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:

Mikhail Melomed	04/22/2016 Date

A High Resolution Metabolomics Study of Prostate Cancer

By

Mikhail Melomed

MSPH

Epidemiology

Dr. Veronika Fedirko

Faculty Thesis Advisor

A High Resolution Metabolomics Study of Prostate Cancer

By

Mikhail Melomed

BS Applied Biology

Georgia Institute of Technology

2012

Faculty Thesis Advisor: Dr. Veronika Fedirko, PhD, MPH

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 Science in Public Health

in Epidemiology

2016

Abstract

A High Resolution Metabolomics Study of Prostate Cancer

By Mikhail Melomed

Background: Although it has been the subject of a multitude of sustained research endeavors, prostate cancer continues to be a major cause of morbidity and mortality in men. Examination of the metabolic profiles of men with prostate cancer could contribute to the identification of novel biomarkers of prostate cancer, thereby enhancing public health interventions and clinical practices.

Study Design: Cases (n=113) and controls (n=258) were selected from an existing, community-based case-control study conducted in the Piedmont Triadarea of North Carolina. Cases were over 50 years old, spoke English, and were newly diagnosed with prostate cancer. Cases and controls were frequency matched by age and race.

Results: Metabolomics analysis yielded 17697 metabolites, and 4485 metabolites with a median coefficient of variation value > 30 remained for statistical analysis. Of these, 27 metabolites, including aspartic acid, were associated with the presence of prostate cancer compared to matched controls.

Conclusion: The metabolic pathway of aspartic acid is significantly disturbed in men with prostate cancer. This finding suggests that aspartic acid may play a role in the disease mechanisms of prostate cancer.

A High Resolution Metabolomics Study of Prostate Cancer

By

Mikhail Melomed

BS Applied Biology

Georgia Institute of Technology

2012

Faculty Thesis Advisor: Veronika Fedirko, PhD, MPH

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 Science in Public Health

in Epidemiology

2016

Acknowledgements

I would like to acknowledge and thank my thesis advisor, Dr. Veronika Fedirko, for her encouragement, guidance, and patience. Dr. Fedirko's enthusiasm for metabolomics is utterly infectious, and she has opened my eyes to a realm of scientific study that reconciles my passions for biology and epidemiology.

I would also like to acknowledge and thank Dr. Yury O. Chernoff, Dr. Denis Kiktev, and Dr. Katy Bruce for serving as my colleagues, mentors, and friends. My journey into a world described by learning, research, and teaching began in their presence.

Finally, I would like to acknowledge and thank my father, Mikhail S. Melomed, my mother, Luda Melomed, my sister, Natasha Melomed, and my significant other, Dr. Aakanksha Angra, for their unconditional love and support.

This work is dedicated to my late grandmother, Maria Kochura. In sharing her love of flowers with me, she instilled in me an admiration for the natural world and all of the life that dwells therein.

Table of Contents

Chapter I: Background 1
1.1: Descriptive Epidemiology 1
1.2: Prostate Cancer Carcinogenesis 2
1.3: Prostate Cancer: Clinical Overview and Pathogenesis 5
1.4: Risk Factors 6
1.5: Metabolomics 7
1.6: The Role of Metabolomics in Biomarker Discovery 10

Chapter II: Materials and Methodology 11
2.1: Study Subjects 11
2.2: Data Collection 12
2.3: Metabolomics 12
2.4: Statistical Analysis 13

Chapter III: Results 15
3.1: Baseline Characteristics 15
3.2: Metabolites Distinguishing Prostate Cancer Cases from Controls 15
3.3: Association of Aspartic Acid with Prostate Cancer Risk 16
3.4: Stratified Analyses by Potential Effect Modifying Factors 16

Chapter IV: Discussion 17
4.1: Biological Plausibility of Findings 17
4.2: Our Work and Available Evidence 18
4.3: Strengths and Limitations 19
4.4: Public Health Implications 19
4.5: Future Directions 20
4.6: Conclusion 20

References 22

Tables 26

Figures and Figure Legends 29

Chapter I: Background

1.1: Descriptive Epidemiology

Prostate cancer is the most frequently diagnosed type of cancer and is currently the second most common cause of cancer death in American men [1]. According to the American Cancer Society (ACS), in 2011, 240,890 men were diagnosed with prostate cancer and 33,720 men died because of it [1]. The ACS estimates that 16.2% of American men who are alive today will be diagnosed with prostate cancer, and an estimate 3.0% of these men will die because of it [1].

The high morbidity and mortality associated with prostate cancer underlies its status as a major public health concern. Prostate cancer is most commonly diagnosed in men over the age of 65 [1]. Americans who identify as African-American or of Black African heritage exhibit higher incidence of and mortality due to prostate cancer, as compared to Americans of European descent [1].

For men diagnosed with local or regional prostate cancer, estimated 5, 10, and 15-year survival rates are 100%, 95%, and 82%, respectively [1]. It is estimated that less than one-third of men who are diagnosed with metastatic prostate cancer survive 5 years after diagnosis [1].

The successful implementation of prostate cancer screening programs is imperative, given that prognosis tends to be best when this condition is detected early. Accepted prostate screening methods include the prostate-specific antigen test (PSA) [1] and magnetic resonance imaging (MRI) [2]. Although an endorectal MRI provides

clinicians with better visualization of the zonal anatomy of the prostate, its use is less common than the PSA because few men present with locoregional or metastatic disease [3]. A variety of prostate cancer diagnostic methods are currently being studied, including the use of serological prostate cancer biomarkers with cancer antigen arrays [4], transrectal ultrasound and contrast enhanced transrectal ultrasound [5], and the use of computer-aided bone scan evaluation systems to render a diagnosis based on bone scan index [6].

Benign prostate hyperplasia (BPH) is a non-cancerous enlargement of the prostate, whereas adenocarcinoma is cancer of the glandular epithelium. Elevated levels of PSA may be indicative of either BPH or prostate cancer. Such a screening method is of limited utility because it can only help to identify a potential prostate cancer case; unfortunately, it does not identify cases that are at an elevated risk of their prostate cancer progressing. Thus, there is a need to develop, validate, and implement screening and diagnostic methods that are highly sensitive, highly specific, and capable of identifying men in whom prostate cancer will progress.

1.2: Prostate Cancer Carcinogenesis

During a human's lifespan, genetic control systems regulate cell birth and death. Such control systems modulate cell birth and death rates by responding to growth signals, growth-inhibiting signals, and death signals. The rates of cell birth and death determine both the size of an adult human and the rate at which that size is achieved. Some adult tissues require continuous cellular proliferation to facilitate constant tissue renewal. However, in many other adult tissues, cellular proliferation is a process employed by

stable cells to support healing processes. When the mechanisms responsible for maintaining normal cellular growth rates malfunction, excess cellular division results and cancer is said to occur.

The aberrant cellular behavior that typifies cancer is a manifestation of genetic damage. Thus, the induction of cancer may occur through mutations in three classes of genes: proto-oncogenes, tumor-suppressor genes, and caretaker genes. Proto-oncogenes are normally responsible for promoting cellular growth. Proto-oncogenes become oncogenes by four mechanisms: point mutation, in which a change in a single base pair activates proto-oncogenic products; chromosomal translocation, which results in the production of chimeric proteins; chromosomal translocation, which results in re-assignment of different promoters to growth-regulatory genes; and the over-production of proto-oncogenic proteins through DNA amplification.

Tumor-suppressor genes are normally responsible for restrained cellular growth. Thus, mutations that deactivate tumor-suppressor genes facilitate aberrant cellular division. Caretaker genes are normally responsible for ensuring genomic integrity. Inactivation of caretaker genes increases the rate at which cells acquire mutations. Such mutations disrupt cellular growth control and lead to cancer. Proto-oncogenes, tumor-suppressor genes, and caretaker genes encode proteins that are responsible for regulation of the cellular cycle, cellular death by apoptosis, and repair of damaged DNA.

Advances in the fields of molecular biology, genetics, epigenetics, and biotechnology have improved our understanding of the events associated with associated with the initiation and progression of prostate cancer [7]. Multiple genes have now been identified and are thought to be germane to prostate cancer carcinogenesis. Contemporary

models of prostate cancer progression often incorporate mechanisms that describe the role of inflammation in the development of preneoplastic and neoplastic lesions [7]. An alternative pathway to prostate cancer -- in addition to aneuploidy, loss of heterozygosity, and genetic mutation – is thought to include abnormal methylation of growth regulatory genes or caretaker genes [7].

Previous studies have shown that the mechanism of growth control for prostate cancer cells is likely to involve the biological interaction of these cells and the microenvironment at the secondary site, which influences the occurrence of growth into metastasis [8]. Seven tumor suppressor genes that act to suppress metastasis without affecting primary tumor growth have been identified; three of these genes, in particular, function as metastasis suppressor genes of prostate cancer and include the following: KAI1, CD44, and MAPK kinase 4 [8]. Many metastasis suppressor genes play important roles in cellular growth control, cellular adhesion, and cytoskeletal reorganization, suggesting the possibility of common mechanism of metastatic suppression [8].

The initiation and progression of prostate cancer is highly associated with proto-oncogenes [9]. Large sets of genes, which are thought to play a proto-oncogenic role in prostate cancer, have been identified; however, few have been characterized in the molecular progression of the disease [10]. Examples of proto-oncogenes with known roles in the molecular progression of prostate cancer include genes responsible for growth factor receptors, proteases, and transcription factors and coactivators [10].

Growth factors implicated in prostate cancer carcinogenesis include insulin growth factor (IGF), the Wnt signaling pathway, the Her2/neu protein, epidermal growth factor receptor (EGFR), and phosphoinositide-3 kinase (PI3K) [10]. Proteases with

known roles in prostate cancer carcinogenesis include members of the homologous MMP family; in particular: MMP 2, 7, 9, and MT1MMP [10]. Known transcription factors and coactivators with roles in prostate cancer include TMPRSS2-ERG, MYC, and homologous proteins SRC-1, SRC-2, and SRC-3 of the p160 SRC family [10].

A growing body of evidence suggests that prostate cancer is largely regulated by posttranslational modifications (PTMs) and epigenetic alterations [9]. PTMs play important roles in cellular functions, gene regulation, tissue development, diseases, malignancy, and drug resistance [11]. Examples of PTMs that have been shown to play a role in prostate cancer include ubiquitination and SUMOylation [9]. Both ubiquitination and SUMOylation pathways can be differentially modulated by a variety of stimuli and stressors to produce sustained oncogenic potentials [9].

1.3: Prostate Cancer: Clinical Overview and Pathogenesis

Prostate cancer has been a recognized disorder since it was first described by ancient Egyptians, and surgical procedures designed to remove the prostate were developed over 100 years ago [12].

Men with elevated levels of PSA are typically biopsied to assess for the presence of prostate cancer [12]. After a biopsy is performed, histopathological grading of prostate tissue is facilitated by Gleason scoring, which allows pathologists to classify tumors from a scale of 1-to-5 (i.e. from least to most differentiated) [12]. The diagnosis of prostate cancer is predicated upon the status of primary tumors and measured by several spectra: T1-4, ranging from organ-confined to fully invasive; N0-1, a measure of lymph node

involvement; and M0 and 1a-c, which describes the presence/degree of distant metastases [12].

In men, the prostate is a small tissue that surrounds the urethra at the base of the bladder, and it functions to produce components of seminal fluid [12]. The human prostate is defined in terms of zonal architecture and is comprised of central, periurethral transition, peripheral zones, and an anterior fibromuscular stroma [12]. The majority of prostate carcinomas occur in the outermost peripheral zone [12].

The human prostate contains a pseudo-stratified epithelium, comprised of three differentiated cell types: luminal, basal, and neuroendocrine [12]. Prostatic intraepithelial neoplasia (PIN) is a precursor to prostate cancer, although this relationship remains to be demonstrated conclusively [12]. In general, the histological presentation of PIN is characterized by the appearance of luminal epithelial hyperplasia, reduction in basal cells, enlargement of nuclei and nucleoli, cytoplasmic hyperchromasia, and nuclear atypia [12]. Human prostate cancer exhibits significant phenotypic heterogeneity; however, over 95% of prostate cancers are classified as adenocarcinoma, which has a luminal phenotype [12].

Common secondary sites of metastasis for prostate cancer include the liver, lung, and pleura; however, if prostate cancer does metastasize, it invariably travels to the bone and forms osteoblastic lesions [12].

1.4: Risk Factors

In 2014, the Continuous Update Project (CUP) Report issued by World Cancer Research Fund International/American Institute for Cancer (WCRF/AICR) compiled available evidence concerning the association between diet, nutrition, physical activity,

and prostate cancer. According to CUP, there is strong evidence to support an association between increased body fatness and advanced prostate cancer risk [13]. The same report also found limited evidence to support an association between the consumption of dairy products, adherence to diets high in calcium, low plasma alphatocopherol concentrations, and low plasma selenium concentrations and an increased risk for prostate cancer [13].

In addition to dietary and lifestyle risk factors, there is a growing body of evidence describing the association between genetic factors and the risk for prostate cancer. Genetic studies have shown that strong familial predisposition plays a role in 5-to-10% of prostate cancers; for example: BRCA1 or BRCA2 genes, which are known to increase risks for breast and ovarian cancer in women, may also increase the risk of prostate cancer in men (Brawley, 2012). Over 100 common single nucleotide polymorphisms (SNPs) have been identified by genome-wide association studies (GWAS), and these SNPs have been shown to be associated with prostate cancer [14]. Because most of these SNPs reside in noncoding genomic regions, it is thought that their role in prostate cancer etiology is long-range regulation of gene expression (i.e. long-range chromatin interactions) [14]. Although GWAS have identified multiple SNPs with an increased risk for prostate cancer, the identification of functional SNPs poses a considerable challenge because risk SNPs are not necessarily causative [14].

1.5: Metabolomics

Endogenous and environmental factors have been shown to be correlated with increased risk of prostate cancer [1]. However, for most of them the existing

epidemiologic evidence is limited. For example: there is limited evidence to suggest that men who consume high quantities of dairy products or have diets high in calcium are at an elevated risk for developing prostate cancer (WCRF Continuous Update Project). Also, low plasma alpha-tocopherol and selenium concentrations were suggested to be associated with increased risk (WCRF Continuous Update Project).

Whether endogenous or environmental, factors such as those described above are varied in their physical states and are dauntingly numerous [1]. Metabolomic profiling could facilitate the identification and characterization of environmental and endogenous factors associated with the physiopathology of prostate cancer. Furthermore, metabolomic profiling may help elucidate the regulatory circuits that modulate prostate cancer pathways, leading to both a better understanding of its etiology and toward the development of biomarkers of risk and progression.

As an emerging technology, metabolomics has enabled investigators to study the complex interactions of nutrients in individuals having unique genomes, dietary histories, and exposure to various environmental and behavioral factors [15]. Metabolomics is the application of small-molecular chemical profiling to complex biosystems, integrating diet and nutrition [15]. Nutritional metabolomics is capable of yielding nutritional models based on the metabolomic profile, genome, epigenetics, and health phenotyping of an individual [15]. Moreover, nutritional metabolomics has been applied to the process of discovering novel biomarkers of nutritional exposure, status, and impact on disease [15].

The rise of metabolomics reflects a paradigm shift from the population average focus of comparable biochemical studies [15]. This paradigm shift is the result of three conceptual developments; they are: (1) the exposome, which captures information

detailing cumulative life exposures; (2) predictive health predicated upon the use of nutrition to optimize vitality and well-being; and (3) the development of models that span multiple interacting functional networks [15].

Small-molecular-weight metabolites within organisms are presently beyond practical detection methods; however, human metabolic databases hold information on approximately 2,500 metabolic intermediates, hormones, and other molecules involved in signaling [15]. Furthermore, over 1,000 metabolic components related to the ingestion of pharmacologic agents and 3,500 metabolomic components related to the consumption of food have been identified [15].

Mass spectrometry is currently considered the best method to detect metabolites, which can vary widely in their physical and chemical composition [15]. High-resolution mass spectrometry enables investigators to measure and differentiate large numbers of chemical substances, based on mass resolution and mass accuracy [15]. In turn, this information allows for the prediction of chemical elemental composition through mass/charge (*m/z*) values, which are used in human metabolite databases to map metabolism [15]. Ultimately, high-resolution mass spectrometry enables metabolome-wide association studies of nutrition and disease – and these studies are further enhanced by chemical and metabolic databases, bioinformatic methods, and various computational approaches [15].

By examining the metabolic profiles of men with prostate cancer, it may be possible to contribute to the identification of novel, modifiable risk factors and biomarkers of this disease. This could not only enhance the foundations of numerous

public health policies, but could also lead to improvements in clinical practices designed to screen for, diagnose, treat, and manage prostate cancer.

1.6: The Role of Metabolomics in Biomarker Discovery

The discovery of novel biomarkers is a formidable challenge, especially when considered against the backdrop of the combinatorics inherent to the human genome. For instance: Sarcosine, a compound first isolated in 1847, is an intermediate product in the synthetic pathway leading to the formation and degradation of amino acid glycine [16]. In 2009, a seminal publication describing the association between changes in levels of sarcosine and prostate cancer progression was published by Sreekumar *et al.* in Nature [16]. Since then, sarcosine has been the subject of investigation as a new biomarker for prostate cancer. Issaaq *et al.* showed that sarcosine is a metabolite with levels that increase throughout the progression of prostate cancer and its metastatic process [17]. Moreover, Cavaliere *et al.* showed that sarcosine could be detected in urine [18].

However, a 2015 case control study of sarcosine as a potential prostate cancer biomarker (N=497) showed that sarcosine levels overlapped between prostate cancer cases (median 15.8uM, range 6.2-42.5uM) and controls (median 16.2uM, range 6.4 to 53.6uM) [19]. Ankerst *et al.* concluded that serum sarcosine is not an eligible marker for the detection of prostate cancer.

As the search for novel biomarkers of prostate cancer continues to pique the interests and resources of academic, medical, and industry researchers, it is clear that a multi-disciplinary approach to validating a prospective biomarker is highly necessary.

Chapter II: Materials and Methodology

2.1: Study Subjects

Data was obtained from an existing community-based case-control study of incident prostate cancer, which was conducted in the Piedmont Triad area of North Carolina. The Committee for Human Research at Wake Forest University, Winston-Salem, North Carolina, granted approval of the research protocol. Cases for the study (n=113) were comprised of black and white men for whom prostate cancer diagnosis was pathologically documented. Cases were over 50 years old, spoke English, and were newly (and for the first time ever) diagnosed with prostate cancer. Cases were identified and recruited from all cases diagnosed with prostate cancer during the study period, in area urology and radiation oncology practices within days of diagnosis. Cases were studied prior to initiation of treatment for prostate cancer. Controls (n=258) were recruited from the same geographic locality as cases and were randomly selected using the Polk Directory. Controls were frequency matched to cases by age and race and did not have a history of prostate cancer. Participants were excluded due to a history of previous cancers (exception: non-melanoma skin cancer), current prostate disease, previous prostate surgery, active tuberculosis, or current liver or kidney disease. Of 113 case subjects who were enrolled into the study, 108 were used in analyses. Of 258 control subjects who were enrolled into the study, 256 were used. Attrition was due to missing biosamples.

2.2: Data Collection

Study participants attended a four-to-five hour study visit at the General Clinical Research Center at Wake Forest University. The study visit included informed consent procedures, interview, completion of a medical/lifestyle questionnaire and Block Food Frequency Questionnaire, anthropometrics, and providing fasting blood, spot, and timed urine samples. Participants were compensated for their time and effort with a payment of $50. Tumors were staged by the TNM system and pathology information on cases was retrieved from hospital tumor registries.

2.3: Metabolomics

Plasma samples were extracted and analyzed by liquid chromatography-high-resolution mass spectrometry (LC-FTMS) as previously described [20]. Briefly, 120-µl aliquots of plasma were treated with acetonitrile (2:1, v/v), containing internal standard mix, and centrifuged at 14,000×*g* for 5 min at 4°C and maintained at 4°C until injection. Data were collected by a Thermo LTQ-FT mass spectrometer (Thermo Fisher, San Diego, CA) from *m/z* 85 to 850 over 10 min with each sample analyzed in triplicate. Peak extraction and quantification of ion intensities were performed by an adaptive processing software package (apLCMS) designed for use with LC-FTMS data [21]. The package output included tables containing *m/z* values, retention time, and integrated ion intensity for each *m/z* feature. The metabolite values were averaged for triplicates, and the data were log-transformed, median centered, scaled to have unit variance, and quantile normalized prior to statistical analyses. Data were also subjected to quality assessment, including exclusion of data for technical replicates with overall Pearson correlation (r)

<0.70 [22]. Extraction of mass spectral data initially yielded 17697 metabolites. Of these, 4485 metabolites had median coefficient of variation values greater than 30 and were used for subsequent analysis. For the remaining metabolites, missing values were replaced with values equal to half of the minimum value of that metabolite's relative abundance across all samples.

2.4: Statistical Analysis

Descriptive characteristics of cases and controls were tabulated and analyzed by using the *t*-test for continuous variables Chi-square test for categorical variables.
The differential expression of plasma metabolites between prostate cancer cases and controls was determined using t tests and visualized using Manhattan plots. False discovery rate (FDR) was computed using Benjamini-Hochberg method [23]. Since our analyses are exploratory, we used a more conservative approach that avoided FDR error by including all metabolites that were significant with FDR = 0.2 (raw P value < 0.05) and then performing statistical testing of these metabolites for pathway enrichment. The 27 significant metabolites identified using the LIMMA package in R (Linear Models for Microarray Data) in Bioconductor (FDR = 0.2) were depicted by a heat map and subjected to pathway analysis after global network analysis using mummichog [23]. For global network analysis, 27 significant metabolites were correlated with other metabolites at an absolute correlation threshold of 0.3, and a correlation FDR threshold of 0.1.

To complement univariate analyses, we also performed unconditional logistic regression to investigate the association between a metabolite of interest (e.g., aspartic acid) and prostate cancer risk, adjusting for matching factors (age and race) and potential

confounders. In all models, a metabolite of interest was either standardized by 1 standard deviation or divided into tertiles (33.3% and 66.6%) based on the distribution of relative abundance readings for aspartic acid among the controls. Covariates of interest included body mass index, education status, smoking status, drinking status, physical activity level, and income, and family history of prostate cancer.

Preliminary unconditional logistic regression models contained the aforementioned outcome variable, exposure variable, frequency matching variables, covariates, and interaction terms. Collinearity was assessed using a SAS-L macro written by Mathew Zack, Jim Singleton, and Kristin Wall. Chunk testing using the likelihood ratio test was used to compare a full unconditional logistic regression model containing all exposure, covariates, and interaction terms to a reduced model with interaction terms removed. All interaction terms were found to be non-significant (P>0.26) and were removed. Confounding variables were identified if their removal from the model resulted in an effect estimate change of greater than 10%. After removal of non-confounding variables, a final model was established and included aspartic acid, a metabolite of interest, as the exposure, age and race as matching variables, and body mass index (BMI), drinking status, physical activity, and family history of prostate cancer as confounders. Missing data for BMI was imputed using median values of BMI for cases and controls.

Analyses were performed using the following: RStudio software, version 0.99.4466 (RStudio Inc., Boston, MA); SAS software, version 9.1 (SAS Institute, Cary, NC); METLIN, v.c1.1 beta (Scripps Center for Metabolomics); and Python software, version 3.3.5. R packages dpylr, stringr, and xlsx were used to facilitate RStudio analysis.

Chapter III: Results

3.1: Baseline Characteristics

The demographics and lifestyle data of the study participants are summarized in Table 1. No significant differences were observed between the two groups in terms of body mass index, education, smoking status, drinking status, physical activity level, and income (Table 1). Cases were more likely than controls to have a family history of prostate cancer (Table 1).

3.2: Metabolites Distinguishing Prostate Cancer Cases from Controls

Metabolites that were significantly associated with prostate cancer were primarily those with a mass-to-charge ratio (m/z) between 200 and 800 (Figure 1, A). Furthermore, metabolites significantly associated with prostate cancer had a chromatographic retention time of between approximately 40-60 units (Figure 1, B).

The average intensities of the 27 significant metabolites are depicted in the heat map (Figure 2, A); however, a clear differential expression between cases with prostate cancer and frequency-matched controls is not exhibited. Representative plots for select significant metabolites comparison between prostate cancer cases and control groups are included in Figure 2, A-D as examples (Figure 2, B).

A network analysis is visualized by a graph that is representative of 27 significant metabolite features (Figure 4, A). Pathways in which the aforesaid metabolites play a role in are displayed (Figure 4, B). We decided to focus on 1 of the 27 significant metabolites,

which has a confirmed mass-to-charge ratio; that metabolite was aspartic acid (m/z = 134.04).

3.3: Association of Aspartic Acid with Prostate Cancer Risk

The age- and race-adjusted odds ratio (OR) comparing men in the highest tertile of aspartic acid relative to the lowest tertiles was 0.31 (95% CI: 0.18, 0.57; *P* for trend < 0.0001) (Table 2). When aspartic acid was standardized by 1 standard deviation, participants who were frequency matched by age and race and exposed to increasing levels of aspartic acid were at a decreased risk of developing prostate cancer (OR=0.66; 95% CI: 0.52, 0.84) (Table 2).

The multivariable adjusted OR comparing comparing men in the highest tertile of aspartic acid relative to the lowest tertile was 0.32 (95% CI: 0.17, 0.57; *P* for trend < 0.0001) (Table 2). When aspartic acid was standardized by 1 standard deviation and body mass index, drinking status, physical activity, and family history of prostate cancer were adjusted for, participants exposed to increasing levels of aspartic acid were at a decreased risk of developing prostate cancer (OR=0.64; 95% CI: 0.50, 0.82) (Table 2).

3.4: Stratified Analyses by Potential Effect Modifying Factors

No evidence was found to support different effects by BMI, age, alcohol, drinking status, and physical activity (Table 3). Although there were suggestions of potentially stronger effects among normal weight and those who exercise, our findings were not statistically significant (all $P > 0.13$) (Table 3).

Chapter IV: Discussion

4.1: Biological Plausibility of Findings

The findings from this study suggest that higher aspartic acid exposure may be associated with lower risk for prostate cancer, and this is biologically plausible for several reasons. Aspartic acid-phosphorylated proteins facilitate a wide variety of cellular functions, including metabolism, protein folding, and cytoskeletal mobility [24]. Furthermore, in human prostate cancer progression, aspartic acid phosphorylation plays a role in the three states of cancer (e.g. non-tumorigenic, tumorigenic, and metastatic cells) [24]. Therefore, it is possible that aspartic acid may exert its effect on prostate cancer pathogenesis through pathways involving aspartic acid-phosphorylated proteins.

In addition to aspartic acid-phosphorylated proteins, the effect of aspartic acid on prostate cancer pathogenesis may involve other classes of proteins, such as chaperones. Protease inhibitors (PIs) are cellular chaperones that function to target or inhibit protein-digesting enzymes, which are also known as proteases [25]. Proteases have been shown to play a wide variety of roles in biological processes, including apoptosis [25]. Classification of PIs is based on the amino acid composition of PIs where protease-PI enzymatic reactions occur [25]. With regard to carcinogenesis in humans, PIs have been shown to prevent ongoing cellular processes begun by carcinogens by modulating metastasis [25]. Moreover, PIs have the potential to reduce carcinogen-induced gene amplification to normalized levels [25]. Members of the aspartyl PI family are known to be found in sunflower, barley and cardoon flowers, and in potato tubers [25].

4.2: Our Work and Available Evidence

Our findings indicate that higher levels of aspartic acid exposure result in a reduced risk of prostate cancer, and that the trend is statistically significant. However, our study did not show statistically significant interaction with aspartic acid exposure and factors that ought to biologically influence prostate cancer carcinogenesis, such as body mass index and physical activity.

Because of paucity in epidemiological evidence describing the association between exposure to aspartic acid and prostate cancer, an epidemiologic consensus regarding this particular observation has not yet been achieved. For example, aspartame, a methyl ester derivative of aspartic acid, has been examined in numerous epidemiological studies as a potential risk factor for cancer. Epidemiologic studies in humans on aspartame intake are sparse and have not demonstrated an association between this synthetic sweetener and cancer risk [26]. In another example, a metabolomics study conducted in 2011 collected and analyzed plasma samples from 200 patients, each of which were diagnosed with one of five types of cancer – including prostate cancer [27]. By measuring plasma free amino acids (PFAA) levels with high-performance liquid chromatography-electrospray ionization-mass spectrometry and performing univariate and multivariate analyses, investigators showed significant differences in the PFAA profiles of cases and controls with prostate cancer; however, aspartate was excluded due to its instability in blood [27]. Thus, evaluating our results in the context of previous, similar studies (e.g. studies that evaluate the association between aspartic acid and prostate cancer) is difficult and highlights the need for further research.

4.3: Strengths and Limitations

This study has some strengths and limitations. A strength of this study is its data collection procedures, which allowed for collation of high quality information and biological samples. Another strength of the study was its community-based design and the studying of cases both within days of diagnosis and prior to the initiation of treatment. As with most case-control studies, there is a potential for this study to be affected by recall biases and ambiguous temporal relationships. Because cases were studied immediately after prostate cancer diagnosis and before the initiation of treatment, the influence on recall bias on effect estimates is likely mitigated. Furthermore, because health conditions that were included in this study likely precede prostate cancer (e.g. sexually transmitted diseases, lifetime accumulation of aspartic acid), the internal validity of the findings presented herein is likely to be minimally threatened. Self-reported information that is sensitive in nature (e.g. alcohol drinking status) may have been subject to under-reporting; however, this likely biased associations toward the null. Yet another limitation of this study is a lack of prostate biopsies on controls, which would have ruled out sub-clinical prostate cancer and helped to reduce possible case-control status misclassification. If, however, some controls did have prostate cancer, our effect estimates would have likely been biased toward the null.

4.4: Public Health Implications

If further studies were to support our own findings, the public health implications could be powerful. Aspartic acid is a semi-essential amino acid in humans and is found in multiple dietary sources, including animal sources, vegetable sources, dietary

supplements, and synthetic sweeteners such as aspartame. If an epidemiologic consensus were reached on the direction and magnitude of the effect describing the association between aspartic acid and prostate cancer, public health agencies and organizations might seek to enact regulatory measures concerning the consumption of dietary items containing this amino acid. In turn, this might engender changes sustainable, population-level changes in prostate cancer health outcomes in men, leading to potential reductions in morbidity and mortality associated with this disease. Additionally, characterization of aspartic acid in pathways that are important in the prostate cancer disease mechanism may contribute to our collective clinical capacity to more accurately diagnose the disease.

4.5: Future Directions

Future iterations of a study such as the one presented herein might seek to improve generalizability of our findings by including men from racial and ethnic backgrounds and geographic localities not represented among the cases and controls in our study. Additionally, it may be prudent to explore the associations between other known metabolites (e.g. amino acids) and prostate cancer. Finally, a prospective study of the association between aspartic acid and prostate cancer may provide useful temporal information.

4.6: Conclusion

In conclusion, our exploratory metabolomics analyses demonstrated that the metabolic pathway of aspartic acid is disturbed in adult men with prostate cancer. This

suggests that the pathway containing aspartic acid may be an integral component of the disease mechanism underlying prostate cancer. These preliminary findings also suggest that further exploration of the causal links between amino acid metabolism and prostate cancer is warranted and that these pathways could be exploited in the development of therapeutic targets for prostate cancer treatment.

References

1. Brawley, O.W., *Prostate cancer epidemiology in the United States.* World J Urol, 2012. **30**(2): p. 195-200.
2. Nam, R.K., et al., *A Pilot Study to Evaluate the Role of Magnetic Resonance Imaging for Prostate Cancer Screening in the General Population.* J Urol, 2016.
3. Carroll, P.R., F.V. Coakley, and J. Kurhanewicz, *Magnetic resonance imaging and spectroscopy of prostate cancer.* Rev Urol, 2006. **8 Suppl 1**: p. S4-s10.
4. Adeola, H.A., et al., *Novel potential serological prostate cancer biomarkers using CT100+ cancer antigen microarray platform in a multi-cultural south african cohort.* Oncotarget, 2016.
5. Li, H., et al., *Prostate cancer: a comparison of the diagnostic performance of transrectal ultrasound versus contrast enhanced transrectal ultrasound in different clinical characteristics.* Int J Clin Exp Med, 2015. **8**(11): p. 21428-34.
6. Uemura, K., et al., *Prognostic value of a computer-aided diagnosis system involving bone scans among men treated with docetaxel for metastatic castration-resistant prostate cancer.* BMC Cancer, 2016. **16**(1): p. 109.
7. Gonzalgo, M.L. and W.B. Isaacs, *Molecular pathways to prostate cancer.* J Urol, 2003. **170**(6 Pt 1): p. 2444-52.
8. Kauffman, E.C., et al., *Metastasis suppression: the evolving role of metastasis suppressor genes for regulating cancer cell growth at the secondary site.* J Urol, 2003. **169**(3): p. 1122-33.

9. Chen, Z. and W. Lu, *Roles of ubiquitination and SUMOylation on prostate cancer: mechanisms and clinical implications.* Int J Mol Sci, 2015. **16**(3): p. 4560-80.

10. Dasgupta, S., S. Srinidhi, and J.K. Vishwanatha, *Oncogenic activation in prostate cancer progression and metastasis: Molecular insights and future challenges.* J Carcinog, 2012. **11**: p. 4.

11. Al-Delaimy, W.K., et al., *A prospective study of smoking and risk of breast cancer in young adult women.* Cancer Epidemiol Biomarkers Prev, 2004. **13**(3): p. 398-404.

12. Shen, M.M. and C. Abate-Shen, *Molecular genetics of prostate cancer: new prospects for old challenges.* Genes Dev, 2010. **24**(18): p. 1967-2000.

13. *Diet, Nutrtion, Physical Activity, and Prostate Cancer*, in *Continuous Update Project Report*. 2014, World Cancer Research Fund International/American Institute for Cancer Research.

14. Du, M., et al., *Chromatin interactions and candidate genes at ten prostate cancer risk loci.* Sci Rep, 2016. **6**: p. 23202.

15. Jones, D.P., Y. Park, and T.R. Ziegler, *Nutritional metabolomics: progress in addressing complexity in diet and health.* Annu Rev Nutr, 2012. **32**: p. 183-202.

16. Cernei, N., et al., *Sarcosine as a potential prostate cancer biomarker--a review.* Int J Mol Sci, 2013. **14**(7): p. 13893-908.

17. Issaq, H.J., T.J. Waybright, and T.D. Veenstra, *Cancer biomarker discovery: Opportunities and pitfalls in analytical methods.* Electrophoresis, 2011. **32**(9): p. 967-75.

18. Cavaliere, B., et al., *Sarcosine as a marker in prostate cancer progression: a rapid and simple method for its quantification in human urine by solid-phase microextraction-gas chromatography-triple quadrupole mass spectrometry.* Anal Bioanal Chem, 2011. **400**(9): p. 2903-12.

19. Ankerst, D.P., et al., *A case control study of sarcosine as an early prostate cancer detection biomarker.* BMC Urol, 2015. **15**: p. 99.

20. JM Johnson, F.S., M Reed, J Pohl, and DP Jones, *A rapid LC-FTMS method for the analysis of cysteine, cystine and cysteine/cystine steady-state redox potential in human plasma.* Clin Chim Acta 2008. **396**(1–2): p. 43–48.

21. T Yu, Y.P., JM Johnson, and DP Jones, *apLCMS—adaptive processing of high-resolution LC/MS data. .* Bioinformatics 2009. **25**(15): p. 1930–1936.

22. J Lee, J.P., MS Lim, et al., *Quantile normalization approach for liquid chromatography-mass spectrometry-based metabolomic data from healthy human volunteers. .* Anal Sci 2012. **28**(8): p. 801–805.

23. Benjamini Y, H.Y., *Controlling the false discovery rate: a prac-tical and powerful approach to multiple testing.* J R Statist Soc B 1995. **57**: p. 289-300.

24. Lapek, J.D., Jr., et al., *Evidence of histidine and aspartic acid phosphorylation in human prostate cancer cells.* Naunyn Schmiedebergs Arch Pharmacol, 2015. **388**(2): p. 161-73.

25. Shamsi, T.N., R. Parveen, and S. Fatima, *Characterization, Biomedical and Agricultural applications of Protease Inhibitors: A review.* Int J Biol Macromol, 2016.

26. Aune, D., *Soft drinks, aspartame, and the risk of cancer and cardiovascular disease.* Am J Clin Nutr, 2012. **96**(6): p. 1249-51.

27. Miyagi, Y., et al., *Plasma free amino acid profiling of five types of cancer patients and its application for early detection.* PLoS One, 2011. **6**(9): p. e24143.

Tables

Table 1. Characteristics of Men Diagnosed with Incident Prostate Cancer and Controls, Piedmont Triad Area, North Carolina, U.S., 1994-1996

Characteristics	Cases (n=108)	Controls (n=256)	P Value
Age	**66.85 (7.64)**	**66.08 (7.57)**	**0.35**
Race (%)			
White	**237 (85.56)**	**101 (84.87)**	
Black	**40 (14.44)**	**18 (15.13)**	**0.87**
BMI	**26.85 (3.85)**	**27/25 (3.51)**	**0.31**
Education (%)			
No College	**184 (65.95)**	**81 (67.5)**	
College	**95 (34.05)**	**39 (32.50)**	**0.16**
Smoking Status (%)			
Light	**79 (28.62)**	**35 (29.17)**	
Moderate	**171 (61.96)**	**69 (57.50)**	
Vigorous	**26 (9.42)**	**16 (13.33)**	**0.47**
Drinking Status (%)			
Never	**98 (35.64)**	**42 (35.59)**	
Past	**63 (22.91)**	**34 (28.81)**	
Current	**114 (41.45)**	**42 (35.59)**	**0.39**
Physical Activity Level (%)			
Light	**26 (9.63)**	**17 (14.41)**	
Moderate	**165 (61.11)**	**64 (54.24)**	
Vigorous	**79 (29.26)**	**37 (31.36)**	**0.29**
Income (%)			
Less than $20,000	**36 (13.58)**	**24 (20.69)**	
$20,000 to $50,000	**127 (47.92)**	**58 (50.00)**	
More than $50,000	**97 (36.60)**	**33 (28.45)**	
Don't Know	**5 (83.33)**	**1 (0.86)**	**0.19**
Family History (%)			
Yes	**42 (18.42)**	**186 (81.58)**	
No	**26 (32.10)**	**55 (67.90)**	**0.01**

Table 2. Multivariable-Adjusted Associations of Incident Prostate Cancer with Aspartic Acid

Exposure	n (Cases/Controls)	Matched OR (95% CI)*	Multivariable OR (95% CI)**
Tertiles			
1	83 (47/36)	1.0 (Reference)	1.0 (Reference)
2	95 (59/36)	0.80 (0.43, 1.47)	0.81 (0.44, 1.50)
3	186 (150/36)	0.31 (0.18, 0.57)	0.32 (0.17, 0.57)
***P* Trend**		<.0001	<.0001
Count per 1 SD	364 (256/108)	0.66 (0.52, 0.84)	0.64 (0.50, 0.82)

*Frequency matched variables: age, race

**Adjusted variables: body mass index, drinking status, family history of prostate cancer, physical activity

Table 3. Multivariable-Adjusted Associations* of Aspartic Acid with Incident Prostate Cancer According to Body Mass Index, Age, Drinking Status, and Physical Activity

Aspartic Acid per 1 SD	*n* (Cases/Controls)	Multivariable** OR (95% CI)	*P* Value for Interaction
BMI			
Normal	**86 (55/31)**	**0.42 (0.22, 0.79)**	
Overweight	**275 (200/75)**	**0.69 (0.52, 0.91)**	**0.23**
Age*			
<66.5	**180 (122/58)**	**0.58 (0.41, 0.84)**	
≥66.5	**184 (134/50)**	**0.68 (0.48, 0.97)**	**0.2**
Drinking status			
Never	**132 (94/38)**	**0.60 (0.39, 0.91)**	
Past or Current	**232 (162/70)**	**0.69 (0.52, 0.93)**	**0.13**
Physical Activity			
Light	**109 (73/36)**	**0.72 (0.47, 1.11)**	
Moderate	**206 (152/54)**	**0.66 (0.48, 0.92)**	
Vigorous	**40 (24/16)**	**0.45 (0.21, 1.00)**	**0.13**

***Age variable is split by median among controls**

****Adjusted variables: body mass index, drinking status, family history of prostate cancer, physical activity**

Figures and Figure Legends

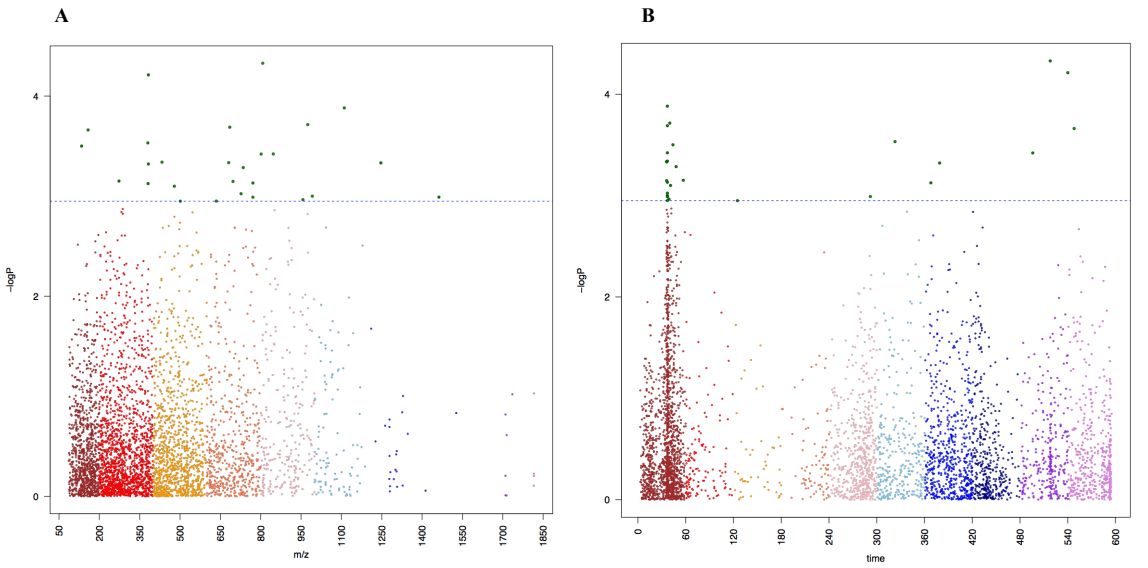
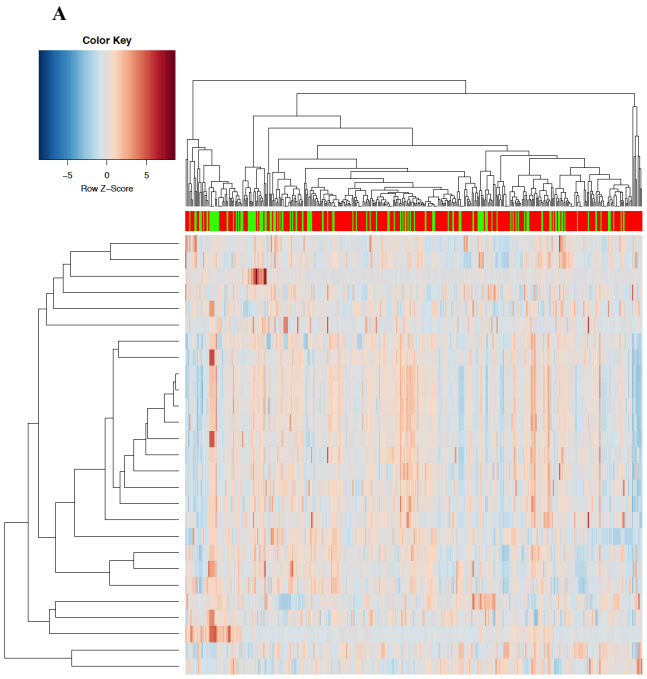


Figure 1. Metabolites that were significantly associated with prostate cancer. **A,** Type 1 Manhattan plot showing the negative log *P* (-log *P*) for each metabolite (m/z feature) as a function of the m/z (mass/charge). **B,** Type 2 Manhattan plot showing the –log P for each metabolite as a function of chromatographic retention time. m/z features above the dashed horizontal line are significant after FDR adjustment. The *blue dashed line* indicates FDR of 0.1 (Benjamini-Hochberg correction).



B

m/z	Putative Metlin Match
382.2705	Sphinganine-phosphate/PE-Cer(d15:1)
380.2431	Unknown
801.5817	Unknown
381.2585	Prostaglandin metabolite
382.2037	Unknown
807.3752	Unknown
974.7502	Ganglioside/Galabiosylceramide
726.8227	Unknown
770.851	Unknown
634.8755	Halogenated fatty acid
770.936	Unknown
1110.7869	Sphingolipid
684.8413	Unknown
991.7868	Triradylglycerols
846.8031	Unknown
432.9164	Unknown
956.8041	Phosphocholine metabolite (PC[24:0])
1246.76	Unknown
734.981	Unknown
696.8442	Unknown
680.8706	Unknown
272.9595	Unknown
1462.492	Unknown
158.0744	Unknown
500.9359	Inositol phosphate
478.8549	Unknown
134.0441	Aspartic acid

Figure 2. A, Heat map generated using 1-way hierarchal clustering. Metabolite intensities of the significant metabolites that were differentially expressed between prostate cancer cases and controls. Each *column* represents a participant, and each *row* represents a metabolite feature. The top 27 metabolites (raw $P < 0.05$) are shown. *Blue* hues indicate lower intensities, and *red* hues indicate higher intensities. **B,** Significant metabolite features by m/z and putative identity (METLIN Metabolomics Database).

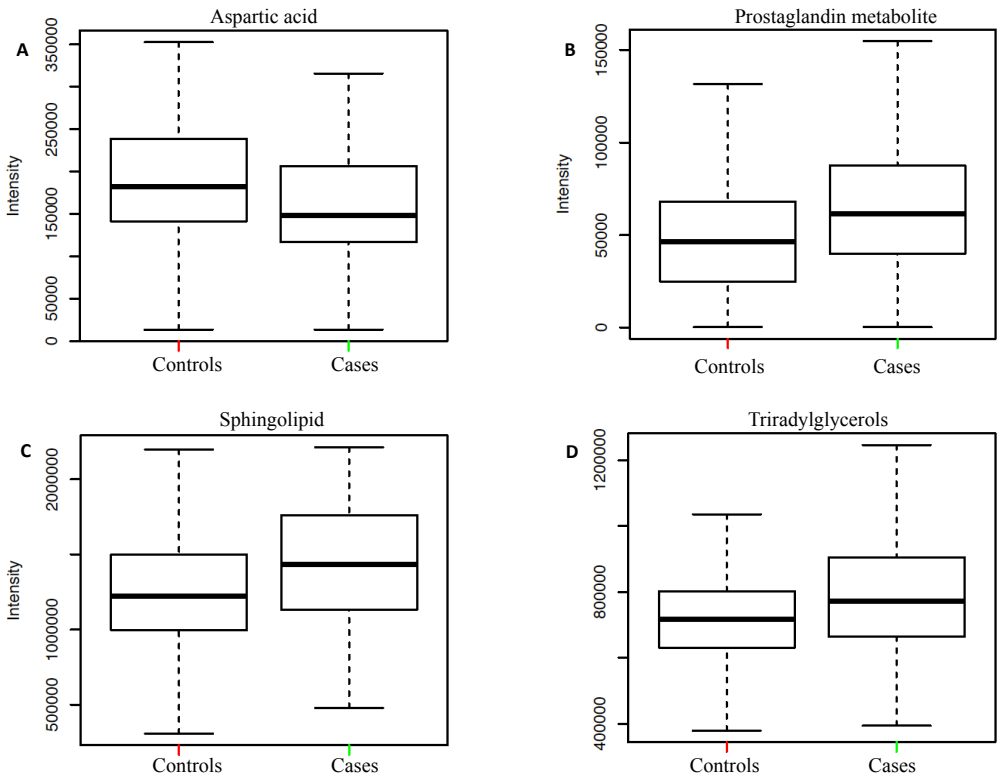
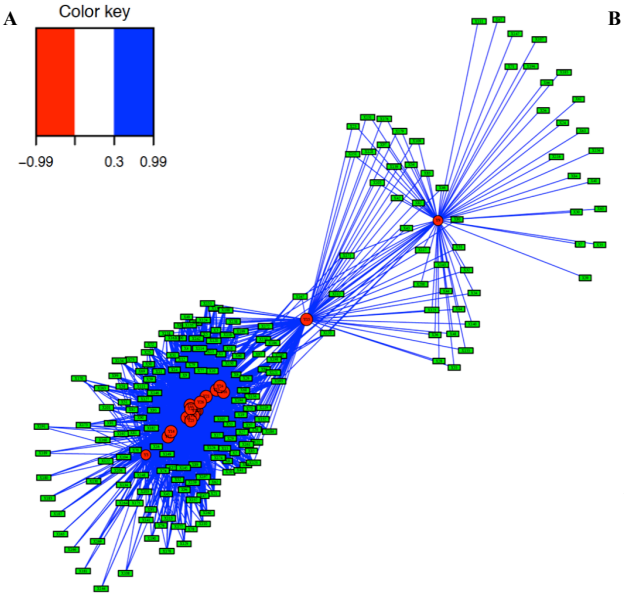


Figure 3. A-D, Examples of significant metabolite features shown in box blots.

A



B

Pathways	P Value
Vitamin E metabolism	0.00034
De novo fatty acid biosynthesis	0.00037
Fatty Acid Metabolism	0.00041
Fatty acid activation	0.00051
Glycosphingolipid metabolism	0.00059
Leukotriene metabolism	0.00073
Vitamin B5 - CoA biosynthesis from pantothenate	0.00085
Caffeine metabolism	0.00103
Polyunsaturated fatty acid biosynthesis	0.00153
Fatty acid oxidation, peroxisome	0.00153
CoA Catabolism	0.00153
Di-unsaturated fatty acid beta-oxidation	0.00153
Beta-Alanine metabolism	0.00158
Phytanic acid peroxisomal oxidation	0.00793
Urea cycle/amino group metabolism	0.00802
Omega-6 fatty acid metabolism	0.00896
O-Glycan biosynthesis	0.00896
Glycosphingolipid biosynthesis - neolactoseries	0.00896
Saturated fatty acids beta-oxidation	0.00896
Blood Group Biosynthesis	0.00896
Alkaloid biosynthesis II	0.00896
N-Glycan biosynthesis	0.01481
Putative anti-Inflammatory metabolites formation from EPA	0.02613
Histidine metabolism	0.02684
Keratan sulfate degradation	0.03675
Glutamate metabolism	0.0433
Lysine metabolism	0.04738

Figure 4. A, Network analysis of 27 significant metabolite features (correlation threshold=0.3, FDR threshold=0.2). Graphics generated using RStudio software, version 0.99.4466 and package xmsPANDA. **B,** Pathway enrichment of 27 significant metabolite features. Pathways identified with mummichog version 1.0.3.