Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Robert Yardley Eagar Date

Association of BMI and Obesity Related Traits in Relation to the Metabolomic Profile of Blood Plasma

By

Robert Yardley Eagar
Master of Public Health

Epidemiology

_____

Yan V. Sun, PhD
Committee Chair

Association of BMI and Obesity Related Traits in Relation to the Metabolomic Profile of Blood Plasma

By

Robert Yardley Eagar

B.S.
Georgia Institute of Technology
2013

Faculty Thesis Advisor:

Yan V. Sun, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2016

Abstract

Association of BMI and Obesity Related Traits in Relation to the Metabolomic Profile of Blood Plasma
By Robert Yardley Eagar

Background: Approximately 35% of American adults currently suffer from obesity, which is known to cause numerous negative health outcomes, including type 2 diabetes and cardiovascular diseases. Prior studies have reported changes in metabolite profile associated with obesity and overweight. Modern high throughput metabolomic studies have recently increased the ability to detect changes in the metabolite profile, allowing for the garnering of novel associations and verifying changes in amino acid concentrations with overweight and obesity.

Methods: Data utilized in this project was obtained through the Emory Twin Study. Participant data including a medical history and complete physical exam were paired with blood plasma metabolomic data, including over 20,000 metabolomic markers obtained through high-performance liquid chromatography – mass spectrometry. A twin specific linear mixed effects regression model was fit to identify biomarkers associated with BMI and waist-to-hip ratio (WHR). Metabolomic features were annotated via the Metlin database and Mummichog pathway analysis.

Results: Among 92 twin pairs and 3 singletons, WHR was directly associated (Bonferroni corrected p-value <0.05) with three metabolomic features: glutamate, acoric acid, and quinoclamine. The glutathione metabolism pathway was also significantly associated with WHR. BMI had a significant direct association with mevalonic acid (MVA) and an undefined porphyrin and a significant inverse association with 2,3-dinor Prostaglandin E1. Further significant associations with BMI were not conclusively annotated.

Discussion: Metabolites associated with WHR were significant between twin pairs only, suggesting greater significance of genetic factors than environmental factors. The glutathione metabolism pathway, which contains glutamate, contributes to oxidative stress. Acoric acid and quinoclamine are not associated with normal human metabolism and demonstrate the need for a targeted approach to verification of annotation. Of the metabolites significantly associated with BMI, MVA was the only metabolite with a clear metabolic pathway association. MVA, which is known to be directly associated with in vivo cholesterol synthesis, was directly associated with BMI within twin pairs, suggesting environmental effects alter the relationship between MVA and BMI. Future studies should aim to ascertain the identities of significantly associated metabolites and the evaluation of the causal relationship between metabolomic markers and obesity.

Association of BMI and Obesity Related Traits in Relation to the Metabolomic Profile of Blood Plasma

By

Robert Yardley Eagar

B.S.
Georgia Institute of Technology
2013

Faculty Thesis Advisor:

Yan V. Sun, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2016

Table of Contents

Background 1

Methods 8

Data Collection 8

Specific Aims 9

Statistical Methods 9

Results 12

Discussion 14

References 18

Tables 23

Figures 26

List of Tables

Table 1. Phenotypic Characteristics of Twins, ETS 23

Table 2. Significant Metabolomic Features Associated Between Pairs with WHR, Bonferroni p<0.05 23

Table 3. Annotation of Significant Metabolomic Features Associated with WHR 23

Table 4. Significant Metabolomic Features Associated Within Pairs with BMI, Bonferroni p<0.05 24

Table 5. Annotation of Significant Metabolomic Features Associated with BMI 24

Table 6. Cross Reference of Significant Metabolomic Features Associated with WHR and BMI 25

List of Figures

Figure 1. Q-Q Plot of LME Regression of WHR Between Pair Effects for Negative Charge Ions 26

Figure 2. Q-Q Plot of LME Regression of WHR Within Pair Effects for Negative Charge Ions 26

Figure 3. Manhattan Plots of Association Strength by M/Z, WHR 27

Figure 4. Manhattan Plots of Association Strength by M/Z, BMI 28

Figure 5. Mummichog Output Pathway, WHR Between Twin Pairs, Positive Ion 28

Background

Issues associated with obesity have grown drastically in the developed world during the end of the twentieth century and have shown no signs of abatement at the beginning of this century, with an estimated 34.9% of America adults currently suffering from obesity (1). Carriage of excess fat and body weight associated with obesity is known to lead to numerous negative health outcomes, ranging from cardiovascular disease to type 2 diabetes and certain types of cancer (2). Since heart disease, a disease associated with obesity, remains the leading cause of death in the United States with over 611,000 associated mortalities in 2013 (3), identifying potential biomarkers associated with obesity may provide significant insights into targeted pathways for pharmaceutical intervention that may reduce the burden of cardiac illness.

Metabolite biomarkers in the form of small molecules in blood plasma provide an attractive target for early diagnosis without developing clinical symptoms, but for an easily diagnosed illness such as overweight or obesity, metabolite biomarkers provide a more useful insight into the metabolomic pathways of disease etiology. Such biomarkers have already been successfully evaluated as diagnostic tools for various diseases and infections, including lung adenocarcinoma and hepatitis (4, 5). Metabolite biomarkers can provide identification of diseases at an early stage, offering great clinical utility by allowing for an earlier course of treatment (6). Further, small metabolite biomarkers may provide insight as to potential therapeutic targets to allow for a reduction in the overall burden of disease (7).

Findings of early targeted studies have laid the foundation for field of metabolomics. Focusing on diabetic and obese subjects and concentrations of 20 plasma amino acids, Felig et al. demonstrated significant increases in the concentration of valine, isoleucine, tyrosine, and phenylalanine and a reduction in the concentration of glycine in 10 obese subjects paired against age and sex matched controls, further noting that the concentration elevations in obese individuals were directly correlated with serum insulin (8). Weight loss in three obese subjects after therapeutic starvation resulted in a reduction in the concentration of branched chain amino acids valine, leucine, and isoleucine as well as phenylalanine and tyrosine with a concordant increase in concentration of glycine. Of the three subjects, one reverted to original weight and had an accompanying revision of his amino acid profile, supporting the importance in current obesity to metabolite profile.

Modern metabolomic methods have advanced well beyond the measurement of 20 amino acids. Current technologies employed focus on several critical areas in the measurement of low molecular weight compounds: separation, identification, and quantification (9). Separation, after initial removal of high molecular weight compounds through precipitation, centrifugation, and filtration, relies upon chromatographic techniques. A mixture of methods of separation have been deployed in modern studies, including gas chromatography (GC), high performance liquid chromatography (HPLC), ultra-performance liquid chromatography (UPLC), and capillary electrophoresis (CE). Separation methods largely rely on the principles of separation by molecular mass, charge, and charge or dipole moment. As a fundamental rule, larger molecules will be

retained in a chromatography column for longer periods, but differences in specific column and solvent affinity will alter trends in elution (retention) time through a column. Identification and quantification rely on the utilization of spectroscopic techniques. As a standalone technology to quantify large metabolites, nuclear magnetic resonance (NMR), particularly proton NMR (^{1}H NMR), offers the only non-destructive means to analyze metabolomic samples. However, the sensitivity of ^{1}H NMR is relatively poor when used as a broad spectrum analytical tool, showing far more effectiveness at characterizing the structural nature of metabolites in a targeted approach. In higher throughput studies, mass spectrometry (MS) is the analytical tool of choice. By ionizing eluded compounds, measuring time of flight, and imputing signal strength, MS can provide concentration data on a vast wealth of metabolite simultaneously. Modern metabolomics studies employ chromatographic methods with MC to provide high-throughput data that may be annotated via internal standards in targeted studies, or a mixture of internal standards and wider annotation via metabolite databases in non-targeted studies.

One recent review regarding the changes in metabolite profile focused on patients undergoing Roux-en-Y gastric bypass (RYGB), providing further insight into starvation studies that are no longer within the realm of normal physiological conditions. In one study, blood serum samples were analyzed for 10 obese diabetic patients who underwent RYGB prior to surgery and 12 months post-operation (10). Metabolites were measured via ^{1}H NMR and GC-MS. Findings in this study support those of the 1969 Felig et al. study that deployed therapeutic starvation instead of gastric bypass surgery: reductions in the concentration of leucine, isoleucine, valine, and glucose were observed (8). Further,

the RYGB study also observed significant reductions in lactate and contrasting increases in phosphatidylcholine. These findings reiterate the potential for changes in obesity status to alter the metabolomic profile of subjects.

Larger targeted studies have provided new insights. Participants in the KORA (Cooperative Health Research in the Region of Augsburg) study were studied to ascertain the association of fate free mas index (FFMI) with serum metabolites (11). Metabolite profiles were obtained from 965 participants in the KORA S4 study and 890 weight-stable subjects in the follow-up KORA F4 study. Targeted metabolites consisted of 190 serum metabolites, including amino acids, acylcarnities, phosphatidylcholines, sphingomyelins, and hexose. FFMI was directly associated with higher concentrations of branched chain amino acids. Higher ratios of branched chain amino acids were detected with respect to glucogenic amino acids with increased FFMI. Phosphatidylcholines also decreased in chain length / saturation with increasing FFMI, but these correlations did not exist among obese subjects, likely due to significant changes in skeletal muscle metabolism.

Findings surrounding phosphatidylcholines have extended to other studies as well. A study authored by Wahl, et al. evaluated 80 obese and 40 normal weight children, ages 6-15 years old, in a targeted study of 163 metabolites by liquid chromatography – mass spectrometry (LC/MS) (12). Obesity was found to decrease concentration of lysophosphatidylcholines, associated with proinflammatory and proatherogenic conditions; glutamine, methionine, and proline. Increases were noted for acylcarnities,

which are involved in fatty acid oxidation and organic acid metabolism. Opposite findings regarding the association between concentration of lysophosphatidylcholines and obesity were documented in a non-targeted study of 14 health MZ twins in an untargeted LC/MS study, which also reported decreases in ether phospholipids, known for their antioxidant properties (13).

While the data garnered from targeted metabolomic studies provides easily quantifiable concentrations of select metabolites that can be compared to phenotypic traits, there are limits to a targeted approach. Targeted studies must identify metabolites of interest prior to the collection of data. Such selection can be clinically justified from known associations, but expanding current knowledge of metabolic correlations requires identifying novel metabolite associations. In this regard, untargeted studies, which review the entire metabolic profile of a subject within the limitations of chromatographic and spectroscopic techniques, provide invaluable insight into metabolomics. Untargeted studies rely on determining the identity of metabolites correlated with phenotypic traits of interest via select internal standards and annotation via pathway analysis and online databases of metabolites. Overall, targeted metabolomic studies currently provide more accurate data without concern of misidentification, while untargeted studies can provide a significantly larger breadth of information at the potential cost of acuity. Advancements in the field will eventually see the combination targeted and untargeted approaches to improve accuracy of detection and continue to increase the number of metabolites analyzed (14).

As such, untargeted metabolomic studies have succeeded in providing insight regarding the etiology of disease. These findings are particularly notable with type 2 diabetes, which is the most prevalent disease burden associated with overweight and obesity (15). In an untargeted study of 447 fasting plasma metabolites, 42 metabolic byproducts from carbohydrates, lipids, and proteins has been found to be significantly associated with type 2 diabetes (16). While numerous diabetes associated metabolites are common with obesity, mouse model studies have demonstrated differences in biomarkers between the exclusively obese and those also affected by diabetes (17). In an effort to address early childhood obesity risk, cord blood metabolites have been evaluated to identify differences in children with rapid postnatal weigh gain (18). Among 16 altered biomarkers, reduced levels of tryptophan associated metabolites, serotonin, tryptophan betaine, and tryptophyl leucine were present in children in the top quartile of postnatal weigh gain and mid-childhood BMI exceeding the 85th percentile.

Combining study participants from three studies in the United States and China, non-targeted metabolomic assays of 317 metabolites in blood samples from 947 participants have yielded confirmatory evidence of BMI associations with metabolites as well as several novel metabolite correlations (19). Non-targeted metabolomic data was collected using LC-MS and GC-MS before comparison to a chemical reference library of 2,500 standards. Significant associations were found for a total of 37 metabolites with BMI, including 12 amino acids, 19 lipids, and 6 other compounds. Of note, 18 new metabolite BMI associations were identified, including seven that were highly correlated with the amino acids valine, tyrosine, phenylalanine, leucine, or isoleucine. These findings

suggest an undirected metabolomic study among adults will yield obesity related biomarkers that may be directed towards better understanding the pathophysiology of obesity, and that continued evaluation of the association of obesity related traits and aforementioned amino acids is of specific interest to this study.

A recent twin study utilizing ^{1}H NMR to analyze the concentration of 56 metabolites among 286 twins (20). Further, this study utilized between twin and within twin modeling in an effort to differentiate between environmental and genetic influence in metabolomic profile. Waist circumference was found to be positively correlated with branched chain amino acids, phenylalanine, tyrosine, alanine, and pyruvate, while glycine and citrate were found to be negatively correlated. Environmental and genetic contributions to changes in phenotypic association with metabolite concentration were found to be largely similar, with the exemption of phenylalanine, which was found to be more strongly associated with genetic factors than environmental factors. This ability to differentiate between environmental and genetic drivers behind phenotypic associations with metabolite concentrations the hallmark benefit of twin studies. This project utilizes a twin modeling approach to isolate changes in metabolite profiles between twin pairs, which carry genetic implications, and within twin pairs, which carry environmental and exposure implications due to shared genetic background.

Methods

Data Collection

This project utilized data collected as part of the Emory Twin Study (ETS) of cardiovascular disease, which recruited 307 male monozygotic and dizygotic twin pairs from the Vietnam Era Twin (VET) Registry. All participants in the VET Registry were born between 1946 and 1956 and served in the U.S. military during the time of the Vietnam War (1964-1975). Details regarding the construction of the Registry are well documented (21). All participants in the ETS signed an informed consent and the study was approved by the Emory Institutional Review Board.

A medical history was obtained from all twins, and a complete physical exam was performed, including measures of height, waist circumference, and hip circumference (22). Metabolomic data was collected from the fasting blood samples of participants in this cohort, including over 20,000 metabolomic markers. Blood plasma metabolites were evaluated for relative abundance using high-performance liquid chromatography – mass spectrometry (HPLC-MS) (23). Processed metabolomic data was evaluated for 92 twin pairs and 3 singletons.

Blood plasma aliquots were treated with acetonitrile to precipitate protein and an internal standard before being centrifuged at 13,000 g for 10 minutes at 4°C. Supernatant was transferred to autosampler vials. Anion exchange (AE) columns were equilibrated to initial conditions for 1.5 minutes before sample injection. MS data were collected from a Thermo LTQ-Velos Orbitrap mass spectrometer (Thermo Fisher, San Diego,CA). A 10

minute gradient was utilized to collect data within the mass/charge ratios (*m/z*) of 85-2000 in positive ionization mode. Using a dual-column procedure, three technical replicates were run for each sample. Adaptive processing software (apLCMS) was used to process LC-MS data to align and extract peaks and perform quantification of metabolites (24). Data filtering, normalization, diagnostics, and summarization were performed using the computer package MSPrep (25). Missing data were imputed from half the minimum observed value for each metabolite across all samples and raw abundance values were log transformed. Batch effect was corrected using the ComBat algorithm within MSPrep (26).

Specific Aims

The objective of this project was to determine the association of concentrations of small metabolites in blood plasma with the obesity related factors of BMI and waist-to-hip ratio (WHR). This study will utilized linear mixed effects regression to evaluate metabolic differences in blood plasma associated with obesity (27). With the goal of identifying biomarkers that may be utilized as risk predictors for obesity, BMI and WHR were treated as dependent variables in this study.

Statistical Methods

The general form of the multiple regression model fitted is $E(Y_{ij}) = \beta_0 + \beta_C Z + \beta_w(X_{ij} - \bar{X}_i) + \beta_B \bar{X}_i$ (27). In this study, Y_{ij} represented metabolite concentration, X_{ij} represented obesity related factor level, and $\bar{X}_i$ represented the mean value of $X$for twin pair i. Z represented the covariate matrix. β_0 represented the intercept. β_C represented

the correlation coefficient of covariates . Current smoking and age were controlled for in the linear regression model. The within-pair coefficient β_w represented the expected change in Y for a unit change in the difference between the individual X and mean value of $X$for the twin-pair. This estimate provided for an understanding for the difference in outcome that may be explained by the difference in X between co-twins, which can be explained by unshared environmental factors (the exposome). The between-pair coefficient β_B gave the expected change in Y for a unit change in mean value of X, controlling for individual difference from the average. This estimate demonstrated the difference in outcome that is explained by between-pair difference in X, or the population level effect of common genetic, maternal, and shared environmental effects.

All variables except smoking status, which was coded in a binary fashion (current vs non-current smoker), were treated as continuous in this model. Normal distributions were verified for all continuous phenotypic variables by measure of skewness and visual inspection of distribution. Predicted mean for each twin pair and predicted difference for each individual was calculated after a random split of subject IDs. Phenotypic variable distributions were further verified for normal distribution after splitting. Metabolite concentrations were measured by log transformed peak intensity.

A Bonferroni-corrected p value of 0.05 as applied to adjust for multiple testing of 12,527 ($p < 4.09\times10^{-6}$) negatively charged ion features and 7,508 positively charged ion features ($p < 6.83\times10^{-6}$) (28). Metabolite identities were determined by referencing against the Metlin database with a 10ppm tolerance, with positively and negatively charged adducts

selected for respective features. Pathway and network analysis was performed using Mummichog, a software product that bypasses traditional identification before network analysis in favor of utilizing network enrichment algorithms to identify metabolites (29). All features with a significant raw association ($p < 0.05$) with BMI and WHR were selected for input to Mummichog. The respective MS mode was selected for each batch of input features, and a 10ppm tolerance for instrumentation was selected. All remaining options were allowed to remain as defaults.

All statistical analyses were performed in the R statistical environment version 3.2.3 (http://www.r-project.org/).

Results

The average participants in the ETS for whom metabolomic data was collected and analyzed was 55.84 (3.25) years old at the time of data collection (**Table 1**). 31.6% of participants were current smokers, while the average participant had a BMI of 29.48 (4.69) and a WHR of 0.95 (0.06). Age, BMI, and WHR were found to be approximately normally distributed by measures of skew and visual analysis of scatter plots. Participants were further randomly split into two groups for statistical modeling. The phenotypic characteristics of each randomly split group were very similar to the larger study cohort, with the exception of a higher proportion of current smokers in split 1. Since current smoking was a binary control variable, this difference was not of concern to modeling.

Q-Q plots were created for each statistical model to evaluate global inflation (see sample **Figure *1*** and **Figure *2***). No extremely abnormal tail deviance was observed in Q-Q plots, which supported the validity of the LME model utilized in this project. The inflation factor varied from 0.89 to 1.04 for each regression, representing moderate deflation to minimal inflation. The lack of significant global inflation further supported the use of the LME model.

WHR was significantly associated with three metabolite features at a Bonferroni corrected p-value <0.05 (**Table 2**). Manhattan plots were constructed for each association and ion charge (**Figure *3***). Annotation of these metabolomic features was performed using the Metlin Database (**Table *3***). Concentrations of glutamate, acoric

acid, and quinoclamine were found to be directly associated with WHR. All significant associations were between twin pairs. These metabolic features were not significantly associated with BMI (Bonferroni p-value ~1) (**Table 6**). Analysis via Mummichog indicated the glutathione metabolism pathway was significantly associated (p-value 7×10^{-5}) among positive ion adducts with WHR (**Figure *5***). Further pathway analysis did not reveal clear associations between other metabolic pathways and WHR.

BMI was significantly associated with six metabolite features at a Bonferroni corrected p-value <0.05 (**Table *4***). Manhattan plots were constructed for each association and ion charge (**Figure *4***). These metabolic features were not significantly associated with WHR (Bonferroni p-value ~1) (**Table 6**). Annotation of these metabolomic features was performed using the Metlin Database (**Table *5***). All significant features were within twin pairs. Concentrations of mevalonic acid and an undefined porphyrin were found to be directly associated with BMI, but precise identification of the porphyrin is not possible from MS data due to numerous porphyrins sharing the same molecular formula but presenting different structural forms. Concentrations of 2,3-dinor Prostaglandin E1 was inversely associated with BMI. Annotation of the metabolite with m/z 260.9158 as thallium is demarcated by being the only ±10ppm adduct in the Metlin database, while the metabolite with m/z 411.2385 has a 0ppm delta match with a grayanotoxin I adduct and a 2ppm delta match with a variety of peptide fragments sharing the same molecular formula bur presenting different structural forms. Analysis via Mummichog did not yield clear metabolic pathway associations with BMI.

Discussion

The three metabolites that were found to be significantly associated with WHR were found to have only a significant association between twin pairs. This finding suggests that variations in glutamate, acoric acid, and quinoclamine with WHR are controlled more by genetic factors than by environmental factors or exposures. Glutamate further falls on the glutathione metabolism pathway. Deficiency in glutathione contributes to oxidative stress, which plays a significant role in aging and the etiology of both chronic and infectious disease(30). The direct association of WHR with glutamate is, however, not a de facto indicator of a protective effect of increased WHR against oxidative stress, as increase concentrations of glutamate may be indicative of controls on the citric acid cycle where the transamination of α-ketoglutarate gives glutamate, which can be used in glycolysis and glycogenesis processes. Acoric acid and quinoclamine have not been previously associated with biological pathways. Acoric acid is a naturally occurring analogue of vasopressin, but has no normal function in human metabolism. It is a documented hemostatic, but is not usually ingested (31). Quinoclamine is a commonly used herbicide in liverwort control, but literature does not demonstrate any associations between quinoclamine and overweight (32). Both acoric acid and quinoclamine could thus by false annotations. Further review through the use of targeted standards would be suitable to verify these associations with WHR.

The limitations of annotation of untargeted metabolomics data is also relevant to finding of metabolites associated with BMI. Of the seven statistically significant associations noted, all of which were within twin pairs, only mevalonic acid has a clear association

with a metabolic pathway. Three annotations were extremely suspect or limited in value by nature. M/z +619.2557 is likely an undefined porphyrin, but MS is limited in that it can only deliver a mass charge ratio and relative abundance. Differentiating between different structural forms of metabolites would require the utilization of additional technologies, such as IR spectroscopy and ^{1}H NMR, which were not applied in this study. Similarly, m/z -411.2385 was annotated as either an undefined peptide, due to numerous peptide chains sharing the same molecular formula as its adduct. However, this feature may alternatively correspond to grayanotoxin I, a rhododendron born poison associated with mad honey, but the data are insufficient to support a conclusion that humors the presence of a rare neurotoxin in measurable concentration in numerous study participants (33). The annotation of m/z -260.9158 is suspect for its larger 8ppm delta from a known adduct in the Metlin database and the further lack of metabolic justification for the presence of measurable quantities of thallium in blood serum of study participants.

Mevalonic acid (MVA) was found to have a direct correlation with BMI within twin pairs, suggesting an environmental effect modifies the relationship between MVA and BMI. Early studies into the relationship between MVA concentration in blood plasma focused on its mechanism in cholesterol control. MVA, an obligate precursor of cholesterol, was found to vary significantly in concentration throughout the day and with diet (34). MVA has been previously documented as being directly associated with obesity as a marker of increased in vivo cholesterol synthesis, but the role of environmental factors in altering MVA concentration has not been reported (35). However, elevated concentrations in overweight may be of critical importance in

understanding the etiology of atherosclerosis. Several placebo controlled randomized prospective trials have demonstrated that lowering plasma cholesterol does not lead to reduced risk of cardiovascular disease, raising the "mevalonate hypothesis" (36). The mevalonate hypothesis posits the stimulation of endothelial cells by the MVA pathway results in the production of both cholesterol and free radicals, but that only the latter are lead to arthrosclerosis through the production of oxidized cholesterol. Further analysis that controls for use of statin drugs should be employed in the analysis of this pathway, as MVA may provide a future target to protect against atherosclerosis. Additionally, untargeted pathway analysis of MVA utilizing raw feature abundance may yield additional findings that are masked due to the relative insignificant association of relevant pathway metabolites with obesity and overweight.

This project utilized a cross-sectional analysis of the metabolomic profile associated with obesity related traits. The findings of this study are thus limited to identifying metabolomic features that are correlated with obesity, but cannot provide insight as to the causal relationship between changes in serum metabolite profile and obesity. Future follow-up studies that track the metabolomic profile of individuals an extended number of years may offer additional insights regarding how increased weight alters metabolite profile and may provide metabolite markers that warn of an increased likelihood to become obese or overweight. Such precursors of obesity and overweight may be useful in clinical prevention, while changes in metabolite profile due to obesity may be suitable for further metabolic pathway analysis to understand the effects of obesity on the body and provide insights as to possible pharmacologic targets.

The annotation of high-throughput untargeted metabolomics data has improved substantially with the availability of large reference databases. However, the high-resolution assays detect a large number of features not being uniquely mapped to known chemicals. The untargeted approach utilized in this study yielded eight statistically significant features associated with the obesity related traits of WHR and BMI. However, reliable annotation and pathway analysis yielded viable annotation for only a quarter of these significant features, highlighting the ongoing difficulties of analyzing LC-MS data without combining further analytical techniques, such as the use of additional targeted standards or pairing LC-MS with NMR for structural analysis (14). While the employment of additional analytical tools will resolve several of the issues faced by this study, some associations will remain difficult to resolve, such as determining associations of peptide features, due to the numerous sequences that have similar molecular weights, but future studies may aim to quantify the association of general protein synthesis pathways with obesity related traits as well as replicate the findings of significance of the peptide adducts identified in this study.

References

1. Ogden CL, Carroll MD, Kit BK, et al. Prevalence of childhood and adult obesity in the United States, 2011-2012. *JAMA* 2014;311(8):806-14.
2. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults--The Evidence Report. National Institutes of Health. *Obes Res* 1998;6 Suppl 2:51S-209S.
3. Health, United States, 2014: With Special Feature on Adults Aged 55-64. National Center for Health Statistics, 2015.
4. Fahrmann JF, Kim K, DeFelice BC, et al. Investigation of metabolomic blood biomarkers for detection of adenocarcinoma lung cancer. *Cancer Epidemiol Biomarkers Prev* 2015;24(11):1716-23.
5. Lian JS, Liu W, Hao SR, et al. A serum metabolomic analysis for diagnosis and biomarker discovery of primary biliary cirrhosis and autoimmune hepatitis. *Hepatobiliary Pancreat Dis Int* 2015;14(4):413-21.
6. Cheng ML, Wang CH, Shiao MS, et al. Metabolic disturbances identified in plasma are associated with outcomes in patients with heart failure: diagnostic and prognostic value of metabolomics. *J Am Coll Cardiol* 2015;65(15):1509-20.
7. Corona G, Rizzolio F, Giordano A, et al. Pharmaco-metabolomics: an emerging "omics" tool for the personalization of anticancer treatments and identification of new valuable therapeutic targets. *J Cell Physiol* 2012;227(7):2827-31.
8. Felig P, Marliss E, Cahill GF, Jr. Plasma amino acid levels and insulin secretion in obesity. *N Engl J Med* 1969;281(15):811-6.

9. Zhang A, Sun H, Wang P, et al. Modern analytical techniques in metabolomics analysis. *Analyst* 2012;137(2):293-300.

10. Lopes TI, Geloneze B, Pareja JC, et al. Blood Metabolome Changes Before and After Bariatric Surgery: A (1)H NMR-Based Clinical Investigation. *OMICS* 2015;19(5):318-27.

11. Jourdan C, Petersen AK, Gieger C, et al. Body fat free mass is associated with the serum metabolite profile in a population-based study. *PLoS One* 2012;7(6):e40009.

12. Wahl S, Yu Z, Kleber M, et al. Childhood obesity is associated with changes in the serum metabolite profile. *Obes Facts* 2012;5(5):660-70.

13. Pietilainen KH, Sysi-Aho M, Rissanen A, et al. Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects--a monozygotic twin study. *PLoS One* 2007;2(2):e218.

14. Cajka T, Fiehn O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal Chem* 2016;88(1):524-45.

15. Must A, McKeown NM. The Disease Burden Associated with Overweight and Obesity. In: De Groot LJ, Beck-Peccoz P, Chrousos G, et al., eds. *Endotext*. South Dartmouth (MA), 2000.

16. Menni C, Fauman E, Erte I, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* 2013;62(12):4270-6.

17. Giesbertz P, Padberg I, Rein D, et al. Metabolite profiling in plasma and tissues of ob/ob and db/db mice identifies novel markers of obesity and type 2 diabetes. *Diabetologia* 2015;58(9):2133-43.

18. Isganaitis E, Rifas-Shiman SL, Oken E, et al. Associations of cord blood metabolites with early childhood obesity risk. *Int J Obes (Lond)* 2015;39(7):1041-8.

19. Moore SC, Matthews CE, Sampson JN, et al. Human metabolic correlates of body mass index. *Metabolomics* 2014;10(2):259-69.

20. Bogl LH, Kaye SM, Ramo JT, et al. Abdominal obesity and circulating metabolites: A twin study approach. *Metabolism* 2016;65(3):111-21.

21. Tsai M, Mori AM, Forsberg CW, et al. The Vietnam Era Twin Registry: a quarter century of progress. *Twin Res Hum Genet* 2013;16(1):429-36.

22. Vaccarino V, Khan D, Votaw J, et al. Inflammation is related to coronary flow reserve detected by positron emission tomography in asymptomatic male twins. *J Am Coll Cardiol* 2011;57(11):1271-9.

23. Zhao J, Zhu Y, Hyun N, et al. Novel metabolic markers for the risk of diabetes development in American Indians. *Diabetes Care* 2015;38(2):220-7.

24. Yu T, Park Y, Johnson JM, et al. apLCMS--adaptive processing of high-resolution LC/MS data. *Bioinformatics* 2009;25(15):1930-6.

25. Hughes G, Cruickshank-Quinn C, Reisdorph R, et al. MSPrep--summarization, normalization and diagnostics for processing of mass spectrometry-based metabolomic data. *Bioinformatics* 2014;30(1):133-4.

26. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8(1):118-27.

27. Carlin JB, Gurrin LC, Sterne JA, et al. Regression models for twin studies: a critical review. *Int J Epidemiol* 2005;34(5):1089-99.

28. Benjamini Y, Yekutieli D. Quantitative trait Loci analysis using the false discovery rate. *Genetics* 2005;171(2):783-90.

29. Li S, Park Y, Duraisingham S, et al. Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* 2013;9(7):e1003123.

30. Wu G, Fang YZ, Yang S, et al. Glutathione metabolism and its implications for health. *J Nutr* 2004;134(3):489-92.

31. Sindet-Pedersen S, Ramstrom G, Bernvil S, et al. Hemostatic effect of tranexamic acid mouthwash in anticoagulant-treated patients undergoing oral surgery. *N Engl J Med* 1989;320(13):840-3.

32. Altland JE, Wehje G, Gilliam CH, et al. Liverwort (Marchantia Polymorpha) Control with Quinoclamine. *Weed Technology* 2007;21(2):483-8.

33. Jansen SA, Kleerekooper I, Hofman ZL, et al. Grayanotoxin poisoning: 'mad honey disease' and beyond. *Cardiovasc Toxicol* 2012;12(3):208-15.

34. Parker TS, McNamara DJ, Brown C, et al. Mevalonic acid in human plasma: relationship of concentration and circadian rhythm to cholesterol synthesis rates in man. *Proc Natl Acad Sci U S A* 1982;79(9):3037-41.

35. Cummings MH, Watts GF, Pal C, et al. Increased hepatic secretion of very-low-density lipoprotein apolipoprotein B-100 in obesity: a stable isotope study. *Clin Sci (Lond)* 1995;88(2):225-33.

36. Keizer HG. The "Mevalonate hypothesis": a cholesterol-independent alternative for the etiology of atherosclerosis. *Lipids Health Dis* 2012;11:149.

Tables

Table 1. Phenotypic Characteristics of Twins, ETS

	Overall (n = 187)	Split 1 (n = 93)	Split 2 (n = 94)
Age (years)	55.84 (3.25)	56.00 (3.27)	55.87 (3.25)
Current Smoker	59 (31.6%)	34 (36.6%)	25 (26.6%)
BMI (kg/m^2)	29.48 (4.69)	29.25 (4.58)	29.72 (4.80)
Waist Circumference (cm)	99.76 (11.92)	98.94 (11.44)	100.57 (12.39)
Hip Circumference (cm)	104.55 (8.87)	104.20 (8.12)	104.88 (9.58)
WHR	0.9528 (0.06258)	0.9480 (0.06713)	0.9577 (0.06434)
mean (SD) or n (%); 1 subject missing BMI measure			

Table 2. Significant Metabolomic Features Associated Between Pairs with WHR, Bonferroni $p<0.05$

Ion	M/Z	Retention	T-Score	SE	Bonferroni p-value	Raw p-value
+	148.0605	134.2859	5.096	1.125	0.014	1.84×10^{-6}
+	286.2011	71.7878	5.266	0.761	0.008	1.03×10^{-6}
-	243.9574	141.9256	4.964	0.778	0.039	3.15×10^{-6}

Table 3. Annotation of Significant Metabolomic Features Associated with WHR

Ion	M/Z	Retention	Suspect Identity	Adduct	Formula	Δ ppm
+	148.0605	134.2859	Glutamate	$[M+H]^+$	$C_5H_9NO_4$	0
+	286.2011	71.7878	Acoric acid	$[M+NH_4]^+$	$C_{15}H_{24}O_4$	0
-	243.9574	141.9256	Quinoclamine	$[M+K-2H]^-$	$C_{10}H_6ClNO_2$	0

Table 4. Significant Metabolomic Features Associated Within Pairs with BMI, Bonferroni $p<0.05$

Ion	M/Z	Retention	T-Score	SE	Bonferroni p-value	Raw p-value
+	149.0807	129.7626	5.211	0.053	0.009	1.20×10^{-6}
+	619.2557	598.0935	-6.168	0.048	<0.001	1.99×10^{-8}
-	260.9158	14.6074	5.300	0.043	0.015	8.34×10^{-7}
-	325.2019	11.2376	5.376	0.040	0.008	6.07×10^{-7}
-	411.2385	12.3239	5.317	0.039	0.010	7.75×10^{-7}

Table 5. Annotation of Significant Metabolomic Features Associated with BMI

Ion	M/Z	Retention	Suspect Identity	Adduct	Formula	Δ ppm
+	149.0807	129.7626	Mevalonic acid	$[M+H]^+$	$C_6H_{12}O_4$	0
+	619.2557	598.0935	Undefined porphyrin	$[M+H-2H_2O]^+$	$C_{36}H_{38}N_4O_8$	0
-	260.9158	14.6074	Thallium	$[M+Na-2H]^-$	Tl	8
-	325.2019	11.2376	2,3-dinor-PGE1	$[M-H]^-$	$C_{18}H_{30}O_5$	0
-	411.2385	12.3239	Undefined peptide	$[M-H_2O-H]^-$	$C_{23}H_{34}N_4O_4$	2
-	411.2385	12.3239	Grayanotoxin I	$[M-H]^-$	$C_{22}H_{36}O_7$	0

Table 6. Cross Reference of Significant Metabolomic Features Associated with WHR and BMI

Ion	M/Z	Retention	Significant	WHR				BMI			
				Within Pairs		Between Pairs		Within Pairs		Between Pairs	
				T-Score	P-value	T-Score	P-value	T-Score	P-value	T-Score	P-value
+	148.0605	134.2859	BP, WHR	1.6596	0.100	5.0962	1.81×10^{-6}	2.2220	2.88×10^{-2}	3.4605	8.19×10^{-4}
+	149.0807	129.7626	WP, BMI	-1.1151	0.267	0.3279	0.744	5.2117	1.20×10^{-6}	-3.9972	1.29×10^{-3}
+	286.2011	71.7878	BP, WHR	-0.0626	0.950	5.2359	1.03×10^{-6}	1.4852	0.141	3.2126	1.81×10^{-3}
+	619.2557	598.0935	WP, BMI	0.5495	0.584	-0.4209	0.675	-6.1682	1.99×10^{-8}	-4.3906	3.02×10^{-5}
-	243.9574	141.9256	BP, WHR	2.0212	0.046	4.964	3.15×10^{-6}	1.3084	0.194	3.7712	2.87×10^{-4}
-	260.9158	14.6074	WP, BMI	0.3847	0.701	-2.081	4.02×10^{-2}	5.2996	8.34×10^{-7}	-3.8854	1.92×10^{-4}
-	325.2019	11.2376	WP, BMI	-0.8870	0.377	0.343	0.732	5.3759	6.07×10^{-7}	-3.856	2.13×10^{-4}
-	411.2385	12.3239	WP, BMI	-0.9154	0.362	0.3473	0.729	5.317	7.75×10^{-7}	-4.117	8.35×10^{-5}

*BP – Between Pairs, WP – Within Pairs

Figures

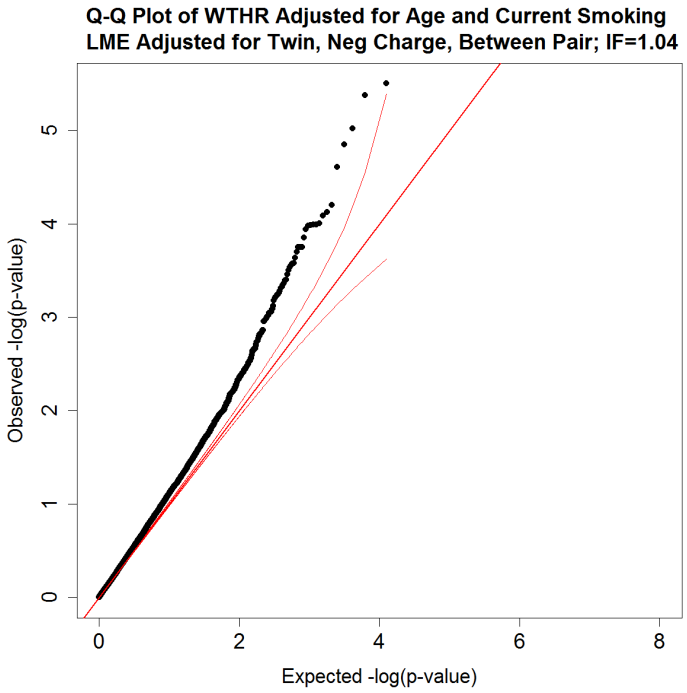


Figure 1. Q-Q Plot of LME Regression of WHR Between Pair Effects for Negative Charge Ions

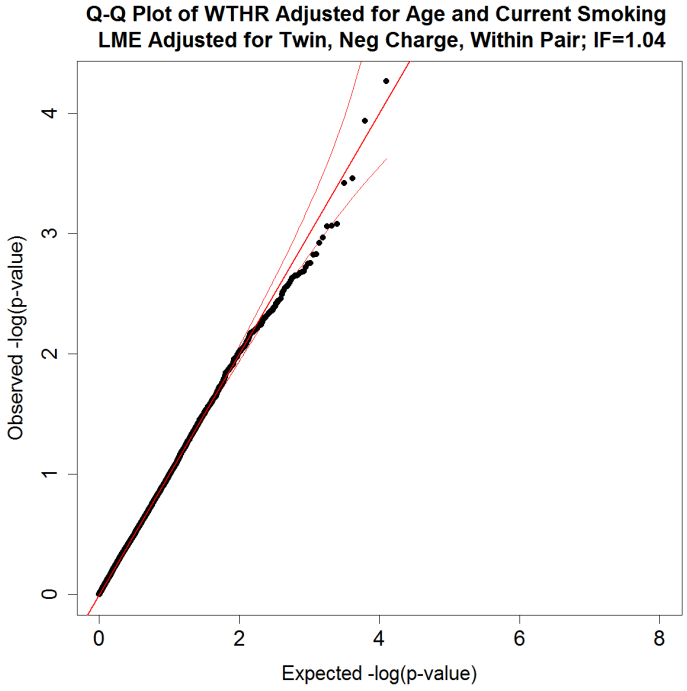
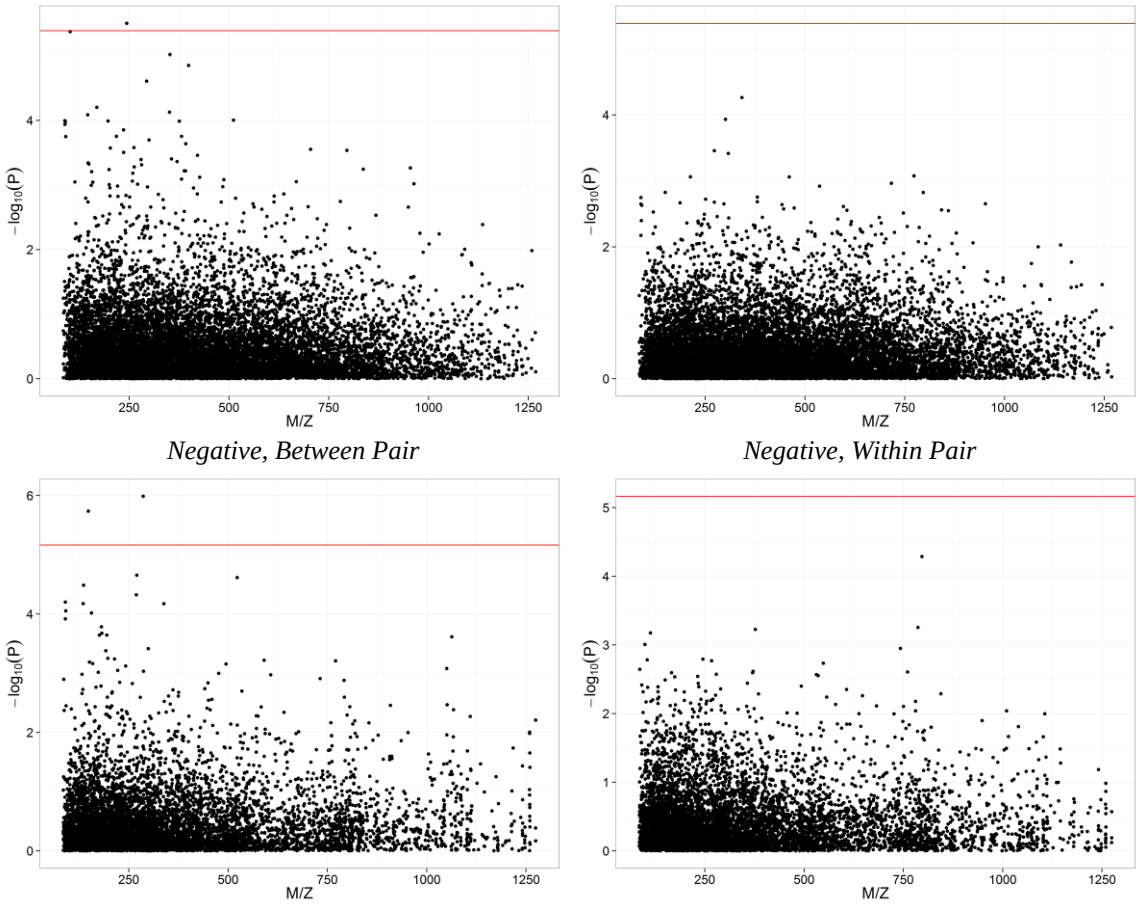


Figure 2. Q-Q Plot of LME Regression of WHR Within Pair Effects for Negative Charge Ions



Negative, Between Pair

Negative, Within Pair

Positive, Between Pair

Positive, Within Pair

Figure 3. Manhattan Plots of Association Strength by M/Z, WHR

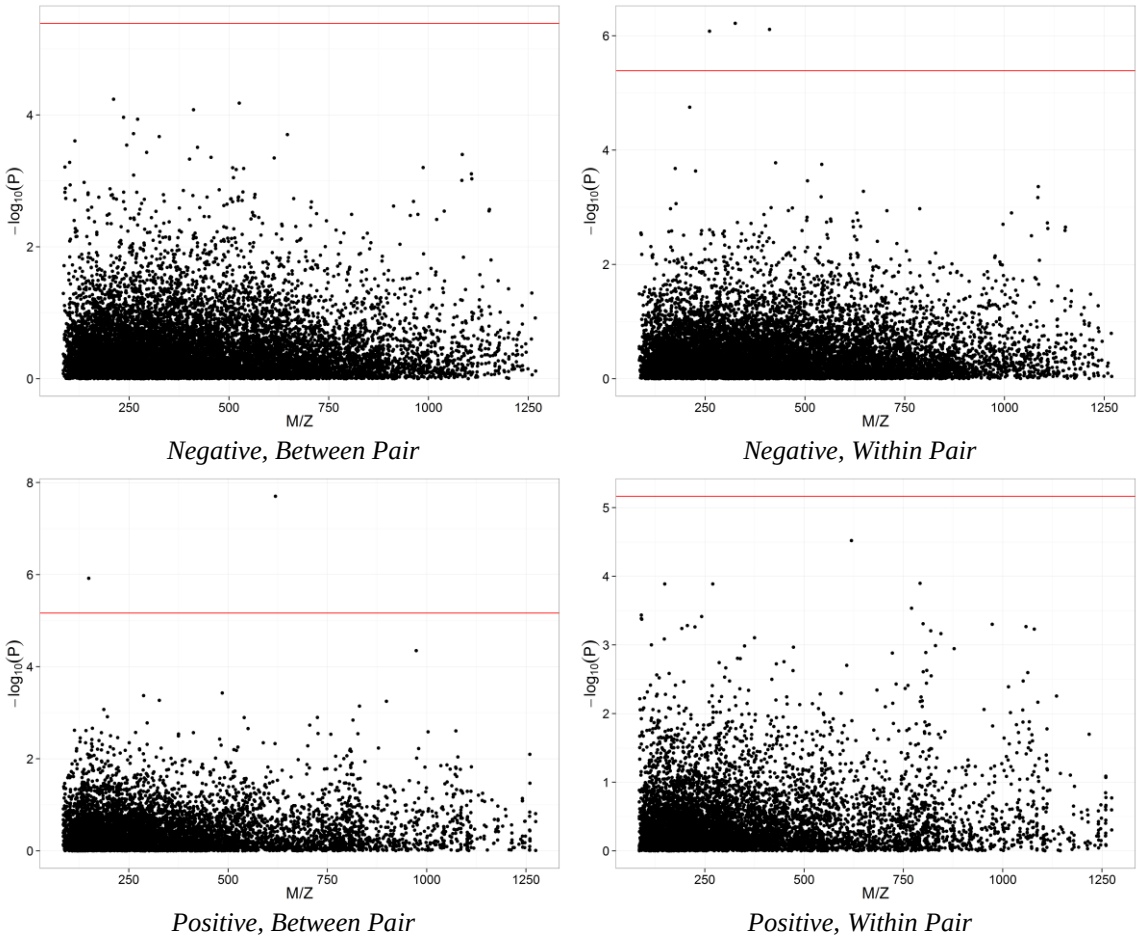


Figure 4. Manhattan Plots of Association Strength by M/Z, BMI

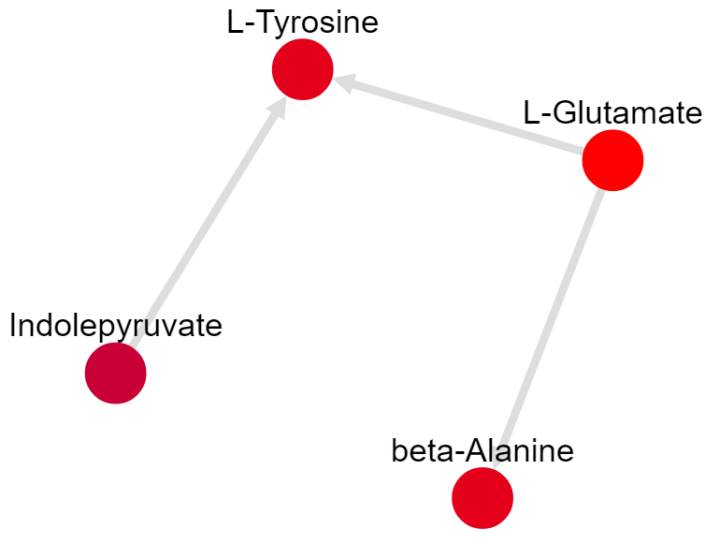


Figure 5. Mummichog Output Pathway, WHR Between Twin Pairs, Positive Ion