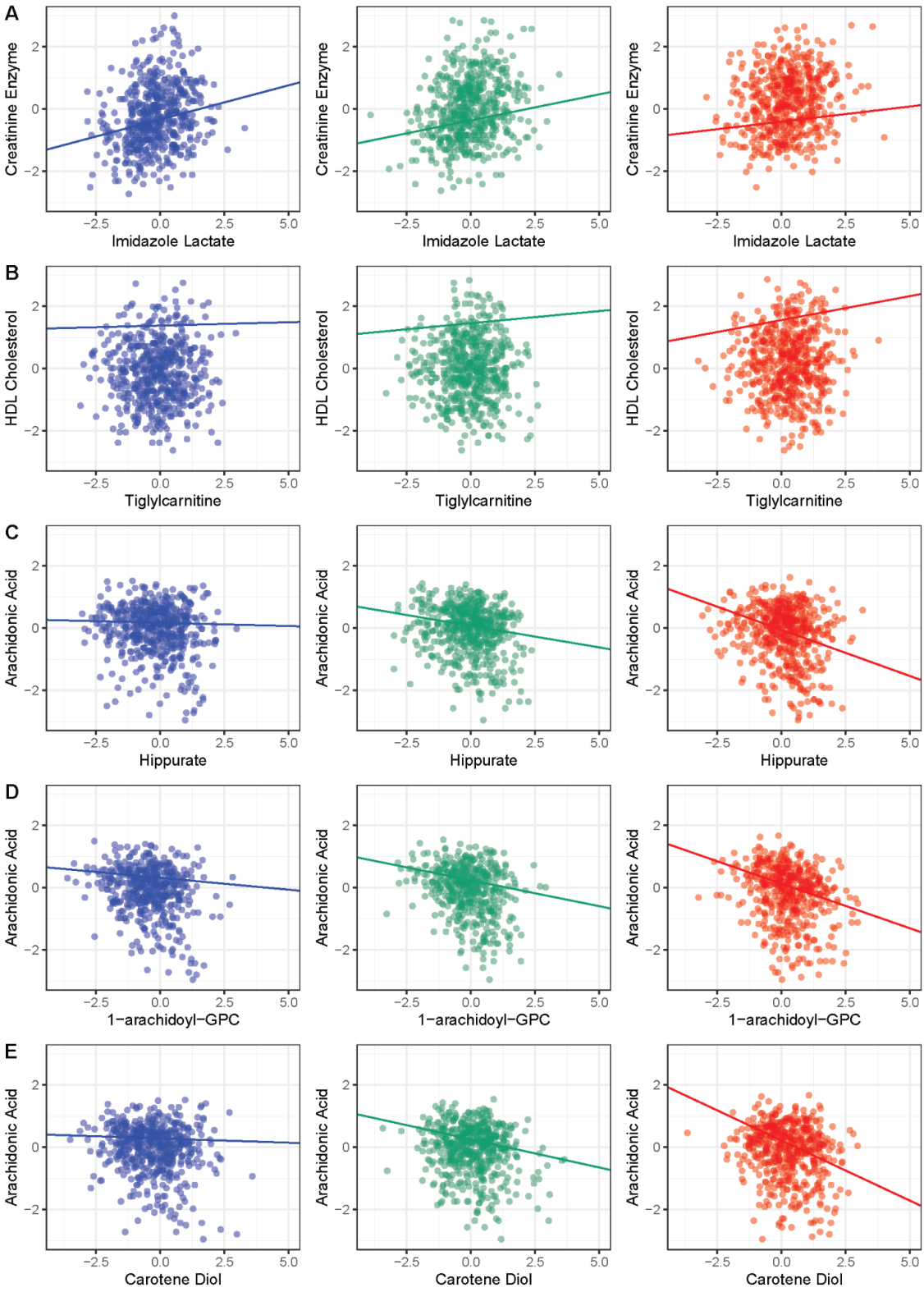

Table 4. List of gut taxa and/or genes associated with each of the five chosen metabolites. These are drawn from previously done work by Diener et. al

Metabolite	Taxa	Gene (rsid)
Microbe and Genetic		
carotene diol	Actinomycetaceae, Bifidobacteriaceae, Eggerthellaceae, Enterobacteriaceae, Erysipelotrichaceae, Lachnospiraceae, Ruminococcaceae, Streptococcaceae	SCARB1 (rs10846744), BCO1 (rs13331438, rs11643509)
indolepropionate	Actinomycetaceae, Bacteroidaceae, Christensenellaceae, Eggerthellaceae, Enterobacteriaceae, Erysipelotrichaceae,	unknown (rs4988235, rs34655000, rs9302387), ACSMA2A (rs6497490)
Genetic Only		
tiglylcarnitine	-	PPM1K (rs9995984), SLC22A5 (rs581968), HSD17B10 (rs1264007, rs201378370)
imidazole lactate	-	KYAT3 (rs2810876), LDHA (rs4150678), SLC6A13 (rs7969761, rs10774020), GOT2 (rs154435, rs66609725, rs11076256), SLC16A3 (rs35121878), HPS5 (rs2305564), CCBL2 (rs1206228892)
Microbe Only		
hippurate	Actinomycetaceae, Bacteroidaceae, Christensenellaceae, Eggerthellaceae, Enterobacteriaceae, Erysipelotrichaceae, Family_XIII, Lachnospiraceae, Ruminococcaceae, Streptococcaceae, Veillonellaceae	-
isoursodeoxycholate	Actinomycetaceae, Bacteroidaceae, Christensenellaceae, Defluviitaleacea, Eggerthellaceae, Enterobacteriaceae, Erysipelotrichaceae, Family_XIII, Lachnospiraceae, Marinifilaceae, Ruminococcaceae, Streptococcaceae, Tannerellaceae, Veillonellaceae	-
1-arachidoyl-GPC	Ruminococcaceae	-

References

1. Bagheri, B., Radmard, N., Faghani-Makrani, A. & Rasouli, M. Serum Creatinine and Occurrence and Severity of Coronary Artery Disease. *Mediev. Archaeol.* **73**, 154–156 (2019).
2. Wang, Y. & Ashokan, K. Physical Exercise: An Overview of Benefits From Psychological Level to Genetics and Beyond. *Front. Physiol.* **12**, 731858 (2021).
3. Lin, Z. *et al.* Genetic Association of Pulmonary Surfactant Protein Genes, SFTPA1, SFTPA2, SFTPB, SFTPC, and SFTPD With Cystic Fibrosis. *Front. Immunol.* **9**, 2256 (2018).
4. Behl, S. *From Clinically Tested to Newly Discovered: Validation of Breast and Ovarian Cancer Predisposing Variants*. (2020).
5. Fava, C. *et al.* Prediction of blood pressure changes over time and incidence of hypertension by a genetic risk score in Swedes. *Hypertension* **61**, 319–326 (2013).
6. Pemmasani, S. K., Atmakuri, S. & Acharya, A. Genome-wide polygenic risk score for type 2 diabetes in Indian population. *Sci. Rep.* **13**, 11568 (2023).
7. Bebo, A. *et al.* Coronary heart disease and ischemic stroke polygenic risk scores and atherosclerotic cardiovascular disease in a diverse, population-based cohort study. *PLoS One* **18**, e0285259 (2023).
8. Fang, P., Kazmi, S. A., Jameson, K. G. & Hsiao, E. Y. The Microbiome as a Modifier of Neurodegenerative Disease Risk. *Cell Host Microbe* **28**, 201–222 (2020).
9. Sanz, Y., Olivares, M., Moya-Pérez, Á. & Agostoni, C. Understanding the role of gut microbiome in metabolic disease risk. *Pediatr. Res.* **77**, 236–244 (2015).
10. Wang, D. D. *et al.* The gut microbiome modulates the protective association between a

Mediterranean diet and cardiometabolic disease risk. *Nat. Med.* **27**, 333–343 (2021).

11. Guinane, C. M. & Cotter, P. D. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therap. Adv. Gastroenterol.* **6**, 295–308 (2013).
12. Schrodt, C. *et al.* The gut microbiome and depression: a review. *Nutr. Neurosci.* 1–7 (2022).
13. Bye, W., Ishaq, N., Bolin, T. D., Duncombe, V. M. & Riordan, S. M. Overgrowth of the indigenous gut microbiome and irritable bowel syndrome. *World J. Gastroenterol.* **20**, 2449–2455 (2014).
14. Leong, K. S. W., Derraik, J. G. B., Hofman, P. L. & Cutfield, W. S. Antibiotics, gut microbiome and obesity. *Clin. Endocrinol.* **88**, 185–200 (2018).
15. Diener, C. *et al.* Genome-microbiome interplay provides insight into the determinants of the human blood metabolome. *Nat Metab* **4**, 1560–1572 (2022).
16. Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* **8**, 1844–1865 (2016).
17. Wilmanski, T. *et al.* Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nat Metab* **3**, 274–286 (2021).
18. Zubair, N. *et al.* Genetic Predisposition Impacts Clinical Changes in a Lifestyle Coaching Program. *Sci. Rep.* **9**, 6805 (2019).
19. Hayes, A. F. *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach*. (Guilford Publications, 2017).
20. Kawashima, H. Intake of arachidonic acid-containing lipids in adult humans: dietary surveys and clinical trials. *Lipids Health Dis.* **18**, 101 (2019).
21. Tallima, H. & El Ridi, R. Arachidonic acid: Physiological roles and potential health benefits

- A review. *J. Advert. Res.* **11**, 33–41 (2018).

22. Lees, H. J., Swann, J. R., Wilson, I. D., Nicholson, J. K. & Holmes, E. Hippurate: the natural history of a mammalian-microbial cometabolite. *J. Proteome Res.* **12**, 1527–1546 (2013).
23. Eggersdorfer, M. & Wyss, A. Carotenoids in human nutrition and health. *Arch. Biochem. Biophys.* **652**, 18–26 (2018).
24. Pallister, T. *et al.* Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. *Sci. Rep.* **7**, 13670 (2017).
25. Rodriguez-Concepcion, M. *et al.* A global perspective on carotenoids: Metabolism, biotechnology, and benefits for nutrition and health. *Prog. Lipid Res.* **70**, 62–93 (2018).
26. Rutting, S. *et al.* Dietary ω-6 polyunsaturated fatty acid arachidonic acid increases inflammation, but inhibits ECM protein expression in COPD. *Respir. Res.* **19**, 211 (2018).
27. DiNicolantonio, J. J. & O'Keefe, J. H. Importance of maintaining a low omega-6/omega-3 ratio for reducing inflammation. *Open Heart* **5**, e000946 (2018).
28. Mineo, C. & Shaul, P. W. Novel biological functions of high-density lipoprotein cholesterol. *Circ. Res.* **111**, 1079–1090 (2012).
29. Li, S., Gao, D. & Jiang, Y. Function, Detection and Alteration of Acylcarnitine Metabolism in Hepatocellular Carcinoma. *Metabolites* **9**, (2019).
30. Zhao, S. *et al.* The Association Between Acylcarnitine Metabolites and Cardiovascular Disease in Chinese Patients With Type 2 Diabetes Mellitus. *Front. Endocrinol.* **11**, 212 (2020).
31. Zheng, D.-M. *et al.* Medium & long-chain acylcarnitine's relation to lipid metabolism as potential predictors for diabetic cardiomyopathy: a metabolomic study. *Lipids Health Dis.*

20, 151 (2021).

32. da Luz, P. L., Favarato, D., Faria-Neto, J. R., Jr, Lemos, P. & Chagas, A. C. P. High ratio of triglycerides to HDL-cholesterol predicts extensive coronary disease. *Clinics* **63**, 427–432 (2008).
33. Coote, J. G. & Hassall, H. The degradation of L-histidine, imidazolyl-L-lactate and imidazolylpropionate by Pseudomonas testosteroni. *Biochem. J* **132**, 409–422 (1973).
34. Emes, A. V. & Hassall, H. The degradation of L-histidine in the rat. The formation of imidazolylpyruvate, imidazolyl-lactate and imidazolylpropionate. *Biochem. J* **136**, 649–658 (1973).
35. Gounden, V., Bhatt, H. & Jialal, I. Renal Function Tests. in *StatPearls* (StatPearls Publishing, 2022).