METABOLOMIC SIGNATURE OF HYPERTENSION

By

Waqas T. Qureshi MD

A thesis submitted to Graduate Faculty of

WAKE FOREST UNIVERSITY GRADUATE SCHOOL OF ARTS AND SCIENCES

In Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

Clinical & Population Translational Sciences Program

May 2015

Examining Committee Members are:

Carl D. Langefeld, PhD (Chairman)

David M. Herrington, MD MHS

Lynne E. Wagenknecht, DrPH

Fang-Chi Hsu, PhD
ACKNOWLEDGEMENTS

I will like to thank the following for their support and guidance. Without them this submission would not have been possible.

First, I will like to thank my mentor and advisor, Dr. David Herrington for his immense interest and enthusiasm that he took in my work. He crafted my whole journey through the woods of science in a charismatic way. My words cannot express my appreciation.

Then I will like to thank Dr. Lynne E. Wagenknecht, who was always there to help me and guide me through writing this thesis. She has been very patient, thoughtful, and instrumental in providing me the data for my thesis work.

I will also like to thank Dr. Elsayed Z. Soliman, Dr. Alain Bertoni, Dr. Dalane Kitzman, Dr. Gregory Hundley, Dr. Joseph Yeboah, and Dr. Carlos Rodriguez who were always there to support and provide input during these years.

I will also like to thank Randi Sullivan, Allie Richardson, Karen Blinson and Georgia Saylor for their friendship and guidance that helped me navigate through difficult times while working this thesis.

Finally, I will like to thank my beautiful wife Fatima, my parents and my daughters Sara and Zaina for being patient and providing me their unrelenting love.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................... ii

ABSTRACT ................................................................................................................................... vi

MAIN BODY

1. CHAPTER I ................................................................................................................................. 1
   a. INTRODUCTION .................................................................................................................. 1
   b. HYPERTENSION – a metabolic disease ............................................................................. 5
   c. METABOLOMICS ............................................................................................................... 8
   d. META - ANALYSIS OF HUMAN METABOLOMICS STUDIES OF HYPERTENSION .......... 11
   e. KNOWLEDGE GAP .......................................................................................................... 18
   f. TABLES
      i. Table 1.1 ......................................................................................................................... 20
   g. FIGURES
      i. Figure 1.1 ......................................................................................................................... 21
      ii. Figure 1.2 ......................................................................................................................... 22
      iii. Figure 1.3 ......................................................................................................................... 23
      iv. Figure 1.4 ......................................................................................................................... 24
      v. Figure 1.5 ......................................................................................................................... 25
      vi. Figure 1.6 ......................................................................................................................... 26
      vii. Figure 1.7 ....................................................................................................................... 27
      viii. Figure 1.2 ...................................................................................................................... 28
   h. CONCEPTUAL MODEL ..................................................................................................... 29

2. CHAPTER II ................................................................................................................................ 30
   a. ABSTRACT .......................................................................................................................... 31
   b. INTRODUCTION ............................................................................................................... 32
3. CHAPTER III
   a. METABOLITE EXTRACTION, SAMPLING and HANDLING ...
   b. METABOLITE SEPARATION
c. METABOLITE DETECTION .................................................. 70

d. METABOLOMIC DATA ANALYSIS ....................................... 72

e. CONCLUSION ................................................................. 78

f. TABLES

i. TABLE 3.1 ............................................................................. 80

ii. TABLE 3.2 ............................................................................ 81

g. FIGURES LEGENDS ............................................................. 82

i. FIGURE 3.1 ........................................................................... 83

ii. FIGURE 3.2 ........................................................................... 84

iii. FIGURE 3.3 ......................................................................... 85

iv. FIGURE 3.4 .......................................................................... 86

v. FIGURE 3.5 ........................................................................... 87

vi. FIGURE 3.6 .......................................................................... 88

4. BIBLIOGRAPHY ............................................................... 89

CURRICULUM VITAE ............................................................... 102
ABSTRACT

Metabolomics is a comprehensive study of small molecular end products of metabolism (metabolites) in living organisms. These metabolites are directly modified by the metabolic processes that are governed by organism's genome, environment and diet. With advancement in technology, a large number of metabolites can be measured in various body fluids. The changes in metabolome precede disease phenotypes and thus, provide the niche for use of metabolomics in early detection and prevention of diseases. Several lines of evidence suggest that hypertension is closely associated with metabolic diseases and derangements. This thesis work was aimed at literature review of current metabolomic studies of hypertension and identified a knowledge gap regarding metabolomic signature of prevalent and incident hypertension. This was addressed by performing a metabolomic analysis of multi-ethnic cohort of Insulin Resistance Atherosclerosis Study. The advanced techniques utilized for sampling, extraction, detection and analyzing a metabolome were also then reviewed. Understanding these metabolic perturbations in hypertension may lead to development of effective prevention strategies for hypertension which currently inflicts ~900 million people worldwide.

Keywords: Metabolomics, metabolome, metabolic analysis, metabolites, nuclear magnetic resonance imaging, mass spectrometry, hypertension
CHAPTER I

“The greatest danger to a man with high BP lies in its discovery, because then some fool is certain to try and reduce it” – 1931 BMJ Editorial

1.1 INTRODUCTION

Even though we have come a long distance from the basic understanding of clinical impact of hypertension since 1931, hypertension remains the most prevalent risk factor of cardiovascular disease. The definition of hypertension has been ever evolving. It is described as a “progressive cardiovascular syndrome arising from complex and interrelated etiologies. Early markers of syndrome are often present before elevation of blood pressure is observed; therefore, hypertension cannot be classified solely by discrete blood pressure thresholds. Progression is strongly associated with functional and structural cardiac and vascular abnormalities that damage the heart, kidneys, brain, vasculature and other organs and lead to premature morbidity and death”.  However, much of epidemiological studies have defined hypertension by systolic blood pressure ≥ 140 mmHg, and/or diastolic blood pressure ≥ 90 mmHg, without being on an antihypertensive medication, or being on antihypertensive medication. Self – report of hypertension has been also used in various epidemiological studies.  

1.1.2 Epidemiology of Hypertension

In 2005, estimated 1 billion people suffered from hypertension worldwide. In 2014, estimated 79 million of United States adults (≥20 year old) suffered from hypertension. Its period prevalence has not shown any signs of improvement from 1999 – 2010. It directly and indirectly contributes to a quarter of ischemic strokes, three quarters of incident heart failure and almost half of cardiovascular related mortality. A 2009 US based study estimated $47.5 billion in direct costs and $100 billion in indirect costs for hypertension and hypertension related conditions. These costs are projected to increase to $343 billion by 2030.
From 1880 till 1980, a downward trend in the incidence of hypertension was observed.\textsuperscript{14} However, since 1980s there is an upward trend in diagnosis of hypertension\textsuperscript{15} which is now stable since 2000.\textsuperscript{16} The reason for this upward trend is not clear and is not well understood. There is also a difference in prevalence of hypertension by sex and race, disparately affecting men (33.3\%) more than women (30.4\%)\textsuperscript{6} and blacks (39.9\%) more than whites (30.3\%).\textsuperscript{17} However, the separation in prevalence observed between men and women narrows steeply after the age of 50.\textsuperscript{6} The effects of aging, sex, race and temporal trends are poorly understood. One of the current hypotheses for pathogenesis of hypertension is that these changes are observed due to differences in metabolism. The following sections will briefly review the literature implicating metabolic diseases and metabolism as a plausible biological mechanism underpinning hypertension.

\textbf{1.1.2 Pathogenesis of Hypertension}

Metabolic diseases such as diabetes\textsuperscript{18}, hyperparathyroidism\textsuperscript{19}, insulin resistance\textsuperscript{20,21} and obesity\textsuperscript{22,23} are strongly associated with hypertension. The common pathogenic pathway frequently suggested involves inflammation, age-related degenerative changes and genes. Since, metabolites play a major role in these diseases. Metabolites are small molecules (smaller than 1000 Da in size), which are breakdown products of carbohydrates, proteins, lipids, drugs, environmental pollutants, organic and inorganic molecules that are formed as a result of cellular metabolism. It is plausible that these metabolites may mediate the relationship of these diseases with hypertension. On the other hand, these might be bystanders and not the actual causative factors responsible for development of hypertension. Nevertheless, evidence regarding association of these metabolites with hypertension is invaluable to understand the causal pathways involved in the genesis of hypertension.
1.1.2.1 Inflammatory Hypothesis

Since 1970s, a great deal of research literature has been devoted to understand the relationship between inflammation and hypertension. Initial studies showed a compelling evidence of the role of adaptive immunity, mediated by T cells in the genesis of hypertension. In brief, these animal studies showed that the loss of T cells by thymectomy led to a blunted to hypertension causing agents; while experimentally restoring activated T cells led to development of hypertension. Apart from T-cells, white blood cells as a whole are also implicated in pathogenesis of hypertension. Silica, a toxin to monocytes (a type of white blood cells) has shown to reduce incidence of hypertension in Lyon rats. Monocytes are activated by angiotensin II (a hormone that is vital for development for hypertension) and increase the secretion of IL-1β. This chemokine then acts on T-cells and transforms them into activated T cells. These changes have been observed in hypertensive patients and have been confirmed by studying several inflammatory biomarkers such as TNFα, IL-1β, IL-6, and IL-8. In a more recent study, transfer of these activated T cells led to blunting of endothelial dependent vasodilation which is a marker of poor vascular health and a precursor of hypertension. Yet another study linked C-reactive protein (biomarker of inflammation) with hypertension in obese individuals. At cellular level, NF-κB mediates the effects of inflammation leading to endothelial dysfunction and hypertension. Abnormalities in NF-κB are found to occur in hypertension. Interestingly, many of these changes occur with aging and accelerated vascular aging is associated with higher degree of inflammation in the body. Even though age is considered an essential confounder in most of the analyses, it is still not clear that what happens with aging which cannot be attributable to known risk factors of hypertension.
1.1.2.2 Aging Hypothesis

Vascular aging and stiffness have been postulated as a precursor of hypertension. Accelerated vascular aging has been observed in metabolic diseases such as diabetes\(^46\) and metabolic syndrome.\(^47\) Arterial aging is shown to retard with better control of blood pressure and metabolic diseases such as diabetes suggesting a role of metabolites in arterial aging.\(^48\) Apart from inflammation caused by metabolic diseases,\(^49-52\) higher degree of oxidative stress is also attributable to vascular aging.\(^53\) Oxidation and formation of reactive oxygen species is a part and parcel of many metabolic pathways and occurs as a result of normal body metabolism. However, excess oxidation may occur in certain metabolic diseases.\(^54-57\) Oxidative stress can lead to telomere dysfunction,\(^58-60\) which has shown to be implicated in development of hypertension. A comprehensive gene co-expression network analysis implicated genes related to leptin, superoxide dismutase 1 (SOD1), mitogen activated phosphokine, STAT3 (gene involved in cellular apoptosis), and ICAM1 involved in development of hypertension.\(^56\) Aging leads to increased oxidative stress and hence causes activation of these genes.\(^61\)

1.1.2.3 Genetic Hypothesis

Initial discovery of hypertension as a heritable disease (estimated 30 – 50% heritability)\(^62\) prompted large genome wide association studies (GWAS) which led to identification of several loci associated with risk of hypertension.\(^63, 64\) However, further findings from Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and the Global Blood Pressure Genetics (Global BPgen) GWAS, these loci explained only 2% of genetic influence on blood pressure, with an estimated effect of just 1 mmHg. On the other hand, some studies failed to find a genome wide association with hypertension.\(^65, 66\)

Nevertheless, there exists some evidence regarding monogenic Mendelian inheritance of hypokalemic hypertension.\(^67\) In hypokalemic hypertensive patients,
genes associated with steroid (CYP17A1) and alcohol metabolism (ALDH2) were demonstrated to be associated with hypertension. However, definitive metabolic pathways thought to be associated with hypertension related to these genes were not identified in this study.

Since, the contribution of genetics is relatively small in development of hypertension; it has been suggested that players further downstream in genetic flow of information play an important role in pathogenesis of hypertension. Furthermore, data from network analysis of various genes suggests involvement of metabolic pathways common to many genes in development of hypertension which are shown in Figure 1.1.

1.2. HYPERTENSION – A metabolic Disease

The biological dogma since 1950s declares that biological information transfers from genome (DNA) to transcriptome (mRNA) which translates into proteome. There are about 2 x 104 genes that lead to transcription of more than 1 million transcripts which in turn produce >1 million proteins. These proteins then modified proteins. Some of these proteins are acted upon by the enzymes which are proteins themselves and cause changes in the metabolites at cellular levels. Eventually these changes signify and lead to a particular phenotype. The nutritional status and lifestyle also affects these interactions. A graphical depiction of this process is shown in Figure 1.2. However, in biological systems this is not as simple as the cartoon shows. There are multiple interlinking pathways which are complex and are affected by the intermediate –omics.

We review here further indirect evidence from transcriptomic, proteomic and dietary studies where metabolic pathways are linked with hypertension. A brief overview of these studies is given below:
1.2.1 Evidence from Transcriptomic Studies

As we go downstream of the genetic information, transcription occurs. Transcription in contrast to genes is a dynamic process. There is a dearth of information available from transcriptomics regarding pathogenesis of hypertension. In 2003, Philip-Couderc et al. for the first time showed that components of electron transport chain (cytochrome c oxidase subunit I, subunit VII, NADH dehydrogenase and ATP synthase F0 subunit) as well as voltage-dependent anion-selective channel protein 2 (VDAC 2) are associated with hypertension in obese dogs. The electron transport chain proteins are coded by mitochondrial DNA and thus, this study suggested involvement of mitochondrial transcriptome in hypertension.

In an experimental rat study, eleven congenic strains with introgressed segments spanning <81.8 kb to <1.33 Mb were developed by introgressing genomic segments of RNO9 from the Dahl salt-resistant (R) rats onto the genome of the Dahl salt-sensitive (S) rats. The congenic strain with the shortest introgressed segment spanning <81.8 kb significantly lowered BP of the hypertensive S rat by 25 mmHg and significantly increased its mean survival by 45 days.

In another interesting transcriptomic study, transcriptome of nucleus tractus solitarii (a collection of neurons in brain that regulates blood pressure) of Wistar Kyoto rats was examined. One group underwent exercise training and the was untrained. Mapping transcriptome using KEGG pathways suggested that 5 genes were differentially expressed in exercise trained rats from untrained rats and are associated with metabolic pathways. The study showed that protein tyrosine phosphatase and histamine related metabolic pathways were important in determining blood pressure.

1.2.2 Evidence from Proteomic Studies

Proteins are formed downstream of transcription in the genetic flow of information. The evidence from proteomics studies is still limited and is underdevelopment. A
small pilot proteomic analysis demonstrated involvement of proteins in lipid metabolism and transport in hypertension. Another proteomic study showed that mitochondrial chaperones, pyruvate dehydrogenase complex subunits, enzymes of Kreb’s cycle, and ketone body oxidation proteins were found to be increased in hypertensive rats. In another study of platelet proteome, angiotensin –II related hypertension was associated with protein degradation in platelets. These proteins were mainly cytoskeletal proteins and the study suggested that these can be used as a biomarker for development of hypertension.

1.2.3 Evidence from Dietary Studies

A link between diet and hypertension has been extensively studied since early 20th century. Sodium restriction and rice diet have been used to treat hypertension since 1940s. However, it was not until 1980s, when various constituents of diet other than sodium were evaluated as risk factors for hypertension. Diet with more vegetables and less fat was found to be associated with lower risk of hypertension. Vegetarian and low fat diet was increasingly used as a therapy for hypertension in 1980s. The initial evidence was limited to antihypertensive effects of low fat diets which extended to low carbohydrate and high protein diets.

The dietary recommendations from various nutritional societies have been ever evolving and the most recent 2013 American Heart Association/American College of Cardiology guidelines recommends with highest level of recommendation (Level I, Evidence level A) that foods with high omega 3 fatty acid content (olives) can reduce systolic/diastolic blood pressure by 6-7/2-3 mmHg and DASH diet (diet with low sodium) by 5-6/3 mmHg. This is backed by the recent slew of literature showing consistent results with reduction in cardiovascular risk and hypertension. Individual dietary factors have also been studied in this regard and some omega 3 fatty acids such as eicosapentanoic acid are found to be associated with increased systemic arterial compliance, arterial inflammation, and insulin resistance.
However, until now no trials that have attempted to supplement pure formulations of these individual nutrients has shown to improve blood pressure. These lines of evidence thus suggest further confirming the relationship of mitochondrial dysfunction with hypertension. Metabolome is directly influenced by transcriptome and is the next downstream step in the biological systems (Figure 1.2). As it known that the main metabolites influenced by mitochondrion are fatty acids, acylcarnitines, and amino acids, the next logical step is to evaluate the role of metabolites in pathogenesis of hypertension.

1.3. METABOLOMICS

As identified earlier from studies given above, there is a strong indication towards studying various metabolites in the context of hypertension. Additionally, it is important to understand the types of metabolites deranged in hypertensive individuals as this may help in developing possible therapeutic strategies to prevent hypertension.

In a way, metabolites are the eventual downstream products of genome, transcriptome and proteome. However, metabolism is not just affected by genes but also by diet and environment. This places metabolome at the juncture of interaction of diet, gene, and environment (Figure 1.3). This also places it in proximity to a specific phenotype. In other words, a metabolomic signature can define specific phenotype of disease. Information about metabolome may not only provide information regarding the diagnosis of a particular phenome but may also influence our understanding of various metabolic diseases and novel unexplored pathways that may carry important therapeutic and preventive avenues. Current epidemiological studies have reached a standstill where identifying new risk factors may not provide incremental diagnostic or prognostic value due to presence of these risk factors in the pathways of already known risk factors. Hence, exploration of metabolome may provide orthogonal risk factors. With the advancement in high
throughput technologies, metabolites in an organic sample may be determined with high accuracy.

1.3.1 Definition

A new field emerged in 1998, when metabolomics was defined by Nicholson et al. as “quantitative measurement of the dynamic multi-parametric metabolic response of living systems to pathophysiologic stimuli or genetic modification’. Metabolomics enables assessment of a broad range of endogenous and exogenous metabolites with potential impact on the investigation of physiologic status, diagnosis of disease, discovery of biomarkers and identification of perturbed biochemical pathways”.

The field has rapidly expanded and mainly can be broadly divided into:

1. Untargeted metabolomics
2. Targeted metabolomics

Untargeted metabolomics is a comprehensive agnostic chemometric approach to evaluate small metabolite molecules that include unknown and known metabolites. Targeted metabolomics on the other hand is a quantitative analysis of a set of chemically known annotated metabolites. In targeted metabolomics, already known pathways are explored and usually a prior hypothesis is present. On the other hand, the advantage of untargeted metabolomics is that it helps in identifying novel metabolites and pathways without a prior hypothesis.

1.3.1.1 Advantages

Metabolomic profiling provides a prospect of identifying novel biomarkers. Changes in metabolome can help us understand various biological processes and this information can be integrated with complements of genome, proteome and transcriptome to provide a system biology approach towards a particular phenotype. The current paradigm suggests that biochemical changes in the body precede the development of subclinical disease and hence, clinical disease.
Information obtained by complement of metabolites may prove to be a piece missing from our explanation of many cardiometabolic diseases.

1.3.1.2 Limitations

Despite the important position of this field, it faces several hindrances. Due to the dynamic nature of metabolome, along with many sources of variation such as intra- and inter-sample variation, experimental conditions, methods of sample preparation and separation, the reproducibility of metabolomic experiment is an issue. In silico and manual quality control is performed to enhance reproducibility and accuracy but these methods are still limited. Various methods of chromatography have been utilized to improve metabolite detection with high accuracy; however there is no single mode of chromatography that can comprehensively separate the whole metabolome in a single run. These issues become paramount in large scale epidemiological studies where standardization of methods and risks of instrumental/temporal drifts are not trivial.\textsuperscript{96, 97} Even if the samples are evaluated at a central laboratory, results obtained from different time periods cannot be directly compared.\textsuperscript{98} This occurs even if same methods for detection of metabolites are employed. Finally, once metabolites have been detected, many of these metabolites are still unknown and their biological role is not well characterized.

Although metabolomics share many of the features of its predecessors –omics fields, it differs from genomics and proteomics in a much more fundamental way. First, genomics and proteomics commonly utilize a methodology to amplify the signal by some kind of PCR related amplification. Metabolites cannot be amplified by any method. Secondly, the structure of proteins and genes is usually a polymer of already known simplified subunits (aminoacids, and DNA). Metabolites differ widely in their structure and shape. Moreover, the dynamic nature of metabolites increases the variance in development of predictive models and hence, statisticians with
experience in metabolomics may play an important role in analyzing information from the data. Once the results are available, a comprehensive understanding of metabolic pathways is also important which demands the involvement of a biochemist in interpretation of findings. Thus, this field provides the next step in the intellectual advancement of science that requires collaboration of scientists from different fields at multiple levels.

1.4. META-ANALYSIS OF HUMAN METABOLOMIC STUDIES OF HYPERTENSION

There are a number of targeted and untargeted metabolomic studies performed in humans that have implicated various metabolites in pathogenesis of hypertension. Knowing about these metabolites and their related pathways may help in identifying patients at risk of hypertension. Knowing about risk of hypertension may facilitate management of individuals at high risk of hypertension. We performed a detailed review of both targeted and untargeted metabolomic studies that evaluated the risk for hypertension associated with metabolites.

1.4.1 METHODS

We performed a systematic review and meta-analysis of published studies in literature.

Types of studies

All observational or experimental untargeted metabolomic human studies assessing the relationship of metabolites with hypertension that have provided some form of effect size were identified. No restrictions were imposed for language, publication status or published date. Targeted metabolomic studies were excluded.

Types of participants

Humans of any age, sex or ethnicity were considered to be included in this meta-analysis.
**Type of outcome**

For a study to be included, it had to report either hypertension or blood pressure as an outcome.

**Source of data collection**

We used electronic searches of scientific literature databases such as MEDLINE (as PubMed), EMBASE, ISI Web of Knowledge, and Google Scholar (Table 1.1). Moreover, we searched for various conference abstracts from their websites as well as contacted authors of known metabolic studies and study groups to identify unpublished and published work in any format. This search was performed from date of inception until December 2014.

Search terms using key subjects in Boolean combinations were “metabolome”, “metabolomics”, “metabolome AND analysis”, metabolome AND hypertension”, “metabolome AND blood pressure”, “metabolomics AND hypertension”, “metabolomics AND blood pressure”, “metabolome AND pre-eclampsia”, “metabolomics AND pre-eclampsia”, “metabolome AND eclampsia”, “metabolomics AND eclampsia”, “metabolic profiling”, “metabolic profiling AND hypertension”, “metabolic profiling AND blood pressure”, “metabolic profiling AND pre-eclampsia”, and “metabolic profiling AND eclampsia”. The full text of original articles and abstracts was reviewed. The reference lists of original articles as well as reviews were scrutinized for identifying additional articles. Studies only qualified if they used a well-established metabolomic analysis methodology to identify and annotate individual metabolites. **Figure 1.4** shows the flow diagram of study selection. We identified 5 studies that fulfilled the above given study criteria.

**Statistical analysis**

The primary analyses were to quantify pooled associations of various metabolic groups (lipids, carbohydrates, amino acids and others) with hypertension. Positive and negative associations were assessed separately. The effect sizes were calculated
as odds ratios for each metabolite. The effect sizes were obtained from the fully adjusted models, if available. Summary statistic was pooled odds ratio which was calculated in a random effects model due to heterogeneity present in the studies. The reason for using summary statistic was to identify a metabolomic signature associated with hypertension; despite the metabolites were different in nature. Heterogeneity was calculated using $I^2$ statistic (with 95% confidence interval). $I^2$ statistic represented the percentage of total variation in the pooled odds ratio that can be attributed to heterogeneity in study characteristics rather than chance. Heterogeneity was graded as low (<25%), moderate (25 – 75%) and high (>75%). We also estimated the risk of bias by subjectively evaluating the funnel plot of standard error of logarithm of odds ratio for asymmetry. We performed these analyses using Comprehensive Meta-Analysis Software v. 2.2.064 (Biostat Inc, Englewood, New Jersey, USA) and p-value of <0.05 was considered significant.

1.4.2 RESULTS

In this meta-analysis of 5 untargeted metabolomic studies, we evaluated the association of various metabolites with hypertension. One study used urinary metabolome, while 3 studies used plasma and 1 study used serum metabolomics. Incident hypertension was evaluated in one study while other studies were cross-sectional in design. We found that the pooled odds ratio of all the positively associated metabolites was 3.35, 95% CI 2.38 – 4.70, p <0.001. The pooled odds ratio of negatively associated metabolites was 0.15, 95% CI 0.08 – 0.27, p <0.001. The forest plots for positively and negatively associated metabolites are shown in Figure 1.5 and Figure 1.6, respectively. There was bias Heterogeneity $I^2$ statistic = 95%. The funnel plot was symmetric and by visualization no bias could be detected (Figure 1.8).

The metabolites positively associated with hypertension included; Alanine, methylconitate, androgen sex steroids, lactic acid, acetone, phenyllactate,
hexonylcarnitine. The metabolites negatively associated with hypertension were mainly amino acids and their derivatives (phenylalanine, valine, methylhistidine, isovalerate, lysine, phenylglyoxylic acid, 2-hydroxyglutarate, hydoxyproline, glucronate, gentisate), Kreb’s cycle intermediates (fumarate, lactose,), gut microbial end-products (formate, hippurate) and carnitines.

1.4.3 SYSTEMATIC REVIEW OF METABOLOMIC STUDIES

A review of the analyzed metabolomic studies as well as metabolomic studies that did not qualify for meta-analysis is given below:

1.4.3.1 Targeted Metabolomic Studies

One of the first study targeted metabolomic study showed that differences in lipoprotein particle composition can very well discriminate normal blood pressure (<130mmhg) from pre hypertension (131 – 149 mmHg) and hypertension (>149 mmHg). Among the lipoproteins, very low density lipoproteins (VLDL) and low density lipoproteins (LDL) tended to provide this information more than other lipoprotein particles. However, interestingly these particles were not able to differentiate between prehypertension and hypertension groups suggesting that the metabolic derangement might have already occurred in the prehypertension range. These findings agree with many epidemiological studies that have shown lipoprotein composition and particle size to be directly related with arterial stiffness.

Another cross-sectional study showed association of serum fatty acids with hypertension. This study identified oleic acid, nonanoic acid, ecosanoic acid, hexaenoic acid and hepatonic acid to be directly associated with hypertension. Serum fatty acids are proposed to increase vascular tone by increasing the sympathetic tone. Also, they may influence cell membrane phospholipids that directly affect interaction of sodium and calcium influx and efflux leading to increase vascular tone.
The earliest targeted metabolomic studies aimed at identifying relationship of hypertension and metabolites were performed in women with pre-eclampsia. Although pregnant women are expected to have a different metabolome, pre-eclampsia is a considered a subclass of hypertensive disorders. Evidence also suggests metabolic perturbations involved in pathogenesis of pre-eclampsia and eclampsia. 110

Since these disorders closely resemble essential hypertension, we have also reviewed some of the targeted metabolic profiling studies in patients with these conditions. In a nested case-control study of 60 pre-eclampsic women, plasma samples analyzed using ultraperformance liquid chromatography – mass spectrometry detected a metabolic signature of 14 metabolites predicting pre-eclampsia with receiver operating characteristic curve area under the curve (AUC)=0.94. The combination of 5 – hydroxytryptophan (amino acid), methylglutaric acid (dicarboxylic acid) were associated with lower risk of hypertension while, oleic acid (fatty acid), docosahexaenoic acid (fatty acid), monosaccharides (carbohydrate), oxolan-3 (fatty acid), 2-oxovaleric acid (fatty acid), acetoacetic acid (ketone body), decanoylcarnitine (carnitine), hexadecenoyleicosatetraenoyl-sn-glycerol (lipid), di-octadecadienoyl-sn-glycerol (lipid), sphingosine – 1 – phosphate (phospholipid), sphinganine – 1 – phosphate (phospholipid) and vitamin D3 derivatives (steroids) was associated with 23 times increased odds of pre-eclampsia in this study. 111 Most of these derangements were related to lipid and phospholipid metabolism thus suggesting involvement of lipid oxidation and breakdown in development of pre-eclampsia. A similar analysis of in a case-control study of 41 women, serum liquid chromatography – mass spectroscopy (LC-MS) system showed that hydroxyhexanoylcarnitine, alanine, phenylalanine and glutamate were present in significantly higher amounts in pre-eclampsia cases. 112 A combination of these metabolites predicted pre-eclampsia accurately [AUC (95% CI) 0.82 (0.80 – 0.85)]. These results were replicated by using nuclear magnetic
resonance based untargeted metabolomic analysis of sera of 30 cases and 59 controls where investigators were able to differentiate between early – versus late-presentation of pre-eclampsia based on glycerol, carnitine, methylhistidine and acetone. The investigators reported the 100% specificity and 76.6% sensitivity of using this metabolomic signature as predictor of late onset pre-eclampsia.

1.4.3.2 Untargeted Metabolomic Studies

By far, the largest piece of evidence regarding involvement of small metabolites in pathogenesis of hypertension comes from an untargeted urinary human metabolomics study aimed to identify metabolites related with hypertension. This study was INTERnational collaborative study on Macro/micronutrients And blood Pressure (INTERMAP) that enrolled 4630 men and women ages 40-59 years randomly selected from 17 centers in United States, United Kingdom, Peoples’ Republic of China, and Japan. Twenty-four hour urine samples were collected and $^1$H-nuclear magnetic resonance ($^1$H-NMR) spectroscopy was performed. This cross-sectional study observed that alanine was directly associated with hypertension, while formate and hippurate were inversely associated with hypertension. Alanine alone was responsible for 2.69 mmHg of elevation of systolic blood pressure for each standard unit increase in its concentration. For each standard deviation increase in formate and hippurate, a reduction of 1.19 and 2.1 mmHg of systolic blood pressure occurred. Formate is a byproduct of fermentation of dietary fiber produced by gut microbiome, while hippurate is formed by anaerobic degradation of aromatic compounds. Alanine is a degraded product of proteins present in meat and hence was found to be positively associated with elevated blood pressure. Zheng et al in black subset of Atherosclerosis Risk in Communities (ARIC) study cohort demonstrated that a standard deviation increase in 4–hydroxyhippurate was independently associated with 17% higher risk of incident hypertension over a period of 10 years. They also identified a principal component of sex steroids,
branched chain amino acids and other amino acid which was weakly associated with incident hypertension (p trend = 0.03). The sex steroid metabolites included 5α – androstan- 3ß, 17ß diol sulfate, androsterone sulfate, and epiandrosterone sulfate. These were individually associated with 17%, 15% and 14% increased risk of hypertension after adjusting for age, sex, leisure-time physical activity, alcohol intake, current cigarette smoking status, prevalent diabetes mellitus, body mass index and glomerular filtration rate. The differences in blood pressure observed between blacks and whites were only studied in one study where the differences in urinary metabolites failed to explain the elevated blood pressure in blacks versus whites. Rather a difference in diet explained most of the difference observed for blood pressure.

Another study used a novel ¹H –NMR – based algorithm called adaptive intelligence binning algorithm to identify bin edges in existing bins of NMR spectrum. This algorithm enables identification of low intensity metabolites. This study identified α1 – glycoprotein and choline to be associated with hypertension. A recent case-control plasma ¹H – NMR spectroscopy study of 157 hypertensive patients and 99 controls in Uygur patients, Zhong et al demonstrated 12 metabolites to be significantly different between both groups. Compared to controls, hypertensive patients were more likely to have higher levels of VLDL, LDL, lactic acid and acetone but less likely to have valine, alanine, pyroracemic acid, inose, p – hydroxyphenylalanine, and methylhistidine. A much smaller urinary gas chromatography mass spectrometry (GC-MS) metabolomic study of 25 black South African males enrolled in Sympathetic activity and Ambulatory Blood Pressure in Africans (SABPA) cross-sectional study found elevated levels of lactate, fumarate, 4-hydroxyphenyllactate, and 2-hydroxyvaleric acid. This study also carried out liquid chromatography mass spectrometry (LC-MS), which showed decreased levels of methyluric acid. The study suggested that all these changes might be related to increased NADH/NAD+ ratio, which resulted in decreased Kreb’s cycle turnover and
hence, relative increase in all the metabolites feeding into it such as those that were found to be elevated in the prior studies. Even though current literature did not support altered mitochondrial metabolism due to chronic alcohol as a major cause of hypertension, the authors of study generated hypothesis that chronic alcohol consumption may be a major driver of these observed metabolite differences due to their involvement in metabolic pathway of alcohol metabolism. A schema of these derangements with affected metabolites as bold is given in **Figure 1.8**.

Another LC-MS analysis of venous effluent from kidneys of renal artery stenosis in 22 individuals showed that glutarate, hydroxyproline, glucuronate, gentistate, glycerol, lactose, 4:0 carnitine, 9:0 carnitine, homogentisate, indoxyl sulfate, 1-methylnicotinamide and 3-ureidopropionic acid were elevated in hypertensives.\textsuperscript{121} These metabolites mainly represent the amino acid intermediaries and fatty acid synthesis related intermediates. More evidence regarding mitochondrial dysfunction in hypertensives comes from an electron spray ionization tandem mass spectrometry study of urinary metabolites of 202 white and black individuals which showed association of long chain acylcarnitines with systolic blood pressure ($R^2 = 0.38; \beta = 0.17; p = 0.005$).\textsuperscript{122}

### 1.5 KNOWLEDGE GAP

Based on the above literature review, we identified several knowledge gaps in the literature which include:

1. Metabolome of a multi-ethnic population has not been examined in the past thus raising issues of generalizability
2. There was only one study that has evaluated relationship of metabolites with incident hypertension. Other studies have only evaluated cross-sectional association of hypertension and metabolites
3. Many studies are limited by their small sample size
4. The largest studies were both $^1$H – NMR spectroscopy based and hence are limited due to inability to evaluate the lipid fraction of metabolome in detail. This fraction is mostly composed of fatty acids and has been shown in the past to be a major driver of insulin resistance as well as hypertension. It is not known if particular metabolites in a specific combination are associated with hypertension independent of traditional risk factors.

Having these limitations and vast knowledge gaps identified from the above studies, we aimed to examine the association of metabolites including fatty acids, amino acids, bile acids, acylcarnitines and sterols with incident as well as prevalent hypertension using multivariate analysis in the participants of Insulin Resistance Atherosclerosis Study (IRAS). This will improve upon our knowledge of metabolic pathways and may help in development of effective preventive and therapeutic strategies.
# Table 1.1: Search terms used and number of articles (last accessed 01/26/2015)

<table>
<thead>
<tr>
<th>Search Term</th>
<th>MEDLINE</th>
<th>EMBASE</th>
<th>WebofScience</th>
<th>GoogleScholar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolome</td>
<td>4,800</td>
<td>2,312</td>
<td>5,322</td>
<td>30,800</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>11,478</td>
<td>4,231</td>
<td>15,325</td>
<td>40,000</td>
</tr>
<tr>
<td>Metabolome AND analysis</td>
<td>3,455</td>
<td>2,351</td>
<td>5,423</td>
<td></td>
</tr>
<tr>
<td>Metabolome AND hypertension</td>
<td>40</td>
<td>27</td>
<td>54</td>
<td>19,100</td>
</tr>
<tr>
<td>Metabolome AND blood pressure</td>
<td>137</td>
<td>65</td>
<td>67</td>
<td>23,900</td>
</tr>
<tr>
<td>Metabolomics AND hypertension</td>
<td>96</td>
<td>54</td>
<td>100</td>
<td>20,900</td>
</tr>
<tr>
<td>Metabolomics AND blood pressure</td>
<td>273</td>
<td>167</td>
<td>290</td>
<td>28,100</td>
</tr>
<tr>
<td>Metabolome AND pre-eclampsia</td>
<td>7</td>
<td>3</td>
<td>12</td>
<td>786</td>
</tr>
<tr>
<td>Metabolomics AND pre-eclampsia</td>
<td>19</td>
<td>9</td>
<td>42</td>
<td>784</td>
</tr>
<tr>
<td>Metabolic profiling AND hypertension</td>
<td>151</td>
<td>103</td>
<td>212</td>
<td>45,100</td>
</tr>
<tr>
<td>Metabolic profiling AND blood pressure</td>
<td>307</td>
<td>238</td>
<td>361</td>
<td>66,700</td>
</tr>
<tr>
<td>Metabolic profiling AND pre-eclampsia</td>
<td>23</td>
<td>12</td>
<td>46</td>
<td>16,800</td>
</tr>
<tr>
<td>Metabolic profiling AND eclampsia</td>
<td>23</td>
<td>13</td>
<td>43</td>
<td>3,350</td>
</tr>
</tbody>
</table>
FIGURES

Figure 1.1: Genes implicated in pathogenesis of hypertension with interconnections with cardiovascular disease
Figure 1.2: Relationship of genome with other omics (Adopted with permission from Wang and Gerszten, *Nature*, 2008)
Figure 1.3: System biology network (Adopted with permission from Adamski et al. Curr Opinion Biotech 2013; 24:39-47)
Figure 1.4: Study selection for systematic review
Figure 1.5: Directly associated metabolites with hypertension

<table>
<thead>
<tr>
<th>Study name</th>
<th>Metabolites</th>
<th>Odds ratio</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holmes 2008i</td>
<td>alanine</td>
<td>6.453</td>
<td>1.435</td>
<td>29.024</td>
<td>2.430</td>
<td>0.015</td>
</tr>
<tr>
<td>Holmes 2008ii</td>
<td>methyl nicotinate</td>
<td>2.923</td>
<td>0.741</td>
<td>11.532</td>
<td>1.531</td>
<td>0.126</td>
</tr>
<tr>
<td>Zheng 2013i</td>
<td>hydroxyhippurate</td>
<td>1.180</td>
<td>1.080</td>
<td>1.290</td>
<td>3.652</td>
<td>0.000</td>
</tr>
<tr>
<td>Zheng 2013ii</td>
<td>5α-androstan-3,17β-diol disulfate</td>
<td>1.120</td>
<td>1.002</td>
<td>1.252</td>
<td>1.991</td>
<td>0.046</td>
</tr>
<tr>
<td>Zheng 2013iii</td>
<td>androsterone sulfate</td>
<td>1.120</td>
<td>0.998</td>
<td>1.257</td>
<td>1.922</td>
<td>0.055</td>
</tr>
<tr>
<td>Zheng 2013iv</td>
<td>epiandrosterone sulfate</td>
<td>1.120</td>
<td>0.989</td>
<td>1.269</td>
<td>1.784</td>
<td>0.074</td>
</tr>
<tr>
<td>Zhong 2014i</td>
<td>acetone</td>
<td>31.758</td>
<td>17.126</td>
<td>58.893</td>
<td>10.975</td>
<td>0.000</td>
</tr>
<tr>
<td>Zhong 2014ii</td>
<td>acetylformate</td>
<td>18.969</td>
<td>10.668</td>
<td>33.731</td>
<td>10.021</td>
<td>0.000</td>
</tr>
<tr>
<td>Zhong 2014iii</td>
<td>lactate</td>
<td>6.549</td>
<td>3.959</td>
<td>10.835</td>
<td>7.317</td>
<td>0.000</td>
</tr>
<tr>
<td>SABPA 2014i</td>
<td>lactate</td>
<td>6.041</td>
<td>1.531</td>
<td>23.835</td>
<td>2.568</td>
<td>0.010</td>
</tr>
<tr>
<td>SABPA 2014ii</td>
<td>phenyl lactate</td>
<td>4.517</td>
<td>1.145</td>
<td>17.824</td>
<td>2.153</td>
<td>0.031</td>
</tr>
<tr>
<td>SABPA 2014iii</td>
<td>hexyl carnitine</td>
<td>6.583</td>
<td>1.668</td>
<td>25.973</td>
<td>2.691</td>
<td>0.007</td>
</tr>
<tr>
<td>Rhe 2014i</td>
<td>methyl nicotinate</td>
<td>4.870</td>
<td>1.577</td>
<td>15.043</td>
<td>2.752</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Pooled odds ratio (Random Effects) 3.346 2.381 4.702 6.956 0.000
Figure 1.6: Inversely associated metabolites with hypertension

<table>
<thead>
<tr>
<th>Study name</th>
<th>Metabolites</th>
<th>Odds ratio</th>
<th>Lower limit</th>
<th>Upper limit</th>
<th>Z-Value</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holmes 2008i</td>
<td>formate</td>
<td>0.180</td>
<td>0.040</td>
<td>0.814</td>
<td>-2.227</td>
<td>0.026</td>
</tr>
<tr>
<td>Holmes 2008ii</td>
<td>hippurate</td>
<td>0.259</td>
<td>0.061</td>
<td>1.099</td>
<td>-1.832</td>
<td>0.067</td>
</tr>
<tr>
<td>Zhong 2014i</td>
<td>alanine</td>
<td>0.306</td>
<td>0.191</td>
<td>0.490</td>
<td>-4.931</td>
<td>0.000</td>
</tr>
<tr>
<td>Zhong 2014ii</td>
<td>valine</td>
<td>0.003</td>
<td>0.001</td>
<td>0.007</td>
<td>-13.524</td>
<td>0.000</td>
</tr>
<tr>
<td>Zhong 2014iii</td>
<td>methyhistidine</td>
<td>0.269</td>
<td>0.167</td>
<td>0.433</td>
<td>-5.408</td>
<td>0.000</td>
</tr>
<tr>
<td>Zhong 2014iv</td>
<td>phenylalanine</td>
<td>0.178</td>
<td>0.108</td>
<td>0.292</td>
<td>-6.840</td>
<td>0.000</td>
</tr>
<tr>
<td>SABPA 2014i</td>
<td>fumarate</td>
<td>0.250</td>
<td>0.066</td>
<td>0.949</td>
<td>-2.037</td>
<td>0.042</td>
</tr>
<tr>
<td>SABPA 2014ii</td>
<td>isovalerate</td>
<td>0.220</td>
<td>0.058</td>
<td>0.835</td>
<td>-2.225</td>
<td>0.026</td>
</tr>
<tr>
<td>SABPA 2014iii</td>
<td>lysine</td>
<td>0.169</td>
<td>0.043</td>
<td>0.667</td>
<td>-2.539</td>
<td>0.011</td>
</tr>
<tr>
<td>SABPA 2014iv</td>
<td>Phenylglyoxylic acid</td>
<td>0.176</td>
<td>0.045</td>
<td>0.695</td>
<td>-2.479</td>
<td>0.013</td>
</tr>
<tr>
<td>Rhee 2014i</td>
<td>2-hydroxyglutarate</td>
<td>0.093</td>
<td>0.028</td>
<td>0.307</td>
<td>-3.906</td>
<td>0.000</td>
</tr>
<tr>
<td>Rhee 2014ii</td>
<td>glucuronate</td>
<td>0.143</td>
<td>0.044</td>
<td>0.471</td>
<td>-3.202</td>
<td>0.001</td>
</tr>
<tr>
<td>Rhee 2014iii</td>
<td>hydroxyproline</td>
<td>0.143</td>
<td>0.044</td>
<td>0.471</td>
<td>-3.202</td>
<td>0.001</td>
</tr>
<tr>
<td>Rhee 2014iv</td>
<td>gentisate</td>
<td>0.164</td>
<td>0.053</td>
<td>0.511</td>
<td>-3.121</td>
<td>0.002</td>
</tr>
<tr>
<td>Rhee 2014v</td>
<td>lactose</td>
<td>0.205</td>
<td>0.066</td>
<td>0.639</td>
<td>-7.735</td>
<td>0.006</td>
</tr>
<tr>
<td>Rhee 2014vi</td>
<td>4.0 carnitine</td>
<td>0.267</td>
<td>0.086</td>
<td>0.831</td>
<td>-2.280</td>
<td>0.023</td>
</tr>
<tr>
<td>Rhee 2014vii</td>
<td>9:0 carnitine</td>
<td>0.267</td>
<td>0.086</td>
<td>0.831</td>
<td>-2.280</td>
<td>0.023</td>
</tr>
</tbody>
</table>

Pooled odds ratio (Random Effects) 0.151 0.085 0.268 -6.478 0.000

<table>
<thead>
<tr>
<th>Odds ratio and 95% CI</th>
</tr>
</thead>
</table>

0.1 0.2 0.5 1 2 5 10

Favours Normotension  Favours Hypertension
Figure 1.7: Funnel Plot showing Lack of Bias in the Studies
Figure 1.8: A schema demonstrating possible role of ethanol and its deranged metabolism in pathogenesis of hypertension (Adopted with permission from Deventer et al. Journal of the American Society of Hypertension (2015) 1–11).
CONCEPTUAL MODEL

Life Style Factors

Medications  Smoking  Diet  Drugs  Alcohol

Metabolome

Changes in mitochondrion in endothelium

Serum changes in fatty acids due to deranged metabolism of the whole body

Hypertension Phenotype

Age  Sex  Race/ethnicity

Individual Factors
CHAPTER II

Metabolomic Signature of Prevalent and Incident Hypertension
(Insulin Resistance Atherosclerosis Study)

Word Count: Abstract 239 words; Manuscript 187, Tables 6; Figure 5

Short Running Title: Metabolic profiling and hypertension

Corresponding Author:

Waqas Qureshi MD
Section on Cardiovascular Medicine,
Department of Internal Medicine,
Wake Forest University, Winston Salem NC 27157
Phone Number: 336 – 716 – 7427
Fax Number: 336 – 716 – 9188
Email: WQureshi@WakeHealth.Edu
ABSTRACT

BACKGROUND

Metabolic perturbations commonly precede hypertension. Identifying metabolomic signature of hypertension may provide novel information about pathogenesis of hypertension.

METHODS

Insulin Resistance Atherosclerosis Study (IRAS) is a multiethnic cohort enrolled in 1992–94 in whom metabolic profiling for 87 metabolites was performed in plasma of 719 participants. Supervised principal component analysis with orthogonal signal correction was used to identify metabolites associated with prevalent and incident hypertension. The relationship of identified metabolites with baseline hypertension and 5-year incident hypertension was examined in negative binomial regression models adjusted for age, sex, ethnicity, alcohol use, smoking status, and body mass index (BMI).

RESULTS

Hypertension at baseline visit was present in 30% and incident hypertension occurred in 11% of the individuals. Hypertensives were more likely to be older African American and males, with lower insulin-sensitivity and higher BMI. Metabolic profiles with increase in glutamine, lysine and tyrosine but decrease in glycine was independently associated with prevalent hypertension (RR per 1 standard deviation change in metabolic signature; 1.37; 95% CI 1.12 – 1.68, p=0.002). Metabolic profile with increase in glutamine but decrease in alanine and valine was independently inversely associated with 5-year risk of incident hypertension (RR per 1 SD change; 0.49 (0.24, 0.97); p=0.04).

CONCLUSIONS

In conclusion, we identified a metabolomic plasma signature of amino acids which is independently associated with prevalent and incident hypertension. The mechanism of pathogenesis is not entirely clear and requires further research.

Key Words: Metabolic profiling, Hypertension, Epidemiology
INTRODUCTION

Hypertension is a complex cardiovascular disease which currently affects 77.9 million people in United States. There is strong evidence of its association with cardiometabolic conditions such as obesity, diabetes, insulin resistance, and metabolic syndrome. Genetic studies have demonstrated association of metabolic pathway-related genes such as those involved in steroid metabolism (CYP17A1) and alcohol metabolism (ALDH2). However, these studies were limited in informing about the actual metabolic pathway associated with hypertension. Dietary metabolites have also been investigated to understand the association of hypertension with metabolism. However, the etiopathogenesis of hypertension remains obscure and demands additional investigation.

Metabolome is the eventual product of genome that has shown to determine a particular disease phenome. Metabolome consists of small metabolites which are derived from breakdown of consumed macronutrients and micronutrients. These metabolites can be studied by targeted and untargeted approaches. In targeted approach specific metabolites are targeted which are known a priori. Previous targeted metabolomic studies have identified association of fermented products of aromatic compounds, oleic acid, palmitic acid, palmitic acid, nonanoic acid, linoleic acid, and carnitine with hypertension. However, these studies were limited by their cross sectional or case control design and limited racial diversity. These studies also were limited since they correlated the association with hypertension in isolation rather than in combination to identify a particular metabolomic signature. We examined the association of a metabolic signature of fatty acids, acylcarnitines, bile acids and amino acids with prevalent and incident hypertension after 5 years of follow-up in Insulin Resistance Atherosclerosis Study (IRAS) cohort.
METHODS

Study Population and Design

Details of the design, procedures, and methods used in the IRAS, a multi-ethnic and multi-center observational cohort of men and women aged 40 to 69 years, designed to assess the relationship between insulin resistance and atherosclerosis, have been previously described. This epidemiological study was conducted at four United States based centers. Hispanics and non-Hispanic whites were studied in centers in San Luis Valley, Colorado and San Antonio, Texas. African American and non-Hispanic whites were enrolled in Oakland and Los Angeles, California. Sampling strategies were used to enroll sufficient number of participants in different age, gender, ethnic and glucose tolerance groups. Participants using insulin were excluded. The baseline study variables from the study cohort (N = 1,624) were obtained during the years 1992 and 1994. A follow-up examination was conducted after 5 years. Baseline serum was available in a subset of participants without diabetes mellitus (N = 719) was obtained at baseline. The study protocol was approved by institutional review boards at individual centers. Baseline characteristics of individuals that underwent metabolomic profiling and individuals without metabolomic profiling are given in supplementary table 2.1. There were several differences between both groups. Diabetics were not part of the group in which serum metabolomics were performed.

Metabolic profiling

Targeted metabolomic profiling and quantification of 87 metabolites in baseline fasting serum samples was completed by Lipomics Inc (San Francisco, CA) using methods previously described. These metabolites included fatty acids (N = 32), plasmalogens (N = 2), amino acids (N = 16), sterols (N = 11), acylcarnitines (N = 13), ketoacids (N = 2) and bile acids (N = 13). A complete list of metabolites is given in Supplementary Table 2.2. Total lipids from the sera were extracted in the
presence of standardized internal standards using chloroform : methanol (2:1 v/v) using Folch’s method of extraction. The total lipid extract was trans-esterified in 1% sulfuric acid in methanol in a sealed vial under a nitrogen atmosphere at 100°C for 45 min. The resulting samples were neutralized with 6% potassium carbonate and the fatty acid methyl esters (FAME) were re-extracted with hexane and prepared for gas chromatography. Fatty acid methyl esters were separated and by capillary gas chromatography (Agilent Technologies model 6890) equipped with a 30 m DB-88MS capillary column (Agilent Technologies). These were then quantified using mass spectrometry and a flame-ionization detector (FID). Quantitative results were obtained by comparing each fatty acid’s peak and area under it to its internal standard control. All determinations passed internal quality assurance and quality control processes. We chose to use mole percent data (mole%) for fatty acids in all analyses. Mole% data are simply fatty acid composition data, with each fatty acid expressed as a percentage of the total pool of triglycerides. By converting quantitative fatty acid concentrations to mole% concentrations, the effect of changes in total plasma lipid levels (e.g. increased or decreased triacylglycerides) was normalized. For other components of serum, direct quantification by mass spectrometry.

Covariates

Participants self-identified their sex, ethnicity and history of cigarette smoking. Cigarette smoking was dichotomized into current and non-current smoking. Alcohol use was also self-identified and dichotomized into alcohol users and non-users. Body mass index was obtained by measuring height and weights of participants at baseline. Insulin resistance, expressed as SI, was calculated using a computer program; minimal model analysis (MINMOD version 3.0). This computer program uses frequently-sampled intravenous glucose tolerance test (FSIGT) and insulin levels in a non-linear least square estimation technique to
calculate insulin sensitivity. A higher value of SI indicates an increase in insulin sensitivity.

Outcome Variables

Blood pressure was measured three times using a random zero mercury sphygmomanometer as part of baseline and follow-up visits. The average of three blood pressures was used to characterize hypertension. Hypertension was defined as systolic blood pressure $\geq 140$ mmHg, diastolic blood pressure $\geq 90$ mmHg and/or self-reported use of antihypertensive medication. Prevalent hypertension was presence of hypertension at the baseline visit. Incident hypertension was defined as hypertension at visit after 5 years of follow-up in those not hypertensive at baseline.

Power Calculation

Assuming prevalence of hypertension as 30% and variance of relative risk as 0.13; we detected a power of 0.56 for detecting relative risk of 1.2 with a sample size of 719 at an alpha of 0.05. For detecting relative risk of 1.3, the sample size of 719 had a power of 0.89. The plot of power across sample sizes is shown in supplementary figure 2.1.

Statistical Analysis

Continuous variables were expressed as mean with standard deviation, while categorical variables were expressed as proportions (percentages). Baseline characteristics were compared by using either t-tests or Mann-Whitney U tests for continuous and chi-square tests for categorical variables between hypertensives at baseline with non hypertensives at baseline. The subset of non hypertensives at baseline was then divided into hypertensives at 5 year follow up and non hypertensives at 5 – year follow up. These were also compared with appropriate statistical tests. Appropriate variance stabilizing transformations were used for the continuous variables. For the purpose of principal component analysis, mean
centered values were calculated for each metabolite. In order to visualize the data, an unsupervised principal component analysis was performed.

In order to analyze the relationship of metabolites with prevalent hypertension, supervised principal component analysis known as orthogonal projection on latent structure – discriminant analysis (OPLS – DA) was performed. This model separates the systematic variation in the X variable into two parts; first that is linearly correlated to Y and second, which is orthogonal (uncorrelated) to Y with achievement of maximum covariance. A tenfold cross-validation was performed and the OPLS model remained robust. The significance of the analysis was evaluated with a F test using overall cross-validation-analysis of variance (CV-ANOVA).\textsuperscript{141} Variable influence on projection (VIP) was used to summarize the importance of X variables (metabolites) in the OPLS model. VIP was calculated by summing the squares of loading weights of metabolites in the OPLS model for prevalent and incident hypertension.\textsuperscript{142} Loading weights and regression co-efficients had been suggested previously for summarizing information however, VIP offered the most parsimonious model interpretation which uses both loading plots and regression coefficients in its calculation. Metabolites with VIP values >1 and p<0.05 were considered to be important.

Significant metabolites with VIP values >1 obtained from OPLS models of prevalent and incident hypertension were scaled to unit variance. An averaged standardized variable was created from summing standardized values of significant metabolites obtained from each OPLS model and dividing the sum with the total number of significant metabolites.

This averaged standardized variable was termed as metabolomic signature. This approach was used since individual metabolites have different risk compared to when metabolites act together due to complex biological interactions among them.
In multivariable negative binomial regression models adjusted for covariates were then used to examine the strength of relationship between this metabolomic signature and prevalent or incident hypertension. This was achieved by using PROC GENMOD with negative binomial distribution and log link function. Negative binomial regression models were used instead of logistic regression models as logistic regression model overestimates the effect size in a cohort where prevalence of outcome variable is >10%. Relative risks (RR) were expressed with 95% confidence intervals (CI). Model 1 included age, sex, and ethnicity, model 2 included model 1 covariates with alcohol use, smoking and body mass index and model 3 included model 2 covariates and insulin sensitivity. For incident hypertension, we created model 4 in which baseline systolic blood pressure was also adjusted. These adjustments were made based on variables used in a previously published study.\textsuperscript{21} Metabolomic signature was both examined as continuous (per 1 standard deviation) and categorical variable (per quartile). We also made examined individual associations of significant metabolites (per standard deviation) with prevalent and incident hypertension in these models.

Incremental change in areas under the curve (AUC) after addition of metabolomic signature to final model of covariates for incident and prevalent hypertension were estimated using deLong’s method.\textsuperscript{143} Significance between differences in AUCs was estimated using non-parametric method (Mann-Whitney U test). These predictive probabilities were cross–validated 10 folds and there was no significant differences in calculated AUCs. \( P \) – value <0.05 was considered significant for main effects and <0.10 for interactions. All statistical analyses were performed using SAS 9.3 (SAS Inc. Cary, NC) and SIMCA v 13.0 (Orland, FL).
RESULTS

Baseline Characteristics

There were 719 individuals that underwent baseline examination and were included in primary analysis. We then excluded individuals with baseline hypertension (N = 218) leaving 498 participants for secondary analysis (Figure 2.1).

The mean age of the baseline cohort was 55±8 years. There were 56% females, 42% non-Hispanic white, 24% non-Hispanic black and 33% Hispanic whites. There were 79/498 (16%) patients that developed incident hypertension, while there were 218/719 (30%) patients with baseline prevalent hypertension. Compared with non hypertensives at baseline, hypertensives in baseline were older, African American and had higher systolic and diastolic blood pressure, body mass index and lower insulin sensitivity (Table 2.1). Compared with non hypertensives at follow up of 5 years, incident hypertensives were more likely to be older, and have higher systolic and diastolic blood pressure, body mass index and lower insulin sensitivity (Table 2.2).

Metabolomic Signature of Prevalent Hypertension

The OPLS model for prevalent hypertension adjusted for model 1 and model 2 covariates are shown in supplementary figure 2.2 and 2.3. S-plot derived from OPLS model demonstrated that certain metabolites contributed more to correlation and covariance structure of prevalent hypertension than others (overall CV-ANOVA F statistic = 6.87, p = 1.9 x 10^{-5}; Figure2.2). Only four of the metabolites had VIP score <1 and p <0.05 (Figure 2.3) which were glutamine, lysine, tyrosine and glycine. Examining the structure of S-plot showed that metabolic profile with lower concentrations of glycine and higher concentrations of glutamine, lysine, and tyrosine was associated with prevalent hypertension. Thus, a standardized metabolomic signature variable was created from standardized quantitative values
of each of these variables. This standardized metabolomic signature variable was evaluated in univariate and multivariate negative binomial regression models adjusted for model 1, model 2 and model 3 covariates as a continuous variable (per 1 SD change in concentration) and per quartile (Table 2.3). In fully adjusted model, a standard deviation change in metabolomic signature was associated with 37% increased relative risk of prevalent hypertension [RR (95% CI); 1.37 (1.12, 1.68), p=0.002]. When examined as quartiles, the upper quartile of this metabolomic signature was associated with 56% increased risk of prevalent hypertension [RR (95% CI); 1.56 (1.05, 2.33), p =0.02]. Interaction for age, sex, race/ethnicity, were not significant for metabolomic signature and prevalent hypertension (p >0.10), however BMI had weakly significant interaction (p = 0.06).

**Metabolomic Signature of Incident Hypertension**

S-plot derived from OPLS model for incident hypertension demonstrated that certain metabolites contributed more to correlation and covariance structure of incident hypertension than others (overall CV-ANOVA F statistic = 5.10, p = 0.006; **Figure 2.4**). VIP values with 95% confidence intervals are shown in **Figure 2.5**. Three metabolites namely; glutamine, alanine and valine had VIP values >1 and p <0.05. Examining the structure of S-plot showed that metabolomic signature with lower concentrations of higher glutamine and lower concentrations of valine and alanine was associated with incident hypertension. This standardized metabolomic signature variable was evaluated in multivariate negative binomial regression models adjusted for model 1, model 2, model 3 and model 4 covariates as a continuous variable (per 1 SD change in concentration) and per quartile (Table 2.4). In fully adjusted model, a standard unit change in metabolomic signature was associated with 51% decreased relative risk of incident hypertension [RR (95% CI); 0.49 (0.24, 0.97), p =0.04]. Similarly the highest quartile of this metabolomic signature was associated with 60% decreased risk of incident hypertension (0.40 (0.16, 0.96), p = 0.02). Interaction for age, sex, race/ethnicity, BMI were not
significant for metabolomic signature and incident hypertension (p >0.10). The individual associations of metabolite constituents of metabolomic signature of prevalent and incident hypertension are given in table S2.3.

Prediction of prevalent and incident hypertension

To examine the incremental value of metabolomic signature, receiver operating characteristics curves for model 3 with and without addition of metabolomic signature was plotted for prevalent hypertension. Metabolomic signature showed mild incremental discrimination when added to model 3 covariates, there was (AUC 0.712 vs. AUC 0.699, p = 0.1). Similarly, there was a non-significant incremental value of metabolomic signature of incident hypertension over and above the model 4 covariates (AUC 0.635 vs. AUC 0.603 p = 0.1). Cross-validated probabilities were also plotted without changing main effects or significance. Figures S2.4 and S2.5 show the receiver operating curves for both prevalent and incident hypertension.

DISCUSSION

The present targeted metabolomic study extends the scientific literature on the association of serum metabolites with prevalent and incident hypertension. Specifically, a metabolomic signature of glutamine, tyrosine, lysine and glycine was associated with prevalent hypertension and a metabolomic signature of glutamine, valine and alanine was associated with incident hypertension. The metabolomic signature of incident hypertension was robust to adjustment of potential confounders including baseline systolic blood pressure and incrementally but minimally added to predictive model of incident hypertension. These relationships were not modified by age, sex, ethnicity or BMI.

Comparison with Prior Literature

To date, this is the first study that has observed association of a specific metabolic profile with incident hypertension in a well-characterized multi-ethnic cohort. Zhong et al. recently showed in untargeted plasma $^1$H –nuclear magnetic resonance
imaging study of 157 hypertensive Chinese patients that hypertensives had much lower levels of valine, alanine, methylhistidine, and phenylalanine than non-hypertensive individuals. We confirmed their findings in this study. However, the goal of this study was to examine the metabolomic profile (signature) rather than to study individual metabolites. We did not examine individual metabolites in this study. We found significant differences in valine and alanine similar to this study; however, they did not find association of glutamine and glycine with hypertension. This study was limited by using a single ethnic group from a hospital and the matching of controls was not robust. The controls were obtained based on lack of family history of hypertension, which may have led to selection bias.\textsuperscript{119}

Glycine has shown to be effective against hypertension in experimental models but has not been tested in humans yet.\textsuperscript{144, 145} The association of glutamine is novel and has not been implicated in the past. INTERnational collaborative study on Macro/micronutrients And blood Pressure (INTERMAP) examined urinary metabolites by $^1$H-nuclear magnetic resonance ($^1$H-NMR) spectroscopy in 24 – hour collected urine of 4,630 men and women ages 40-59 years randomly selected from 17 centers in United States, United Kingdom, Peoples’ Republic of China, and Japan. In this study, alanine was responsible for 2.69 mmHg of elevation of systolic blood pressure for each standard unit increase in its concentration.\textsuperscript{114}

\textit{Possible Mechanism of Action (Modus Operandi)}

When examining metabolic pathways in Homo sapiens for biosynthesis of amino acids from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database \textsuperscript{146-148} (Figure S2.6), the relationship of amino acid synthesis was closely related with glycolysis and Krebs’ cycle intermediate metabolites. Previous studies have shown elevated pyruvate levels in hypertensive individuals. Elevated pyruvate levels happen as a result of low levels of glycine and serine. They also lead to elevated levels of alanine and lower levels of glutamine which were implicated in
pathogenesis of hypertension in this study. Thus, this suggests central role of pyruvate that may connect all the implicated amino acids in pathogenesis of hypertension. Even though these amino acids are included in diet, much of their plasma levels are governed by metabolic pathway enzymes and genes controlling thus. This is also suggested by a recent analysis of 3086 Dutch participants of Rotterdam study which failed to show relationship of dietary glutamic acid, lysine, tyrosine and cysteine with prevalent or incident hypertension over a follow-up period of 6 years.\textsuperscript{149} Recently in arteries of hypertensive rats, especially aorta the activity of transglutaminase 1 and 2 was found to be decreased.\textsuperscript{150} These enzymes have pleiotropic function and mainly cross link amino acids such as lysine and glutamine (\textbf{Figure S2.7}). The relationship of valine and tyrosine has been established with insulin resistance.\textsuperscript{151, 152} Their relationship with incident diabetes mellitus in Framingham Offspring Study was highly significant and displayed a predictive value for new – onset diabetes mellitus up to 12 years.\textsuperscript{153} However, their relationship with hypertension has not been well studied. In this analysis, we found an independent impact of these metabolites despite adjusting the analysis for insulin sensitivity. We also excluded patients with known diabetes or diabetes diagnosed by glucose tolerance test in order to understand the independent effect of these metabolites with hypertension. Despite this, we still detected significant associations thus this analysis suggests that there are other possible mechanisms that may explain association metabolites with hypertension. Nevertheless, future replication studies are necessary to replicate our findings.

Glycine was another strong negative factor associated with prevalent hypertension. Glycine/arginine polymorphism in the beta2 adrenergic receptor gene has been associated with hypertension.\textsuperscript{154, 155} Glycine has shown to protection to ischemia – reperfusion injury suggesting its role against oxidative stress.\textsuperscript{156} Elevated blood pressure in rat offsprings induced by low protein diet in rat mothers was reversed
by glycine supplementation.\textsuperscript{157} All of these point towards some of the common pathways that are known to be responsible for hypertension.

\textit{Implications}

There are multiple implications of the study. Knowing a specific metabolic profile with low glutamine, high valine and alanine that can predispose to hypertension is important for large scale preventive strategies where proper resource allocation can be used to identify high risk individuals. We observed metabolic profile of incident hypertension pointing towards involvement of glutamine – alanine cycle dysfunction which is a crucial component of gluconeogenesis. This pathway has not yet been manipulated for developing prevention strategies of hypertension and may be the basis of multiple studies that have shown association of diabetes with hypertension. Indeed, we saw higher levels of insulin resistance in hypertensive patients. While we found that different metabolic profiles were associated with prevalent and incident hypertension, we found that increasing levels of glutamine portends a high risk of prevalent hypertension but a lower risk of incident hypertension. Thus, considering it alone in an analysis will give erroneous interpretation. When glutamine is considered with other correlated metabolites, it provides better explanation of this observation; thus suggesting consideration of other metabolites in the analysis when interpreting a single metabolites and its risk associated with a phenotype. An untargeted metabolomic analysis will take this consideration into account as it will capture most of the metabolites associated with the studied metabolite.

\textit{Strengths and Limitations}

The strengths of this analysis include large sample size, multi-ethnic background, excellent follow up, non-diabetes cohort and reliable clinical variables obtained by well-trained research assistants. However, our analysis is not without limitations. We did not perform an agnostic untargeted metabolomic analysis. The number of
metabolites in an untargeted metabolomic analysis may range from several hundreds to thousands. However, previous analyses demonstrated many metabolites which sometimes are not even annotated and have unknown functional significance. The information obtained for antihypertensives was by self-report and may suffer from recall and interviewer bias. However, we used pre-specified questionnaires and well–trained interviewers for this purpose to minimize these biases. Lastly, the study lacks an ancillary replication study that would have provided more validity to our results. However, we performed ten-fold cross-validation of our sample without much change in our results.

Conclusions

In conclusion, we demonstrated that metabolomic signature of amino acids is associated with prevalent and incident hypertension. This metabolomic signature can minimally increase the discrimination over and above known risk factors. Further replication and functional studies are needed to validate these findings and establish functional significance.

ACKNOWLEDGEMENTS

IRAS was supported by grants U01-HL47892, DK-29867 and R01-58329 from the National Heart, Lung, and Blood Institute, and by grant M01-RR-43 from the National Institute of Health. WT Qureshi is funded by Ruth L. Kirschstein NRSA Institutional Training Grant 5T32HL076132-10.

DISCLOSURES

Steve Watkins is an employee of Lipomics Inc.
### Table 2.1: Baseline characteristics of cohort (N = 719)

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Hypertension at baseline (N = 218)</th>
<th>No hypertension at baseline (N = 501)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>58±8</td>
<td>54±8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female</td>
<td>233 (56%)</td>
<td>277 (56%)</td>
<td>0.53</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td>0.008</td>
</tr>
<tr>
<td>White</td>
<td>87 (40%)</td>
<td>217 (51%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>69 (32%)</td>
<td>105 (20%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>62 (28%)</td>
<td>177 (29%)</td>
<td></td>
</tr>
<tr>
<td>Smoking Status</td>
<td></td>
<td></td>
<td>0.18</td>
</tr>
<tr>
<td>Current</td>
<td>24 (11%)</td>
<td>80 (16%)</td>
<td></td>
</tr>
<tr>
<td>Former</td>
<td>92 (42%)</td>
<td>188 (38%)</td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>448 (90%)</td>
<td>193 (88%)</td>
<td>0.60</td>
</tr>
<tr>
<td>Insulin sensitivity index</td>
<td>2.5±2.2</td>
<td>1.6±1.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Systolic Blood pressure (mmHg)</td>
<td>136±18</td>
<td>115±12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>83±11</td>
<td>75±8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>30±6</td>
<td>28±5</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Categorical variables expressed as % and continuous variables as mean ± standard deviations.
Table 2.2: Baseline characteristics in subset of cohort without baseline hypertension (N = 498)

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>Hypertension at 5 years (N = 79)</th>
<th>No hypertension at 5 years (N = 419)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>56±8</td>
<td>53±8</td>
<td>0.01</td>
</tr>
<tr>
<td>Female</td>
<td>44 (56%)</td>
<td>233 (56%)</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td>White</td>
<td>40 (42%)</td>
<td>176 (51%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>16 (19%)</td>
<td>89 (21%)</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>23 (29%)</td>
<td>154 (37%)</td>
<td></td>
</tr>
<tr>
<td>Smoking Status</td>
<td></td>
<td></td>
<td>0.85</td>
</tr>
<tr>
<td>Current</td>
<td>11 (14%)</td>
<td>69 (16%)</td>
<td></td>
</tr>
<tr>
<td>Former</td>
<td>31 (39%)</td>
<td>157 (38%)</td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>73 (92%)</td>
<td>374 (89%)</td>
<td>0.54</td>
</tr>
<tr>
<td>Insulin sensitivity index</td>
<td>1.8±1.2</td>
<td>2.6±2.3</td>
<td>0.003</td>
</tr>
<tr>
<td>Systolic Blood pressure (mmHg)</td>
<td>123±12</td>
<td>114±12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diastolic blood pressure(mmHg)</td>
<td>78±9</td>
<td>74±7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>29±5</td>
<td>27±5</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Categorical variables expressed as % and continuous variables as mean ± standard deviations
Table 2.3: Association of metabolomic signature (glutamine, lysine, tyrosine and glycine) with prevalent hypertension

<table>
<thead>
<tr>
<th>Metabolic signature</th>
<th>Relative Risk (95% CI)</th>
<th>Model 1*</th>
<th>Model 2†</th>
<th>Model 3‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quartile 1</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td></td>
</tr>
<tr>
<td>Quartile 2</td>
<td>1.47 (0.99, 2.18)</td>
<td>1.28 (0.85, 1.93)</td>
<td>1.29 (0.85, 1.95)</td>
<td></td>
</tr>
<tr>
<td>Quartile 3</td>
<td>1.56 (1.05, 2.32)</td>
<td>1.35 (0.90, 2.03)</td>
<td>1.37 (0.91, 2.07)</td>
<td></td>
</tr>
<tr>
<td>Quartile 4</td>
<td>1.92 (1.32, 2.82)</td>
<td>1.57 (1.06, 2.33)</td>
<td>1.56 (1.05, 2.33)</td>
<td></td>
</tr>
</tbody>
</table>

*Model 1: Adjusted for age, sex, race/ethnicity
†Model 2: Adjusted for model 1 covariates and alcohol use, smoking, and insulin sensitivity
‡Model 3: Adjusted for model 2 covariates and body mass index
Table 2.4: Association of metabolomic signature (glutamine, alanine and valine) with incident hypertension

<table>
<thead>
<tr>
<th>Metabolic signature</th>
<th>Relative Risk (95% CI)</th>
<th>Model 1*</th>
<th>Model 2†</th>
<th>Model 3‡</th>
<th>Model 4δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quartile 1</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td></td>
</tr>
<tr>
<td>Quartile 2</td>
<td>0.57 (0.30,1.08)</td>
<td>0.68 (0.34,1.36)</td>
<td>0.70 (0.35, 1.40)</td>
<td>0.75 (0.40, 1.54)</td>
<td></td>
</tr>
<tr>
<td>Quartile 3</td>
<td>0.62 (0.32, 1.18)</td>
<td>0.84 (0.43, 1.67)</td>
<td>0.88 (0.44, 1.76)</td>
<td>0.94 (0.46, 1.92)</td>
<td></td>
</tr>
<tr>
<td>Quartile 4</td>
<td>0.27 (0.12,0.60)</td>
<td>0.37 (0.16, 0.86)</td>
<td>0.39 (0.16, 0.92)</td>
<td>0.40 (0.16, 0.96)</td>
<td></td>
</tr>
<tr>
<td>P – value for trend</td>
<td></td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Per 1 SD increase</td>
<td>0.57 (0.38, 0.85)</td>
<td>0.48 (0.25, 0.92)</td>
<td>0.49 (0.25, 0.95)</td>
<td>0.49 (0.24, 0.97)</td>
<td></td>
</tr>
<tr>
<td>P – value</td>
<td></td>
<td>0.007</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
</tr>
</tbody>
</table>

*Model 1: Adjusted for age, sex, race/ethnicity
†Model 2: Adjusted for model 1 covariates and alcohol use, smoking, and insulin sensitivity
‡Model 3: Adjusted for model 2 covariates and body mass index
δModel 4: Adjusted for model 2 covariates and baseline systolic blood pressure
FIGURE LEGENDS

Figure 2.1: Consortium Diagram of Study Sample

Figure 2.2: Orthogonal projections on latent structures S-plot showing covariance and correlation structure of various metabolites with prevalent hypertension

Figure 2.3: Variable importance on the projection (VIP) plot of metabolites associated with prevalent hypertension

Figure 2.4: Orthogonal projections on latent structures S-plot showing covariance and correlation structure of various metabolites with incident hypertension

Figure 2.5: Variable importance on the projection (VIP) plot of metabolites associated with incident hypertension
Figure 2.1: Consortium Diagram of Study Sample
Figure 2.2: Orthogonal projections on latent structures S-plot showing covariance and correlation structure of various metabolites with prevalent hypertension
Figure 2.3: Variable importance on the projection (VIP) plot of metabolites associated with prevalent hypertension
Figure 2.4: Orthogonal projections on latent structures S-plot showing covariance and correlation structure of various metabolites with incident hypertension
Figure 2.5: Variable importance on the projection (VIP) plot of metabolites associated with incident hypertension
SUPPLEMENTARY MATERIAL

Metabolomic Signature of Prevalent and Incident Hypertension

(Insulin Resistance Atherosclerosis Study)
Table S2.1: Baseline characteristics of individuals with and without serum metabolites

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cohort not studied (N = 908)</th>
<th>Cohort with metabolomics (n = 716)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>57±8</td>
<td>55±8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female</td>
<td>44%</td>
<td>44%</td>
<td>0.51</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>White</td>
<td>34%</td>
<td>42%</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>32%</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>34%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td>Current Smokers</td>
<td>31%</td>
<td>27%</td>
<td>0.16</td>
</tr>
<tr>
<td>Alcohol use</td>
<td>87%</td>
<td>89%</td>
<td>0.09</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>27%</td>
<td>0%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Insulin sensitivity index</td>
<td>1.2±1.5</td>
<td>2.2±2.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hypertension at baseline</td>
<td>55%</td>
<td>39%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Systolic Blood pressure (mmHg)</td>
<td>128±17</td>
<td>122±17</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diastolic blood pressure (mmHg)</td>
<td>78±10</td>
<td>79±9</td>
<td>0.84</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>28±6</td>
<td>30±6</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Categorical variables expressed as % and continuous variables as mean ± standard deviations
Table S2.2: List of all metabolites measured in the serum

<table>
<thead>
<tr>
<th>Fatty acids</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lauric acid (12:0)</td>
<td>Alanine</td>
</tr>
<tr>
<td>Myristic acid (14:0)</td>
<td>Asparagine</td>
</tr>
<tr>
<td>Pentadecanoic acid (15:0)</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Palmitic acid (16:0)</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>Stearic acid (18:0)</td>
<td>Glycine</td>
</tr>
<tr>
<td>Arachidic acid (20:0)</td>
<td>Histidine</td>
</tr>
<tr>
<td>Behenic acid (22:0)</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>Lignoceric acid (24:0)</td>
<td>Leucine</td>
</tr>
<tr>
<td>Myristoleic acid (14:1n5)</td>
<td>Lysine</td>
</tr>
<tr>
<td>Palmitoleic acid (16:1n7)</td>
<td>Methionine</td>
</tr>
<tr>
<td>Vaccenic acid (18:1n7)</td>
<td>Ornithine</td>
</tr>
<tr>
<td>Oleic acid (18:1n9)</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>Eicosaenoic acid (20:1n9)</td>
<td>Proline</td>
</tr>
<tr>
<td>Erucic acid (22:1n9)</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Nervonic acid (24:1n9)</td>
<td>Tyrosine</td>
</tr>
<tr>
<td>Mead acid (20:3n9)</td>
<td>Valine</td>
</tr>
<tr>
<td>Linoleic acid (18:2n6)</td>
<td><strong>Acylcarnitines</strong></td>
</tr>
<tr>
<td>Gama – linolenic acid (18:3n6)</td>
<td>Acylcarinitine acetyl carnitine</td>
</tr>
<tr>
<td>di-homo-gamma-linolenic acid (20:3n6)</td>
<td>Acylcarinitine propionyl carnitine</td>
</tr>
<tr>
<td>Arachidonic acid (20:4n6)</td>
<td>Acylcarinitine butyryl carnitine</td>
</tr>
<tr>
<td>Adrenic acid (22:4n6)</td>
<td>Acylcarinitine valeryl carnitine</td>
</tr>
<tr>
<td>Osbond acid (22:5n6)</td>
<td>Acylcarinitine hex noyl carnitine</td>
</tr>
<tr>
<td>Eicosadienoic acid (20:2n6)</td>
<td>Acylcarinitine octanoyl carnitine</td>
</tr>
<tr>
<td>Alpha-linolenic acid (18:3n3)</td>
<td>Acylcarinitine de canoyl carnitine</td>
</tr>
<tr>
<td>Steridonic acid (18:4n3)</td>
<td>Acylcarinitine do decanoyl carnitine</td>
</tr>
<tr>
<td>Eicosatetraenoic acid (20:4n3)</td>
<td>Acylcarinitine myristoyl carnitine</td>
</tr>
<tr>
<td>Eicosapentaenoic acid (20:5n3)</td>
<td>Acylcarinitine palmitoyl carnitine</td>
</tr>
<tr>
<td>Docosapentaenoic acid (22:5n3)</td>
<td>Acylcarinitine stero yl carnitine</td>
</tr>
<tr>
<td>Docosahexaenoic acid (22:6n3)</td>
<td>Acylcarinitine oleoyl carnitine</td>
</tr>
<tr>
<td>Palmitoelaidic acid (t16:1n7)</td>
<td>Acylcarinitine lin olyl carnitine</td>
</tr>
<tr>
<td>Elaidic acid (t18:1n9)</td>
<td><strong>Bile acids</strong></td>
</tr>
<tr>
<td>Plasmalogen palmitic acid (dm 16:0)</td>
<td>Cholic acid</td>
</tr>
<tr>
<td>Plasmalogen stearic acid (dm 18:0)</td>
<td>Chenodeoxycholic acid</td>
</tr>
<tr>
<td>Plasmalogen oleic acid (dm 18:1n9)</td>
<td>Deoxycholic acid</td>
</tr>
<tr>
<td><strong>Sterols</strong></td>
<td>Lithocholic acid</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Ursodeoxycholic acid</td>
</tr>
<tr>
<td>Desmosterol</td>
<td>Taurocholic acid</td>
</tr>
<tr>
<td>Lanosterol</td>
<td>Taurochenodeoxy cholic acid</td>
</tr>
<tr>
<td>Lathosterol</td>
<td>Tauroli thocholic acid</td>
</tr>
<tr>
<td>7-dehydrocholesterol</td>
<td>Glycocholic acid</td>
</tr>
<tr>
<td>Cholesterol</td>
<td>Glycochenodeoxycholic acid</td>
</tr>
<tr>
<td>b-Sitosterol</td>
<td>Glycodeoxycholic acid</td>
</tr>
<tr>
<td>Campesterol</td>
<td>Glycolithocholic acid</td>
</tr>
<tr>
<td>Stigmasterol</td>
<td>Glycoursodeoxycholic acid</td>
</tr>
<tr>
<td>7alpha-hydroxycholesterol</td>
<td><strong>Bile acids</strong></td>
</tr>
<tr>
<td>Corpostanol</td>
<td>Cholic acid</td>
</tr>
</tbody>
</table>
Table S2.3: Univariate associations of individual metabolites of metabolomic signature with prevalent and incident hypertension (per standard deviation of metabolite)

<table>
<thead>
<tr>
<th>Metabolites</th>
<th>Prevalent hypertension</th>
<th>p-value</th>
<th>Incident hypertension</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>0.67 (0.56, 0.80)</td>
<td>&lt;0.001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>1.25 (1.07, 1.46)</td>
<td>0.005</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lysine</td>
<td>1.24 (1.06, 1.46)</td>
<td>0.006</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Glutamine</td>
<td>1.20 (1.03, 1.41)</td>
<td>0.02</td>
<td>0.79 (0.62, 1.02)</td>
<td>0.07</td>
</tr>
<tr>
<td>Valine</td>
<td>-</td>
<td></td>
<td>1.24 (0.97, 1.57)</td>
<td>0.09</td>
</tr>
<tr>
<td>Alanine</td>
<td>-</td>
<td></td>
<td>1.19 (0.94, 1.51)</td>
<td>0.14</td>
</tr>
</tbody>
</table>
Figure S2.1: Power calculation and sample size

![Graph showing power calculation and sample size](image)
Figure S2.2: Orthogonal projections on latent structures discriminant analysis score plot showing discrimination between baseline hypertensive (green) and non-hypertensive (blue) individuals using model 1 covariates.
Figure S2.3: Orthogonal projections on latent structures discriminant analysis score plot showing discrimination between baseline hypertensive (green) and non-hypertensive (blue) individuals using model 2 covariates.
Figure S2.4: Receiver operating characteristic curve for metabolomic signature added to the full model for prevalent hypertension
Figure S2.5: Receiver operating characteristic curve for metabolomic signature added to the full model for incident hypertension
Figure S2.6: Biosynthesis of amino acids and their relationship with glycolytic and Kreb's cycle intermediates in Homo sapiens (Adopted from KEGG pathways)
Figure S2.7: Cross-linking of amino acids by transglutaminase (Adopted with permission from Nature Reviews Molecular Cell Biology 4, 140-156)

a: model for the direct de novo polymerization of substrates by transglutaminases (TG), in which enzyme catalyzed process drives the assembly of subunits (blue ellipse). Each subunit is shown with a single lysine (Lys) donor (green) and glutamine (Gln) acceptor (red) functionality. The N-glutamyl – lysine linkages are marked as X.

b: model for fortifying a non-covalent reversible assembly by TG-mediated cross linking.

c: stabilization of fibrin clots by ‘spot welding’ is illustrated. Thrombin converts fibrinogen into fibrin by cleaving fibrinopeptide from amino-termini of α and β chains that are in E domains (purple) of the protein. Creating new amino-termini of glycine is important because they complement the polymerization pockets in D regions (yellow). The TG reactive Gln (orange circles) and Lys (pink circles) participates in fibrin to fibrin cross linking and function to anchor blood cells and plasma components to form clots and thrombi. K303 is involved in incorporating plasmin inhibitor into the network. D and E are chromatographic fractions generated by trypsin partial digestion of fibrinogen.
CHAPTER III
METABOLOMIC METHODS AND ADDITIONAL ANALYSES

In chapter II, we demonstrated the strength of association of a metabolomic signature with prevalent and incident hypertension. These metabolites that we studied are determined through a complex sequential procedure of several discrete steps including bio-sampling, separation, metabolite detection and data analysis (Figure 3.1). Each step is critically important to insure valid and interpretable results. Here, we discuss the various important aspects of our metabolomic study as well as general context of a metabolomic study with an emphasis on multivariate analyses. We have also presented ancillary analyses of our data studied by using hierarchical cluster analysis and weighted co-expression network analysis.

3.1 METABOLITE EXTRACTION, SAMPLING and HANDLING

For carrying out a metabolomic study, particular attention should be given the type of bio-sample and sampling techniques, as these determine the types of metabolites that need to be studied (Table 3.1). While collecting bio-samples, care must be paid to avoid degradation and contamination of collected sample. Some necessary considerations for sampling include: timing (e.g. 24-hour sampling versus collection after a 12- to 14-hour fasting period), sudden dietary changes (potentially interfere with the metabolome for over a week), bacterial or fungal contamination (especially for urine samples), medication intake (metabolites of medications may change the metabolome), diurnal variation in the hormones (anabolic and catabolic hormones vary during the day and may alter the metabolome), randomization, and transport and storage conditions of samples.

Serum is generally the preferred blood bio-specimen for metabolomic studies which is used in our study. This is because plasma retains fibrinogen and other coagulation cascade proteolytic enzymes that can continue to be active ex-vivo and alter the metabolome or interfere with metabolite detection. Furthermore, the
ethylenediaminetetraacetic acid (EDTA) typically used as the anti-coagulant in plasma samples can denature chromatograms and proteins in a way that may influence results. Serum has fewer potentially interfering coagulation cascade enzymes and proteins. However, even in serum there remain active enzymes that can continue to modify the metabolome *ex-vivo*. To counteract these enzymatic processes, serum can be collected on ice. It is recommended to use samples stored at -20°C within 7 days and samples stored at -80°C within 1 month. However, in one report, changes in the metabolome were negligible after storage for 2.5 years. In our cohort, the samples were analyzed after a period of 15 years and it is possible that the metabolome might have changed over time. In the case of repeated usage, fewer than three freeze–thaw cycles is advisable.

As reviewed in Chapter II, INTERMAP study which is a landmark urinary metabolomic study collected 24–hour urine from 4,360 participants. This technique has many advantages, such as a noninvasive method of collection, lack of special preparation for collection, almost no sample pre-treatment, and lower protein content, which helps to increase the sensitivity of identifying other metabolites. Urine, like blood or plasma, provides a metabolomic “footprint” of the whole body, not necessarily a single organ. A 24-hour urine sample is a labor intense sampling technique and it requires that participants receive certain instructions and follow specific procedures. Improper procedures may lead to bacterial overgrowth secondary to either contamination or infection which can affect the urine metabolome. Many other types of samples also have been used in metabolomics research (for example; cell lysates, spinal fluid, etc). Although these types of bio-samples are less amenable to large-scale clinical and population research but may provide more reliable information regarding metabolomic changes in the organ of interest.

Prior to sample separation, sample preparation steps are usually required. Standardization of these sample preparation steps is pivotal to avoid random or,
even worse, systematic bias in the results. For example, alcohol-based extraction is a common procedure was used in sample extraction of our study. The metabolites of Kreb's cycle pathway are up-regulated when sample is subjected to boiling ethanol or freeze thaw, but are down-regulated when cold methanol is used. The ideal method should be highly reproducible, simple, rapid, unselective, and include a metabolism-quenching step. The latter is not frequently used in human studies, but is important to consider when performing a metabolomic study because many active enzymes may influence metabolites in the collected sample (especially in plasma or serum samples as in our case). If the enzymes continue to remain active, they may change concentrations of metabolites of interest. Special consideration should be paid to this matter when evaluating metabolites that are not affected by rate-limiting steps, for example, amino – acids in our case. Thus, methanol was used to extract metabolites in our samples.

3.2 METABOLITE SEPARATION METHODS

In our study, we used a technique called chromatography to separate various metabolites. Chromatography is the principal separation technique used to obtain fine resolution of metabolites for metabolomics studies. It plays a key role in obtaining analytical data needed in metabolic profiling. The separation of metabolites is achieved by interactions of analytes with solvent and a variety of stationary phases (eg. solid or liquid chromatography columns, etc. – see below).

The basic theory of chromatography was first described in 1903 when a chromatogram made up of calcium carbonate (stationary phase) was inoculated with a leaf extract (sample) mixed with ethanol (mobile phase). For metabolomics research, a sample (or sample extract) is dissolved in a mobile phase (liquid or gas). This mobile phase is then forced through an immiscible stationary phase. The component of sample that has an affinity for the stationary phase will travel slowly, whereas the component that repels or does not dissolve in the stationary phase is
the first one to reach the other end of chromatography column. By making changes in these phases, metabolites can be separated with a high degree of resolution and a high-quality chromatogram can be obtained.

Hydrophilic interaction liquid chromatography (HILIC) and reverse phase liquid chromatography (RPLC) are frequently used forms of liquid chromatography; their relative advantages and disadvantages are listed in Table 3.2. These techniques are mainly used for separation of secondary neutral metabolites, lipids, flavonoids, and hydrophobic vitamins. Another way to increase resolution of metabolites is utilized in ultra-performance liquid chromatography (UPLC), where the particle size is reduced at the cost of an increase in back pressure. A 10-minute run on UPLC can provide many metabolites at a time. Figure 3.2 shows some metabolites obtained using a high throughput UPLC.

Capillary electrophoresis is different type of separation technique that uses an electric field applied to sub-millimeter channels (capillaries) to separate analytes. The separation usually takes around 20 minutes. This method is useful for separation of nucleic acids, vitamins (hydrophobic), amino acids, coenzymes, and amino acids.

In our study, we used gas chromatography which vaporizes the metabolites without decomposing them. Based on the mass and volatility of the substance, an inert gas like helium or nitrogen is used as a mobile phase. The stationary phase is a layer of liquid on an inert solid surface. The concentration of metabolite of interest is affected directly by vapor pressure of the gas. Thus, it is also known as gas-liquid partition chromatography.

There are several advantages of linking these separation techniques with analytical tools such as NMR and mass spectroscopy. Processing samples with these analytical tools will show peaks; however, if the analyte of interest exists in very small concentrations (as is usually true in biological fluids) other compounds with
similar-sized peaks may be difficult to differentiate from the analyte of interest. Chromatography uses the physical and chemical properties of the mobile and stationary phases to further increase resolution of this detection process.

The interface between the separation apparatus and the detectors is also critical. Snyder and Kirkland note desirable characteristics of such an interface: 162

1. There should be no reduction in chromatographic performance over time.
2. There should be no uncontrolled chemical modification of the analyte during chromatography.
3. The sample transfer to mass spectrometer should be high and efficient.
4. The interface should be reliable, easy to use, and inexpensive.
5. The interface should give a low chemical background.
6. The interface should be capable of operating across a wide variety of chromatography conditions.
7. The interface should not affect the vacuum requirements of the mass spectrometer and should be compatible with all capabilities of the mass spectrometer.
8. The interface should provide quantitative information with reproducibility better than 10%, with low limits of detection. This response should be linear over a wide range of sample sizes.

In our study, flow injection analysis was also performed which is a technique which ionizes the sample and injects it directly into the mass-spectrometer instrument without a preceding separation step. This method is mainly used for targeted metabolomic analysis where the unique mass spectra are known beforehand and are easily identify in the mass spectrum. This analysis method takes a <10 minutes per sample.

**3.3 METABOLITE DETECTION**
Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are the primary analytical technologies of metabolomics. Both should be considered as complementary technologies rather than separate methods for a metabolomic study. Almost all of the high abundance molecules can be detected by NMR spectroscopy. For low-abundance molecules, mass spectrometry is more suitable. We will not review NMR in detail here because in our study MS was used and is out of the scope of this thesis. However it is worth mentioning that NMR utilizes the principle that atomic nuclei with non-zero spin when subjected to a strong magnetic field and irradiated with a characteristic radiofrequency wave emit radiofrequency signals. These signals, which can be detected by an antenna or coil, reflect the number and proximity of the specific nuclear isotopes in question which can be used to infer the quantity and structural features of the compounds in the sample.\textsuperscript{163} Because NMR is nondestructive, a sample can be reused many times. Another advantage is that the sample does not require extensive preparation before analysis. However, NMR is limited by its lower sensitivity; it can only detect high-abundance metabolites with concentrations >100 nmol/L or 5–10 μM in size. Also, if one-dimensional NMR is used, many metabolites may have overlapping peaks, especially those in low abundance and can be missed.

In contrast, MS takes advantage of the fact that the velocity of charged particles in an electronic field vary in a precise manner as a function of mass and charge. This permits positive identification of a wide range of molecules including those that are typically part of the metabolome. Since its invention in 1946, time of flight mass spectroscopy (TOF-MS) has grown significantly in its capabilities. The mechanism of ionization has continued to evolve in analyzing biological samples, to address the fact that organic analytes tend to be fragile and often dissociate and break down when hard ionizing techniques are used. Matrix-assisted laser desorption ionization (MALDI)\textsuperscript{164} is a soft ionization technique that began to be used in TOF – MS in the 1980s and 1990s.
Different ionization techniques now used include electron impact ionization (EI)\textsuperscript{165, 166}, electron spray ionization (ESI)\textsuperscript{167-169}, Atmospheric pressure chemical ionization (APCI)\textsuperscript{170}, desorption electron spray ionization (DESI)\textsuperscript{171}, and desorption/ionization on silicon (DIOS)\textsuperscript{172, 173}. Combinations of two ionization techniques have also been documented (e.g. ESI/APCI, MALDI/DIOS)\textsuperscript{174}, as has simultaneous use of two ionization techniques (multimode ionization; MMI)\textsuperscript{174}

Further modifications of electrodes apply electrostatic force to the ions to improve their resolution before detection. The thermofisher orbitrap and quadrupole ion traps are among the many instruments that use electrostatic or magnetic fields to capture ions in the vacuum system of the mass spectrometer, improving resolution a thousand folds. Due to these improvements, mass spectroscopy has higher sensitivity than NMR and can detect low-abundance metabolites (as low as 1 pmol/L).\textsuperscript{175} However, this sensitivity comes at a cost of low reproducibility. Robust quality control methodology, along with analytical techniques to identify temporal drift or changes in the mass spectrometer, is constantly needed. Internal standards and/or quality control samples are a part of each run to identify errors.

To identify a particular metabolite, its accurate mass, retention time, MS/MS spectrum match, isotope abundance pattern, and fragmentation pattern all should be collected and matched with purified standards under identical conditions.\textsuperscript{176-178} This is a time-consuming process. The Metabolomics Standards Initiative has recommended that investigators “putatively annotate” these metabolites without actually knowing their chemical structures unless it is absolutely necessary.\textsuperscript{176}
3.4 METABOLOMIC DATA ANALYSIS

There are mainly two forms of metabolomic experiments performed: (a) untargeted metabolomics, a comprehensive agnostic chemometrics approach to evaluate small metabolite molecules that include unknown and known metabolites, and (b) targeted metabolomics, a quantitative analysis of a set of chemically known and annotated metabolites. The advantage of targeted metabolomics is that already known pathways are explored by this method and a prior hypothesis is present. On the other hand, the untargeted metabolomics approach can identify novel metabolites and pathways without a prior hypothesis.

Multiple statistical methods have been used to analyze metabolites. For targeted metabolomics, basic statistical tests such as Student’s $t$-test, analysis of variance, and non-parametric tests like the Kruskall–Wallis test may provide adequate statistical means to assess the presence of a signal and its association with a trait of interest. However, many metabolomic signals are highly correlated and thus violate fundamental assumptions of independence for these tests. In such cases, multivariate methods provide an attractive choice. Such analyses are used at various stages of metabolomics—not just to analyze associations in biomarker studies, but also to identify spectral patterns and confirm data robustness of chromatography, mass spectrometry, and NMR spectra. Advanced forms of multivariate analyses are discussed below:

3.4.1 UNSUPERVISED MULTIVARIATE ANALYSES

In our analysis for Chapter II, we initially used unsupervised multivariate analyses to observe the correlational structure of metabolites. For these analyses, no a priori information is utilized to discern various metabolomic patterns. Rather, the correlational structure of the data space is described without regard to its relationship to any external trait or variable. We describe below principal component analysis and network analysis.
3.4.1.1 PRINCIPAL COMPONENT ANALYSIS

An example of unsupervised multivariate analysis is principal component analysis (PCA) which was used in our study. It is a statistical analytical strategy used to summarize multidimensional correlated data. This mathematical technique was first introduced by Pearson in 1901. Hotelling later on developed it further in 1933. The main idea is to reduce the dimensionality of large number of variables into a smaller number of principal components based on the covariance structure of data. The computer softwares have made this easy and they produce multiple principal components (PC) within minutes. The first principal component explains the most variation in the dataset, the second PC explains less than the first PC, the third PC less than the second PC, and so on.

Variance is a measure of spread of data. In simple statistics, variance is a 1-dimensional measure. Variance of X is given by:

\[
var (X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{n - 1}
\]

If there are more data, variation between two or more dimensions become important. Covariance is a two dimensional measure that provides the information about the degree that both dimensions vary from the mean with respect to each other and can be calculated for two dimensions:

\[
var (X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1}
\]

Covariance is measured between two dimensions at a time. If there are more than two dimensions, then a covariance matrix may be created. We can calculate \( \frac{n!}{2(n-2)!} \) different covariance values; where \( n = \text{dimension} \). This covariance matrix can be given as:
$C^{n \times n} = \left[ \text{cov} \left( \text{dim}_i, \text{dim}_j \right) \right]$ for $i$ and $j$ dimensions. Once covariance matrix is constructed, one could derive eigenvectors from it. An eigenvector is a non-zero vector of this covariance matrix, which when multiplied with matrix yields eigenvalue of the vector. The eigenvector gives the direction of the line of best fit for a given set of data, while eigenvalue provides information about variance in the data in that direction. The eigenvector with highest eigenvalue is the first PC as it explains the most variation in the data. The next eigenvector is orthogonal (perpendicular) to the previously calculated PC. Initially, after mean centering the log transformed quantities of metabolites, we examine the score plots. A score is a new variable that summarizes the independent metabolites. The first score explains the largest variation of $X$ (metabolite) variable space. Plots may be constructed between score 1 and score 2, score 2 and score 3 etc. These score plots provide an unsupervised multivariate approach towards individual participants based on their metabolites. Scores of metabolites are weighted averages of these metabolites. These weights are called loadings. Loadings express the dominating correlation structure of metabolites.

Principal component analysis is frequently used as the first analysis to visualize the data. The elbow method is used to select the principal components that account for the largest proportion of variance in the data. Loading plots and score plots are then used to visualize and identify groups of metabolites and samples that may explain observed clustering in the data. These clustered observations are then checked and screened for various hypothetical characteristics. This approach can also help to discern technical or confounding factors that may have biased the observations. Samples in which medications are present, or from individuals with an inherent metabolic defect, typically stand out on such plots. This method also helps to evaluate the internal controls. Ideally, the controls should not show a specific pattern, since a specific pattern away from the center on a given principal component score plot may identify technical problems. For instance, when a linear
orientation of internal controls is evident on a principal component score plot, this may signify a temporal drift in the separation or detection techniques.

3.4.1.2 NETWORK ANALYSIS

Metabolites can be visualized using network analysis. This method of visualizing data in metabolomics is not widely used. However, many genetic studies have used network analyses in order to interpret information hidden in these large datasets.\textsuperscript{180-182} This form of analysis is based on tying relationship of various metabolites based on their mutual correlation and covariance. The strength of these ties can be graded base on assigning them weights. Unweighted analyses are not usually helpful in genetics but an example using our dataset for incident hypertension showed consistent results as observed by using OPLS model. Figure 3.3 shows unweighted network of metabolites observed in our study. Note that new diagnosis of hypertension at 5 year follow up is closely surrounded by alanine, valine and glutamine. The closer a metabolite to another metabolite is, the stronger is the correlation of that metabolite with other metabolite is. However, the links are myriad and such a network is frequently called as “hairpin”.

There are several forms of networks that use interaction patterns between various features. It helps in understanding the internal structure of the biological system. Metabolomics is appropriate for such analyses as metabolites do not function in isolation and are highly correlated. They occur along certain pathways and function in certain ways which can be exploited via network analysis. This is a module based analysis. There are several steps in its construction:

1. A network is constructed using the interaction patterns between features of interest
2. A module or pathway is identified
3. Biologically interesting and plausible module is identified and is related with an external source of clinical data, SNPs, gene ontology, proteomics
4. Robustness of study module is checked across different datasets (e.g. in similar phenotypes or different species)

5. Identify key drivers in the biologically interesting features (e.g. proteins, genes, metabolites) in pathways to be used as biomarkers, therapeutics, or for experimental validation

In contrast to unweighted network analysis, weighted correlation networks have been used in reducing high dimensional data to understand the more discrete modular structure in the data. These analyses have helped to identify important relationships and pathways using gene methylation, RNA-seq, f-MRI datasets. However, these networks have not been used for assessing metabolites. Basically, a network can be represented by adjacency symmetric undirected matrix \( A = [a_{ij}] \), that encodes whether or how a pair of nodes is connected. The adjacency matrix reports the connection strength between two features in weighted networks.

We also performed a weighted co-expression network analysis where we identified 3 modules of fatty acids, sterols and amino acids associated with prevalent hypertension. These modules are outlined in Figure 3.4 and show consistency of our results obtained in chapter II where we found same amino acids (module 2) associated with prevalent hypertension. However, this analysis found other metabolites that were associated with hypertension. Even though these metabolites are previously shown in literature to be associated with hypertension, we did not observe these relationships in OPLS models used in chapter II analyses.

3.4.1.3 CLUSTER ANALYSIS

A cluster analysis can be described as unsupervised learning. The grouping is based on independent variables. There are various types of clustering methodologies. Hierarchical clustering is usually used for small number of independent variables. K-mean clustering is used for huge number of observations and number of clusters is defined a priori. A dendrogram is a visual depiction of clusters and their
membership. Cluster analysis uses distance measurements obtained by correlation in a matrix based on measured characteristics of subjects and helps in identifying various groups of subjects who are more similar to each other than those in other groups. This method can be used to describe groups of subjects closely related with a disease phenotype such as hypertension, without having any a priori assumption about them.

A 2-way cluster analysis can help to identify relationships between subjects as well as metabolites as in our case. We used Ward’s method of hierarchical clustering method to define various clusters. For Ward’s minimum variance method, at each stage we define the distance between two clusters $K$ and $L$ as

$$D_{K,L} = \frac{\sum_j (x_{K,j} - x_{L,j})^2}{\frac{1}{n_K} + \frac{1}{n_L}}$$

where $j$ indexes the cluster’s 87 metabolites. That is, $x_{K,j}$ is the value of the $j$th standardized variable score for cluster $K$ and $n_K$ is the number of original subjects in cluster $K$ at that stage. Put another way, at any stage the distance between two clusters is defined as a function of the sum of squared differences in standardized scores. The distance is calculated between every possible combination of two clusters.

**3.4.1.4 EXAMPLE**

Cluster membership was assigned and a scree plot was created, which helped in identifying 3 as the optimum number of clusters. This was the lowest number of clusters that explained maximum variance in the dataset and is called elbow method. This cluster was associated with 55% increased risk of prevalent hypertension (OR 1.55; 1.11 – 2.19, p = 0.01) compared to others. Differences in the shades of blue and red color show the correlations between metabolites and various subjects (Figure 3.5). Blue color showed negative correlation and red color showed
positive correlation. The strength of the shade of color depended upon the strength of correlation. There are visual differences in the blue cluster metabolites (group at highest risk of hypertension) compared to the others. In participants without baseline hypertension, 2-way cluster analysis revealed a group of participants associated with decreased risk of incident hypertension (OR 0.56; 95% CI 0.32, 0.96, \( p = 0.03 \)) (Figure 3.6). Most of these differences are observed in amino acids and long chain fatty acids. Thus, these additional analyses show that cluster analysis can also help in identifying individuals with hypertension.

3.4.2 SUPERVISED MULTIVARIATE ANALYSES

Supervised analyses attempt to link metabolomic patterns to specific traits or attributes of the sample or the subjects from which the samples were obtained. Thus, unlike unsupervised analyses, supervised analyses depend on a specific prior hypothesis. However, similar to unsupervised analyses, multivariate methods are typically required to address the large number of highly correlated metabolomic signals that are collected. These methods include cluster analysis, partial least squares (PLS) regression, orthogonal projections to latent structures (OPLS) and network analyses. The main principle of these methods is to organize the data into correlated groups that can explain the covariance between metabolites and the dependent variable. The dependent variable can be a continuous or categorical variable. To estimate risk ratios, metabolite concentrations frequently are changed into standardized units and evaluated in multiple regression models adjusted for confounders. Most studies use a false discovery rate or modified Bonferroni’s method to establish a threshold for significance testing. The approaches in chapter II were supervised multivariate analyses and are briefly described below:

3.4.2.1 PARTIAL LEAST SQUARES (PLS) REGRESSION

A statistical method that bears some relation to principal component analysis is partial least square regression. This method may be considered a supervised
multivariate method in which an iterative procedure fits a linear regression model by projecting dependent variable (blood pressure change) and independent variables (metabolites). When a binary categorical variable is used, a variant of partial least square regression called partial least square – discriminant analysis is used. This method decomposes independent variables (metabolites) and dependent variable (blood pressure or hypertension) to maximize the covariance between their respective matrices. Similar to principal component analysis, the output of this model also provides principal components. The first principal component explains the most variation between X and Y. The second component is orthogonal to first principal component and explains lesser variation then the first principal component. A loadings plot provides weights for individual metabolites that are important in the estimation of principal component.

3.4.2.2 ORTHOGONAL PROJECTIONS to LATENT STRUCTURES (OPLS)

Trygg et al. developed another multivariate supervised analytical method for chemometrics which was similar to PLS in 2002. The central concept of OPLS was that it separated the information between dependent and independent variables into predictive and uncorrelated data. In other words, it removes non-correlated systematic variation of independent variables that is not explained by dependent variable (prevalent and incident hypertension in our study). This improves interpretability of the model, however the prediction of the model does not change and is similar to PLS regression model.

3.5 CONCLUSION

In conclusion, metabolomics is a powerful tool that can provide us new ways to interpret causal pathways leading to disease phenotypes. However, due to highly correlative structure of the data multivariate analyses are essential to make meaningful inferences. Nevertheless, a prerequisite for this is the intimate knowledge of the metabolic pathways, robustness of sample preparation, separation
and detection methods. Ultimately, metabolomic studies will provide new strategies for risk prediction, biomarker detection, and identification of novel mechanisms to reduce risk for cardiovascular health care burden.
### Table 3.1: Sample types used in metabolomics studies

<table>
<thead>
<tr>
<th>Sample</th>
<th>Advantage</th>
<th>Disadvantage</th>
<th>Approximate Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plasma/Serum</strong></td>
<td>Minimally invasive, readily available, provides information of metabolic footprint of many metabolic reactions</td>
<td>May not reflect specific tissue level changes (e.g. cardiac changes)</td>
<td>150 – 550 µL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue level metabolic fluxes cannot be measured</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requires deproteination of mass spectrometer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anticoagulant used for plasma collection may interfere with the endogenous metabolites</td>
<td></td>
</tr>
<tr>
<td><strong>Urine</strong></td>
<td>Noninvasive, readily available, provides information of metabolic footprint of many metabolic reactions</td>
<td>May not reflect specific tissue level changes (e.g. cardiac changes)</td>
<td>500 – 1000 µL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tissue level metabolic fluxes cannot be measured</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contains high amount of urea which can damage mass spectrometer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Differences in pH of urine leads to difficulty in evaluation of various metabolites especially when nuclear magnetic resonance spectroscopy was used</td>
<td></td>
</tr>
<tr>
<td><strong>Primary cell culture</strong></td>
<td>Tissue level information can be obtained, metabolic fluxes can be studied</td>
<td>Takes time, expensive, technically demanding and requires complex procedures, phenotype may change over time</td>
<td>100 – 600 µL</td>
</tr>
<tr>
<td><strong>Tissue</strong></td>
<td>Tissue level information can be obtained, metabolic fluxes can be studied, accurate description of phenotype</td>
<td>Takes time, expensive, technically demanding and requires complex procedures, invasive for some of the specific tissues</td>
<td>&gt;20 mg</td>
</tr>
</tbody>
</table>
### Table 3.2: Advantages and disadvantages of liquid chromatography methods

<table>
<thead>
<tr>
<th></th>
<th>RPLC</th>
<th>HILIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td>Well understood retention mechanisms</td>
<td>Retains highly polar analytes not retained by RPC</td>
</tr>
<tr>
<td></td>
<td>Widely applicable</td>
<td>Less interference with matrix components</td>
</tr>
<tr>
<td></td>
<td>Faster equilibrium</td>
<td>Complementary selectivity to RPC</td>
</tr>
<tr>
<td></td>
<td>Wide variety of mobile phases available</td>
<td>Polar compounds retained more than parent compound</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High organic mobile phases promotes enhanced ESI – MS response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Direct injection of precipitate supernatant without dilution is possible</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Facilitates use of lower volume sample</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Increased throughput</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Poor peak shape with basic analytes</td>
<td>Sensitive to sample diluent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mechanism not well understood</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longer equilibration time</td>
</tr>
</tbody>
</table>

RPLC; reversed phase liquid chromatography, HILIC; hydrophilic interaction liquid chromatography
FIGURE LEGENDS

**Figure 3.1** Steps of a general metabolomic study

**Figure 3.2**: Intermediary metabolites detected after 10 – min of liquid chromatography Fourier transformation mass spectrometry of human plasma (Adopted with permission from Annu Rev Nutr. Aug 21, 2012; 32: 183–202)

**Figure 3.3**: A weighted network of metabolites in Insulin Resistance Atherosclerosis Study cohort with prevalent hypertension

**Figure 3.4**: Unweighted network of metabolites in Insulin Resistance Atherosclerosis Study cohort without baseline hypertensive participants

**Figure 3.5**: Two way cluster analysis showing 3 different clusters of subjects with blue cluster at highest risk of prevalent hypertension

**Figure 3.6**: Two way cluster analysis showing 3 different clusters of subjects with green cluster at highest risk of incident hypertension
Figure 3.1 Steps of a general metabolomic study

a. Sample collection and storage
b. Tissue extraction
c. Pool aliquots from each extracted sample in batch
   Quality control (QC) sample preparation
   Mass spectrometry analysis
   Metabolite identification

Following steps involve detailed metabolomic analysis:

- Retention time (min)
- MS/MS spectra
- Mass accuracy
- Chromatographic profiles

f. Comparison of metabolite features between protocols
   - ESI+
   - ESI-

- Percentage of identified metabolites

Data preprocessing-multivariate analysis, e.g., PCA
Figure 3.2: Intermediary metabolites detected after 10–min of liquid chromatography Fourier transformation mass spectrometry of human plasma (Adopted with permission from Annu Rev Nutr. Aug 21, 2012; 32: 183–202)
Figure 3.3: Unweighted network of metabolites in Insulin Resistance Atherosclerosis Study cohort without baseline hypertensive participants
qGLN, glutamine; qTYR, tyrosine; qLYS, lysine; HTNv1, prevalent hypertension; qLANO, lanosterol; qCHOL, cholesterol; q7.DHC, 7-dehydrocholesterol; q7a.HC, 7-alpha hydroxycholesterol; qAC16:0, acylcarnitine 16:0; qAC14.0, acylcarnitine 14:0; qAC6.0, acylcarnitine 6:0; qAC2.0, acylcarnitine 2:0; m16.1n7, palmitoleic acid; m16:0, palmitic acid; m18.1n7, oleic acid; m22.4n6; Dihomo-gamma linoleic acid; m22.6n3, Docosahexaenoic acid
Figure 3.5: Two way cluster analysis showing 3 different clusters of subjects with blue cluster at highest risk of prevalent hypertension
Figure 3.6: Two way cluster analysis showing 3 different clusters of subjects with green cluster at highest risk of incident hypertension
BIBLIOGRAPHY


33. Azevedo A, Barros H. C-reactive protein is an intermediate step between obesity and hypertension. *Archives of internal medicine*. 2006;166:1526-1527; author reply 1527


66. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661-678


76. Crampton JH. Sodium restriction in the diet for hypertension. *Clinics of the Virginia Mason Hospital*. 1946;24:3-6
83. Moore TJ, Conlin PR, Ard J, Svetkey LP. Dash (dietary approaches to stop hypertension) diet is effective treatment for stage 1 isolated systolic hypertension. *Hypertension*. 2001;38:155-158


107. Wakabayashi I, Masuda H. Lipoprotein (a) as a determinant of arterial stiffness in elderly patients with type 2 diabetes mellitus. *Clinica chimica acta; international journal of clinical chemistry*. 2006;373:127-131


148. Tanabe M, Kanehisa M. Using the kegg database resource. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]*. 2012;Chapter 1:Unit1 12


150. Petersen-Jones HG, Johnson KB, Hitomi K, Tykocki NR, Thompson JM, Watts SW. Transglutaminase activity is decreased in large arteries from hypertensive rats, compared to normotensive controls. *American journal of physiology. Heart and circulatory physiology*. 2015;ajpheart 00402 02014


CURRICULUM VITAE
Waqas T. Qureshi MD

Section on Cardiovascular Medicine, Phone: (336) 716-7427
Watlington Building, Medical Center Blvd
WQureshi@wakehealth.edu
Winston-Salem, NC 27157

Education

Wake Forest University, Winston-Salem, NC, May 2015
Candidate for Masters of Science, Area of Specialization: Clinical, Translational and Population Sciences
Dissertation: “Metabolomic Features of Hypertension”
Faculty Advisor: Dr. Lynne Wagenknecht and Dr. David Herrington
GPA: 3.92

Rawalpindi Medical College, Health Sciences University, Lahore, Pakistan, April 2007
Bachelor of Medicine and Surgery
High Honors in Internal Medicine, Pediatrics, Gynecology/Obstetrics, General Surgery, Ophthalmology, Clinical pathology, General pathology, Behavioral Sciences, Physiology, Anatomy, Biochemistry.
Rank 1 among 1542 medical students of Punjab, Pakistan. Won 5 consecutive gold medals for being summa cum laude medical student each year from 2002 – 2007.

Research Interests

- Epidemiology of metabolic heart disease
- Metabolomics
- Epidemiology of Atrial Fibrillation

Positions and Employment

Clinical Research Fellow, Magnetic Resonance Imaging, Department of Radiology, University of North Carolina, NC 2008 – 09
Intern, Henry Ford Hospital/Wayne State University, Detroit, MI 2009 – 10
 Resident, Henry Ford Hospital/Wayne State University, Detroit, MI 2010 - 2012
Chief resident, Henry Ford Hospital/Wayne State University, Detroit, MI 2012 – 13
Instructor Internal Medicine, Wayne State University, Detroit, MI  2012 - 13
Instructor Hospitalist Medicine, Wake Forest University, Winston Salem, NC 2013 - 14
Fellow, Division of Cardiovascular Medicine, Wake Forest University, Winston Salem, NC 2013

Honors
Summa Cum Laude Rank 1 in Medical School for 5 consecutive years, 2002 - 2008
Young Investigator Travel Award by American Society of Nuclear Cardiology, 2010
Junior Resident of the Year Award, Henry Ford Hospital/Wayne State University, Detroit, MI 2010 - 11
Travel Award by American Society of Geriatrics, 2011
Senior Resident of the Year Award, Henry Ford Hospital/Wayne State University, Detroit, MI 2011 - 12
Resident Research Award by Henry Ford Hospital/Wayne State University, Detroit, MI 2011 - 12
Molecular Epidemiology Working Group (MEG) of the AACR Scholar-in-Training Award, 2012
Diplomat of American Board of Internal Medicine, 2012
Jahnigen & Williams Young Investigator Award by American Geriatrics Society, 2013
Travel Award by American Society of Nephrology, 2013
Letter of honor for top reviewer by Editor of Annals of Internal Medicine, 2014
Selected abstract from Top 10 from a subspecialty American Heart Association Meetings, 2014

Research Experience

Research Fellow, Wake Forest University, Winston Salem, NC 2013 - present
Research Advisor: Dr. Lynne Wagenknecht and Dr. David Herrington
- Determine the metabolic profiles of individuals that develop hypertension, diabetes and atherosclerosis
- Presented results at American Heart Association Epidemiology and Nutrition conference San Francisco, CA March 2014 and Annual Scientific Sessions of American Heart Association in Chicago November 2014

Research Fellow, Henry Ford Hospital/Wayne State University, Detroit, MI, Summer 2011
Research Advisors: Dr. Hani Sabbah
• Evaluated the Impact of Boston Scientific Vagal Nerve Stimulator in Dogs on Improvement of Left Ventricular Function and Inflammatory Markers

• Presented results at Annual Sessions of American College of Cardiology in Chicago March 2012.

**Research Fellow, Henry Ford Hospital/Wayne State University, Detroit, MI, 2009 - 2012**

Research Advisor: Dr. Mouaz Al-Mallah

• Collaborated with national and international researchers for development of COronary CT Angiography Evaluation For Clinical Outcomes: An International Multicenter Registry (Clinical trial identifier: NCT01443637) with more than 27,000 participants that have undergone cardiovascular computed tomography angiography

• Collaborate with researchers from Johns Hopkins University for development of a Henry Ford Exercise Testing Project (FIT Project) which is a multi-ethnic retrospective cohort of 69,885 with up to 21 years of follow-up.

• Presented multiple abstracts and published papers from these databases

**Other Experiences**

**Instructor, Rawalpindi Medical College, Health Sciences University, Pakistan, 2005 - 06**

• Collaborated with University of Virginia, VA for implementation of telemedicine in disasters

• Taught basic health workers to triage and manage patients with spinal cord injury

• Organized multiple blood drives and performed a knowledge attitude and practices survey to improve participation of young adults in blood drives

• Presented outcomes of early management of paraplegic patients at American College of Physicians in Michigan 2010.

**Instructor, Life Savers Organization, Non – government non – profit organization, Pakistan 2005-06**

• Taught American Heart Association BLS and ACLS to medical students

**Instructor, Heart File, Non – government non – profit organization, Pakistan 2001 – 2004**

• Implemented a program to improve early identification of hypertension and non-communicable cardiovascular diseases (myocardial infarction and stroke) in a large city of >1 million population.
Leadership in Medicine, Henry Ford Hospital/Wayne State University School of Medicine, workshop 2012

- Collaborated with University of Virginia, VA for implementation of telemedicine in disasters
- Taught basic health workers to triage and manage patients with spinal cord injury
- Organized multiple blood drives and performed a knowledge attitude and practices survey to improve participation of young adults in blood drives
- Presented outcomes of early management of paraplegic patients at American College of Physicians in Michigan 2010.

Business in Medicine Course, University of Michigan, Ann Arbor, MI 2012 - 13

- Learned about health care delivery models, increasing productivity expectations, infrastructure and managing staff in medical practices
- Learned about health cost containment and improving health care efficiencies

Quality Improvement of Resident Morbidity and Mortality Report Initiative, Henry Ford Hospital/Wayne State University, 2012

- Systematically developed morbidity and mortality conference into a fish and bone analysis to evaluate policies, procedures, processes, and people

Stanford University Resident Leadership Course May 2011

- Learned through various modules to deal with conflicts, develop effective teaching skills and improve communication skills

Code Blue Refresher Courses May 2011

- Organized a quality improvement project to improve quality of ACLS codes by residents

Patient diabetes education workshop 2010 – 2012

- Developed a program to educate diabetic population to increase adherence, telemonitoring of point of care blood glucose levels which led to improvement of HbA1c by an average of 1%
- Increased hospital revenue by implementing a single point of care HbA1c testing in high risk inpatients
Volunteer Service

**Member, American College of Cardiology, Accreditation Compliance Work Group**

*02/ 2015 - present*

Provide peer review regarding the scientific rigor and fair balance of ACCF certified educational content, thereby ensuring adherence to the Accreditation Council for Continuing Medical Education (ACCME) Standards and Guidelines including the Standards for Commercial Support, and American Nurses Credentialing Center (ANCC) Content Integrity Standards. Review faculty and planner conflict of interest disclosure information and identify potential conflicts of interest. Conduct the conflict of interest process and recommend appropriate methods of resolution to ACCF staff. Review and approve Certification Request Forms (CRF) for CME/CNE certification of education activities which have been submitted to ACCF staff. Meet annually at ACC meeting and quarterly phone conferences.

**Member, American College of Cardiology, Geriatrics Cardiology Group 2013 - present**

Monthly conference calls

**Member, Multiethnic-cohort of Atherosclerosis, Winston Salem NC, 10/ 2013 - present**

Member of P&P of metabolomics group

Primary interest groups of Nutrition and dietary studies, hypertension and metabolomics, metabolic syndrome and metabolomics

**Senior Member, Internal Medicine Residents Council, Michigan 2011-12**

Internal Medicine Residents Council is a peer elected council of resident physicians. We advocate and collaborate between the residents and program leadership. Through this council we have brought changes in the system based work flows especially better hand offs from ER physicians and early discharges from the floors

**Executive Council Member, Michigan State Medical Society, Michigan 2010 – 13**

I have worked as a delegate initially and then got promoted to executive council member for Wayne County Medical Society of Southeast Michigan as well as Michigan State Medical Society. I have advocated resident physicians at the state level through this platform. One of them being the decrease in the state funding of graduate medical education.

**Vice President, Internal Medicine Residents Council, Henry Ford Hospital/Wayne State University, 2010 – 12**

- Worked as a spokesperson for medical residents
• Led effort to develop formal structure of the council which included developing bylaws

**St. Francis Cabrini Clinic, America’s Oldest Free Clinic Detroit, MI 2010 – 2013**

• Provided volunteer services and care to the patients without insurance
• Managed chronic kidney disease patients with multiple comorbidities in resource poor settings

**Member, Medicos Aides Society, Non – government non – profit organization, Pakistan 2004-07**

• Organized blood drives and collected blood for hemophilia and thalassemia patients; estimated 1200 bags of blood every year

**Peer Reviewed Publications**

Almahmoud MF, O’Neal WT, **Qureshi W**, Soliman EZ. Electrocardiographic versus Echocardiographic Left Ventricular Hypertrophy in Prediction of Congestive Heart Failure in the Elderly. Clinical Cardiology (in press)


**Qureshi W**, Alirhayim Z, Khalid F. Letter: Warfarin in atrial fibrillation with dialysis, is the glass half full? Circulation


Kaatz S, Qureshi W, Lavender RC. Venous thromboembolism: What to do after anticoagulation is started. Cleveland Clinic journal of medicine. 2011;78:609-618


**BOOK CHAPTERS**


Waqas Qureshi. Cardiology Questions for American Board of Internal Medicine. – in press 2013

Barboza J, Cabrera R, Olarte K, Qureshi W. Intern Survival Series, Henry Ford Hospital 2012 - 2013
**ORAL PRESENTATIONS**

**Qureshi W**, O’Neal WT, Al-Mallah MH. Digoxin Use in Atrial Fibrillation Is Associated with Increased Risk of Mortality: A Systematic Review and Meta-Analysis of More than 200,000 Patients at American College of Cardiology 64th Annual Scientific Session and TCT@ACC-i2, March 14-16, 2015 in San Diego, CA.


Ambulgekar N, El-Refai M, Njeim M, **Qureshi W**. (2011, September). Does documentation of daily weights leads to decrease in readmissions in patients with acute decompensated heart failure?. Oral Presentation presented at: American College of Physicians; Grand Rapids, MI.


Singla S, Mittal C, **Qureshi W**. Predictors of Post-Transplant Cardiac Dysfunction in Liver Transplant Recipients. American Transplant Congress 2013
Hassan S, **Qureshi W**, Kuriakose P. Thromboembolism in Von Willebrand Disease, Annual Meeting of American Society of Hematology, 2012

Khalid F, Hassan S, Qureshi W. Erythromelalgia – an uncommon presentation after aspirin withdrawal, society of General Internal Medicine 2012


**PEER REVIEW ONLINE PUBLICATION**


**MEDICAL NEWS**


**MANUSCRIPT REVIEWER EXPERIENCE**

01/2012 – present Reviewer for Atherosclerosis

08/2011 – present Reviewer for Journal of Clinical and Experimental Cardiology

07/2011 – present Reviewer for Annual meeting of Society of General Internal Medicine and served as Chairman of reviewers of clinical vignettes of Mid-West Regional meeting of Society of General Internal Medicine

09/2012 – present Reviewer for Annals of Internal Medicine

12/2012 -01/2013 Reviewer for Meeting abstracts for Annual conference of American College of Physicians
07/2013 – present  Reviewer for Journal of American Heart Association
09/2013 – present  Reviewer for American Journal of Kidney Diseases
09/2013 – present  Reviewer for Biomedical Central Women’s Health Journal
09/2013 – present  Reviewer for American Journal of Cardiology
11/2013 – present  Reviewer for Digestive Disease Sciences
12/2013 – present  Reviewer for Cardiac Diabetology
03/2013 – present Reviewer for Cleveland Clinical Journal of Medicine
05/2014 – present  Reviewer for Journal of American Medical Association (JAMA)

**EDITOR**

07/2013 – present  Journal of Cardiology and Clinical Research
10/2013 – present  Austin Journal of Clinical Cardiology
10/2014 – present  Frontiers of Cardiovascular Epidemiology – Nature Publishing Group

**POSTER PRESENTATIONS**

Al-Mallah MH, Ahmed A, **Qureshi W**. The Frequency and Predictors of Referral to Coronary Angiography Among Patients with Moderate or Severe Ischemia on Stress Testing at American College of Cardiology 64th Annual Scientific Session and TCT@ACC-i2, March 14-16, 2015 in San Diego, CA.

**Qureshi W**, Blaha M, Keteyian S, Brawner C, Al-Mallah MH. No Racial Differences in the Association of Cardiorespiratory Fitness with Incident Atrial Fibrillation: Results from The Henry Ford Exercise Testing Project (The FIT Project) at American College of Cardiology 64th Annual Scientific Session and TCT@ACC-i2, March 14-16, 2015 in San Diego, CA.


Mohammad AF, **Qureshi W**, Soliman EZ. Electrocardiographic versus Echocardiographic Left Ventricular Hypertrophy in Prediction of Congestive Heart Failure in the Elderly at American Heart Association EPI/Lifestyle Scientific Sessions Baltimore, MD Mar 3 – 6, 2015.


**Qureshi W**, Traditional Risk Factors For Atrial Fibrillation In Individuals With Normal Left Atrial Size: The Atherosclerosis Risk In Communities Study AHA epi/npam San Francisco Mar 18 – 21, 2014

**Qureshi W**, Risk of Incident Hypertension and Diabetes Mellitus in Habitual Snorers and Sleep Apneic Individuals; The Multiethnic Study of Atherosclerosis AHA epi/npam San Francisco Mar 18 – 21, 2014


presented at: Annual Meeting of American College of Nuclear Cardiology; Baltimore, MD.


Khalid F, Mittal C, Alirhayim Z, Qureshi W. CKD and endstage dialysis patients National Kidney Week Orlando, Florida 2013

Iani Patsias, Chetan Mittal, Mohammad Elbatta, Kiran Garikapati, Patrick Bradley, Gagandeep Cheema, Aishwarya Kuchipudi, Hassaan Raza Jafri, Zaid Alirhayim, Syed Hassan, Fatima Khalid, Qureshi W. Optimal Duration to Restart Warfarin after Discontinuation of Therapy in Atrial Fibrillation Patients That Developed Gastrointestinal Bleeding ACC San Francisco Mar 10 2013

Zaid Alirhayim, Syed Hassan, Fatima Khalid, Qureshi W. Clinical and economic burden of incidental findings suspicious for cancer on non-contrast chest computed tomography performed during myocardial perfusion imaging AACR Annual Meeting 2013 in Washington, DC

Zaid Alirhayim, Qureshi W. Suboptimal resumption of warfarin in octogenarians with atrial fibrillation that develop gastrointestinal bleeding Annual Scientific Meeting of the American Geriatrics Society 2013, Grapevine, TX

Pablo Buitron, Qureshi W. Simplified Criteria To Risk Stratify For The Mode Of Echocardiography In Patients With Mrsa Bacteremia At Low Risk Of Infectious Endocarditis (IE) Annual Scientific Meeting of American Heart Association 2012, Los Angeles, CA

Patrick Bradley, Qureshi W. Resuming Anticoagulation After An Episode Of Gastrointestinal Bleeding In Patients With Atrial Fibrillation Annual Meeting of Society of Hospitalist Medicine 2013, National Harbor, Maryland


Qureshi W, Fatima Khalid, Adam Greenbaum. Cardiac complications of Sheehan’s Syndrome Annual Meeting of Society of General Internal Medicine 2011

Namo S, Olarte K, Qureshi W, Hector N, Buran G. Utility of Anion Gap as a predictor of Lactic Acidosis in patients in the emergency department SHM 2011 Grapewine, TX


Mittal C, Singla S, Qureshi W, Hassan S, Huang M. Predictors of Post-Transplant Cardiac Dysfunction in Liver Transplant Recipients ATS 2013

Garikapati K, Qureshi W. Outcomes Of Patients With Left Ventricular Diastolic Dysfunction In Adult Hematopoietic Stem Cell Transplantation AHA QCOR 2013


**Laboratory and Computer Skills**

Northern blotting; SDS-PAGE; microinjection; spectrophotometry; mitochondrial separation and electron transport chain assays

Microsoft Word, Excel, PowerPoint; SAS; SPSS; SIMCA; Comprehensive meta-analysis; JMP; RevMan v. 5.0; R-project

**Professional Memberships**

American Heart Association, Fall 2010 – Present

American College of Cardiology, Fall 2010 – Present

American College of Physicians, Fall 2009 – Present

**Selected Honors, Awards, and Fellowships**

1. Young Investigator Travel Award by American Society of Nuclear Cardiology 2010 ($500)
2. Best cardiovascular resuscitation poster at annual scientific sessions of American Heart Association, Orlando, FL 2011
3. Student/Resident/Jahnigen & Williams Award Annual Scientific Meeting of the American Geriatrics Society 2013, Grapevine, TX
5. Molecular Epidemiology Working Group (MEG) of the AACR Scholar-in-Training Award ($1,500)
6. Senior Resident of the Year Award 2011-12
7. Junior Resident of the Year Award 2010-11
8. Best Intern Award 2009
10. Henry Ford Hospital resident research award 2011 – 12 ($1000)
11. Travel Award by American Geriatrics Society 2012 Travel Award ($500)
12. Best Poster presentation in ACP meeting 2011 Resident Research Award 2012 Outstanding Resident Cash Award 2012 ($1500)
13. American Society of Nephrology Research Travel Award ($800)

References

**Lynne E. Wagenknecht DrPH**
Professor and Associate Director,
Division of Public Health Sciences,
Wake Forest University School of Medicine,
Winston Salem, NC 27157
Phone: +13367167652
Email: lwgnkcht@wakehealth.edu

**David M. Herrington MD MHS**
Dalton McMichael Chair in Cardiovascular Medicine
Professor, Section on Cardiovascular Medicine
Vice-Chair for Research, Dept. of Internal Medicine
Medical Center Boulevard \ Winston-Salem, NC 27157
p 336.716.4950
Email: dherring@wakehealth.edu