NOVEL APPROACH TO EXAMINE THE INTERACTIVE ROLE OF DIETARY, LIFESTYLE, AND GENETIC FACTORS ON CARDIOMETABOLIC RISK

Niha Zubair

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Nutrition, Gillings School of Global Public Health.

Chapel Hill
2013

Approved by:

Dr. Linda S. Adair

Dr. Elizabeth Mayer-Davis

Dr. Michelle Mendez

Dr. Karen Mohlke

Dr. Kari North

ABSTRACT

NIHA ZUBAIR: Novel Approach to Examine the Interactive Role of Dietary, Lifestyle, and Genetic Factors on Cardiometabolic Risk
(Under the direction of Linda Adair)

With modernization, cardiometabolic (CM) disease risk has increased in low- and middle-income countries. We sought to understand CM risk in these settings, both in young adults, for whom prevention is an important goal, and in an older population, for whom risk is better established. Differences in the prevalence and patterns of co-occurrence of CM risk factors likely reflect variation in diet, lifestyle, and genetics. Innovative methods are needed to better understand the synergistic effects between these modifiable and non-modifiable factors on CM risk.

We evaluated the patterning of CM risk factors in a young adult population participating in the 2005 Cebu Longitudinal Health and Nutrition Survey (CLHNS) (n = 1,621). Using cluster analysis, we grouped individuals by CM biomarkers and then assessed how diet, adiposity, and environment predicted these CM clusters. Despite the population's youth and leanness, cluster analysis found patterns of CM risk. While measures of adiposity strongly predicted cluster membership, diet and environment also independently predicted clustering.

Next, we aimed to capture the complex relationship between genetics, adiposity, and CM risk. Here we created genetic risk scores for inflammatory and lipid traits; these scores combined the relatively small additive effects of individual SNPs in Filipino women in the 2005 CLHNS (n= 1,649). We found that each genetic risk score explained a greater

proportion of variance in the specified CM trait than any given individual SNP. In addition, we observed that the triglyceride genetic risk score interacted with measures of adiposity to influence triglyceride levels.

Lastly, we used cluster analysis to identify groups of women from the 2005 CLHNS (n= 1,584), who shared similar patterns of genetic risk across multiple CM phenotypes. Here we found five distinct genetic risk clusters. These genetic risk clusters along with measures of adiposity and dietary factors, predicted CM trait levels and patterns in this population.

In conclusion, our results suggest that examining the synergistic influence of modifiable and non-modifiable factors on CM traits and patterns can help provide insight into the etiology of CM diseases, and thus potentially inform targeted prevention efforts.

ACKNOWLEDGEMENTS

This work would not have been possible without the input, advice, guidance and support of many people. Most importantly, I am grateful to my advisor, Linda Adair, who has taught me to think deeply and critically, while encouraging me to follow my research interests. In addition to being a superior scientific mentor, Linda is a passionate, caring, and thoughtful human being.

I also am extremely grateful to all the members of my committee: Dr. Elizabeth Mayer-Davis, for pushing me to articulate the clinical importance of my research. Dr. Michelle Mendez, for taking the time to discuss the pros and cons of the various epidemiological methods. Dr. Karen Mohlke, for her detailed and superior genetic advice, which taught me to critically evaluate my own methods. Dr. Kari North, a genetic epidemiology guru, who always encouraged me to further my understanding for this subject.

Finally, I owe a debt of gratitude to all my UNC friends and colleagues. I honestly could not have asked for a better experience in North Carolina. I would like to give a special shout-out to my Morenas, Carmen and Carolina. I am so lucky to have met you both. Our balance of work and play, along with our meshing of personalities, made for the best experience in the Ph.D. program. Lastly, I would like to thank my parents for encouraging me to pursue a higher education and always supporting (and feeding) me during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BP	Blood pressure
BMI	Body mass index
CM	Cardiometabolic
CRP	C-reactive protein
CVD	Cardiovascular disease
HDL-C	High-density lipoprotein cholesterol
HOMA-IR	Homeostatic model assessment insulin resistance
IR	Insulin resistance
LDL-C	Low-density lipoprotein cholesterol
MetS	Metabolic syndrome
Mlogit	Multinomial logistic regression
OW	Overweight
SES	Socioeconomic status
SNPs	Single nucleotide polymorphisms
TG	Triglycerides
WC	Waist circumference
WHO	World Health Organization

Chapter 1. INTRODUCTION

OVERVIEW

Rapid nutritional and lifestyle changes in developing countries contribute to a growing burden of overweight, visceral adiposity, and associated cardiometabolic (CM) diseases. Eighty percent of global deaths from these diseases occur in low- and middle-income countries. These concerns are especially pertinent for Asians: the World Health Organization (WHO) concluded that the risk of CM-based diseases is elevated for Asians with a body mass index (BMI) greater than 23 kg/m^2, suggesting the use of a lower cut-point for overweight (OW). Research demonstrates that CM risk factors tend to co-occur and may be causally interrelated. Furthermore, differences in the prevalence and patterns of co-occurrence of CM risk factors likely reflect variation in diet, lifestyle, and genetics. However, there is insufficient research on the interplay between these modifiable and non-modifiable factors and how they relate to CM risk patterns. In addition, gene-gene and gene-environment interactions are particularly important in relation to complex traits such as CM diseases and innovative methods are needed to account for potential synergistic effects. Studying the interactive influence of dietary, lifestyle, and genetic factors can help provide insight into the etiology of CM diseases, and thus aid in informing targeted prevention efforts.

SPECIFIC AIMS AND OBJECTIVES

We used cross-sectional data from the 2005 Cebu Longitudinal Health and Nutrition Survey (CLHNS) of Filipino middle-aged women and their young adult offspring. Since 1983 the CLHNS has collected detailed longitudinal data from a cohort of women and their offspring. The Metro Cebu area is the second largest and most rapidly growing urban area in the Philippines. In tandem with increasing modernization, Cebu is experiencing a higher prevalence of OW and associated CM risks, including hypertension, elevated inflammation, and adverse lipid profiles. The demographic and health trends observed in Cebu represent current trends occurring in Asia. The CLHNS is a unique dataset in that it has detailed diet, lifestyle, and genetic data; this combined with the rapid nutrition and lifestyle transition make the CLHNS an ideal dataset for our study.

Here we examined the patterns and determinants of CM risk factors among individuals in this study population. Our previous research identified five profiles of Filipino middle-aged women with similar CM characteristics: (1) Healthy, (2) Elevated blood pressure (BP), (3) Low high-density lipoprotein cholesterol (HDL-C), (4) Insulin resistant, and (5) Elevated C-reactive protein (CRP). We found modifiable risk factors for these five CM patterns, including measures of adiposity and dietary intake. Next we extended this analysis to their young adult offspring, for whom prevention is still an important goal.

In order to further understand the etiology of CM risk in these older women, for whom CM risk is more established, we used genetic risk scores, which combined the relatively small additive effects of individual single nucleotide polymorphisms (SNPs), to better capture the complex relationship between genetics, adiposity, and CM risk.

Lastly, we used a novel application of cluster analysis to identify groups of these women who share similar patterns of genetic risk scores across multiple CM phenotypes. We then examined how these genetic risk clusters related with CM traits and patterns in this population, while accounting for other factors such as age, diet, and anthropometry.

The following describes the specific aims for this study:

Aim 1: Determine biologically relevant patterns of co-occurrence of CM characteristics in young adults and model the determinants of these CM patterns. Previous research used cluster analysis to identify biologically relevant groups of middle-aged Filipino women with similar CM characteristics. Here we used the same method in the young adult offspring; variables used to create clusters included biomarkers representing hypertension, inflammation, insulin resistance, and lipid abnormalities. We modeled the determinants of these CM clusters in young adults, focusing on risk factors such as adiposity and dietary intake.

Aim 2: Develop genetic risk scores to better capture the complex relationship between genetics, adiposity, and CM risk. In this aim we developed genetic risk scores for inflammatory and lipid traits in Filipino women; each score represents a summation of the genetic risk variants associated with a single CM trait. We assessed the ability of these scores to explain the variation in CM traits as opposed to individual genetic variants. We also examined whether measures of adiposity, one of the strongest predictors of CM risk, interacted with the genetic risk scores to synergistically influence trait levels.

Chapter 2. LITERATURE REVIEW

SCOPE OF THE PROBLEM

Developing countries undergoing nutrition and lifestyle changes display an increasing burden of overweight (OW), visceral adiposity, and associated cardiometabolic (CM) diseases; this emphasizes the need for research in these settings.[10-12] These concerns are especially pertinent for Asians. The World Health Organization (WHO) concluded that the risk of CM based diseases is elevated for Asians with a body mass index (BMI) greater than 23 kg/m^2, suggesting the use of a lower cut-point for OW in these populations. At the same BMI, Asians tend to have more percent body fat and central adiposity than other ethnicities.[13] In addition, studies show an increasing prevalence of the metabolic syndrome (MetS) in Asian populations.[14-16]

The demographic and health trends observed in Cebu Longitudinal Health and Nutrition Survey (CLHNS) represent current trends throughout Asia. The wide range of environmental, social, behavioral, and genetic data in the CLHNS can help improve understanding of the predictors of CM risk. Further, this sample provides variation in CM phenotype (e.g. high waist circumference in non-obese). In the Philippines, a country with a population of nearly 90 million, recent surveys identified OW, cigarette smoking, hypertension, high cholesterol levels, type 2 diabetes, and heart disease as emerging public health issues.[12, 17-19] Ischaemic heart disease was one of the top causes of all age mortality in 2002, accounting for 10% of all deaths.[20] According to the WHO, estimated disability adjusted life years from heart disease are higher in the Philippines than in the U.S. or

China.[21] In addition, national survey data in the Philippines found a high prevalence of the following CM risk factors: low levels of high-density lipoprotein cholesterol (HDL-C) in 60.2% of men and 80.9% of women, abdominal obesity in 17.7% of men and 35.1% of women, blood pressure (BP) >130/85 mmHg in 33.3%, hypertriglyceridemia in 20.6%, and fasting glucose >100 mg/dL in 7.1%.[22] A body of literature demonstrates that such CM risk factors tend to co-occur and may be causally interrelated.[22-26] Differences in the prevalence and patterns of co-occurrence of CM risk factors likely reflect variation in diet, lifestyle, and genetics. Studying the interactive influence of dietary, lifestyle, and genetic factors can help provide insight into the etiology of CM diseases, and thus aid in creating targeted prevention efforts, especially for at-risk Asian populations.

WHY STUDY PATTERNS OF CARDIOMETABOLIC RISK FACTORS?

Why not study individual risk factors?

A substantial literature links obesity to insulin resistance, dyslipidemia, vascular dysregulation, and inflammation, and consequently, to elevated risk of CM diseases.[27-29] These factors tend to cluster and together significantly predict CM disease, leading to the definition of the MetS. The term MetS refers to a grouping of CM risk factors with a supposed common underlying pathophysiology.[23] While the MetS definition is frequently used in research and clinical settings[30], there lacks a clear and consistent definition of this term, leading to inconsistencies.

Why not simply diagnose individuals with metabolic syndrome?

While the original concept of MetS has been useful, there exist many concerns with using this definition. One concern with MetS includes the lack of research demonstrating that MetS stems from a common underlying pathophysiology[31-33]; treatment of MetS is no

different that treating the specific CM factors present.[34, 35] Research suggests that CM risk depends not only on the diagnosis of MetS, but the actual clustering of CM risk factors present.[36] [30] Another concern with MetS includes the arbitrary inclusion/exclusion of specific CM risk factors, for example inflammatory factors are typically not included in MetS definitions.[37]

The composite MetS definition ignores the heterogeneity in the patterns of CM risk factor clustering. Simply using this definition could obscure documented differences in the prevalence and patterns of CM risk factors across ethnic, age, and sex groups.[30, 38] Understanding these differences can provide insight into the etiology and treatment options for CM diseases. As an example of the heterogeneity in risk factor patterning across ethnicities, low HDL-C followed by elevated BP are the most prevalent components of the MetS among Filipinos, whereas in the United States abdominal obesity followed by low HDL-C are the most prevalent MetS components.[22] MetS is becoming more common in young adults as rates of obesity increase;[25, 39] therefore understanding the prevalence and patterning of CM risk factors in younger adults may help in the prevention of future CM disease. In addition, many studies show sex differences in the prevalence and patterns of CM risk factors,[40, 41] thus we examined sex differences between CM risk factors in young adults in the CLHNS.

The MetS definition typically includes 5 basic indicators: central obesity, elevated triglycerides (TG), low HDL-C, elevated BP, and elevated fasting plasma glucose [37]. By simply using this definition one fails to include indicators, such as inflammatory markers, important in predicting CM outcomes. Specifically, research shows that elevated C-reactive protein (CRP) levels, often not included in the classic MetS definition, predict cardiovascular

disease (CVD) and type 2 diabetes independent of MetS status.[42] In addition, evidence from western populations shows that such inflammatory factors co-occur with other MetS risk factors.[43] In order to allow for flexibility in the CM biomarkers we examined the clustering of individual CM risk factors rather than apply a MetS definition.

Gaps in cardiometabolic risk factor clustering

Few studies look at how CM risk factors cluster.[15, 44, 45] The substantial variability in exposures and outcomes among individuals in the CLHNS enhances the likelihood that we can identify significant and clinically important relationships of diet and lifestyle on CM risk factors. Our previous research identified five profiles of Filipino middle-aged women with similar CM characteristics: (1) Healthy, (2) Elevated BP, (3) Low HDL-C, (4) Insulin resistant, and (5) Elevated CRP. We found modifiable risk factors for these five CM patterns, including measures of adiposity and saturated fat intake. Differences in the prevalence and patterns of co-occurrence of CM risk factors likely reflect diet, lifestyle, and genetics. Therefore we aimed to study the interactive influence of these factors, with the purpose to provide insight into the etiology of CM diseases.

HOW DO DIETARY, ENVIRONMENTAL, AND ANTHROPOMETRIC FACTORS INFLUENCE CARDIOMETABOLIC RISK?

As low- and middle-income countries undergo the nutrition transition, large shifts in diet and activity patterns coincide with urbanization and economic development.[46] This rapid transition allows us to capture changes one cannot capture so readily in the U.S. These changes include: less physical activity and increased consumption of fat, caloric sweeteners, and meat.[46] Such diet and physical activity changes have been shown to influence CM risk factors. For example, Yao and colleagues found that a diet high in carbohydrates, low in

polyunsaturated fat, and low in fruits and vegetables was associated with an adverse lipid profile, independent of body fatness. They also found an independent beneficial effect of physical activity on HDL-C and fasting insulin.[47] Evidence also suggests that diet and lifestyle can simultaneously affect CM risk. For example, a study conducted in a Gambian population showed that a high fat diet did not result in an atherogenic lipid profile in a lean population with a high level of occupational activity.[48]

Developing countries like the Philippines are experiencing large shifts in diet and activity patterns.[46] The traditional Filipino diet contains high amounts of refined carbohydrates and sodium, accompanied by low amounts of protein. Filipinos typically consume refined white rice at every meal, while consuming little animal foods and fat compared to western populations. Coconut oil, the Filipinos' main source of fat, contains notably high levels of lauric acid and recent studies have shown that lauric acid has a more favorable effect on the total cholesterol to HDL cholesterol ratio than any other fatty acid, either saturated or unsaturated, primarily by increasing HDL-C levels.[49] In regards to physical activity, Filipinos traditionally engaged in high amounts of due to the large physical demands of work.

Other environmental factors include infection and pathogenicity. [50-52] According to a 2004 WHO report, infectious diseases account for more than 30% of all mortality in Southeast Asia.[53] Exposure to a pathogenic environment serves as a primary source of inflammatory stimuli, and results in elevated levels of CRP. Earlier investigations in CLHNS samples showed evidence of the role of exposure to a pathogenic environment in predicting plasma CRP levels.[52]

Excess adiposity is one of the strongest predictors of CM disease and risk.[54-57] The lifestyle changes described above contribute to a growing burden of OW, visceral adiposity, and thus associated CM diseases.[58-61] These concerns are especially pertinent for Asians; compared with Caucasians, Asians have increased visceral adiposity and greater insulin resistance at similar levels of BMI.[62-65]

Prior work in CLHNS found substantial age and secular trends in weight among adult women, notably a nearly 7-fold increase in OW over a 21-year period.[66] This increase is associated with adverse CM profiles, including hypertension, elevated markers of inflammation, and adverse lipid profiles.[52, 67, 68] Waist circumference (WC), a proxy for visceral adipose tissue, is among the best-established predictors of CM risk and past work in the CLHNS and other Asian populations support this notion.[45, 52, 68-70] Research has also demonstrated that increased WC predicts CM abnormalities in both normal weight and OW/obese individuals, highlighting the potential for visceral fat to influence development of CM risk factors independent of overall BMI status.[44]

Although research demonstrates that these dietary, environmental, and anthropometric factors associate with CM risk factors, solely looking at these characteristics, without accounting for genetics, will never provide a comprehensive understanding of the etiology of CM diseases. Therefore we also assessed how genes influence CM risk in this population.

HOW DO GENES AFFECT CARDIOMETABOLIC RISK?

Previous research has found SNPs associated with specific CM risk factors such as: TG, HDL-C, low-density lipoprotein (LDL-C), systolic BP, diastolic BP, glucose, homeostatic model assessment insulin resistance HOMA-IR, and CRP.[1, 2, 5, 7-9, 71, 72] Several

studies have tried to identify underlying genetic risk factors for MetS, but no study has successfully found genetic variants that are shared by all the components of MetS, challenging the view that MetS has a common genetic background.[73-75] MetS is a complex trait with numerous features. Due to its heterogeneity in clustering of CM risk factors, MetS probably results from an interaction of dietary, lifestyle, and genetic factors.[30] This complexity makes it difficult for the identification of replicable genetic associations that might eventually form the basis of clinical predictive tests for MetS.[76] Therefore we found it essential to study the interactions between gene-gene and gene-environment and their associations with CM risk factor patterns.

HOW DO GENE-GENE AND GENE-ENVIRONMENT INTERACTIONS RELATE TO CARDIOMETABOLIC RISK?

Previous studies have shown that gene-gene and gene-environment interactions associate with specific CM risk factors; in this sense "environment" represents any non-genetic measure. In a study looking at pairwise gene-gene interaction, Tam et al. confirmed the associations of two common genetic polymorphisms of *GCK* and *GCKR* and their interaction on fasting plasma glucose in Chinese adults and adolescents.[77] As an example of gene-diet interaction, a recent Japanese study found a significant interaction between the *CDKAL1* polymorphism and dietary energy intake that influenced glucose regulation, possibly through impaired insulin secretion.[78] Recent work in the CLHNS found that central obesity might accentuate the effect of the TG-increasing allele of an *APOA5* variant.[7] Additionally, earlier research in this study population found the first evidence that exposure to a pathogenic environment may modify the genetic influence at the *HNF1A*, *LEPR* and *6q16.1* loci on plasma CRP levels.[1]

These studies, while verifying the notion that interactions between genes, diet, anthropometric, and other environmental factors influence CM traits, only examined pair-wise interactions. Complex diseases are probably caused by interactions between multiple genes and environmental factors,[76] which may explain why many of the pair-wise interactions found are rarely replicated. Therefore in this study we used innovative methods to examine these effects.

WHAT METHODS CAN ACCOUNT FOR POTENTIAL SYNERGISTIC EFFECTS OF GENE-GENE AND GENE-ENVIRONMENT?

Individually, common genetic variants only minimally explain common complex diseases such as CVD, type 2 diabetes, and other related CM conditions.[79] Jointly considering the relatively small effects of these individual single nucleotide polymorphisms (SNPs) may better capture underlying genetic risk associated with these diseases. Recently, genetic risk scores have been implemented to interrogate the impact of multiple SNPs of CM disease.[80-84] In addition, perhaps combinations of genetic variants (vs. a single SNP) interact with environmental factors and better predict CM risk.

Some studies create a genetic risk score by summing up the number of risk alleles pertaining to a single CM phenotype, while other studies similarly construct a genetic risk score, but choose risk alleles associated with a CM disease of interest. The latter method combines risk variants pertaining to multiple phenotypes with the intention of better capturing the intricate relationship between genetics and common complex disease. Still, the majority of these studies find slight to no improvement in classifying at-risk individuals. [82, 83] This approach masks the actual patterning of genetic risk across phenotypes. Perhaps, understanding this heterogeneity in genetic risk clustering may aid in predicting and

preventing CM disease,[79, 85, 86] especially since these diseases themselves display a specific patterning of risk factors including insulin resistance, dyslipidemia, hypertension, and inflammation.[87, 88]

WHY CEBU?

The CLHNS covers a wide range of health-related topics specific for each stage of the life cycle. The CLHNS has followed, since 1983, a cohort of women who reside in Metropolitan Cebu in the central Philippines. Metro Cebu, with a population nearing 2 million, shares many similarities with other large cities in developing and transitional countries of Asia, enhancing the likelihood that results from this location are generalizable to other settings. It is one of the fastest growing and rapidly developing regions of the country and has particular relevance for understanding CM trends.[89]

SUMMARY

This work addresses an important international public health issue: understanding the multifactorial etiology of CM diseases. Recent nutritional and lifestyle changes in developing countries have propagated the burden CM diseases, emphasizing the need for research in these settings.[10-12] This rapid transition combined with detailed diet, lifestyle, and genetic data, make the CLHNS an ideal dataset for this research. Documented differences in the prevalence and patterns of co-occurrence of CM risk factors likely reflect diet, lifestyle, and genetics. Therefore we aimed to study the interactive influence of these factors, with the purpose to provide insight into the etiology of CM diseases, and thus help inform targeted prevention efforts.

Chapter 3. CLUSTERING AND DETERMINANTS OF CARDIOMETABOLIC RISK FACTORS AMONG FILIPINO YOUNG ADULTS

OVERVIEW

With modernization, cardiometabolic (CM) disease risk has increased in low and middle-income countries. To better understand CM disease etiology, we evaluated the patterning of CM risk factors in a susceptible young adult population.

Participants included 1,621 individuals from the 2005 Cebu Longitudinal Health and Nutrition Survey. Using cluster analysis, we grouped individuals by the following biomarkers: triglycerides (TG), high-density and low-density lipoprotein cholesterol (HDL-C, LDL-C), C-reactive protein (CRP), blood pressure (BP), homeostasis model assessment of insulin resistance, and fasting glucose. Using multinomial logistic regression models we assessed how diet, adiposity, and environment predicted CM clusters.

We identified 5 distinct sex-specific clusters: (1) Healthy/high HDL-C (with the addition of high LDL-C in women), (2) Healthy/low BP, (3) High BP, (4) Insulin resistant (IR)/high TG, (5) High CRP. Though we did not identify a specific cluster primarily defined by low HDL-C, over 65% of men and 70% of women had this trait, making low HDL-C the most pervasive CM risk factor. In men and women, decreased intake of saturated fat predicted membership in the High CRP cluster, compared to the Healthy clusters. Men with poorer environmental hygiene were more likely to be in the High CRP cluster, compared to the Healthy clusters (odds ratio 0.74 [95% CI 0.60-0.90] and 0.83 [0.70-0.99]). Adiposity measures were the strongest predictors of membership in the IR/high TG.

Despite the population's youth and leanness, cluster analysis found patterns of CM risk. While adiposity measures predicted clustering, diet and environment also independently predicted clustering, emphasizing the importance of screening lean and overweight individuals for CM risk. Finding predictors of risk in early adulthood could help inform prevention efforts for future CM disease.

BACKGROUND

Low and middle-income countries undergoing rapid nutrition and lifestyle changes display an increasing burden of obesity, visceral adiposity, and associated diseases.[58, 90, 91] These concerns are heightened for Asians and young adults. The risk of CM diseases has been shown to be elevated among Asians at lower levels of BMI, prompting the World Health Organization to recommend the use of a lower BMI cut-point to define overweight in this population.[92] In addition, overweight young adults are likely to remain overweight throughout life and have increased risk of CM diseases, such as cardiovascular disease and type 2 diabetes.[90, 93-95]

A substantial literature links obesity to insulin resistance, dyslipidemia, hypertension, and inflammation, and consequently to elevated risk of CM diseases.[4, 27, 96, 97] These factors tend to co-occur, leading to the definition of the metabolic syndrome (MetS).[23] However, using the MetS definition presents several problems. First, there is a lack of research demonstrating that MetS stems from a common underlying pathophysiology[98-100]: treatment of MetS is no different than treating the specific CM factors present.[34, 35] In addition, objectively evaluating the clustering of CM risk factors, rather than the diagnosis of MetS, is more useful for predicting and preventing disease.[87, 88] Lastly, the inclusion/exclusion of specific CM risk factors in the MetS definition is unfounded. For example, inflammation, as

indicated commonly by elevated C-reactive protein (CRP), is often not included in the classic MetS definition, despite that it predicts CVD and type II diabetes independent of MetS status.[4]

Motivated by the downfalls of applying a uniform MetS definition, we used cluster analysis to identify groups of young adults, from the 2005 Cebu Longitudinal Health and Nutrition Survey (CLHNS), who share similar patterns of CM risk factors. Furthermore, differences in the prevalence and patterns of co-occurrence of these risk factors likely reflect variation in modifiable and non-modifiable characteristics. However, there is a lack of research relating such characteristics to the clustering of CM risk factors, particularly among young adults. Thus we sought to determine how diet, adiposity, environment, and sex related to the clustering of CM risk factors in Filipino young adults.

This study population is ideal for our research question because (1) the majority of participants did not have any clinical disease, (2) Cebu is undergoing a rapid nutrition and lifestyle transition, and (3) the CLHNS includes detailed diet, lifestyle, anthropometric, and biomarker data. By using an at-risk young adult population, we can gain a better understanding of how modifiable and non-modifiable characteristics relate to CM risk factors in young adulthood, which can help inform prevention strategies for future CM disease.

METHODS

Survey design

The present analysis includes young adults (index children) assessed in the 2005 CLHNS (mean age 21 years).[89] Briefly, the CLHNS is a community-based cohort of women and their index children followed since 1983. The original participants included all pregnant women from 33 randomly selected communities of Metro Cebu, who gave birth between

May 1, 1983, and April 30, 1984. Surveys took place immediately after birth, bimonthly for 2 years, in 1991, 1994-5, 1998-99, 2002, and 2005. In 2005, fasting blood was drawn for CVD biomarkers and genetics. Here we use data from the index children still participating in the 2005 CLHNS.

Blood samples were collected on 1,790 individuals. Excluding women who were pregnant at the time of blood draw, we clustered 1,621 (889 men and 732 women) individuals with complete fasting biomarker data and with CRP levels < 10 mg/L (a level representing low-level basal inflammation rather than current/recent illness).[88] Of those clustered, 1,569 individuals with complete diet, socioeconomic, and anthropometric data were included in the multivariate analysis (871 men and 698 women). All data were collected with informed consent, using protocols approved by the institutional review board of the University of North Carolina, Chapel Hill.

Cardiometabolic biomarkers

Fasting plasma CM biomarkers included triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), glucose, insulin, and CRP. Blood samples were collected in participants' homes in the morning after an overnight fast. Venous blood was collected in EDTA anti-coagulant vacutainer tubes. After mixing to inhibit clotting, glucose was measured immediately using the glucose dehydrogenase method (One Touch Ultra Blood Glucose Monitoring System, Lifescan Johnson and Johnson). Blood samples were stored on ice for no more than 2 hours and were then centrifuged to separate plasma.

After separation, samples were frozen and remained frozen at -80 °C until ready for analysis. Total lipid concentrations were measured at the Emory Lipid Research Laboratory

using enzymatic methods with reagents from Beckman Diagnostics on the Beckman Diagnostics CX5 chemistry analyzer (Fullerton, CA). HDL-C was determined using the homogeneous assay for direct determination (Genzyme, Exton, PA). LDL-C was determined using the Friedewald formula, except if triglycerides exceeded 400 mg/dl then LDL-C was directly determined using a homogenous assay (Genzyme, Exton, PA). The Emory Lipid Research Laboratory is a participant in the CDC/NHLBI Lipid Standardization Program to ensure accuracy and precision of the determinations. Plasma insulin was measured using automated Bayer® ADVIA Centaur chemiluminescent methods.[101] CRP concentrations were determined using a high sensitivity immunoturbidimetric method (Synchron LX20, lower detection limit: 0.1 mg/L).

Other cardiometabolic biomarkers included homeostatic model assessment insulin resistance (HOMA-IR), and systolic and diastolic blood pressure (BP). HOMA-IR was calculated as 22.5/ (insulin × glucose). Systolic BP and diastolic BP were measured in triplicate after a 10 minute seated rest using a mercury sphygmomanometer. The mean of the three measurements was used.

We used cutpoints for these biomarkers based on recommendations from the International Diabetes Federation (IDF), the American Heart Association, and other previously recognized and accepted cutpoints (Table 1).[88, 96, 102, 103]

Anthropometry

Body weight, height, and waist circumference (WC) were measured using standard techniques[104]. BMI was calculated as the ratio of weight (kg) to height (m^2). We used cutpoints for Asians to define overweight (OW) as a BMI $\geq$ 23 kg/m^2.[105] Cutpoints for Asians define central adiposity as WC $\geq$ 80 cm for women and WC $\geq$ 90 cm for men[96]; since

less than 8% of individuals have WC above these cutpoints, we used median values (men = 71 cm and women = 66.5 cm) to define at-risk groups.

Dietary data

Dietary data were derived from two 24-hour dietary recalls and the mean intake was used in the analyses. A nutritionist reviewed all dietary recalls immediately after collection. When implausible values were found, interviewers revisited respondents for verification. Energy and saturated fat intakes were calculated using the Philippines Food Composition Tables.[19, 106]

Sociodemographic and lifestyle characteristics

We included the following sociodemographic and lifestyle characteristics in our analysis: household assets, urbanicity, environmental hygiene, graduation status, smoking status, alcohol consumption, and level of physical activity.

The assets score, ranging from 0 to 10, measures household economic status. It reflects the type of lighting used, ownership of house, type of housing material, and ownership of selected assets: television, air conditioner, tape recorder, refrigerator, and motor vehicle. We dichotomized this variable at the median, ≤ 5 assets or > 5 assets. The urbanicity index is comprised of 7 components derived from CLHNS barangay-level survey data.[107] A higher score designates a more urbanized barangay. We dichotomized this variable at the median, ≤ 43 or > 43. The hygiene score measures environmental cleanliness using data on the interviewer's rating of cooking area, immediate area around the house, toilet type, and water source. The score ranges from 0 to 9 with larger values indicating greater cleanliness.[52] High school (HS) graduation status was classified as yes or no. Smoking and alcohol

consumption were assessed as yes or no. The majority of women did not smoke (> 93%) therefore we did not include this covariate in their analysis.

Physical activity was assessed by asking respondents to report time spent in all activities during a typical day. Each activity was assigned a metabolic equivalent (MET) value using the updated Compendium of Physical Activities. We identified minutes/week of moderate to vigorous physical activity (MVPA=METS >3) performed during occupation, leisure time, and household activities to approximate an overall minutes/week of MVPA. The majority of women did not participate in any MVPA (82%), thus MVPA was only included in the analysis of the men. We categorized physical activity: no MVPA, low to medium amounts of MVPA (<sex-specific median of 720 minutes/week), and high amounts of MVPA (≥720 minutes/week).

Cluster analysis

We performed a K-means cluster analysis to identify groups of young adults with similar CM risk factor patterns using SAS PROC FASTCLUS (SAS version 9.2, SAS Institute, Cary, NC). This procedure implements the K-means clustering algorithm (least squares method). K-means clustering uses the Euclidean distance, computed from input variables, to assign cluster membership by minimizing the distance among subjects in a cluster while maximizing the distance between clusters. The procedure first selects cluster seeds, a set of points calculated as a first guess of the cluster means. Next it calculates the Euclidean distance from each subject to each cluster seed; each subject is assigned to the nearest seed to form temporary clusters. The means of each of the temporary clusters are calculated and replace the seed values. Distance calculation and member assignment progress in an iterative fashion until no further changes occur.[108, 109]

Final cluster solutions are sensitive to initial seed values. To remedy this problem and to use a more objective approach to picking a cluster solution we created an algorithm modified from a previous clustering algorithm.[69, 110] This algorithm performed 1,000 iterations of each cluster procedure using randomly generated initial cluster seeds. For each of the 1,000 cluster solutions it calculated the ratio of between-cluster variance to within-cluster variance or $R^2/(1-R^2)$, where R^2, pooled across all variables, represented the ability to predict each input variable from the cluster.[109] We wanted to maximize the ratio of between-cluster variance to within-cluster variance and therefore wanted to find the largest R^2. The algorithm identified the iteration/cluster solution with the largest R^2.[110]

Cluster analysis was conducted separately in the women and men to account for differences in patterns of CM risk by sex. The variables entered into the cluster analysis were chosen to represent hypertension, inflammation, insulin resistance, and lipid abnormalities, and included sample and sex-specific standardized values of TG, HDL-C, LDL-C, systolic BP, diastolic BP, glucose, HOMA-IR, and CRP.

Statistical analysis

We used sex-specific multinomial logistic regression models to examine predictors of cluster membership in young adults. For men and women, the full models included the following covariates: OW status (BMI $\geq$ 23 kg/m^2), high WC (WC $\geq$ 80cm), % energy intake from saturated fat, energy intake, alcohol consumption, household assets, urbanicity, environmental hygiene, and education status; smoking status and level of physical activity were additionally included for men. We used the multivariate nutrient density method to control for confounding and to remove extraneous variation due to total energy intake.[111]

Multicollinearity between % of energy intake from saturated fat and total energy intake was not an issue (correlation coefficient<0.4).

We conducted manual backwards elimination (likelihood ratio test) to test whether each covariate improved model fit. If it did not improve model fit and also did not predict cluster membership the covariate was removed. Throughout our analysis we used P <0.05 (2-sided) as the criterion for significance. Regression analysis was performed with Stata 12.0 (Stata Corporation, College Station, TX, 2006).

RESULTS

Prevalence of cardiometabolic risk

Baseline characteristics are presented in Tables 2 and 3 for men and women respectively. Men had a high prevalence of low HDL-C (60%), while a low prevalence of elevated LDL-C (6%), elevated fasting glucose (3%), elevated HOMA-IR (3%), and elevated CRP (7%). Women had a high prevalence of low HDL-C (68%), while a low prevalence of elevated TG (9%), hypertension (2%), elevated fasting glucose (3%), elevated CRP (8%), and elevated HOMA-IR (4.5%). In comparison to women, men had a higher prevalence of elevated TG and hypertension. While in comparison to men, women had a higher prevalence of low HDL-C, elevated LDL-C, and elevated HOMA-IR.

Cluster analysis

We conducted a series of cluster analyses with 3 to 6 clusters specified, and chose the 5-cluster solution for both men and women because it yielded distinct CM risk factor patterns and each cluster contained approximately ≥ 5% of the sample.[112] The 5-cluster solutions had $R^2 = 0.35$ and $R^2 = 0.36$ in men and women respectively, indicating slightly more than 1/3rd of the variance in CM biomarkers was explained by the clusters. For men we identified the

five clusters as: (1) Healthy/high HDL-C, (2) Healthy/low BP, (3) High BP, (4) Insulin resistant (IR)/high TG, and (5) High CRP. For the women we identified the same five clusters except the first cluster also included LDL-C: (1) Healthy/high HDL-C/high LDL-C. We named the clusters according to what risk factor(s) had the highest/lowest mean relative to other clusters (Figure 1); the term "healthy" represents low Z-scores for the majority of CM biomarkers (except HDL-C). We ordered these clusters such that clusters 1-5 in men and women represented similar CM patterns.

Cardiometabolic patterns in young adult men

Mean z-scores of the CM biomarkers varied markedly by cluster (Figure 1), as did the prevalence of risk factors defined by cutpoints to represent "high risk" (Tables 2 and 3)**.** Men in the Healthy/high HDL-C cluster (n=144, 16%) had the zero prevalence of low HDL-C. Men in the Healthy/low BP cluster (n=315, 35%) had the lowest prevalence of hypertension (0%) and a high prevalence of low HDL-C (73%). Men in the High BP cluster (n=290, 33%) had a relatively high prevalence of hypertension (38%) and low HDL-C (69%). Men in the IR/high TG cluster (n=65, 7%) had highest prevalence of elevated TG (88%), elevated fasting glucose (15%), and elevated HOMA-IR (29%); in addition, these men had a high prevalence of low HDL-C (68%). Lastly, men in the High CRP cluster (n=75, 8%) had the highest prevalence of elevated CRP (80%), and a high prevalence of low HDL-C (75%).

Cardiometabolic patterns in young adult women

Mean z-scores of the CM biomarkers varied markedly by cluster (Figure 1), as did the prevalence of risk factors defined by cutpoints to represent "high risk" (Tables 2 and 3)**.** Women in the Healthy/high HDL-C/high LDL-C cluster (n=158, 22%) had the lowest prevalence of low HDL-C (27%) and a relatively high prevalence of LDL-C (32%); none of

these women had hypertension. Women in the Healthy/low BP cluster (n=252, 34%) had no hypertension and a high prevalence of low HDL-C (86%). Women in the High BP cluster (n=233, 32%) had a relatively high prevalence of hypertension (6%), and low HDL-C (73%). Women in the IR/high TG cluster (n=48, 7%) had highest prevalence of elevated TG (50%), elevated fasting glucose (27%), and elevated HOMA-IR (63%); in addition, these women had a high prevalence of low HDL-C (79%). Lastly, women in the High CRP cluster (n=41, 6%) had the highest prevalence of elevated CRP (95%) and a high prevalence of low HDL-C (73%); none of these women had hypertension.

Multivariable analysis in young adult men

The final multivariate model in the men included the following covariates: high WC, OW status, % of energy intake from saturated fat, energy intake, household assets, smoking status, alcohol consumption, and environmental hygiene (Table 4).

Compared to the Healthy/high HDL-C cluster: being normal weight and not consuming alcohol increased the likelihood of being in the Healthy/low BP cluster; higher WC increased the likelihood of being in the High BP cluster; higher WC, being OW, having more assets, and smoking increased the likelihood of being in the IR/high TG cluster; decreased % of energy intake from saturated fat and lower environmental hygiene increased the likelihood of being in the High CRP cluster.

Compared to the Healthy/low BP cluster: higher WC, being OW, and not smoking increased the likelihood of being in the High BP cluster; higher WC, being OW, and having more assets increased the likelihood of being in the IR/high TG cluster; being OW, decreased % of energy intake from saturated fat, having more assets, not smoking, alcohol

consumption, and decreased environmental hygiene increased the likelihood of being in the High CRP cluster.

Compared to the High BP cluster: being OW, having more assets, and smoking increased the likelihood of being in the IR/high TG cluster; lower WC and decreased environmental hygiene increased the likelihood of being in the High CRP cluster.

Compared to the IR/high TG cluster, lower WC and not smoking increased the likelihood of being in the High CRP cluster.

Multivariable analysis in young adult women

The final multivariate model in the women included the following covariates: high WC, OW status, % of energy intake from saturated fat, energy intake, urbanicity, and HS graduation status (Table 4).

Compared to the Healthy/high HDL-C/high LDL-C cluster: no covariates increased the likelihood of being in the Healthy/low BP cluster; decreased % of energy intake from saturated fat, increased energy intake, and not graduating from HS increased the likelihood of being in the High BP cluster; being OW increased the likelihood of being in the IR/high TG cluster; decreased % of energy intake from saturated fat, increased energy intake, and lower urbanicity increased the likelihood of being in the High CRP cluster.

Compared to the Healthy/low BP cluster: higher WC and being OW increased the likelihood of being in the High BP cluster; higher WC and being OW increased the likelihood of being in the IR/high TG cluster; being OW, decreased % of energy intake from saturated fat, and lower urbanicity increased the likelihood of being in the High CRP cluster.

Compared to the High BP cluster: being OW increased the likelihood of being in the IR/high TG cluster; decreased urbanicity increased the likelihood of being in the High CRP cluster. No covariates distinguished the IR/high TG cluster from the High CRP cluster.

DISCUSSION

Cluster analysis is a useful tool for identifying groups of individuals who share similar CM risk factor patterns. In contrast with the MetS definition, cluster analysis allows for flexibility. For example, we included a measure of inflammation in the cluster analysis, a risk factor not commonly included in MetS definitions, which allowed us to identify a distinct group characterized primarily by elevated CRP levels. In addition, we did not include WC as a criterion for the clustering algorithm, unlike the IDF, which requires elevated WC in the definition.[96] This enabled us to distinguish for which clusters elevated WC (a modifiable risk factor) predicted cluster membership.

By using cluster analysis, we were able to capture the heterogeneity in patterns of CM risk factor clustering. Research has demonstrated that mortality risk is dependent on the actual combinations of CM risk factors, highlighting the importance of understanding these sex differences in the clustering of CM risk factors.[38] While our analysis found relatively similar CM risk clusters among men and women, the predictors of these clusters varied by sex. Perhaps as these young adults age more distinct CM patterns between men and women will emerge.

A high prevalence of low HDL-C, a risk factor for heart disease, has been reported in the Philippines and other Asian populations.[67, 113, 114] This was reflected in the cluster analysis results: over 65% of men and 70% of women, not in the Healthy/high HDL-C cluster, had low HDL-C levels.

Previous work among the mothers in Cebu suggested that saturated fat intake, perhaps from coconut oil, could be protective against low HDL-C levels.[69, 115] However in young adults, we saw saturated fat intake had varying relationships with different CM risk factors. In both men and women, decreased % energy intake from saturated fat predicted membership in the High CRP group when compared to the two Healthy clusters. In addition, a decrease in % saturated fat intake predicted membership in the High BP group in women, compared to the Healthy/high HDL-C/high LDL-C group.

The association of saturated fat intake with healthy CM profiles could reflect the types of saturated fats consumed in this population. Coconut oil, the most common and traditional cooking oil in Cebu, is rich in lauric acid.[116] Lauric acid improves the total cholesterol to HDL-C ratio, more than any other saturated or unsaturated fatty acid, primarily by increasing HDL-C levels.[49] In addition, a replacement of carbohydrates with lauric acid produces a decrease in this ratio.[49] This proves especially relevant in our study population since over half of energy intake comes from carbohydrates, the majority of which are refined rice products. Other studies have found diets rich in coconut oil or in saturated fat do not alter markers of inflammation, fasting glucose, fasting insulin, HOMA-IR, or incident diabetes.[117, 118]

Men with poorer environmental hygiene (increased pathogenicity) were more likely to be in the High CRP cluster compared to the two Healthy clusters. These results support previous research conducted in the CLHNS and reinforce the involvement of pathogen exposure in activating pro-inflammatory pathways.[50-52] But why do we fail to observe this hygiene effect in women? Adiposity relates more strongly with inflammation in women than

in men, thus it is possible the effects of adiposity overwhelmed the effects of the hygiene score in women.[100, 119]

As expected, WC and OW status were the strongest predictors of membership in the IR/high TG cluster, underscoring the adverse health effects of excess visceral adipose tissue, for which WC serves as a proxy.[120] WC is among the best-established predictors of CM risk and past work in the CLHNS and other populations support this notion.[45, 50, 52, 68] Research has also demonstrated that increased WC predicts CM abnormalities in both normal weight and OW individuals, highlighting the potential for visceral fat to influence the development of CM risk factors, independent of BMI.[121]

This population has a low prevalence of overweight (18%). However, among normal weight individuals, CM risk factors were already present: 63% of the sample with BMI<23 kg/m^2 had low HDL-C. Despite leanness, cluster analysis found patterns of CM risk. While measures of adiposity predicted some of these patterns, modifiable factors such as dietary intake and pathogen exposure also independently predicted cluster membership. This emphasizes the importance monitor and screen lean individuals for CM risk and future CM diseases, especially in Asian populations where the risk of CM diseases is elevated at a lower BMI (likely due to increased visceral fat at lower BMIs).[92]

Several limitations warrant mention. A limitation of cluster analysis is that not all individuals within a certain cluster necessarily share all characteristics. For example, in our "Healthy" clusters we found the average Z-scores for CM risk biomarkers were relatively low (except HDL-C), but we cannot ascribe these low values to each individual in the cluster.

Attrition and selection bias are also concerns. Migration of the more educated, urban segment of the original cohort has left us with a sample that is no longer representative of the

population from which it was drawn.[89] The sample was further reduced due to selection criteria. From the full sample of 1,887 young adults in 2005, the multivariate analysis included those that were fasting and not pregnant with complete biomarker, anthropometric, and socioeconomic data, resulting in an analytic sample of 1,621. Comparing baseline socioeconomic characteristics, we found a lower percentage of HS graduates among women excluded vs. those included in the analysis (68% vs. 78% respectively, ANOVA $P<0.05$).

In sum, despite the population's young age, lack of clinical disease, and relative leanness, cluster analysis identified distinct patterns of CM risk factors. By using cluster analysis we made fewer assumptions regarding the underlying etiology and allowed relationships among CM risk factors to emerge from the data themselves. We found sex-specific clustering of CM risk factors and were able to evaluate how diet, adiposity, and environmental factors influenced these patterns. As expected, measures of adiposity predicted specific CM risk patterns. However, diet and environmental factors also independently predicted risk factor clustering. This emphasizes the importance of screening both lean and OW individuals for CM risk, especially in Asian populations where the risk of CM diseases is elevated at lower BMI.[92] Future studies examining how CM risk patterns change longitudinally could provide insight to how CM risk evolves across the life course. Finding modifiable and non-modifiable predictors of CM risk in early adulthood could help inform targeted prevention efforts for future CM disease.

TABLES AND FIGURES

Table 3.1 Criteria for defining elevated cardiometabolic risk

Risk factors	Cutpoint
Triglycerides*	≥ 150 mg/dL
HDL cholesterol*	Males < 40 mg/dL
	Females < 50 mg/dL
LDL cholesterol†	≥ 130 mg/dL
Systolic BP*	≥ 130 mm Hg
Diastolic BP*	≥ 85 mm Hg
Glucose*	≥ 100 mg/dL
HOMA-IR‡	≥ 4.65 mg/dL x μg/mL
CRP§	> 3.0 mg/dL

Cutpoints represent levels at which there is an increased risk of cardiometabolic diseases. *Cutpoints are defined by the IDF.[96] †Cutpoint is defined by the National Cholesterol Education Program.[102] ‡Cutpoint is defined by Stern et al. [103] §Cutpoint is defined by the American Heart Association. [88]

Table 3.2 Characteristics of young adult men in the 2005 CLHNS

	All Men (n= 871)	Healthy/High HDL-C (n= 139)	Healthy/Low BP (n= 312)	High BP (n= 282)	IR/High TG (n= 65)	High CRP (n= 73)
Age, y	21.0 ± 0.0	20.9 ± 0.0	20.9 ± 0.0	21.0 ± 0.0	20.9 ± 0.0	21.0 ± 0.0
Cardiometabolic biomarkers*						
Elevated TG (%)	19.7 ± 1.3	15.3 ± 3.0	11.1 ± 1.8	15.9 ± 2.1	87.7 ± 4.1	20.0 ± 4.6
Low HDL-C (%)	59.6 ± 1.6	0.0 ± 0.0	72.7 ± 2.5	69.3 ± 2.7	67.7 ± 5.8	74.7 ± 5.1
Elevated LDL-C (%)	5.7 ± 0.8	4.9 ± 1.8	2.9 ± 0.9	7.9 ± 1.6	10.8 ± 3.9	6.7 ± 2.9
Hypertension (%)	19.0 ± 1.3	20.8 ± 3.4	0.0 ± 0.0	37.6 ± 2.8	29.2 ± 5.7	14.7 ± 4.1
Elevated fasting glucose (%)	3.1 ± 0.6	1.4 ± 1.0	1.6 ± 0.7	2.1 ± 0.8	15.4 ± 4.5	6.7 ± 2.9
Elevated HOMA-IR (%)	2.5 ± 0.5	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 0.5	29.2 ± 5.7	1.3 ± 1.3
Elevated CRP (%)	7.1 ± 0.9	0.7 ± 0.7	0.0 ± 0.0	0.0 ± 0.0	3.1 ± 2.2	80.0 ± 4.6
Anthropometrics						
Waist circumference (WC; cm)	72.1 ± 0.2	71.2 ± 0.6	69.5 ± 0.3	73.9 ± 0.5	80.2 ± 1.3	71.3 ± 0.7
High WC† (%)	48.1 ± 1.7	45.3 ± 4.2	34.2 ± 2.7	60.1 ± 2.9	81.3 ± 4.9	37.0 ± 5.7
BMI (kg/m^2)	21.0 ± 0.1	20.8 ± 0.2	20.0 ± 0.1	21.6 ± 0.2	24.2 ± 0.5	20.9 ± 0.3
Overweight‡ (%)	19.4 ± 1.3	20.9 ± 3.5	6.4 ± 1.4	26.1 ± 2.6	50.0 ± 6.3	19.2 ± 4.6
Dietary						
Energy (kcal)	2,221.8 ± 35.2	2,330.5 ± 87.4	2,154.0 ± 53.5	2,237.1 ± 66.0	2,376.1 ± 158.1	2,110.6 ± 107.2
Saturated fat (%)	7.8 ± 0.2	8.9 ± 0.5	7.7 ± 0.3	7.5 ± 0.3	8.6 ± 0.6	6.7 ± 0.4
Cigarette smoking (%)	49.3 ± 1.7	46.0 ± 4.2	54.0 ± 2.8	44.2 ± 3.0	60.9 ± 6.1	45.2 ± 5.9
Alcohol drinking (%)	85.2 ± 1.2	89.2 ± 2.6	81.5 ± 2.2	85.9 ± 2.1	85.9 ± 4.4	90.4 ± 3.5
Socioeconomic						
Number of assets	5.2 ± 0.1	5.5 ± 0.2	4.9 ± 0.1	5.2 ± 0.1	6.1 ± 0.3	5.1 ± 0.2
Hygiene score	6.1 ± 0.1	6.5 ± 0.1	5.9 ± 0.1	6.2 ± 0.1	6.4 ± 0.2	5.7 ± 0.2
Urbanicity score	41.2 ± 0.5	43.7 ± 1.1	39.9 ± 0.8	41.0 ± 0.8	42.7 ± 1.6	41.2 ± 1.6
Graduated high school (%)	60.2 ± 1.7	71.2 ± 3.9	53.8 ± 2.8	61.0 ± 2.9	68.8 ± 5.8	56.2 ± 5.8

Data are means ± SE or % ± SE. *Cutpoints are defined using Table 1. †High waist circumference defined as >71cm for men; ‡BMI ≥ $23 kg/m^2$

Table 3.3 Characteristics of young adult women in the 2005 CLHNS

	All Women (n= 698)	Healthy/High HDL-C /High LDL-C (n= 138)	Healthy/Low BP (n= 248)	High BP (n= 228)	IR/High TG (n= 46)	High CRP (n= 38)
Age, y	20.9 ± 0.0	21.0 ± 0.0	20.9 ± 0.0	20.9 ± 0.0	20.9 ± 0.1	20.9 ± 0.1
Cardiometabolic biomarkers*						
Elevated TG (%)	8.6 ± 1.0	3.8 ± 1.5	4.4 ± 1.3	7.7 ± 1.8	50.0 ± 7.3	9.8 ± 4.7
Low HDL-C (%)	67.8 ± 1.7	27.2 ± 3.6	85.7 ± 2.2	72.5 ± 2.9	79.2 ± 5.9	73.2 ± 7.0
Elevated LDL-C (%)	12.3 ± 1.2	32.3 ± 3.7	3.6 ± 1.2	4.7 ± 1.4	20.8 ± 5.9	22.0 ± 6.5
Hypertension (%)	2.0 ± 0.5	0.0 ± 0.0	0.0 ± 0.0	5.6 ± 1.5	4.2 ± 2.9	0.0 ± 0.0
Elevated fasting glucose (%)	3.0 ± 0.6	0.0 ± 0.0	1.6 ± 0.8	1.7 ± 0.9	27.1 ± 6.5	2.4 ± 2.4
Elevated HOMA-IR (%)	4.5 ± 0.8	0.0 ± 0.0	0.8 ± 0.6	0.0 ± 0.0	62.5 ± 7.1	2.4 ± 2.4
Elevated CRP (%)	7.5 ± 1.0	1.3 ± 0.9	1.6 ± 0.8	1.3 ± 0.7	14.6 ± 5.1	95.1 ± 3.4
Anthropometrics						
Waist circumference (WC; cm)	67.9 ± 0.3	66.7 ± 0.6	65.6 ± 0.3	69.0 ± 0.5	76.8 ± 1.7	70.4 ± 1.4
High WC† (%)	48.1 ± 1.9	43.5 ± 4.2	36.0 ± 3.1	56.8 ± 3.3	78.3 ± 6.1	55.3 ± 8.2
BMI (kg/m^2)	20.3 ± 0.1	19.9 ± 0.2	19.3 ± 0.1	20.7 ± 0.2	24.2 ± 0.8	21.0 ± 0.6
Overweight‡ (%)	15.2 ± 1.4	12.3 ± 2.8	6.1 ± 1.5	18.1 ± 2.6	50.0 ± 7.5	26.3 ± 7.2
Dietary						
Energy (kcal)	1,605.6 ± 33.1	1,588.7 ± 72.1	1,601.8 ± 61.1	1,629.8 ± 57.3	1,493.8 ± 88.4	1,683.7 ± 120.4
Saturated fat (%)	8.5 ± 0.2	9.5 ± 0.4	8.6 ± 0.3	8.0 ± 0.3	8.5 ± 0.7	7.5 ± 0.6
Cigarette smoking (%)	6.8 ± 1.0	6.6 ± 2.1	6.5 ± 1.6	6.2 ± 1.6	10.9 ± 4.6	7.9 ± 4.4
Alcohol drinking (%)	55.0 ± 1.9	57.4 ± 4.3	52.2 ± 3.2	56.4 ± 3.3	56.5 ± 7.4	55.3 ± 8.2
Socioeconomic						
Number of assets	5.3 ± 0.1	5.5 ± 0.2	5.2 ± 0.1	5.3 ± 0.1	5.3 ± 0.2	5.1 ± 0.3
Hygiene score	6.2 ± 0.1	6.4 ± 0.1	6.1 ± 0.1	6.0 ± 0.1	6.3 ± 0.2	6.2 ± 0.3
Urbanicity score	41.4 ± 0.5	40.7 ± 1.2	41.6 ± 0.8	40.8 ± 0.9	42.3 ± 2.0	44.9 ± 2.1
Graduated high school (%)	78.3 ± 1.6	86.2 ± 2.9	79.4 ± 2.6	72.7 ± 3.0	76.1 ± 6.4	78.9 ± 6.7

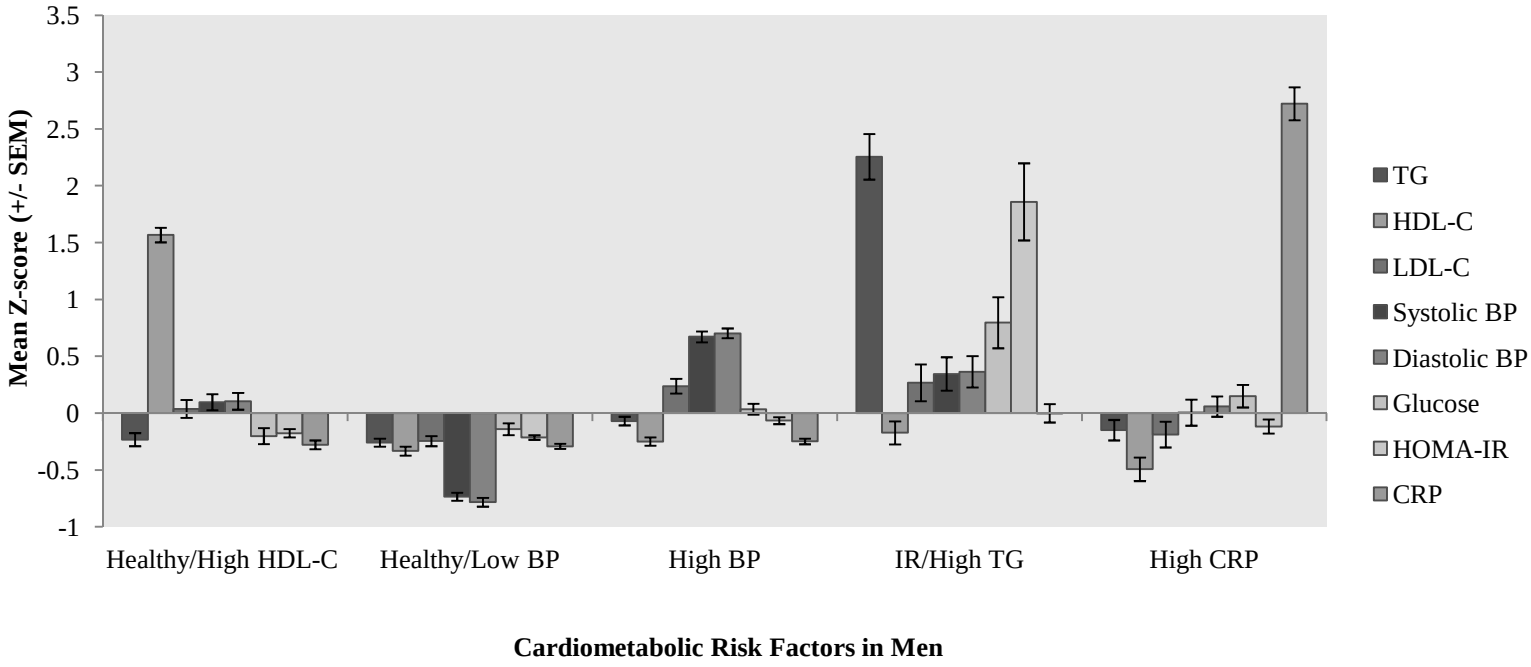
Data are means ± SE or % ± SE. *Cutpoints are defined using Table 1. †High waist circumference defined as >66.5cm for women; ‡BMI ≥ 23kg/m^2

Table 3.4 Predictors of cluster membership

	Predicted male cluster			
Referent male cluster	Healthy/Low BP	High BP	IR/High TG	High CRP
Healthy/High HDL-C	− OW* [0.32 (0.16,0.64)] − Alcohol [0.51 (0.27,0.96)]	+ High WC† [1.87 (1.15,3.04)]	+ High WC† [3.68 (1.62,8.36)] + OW* [2.17 (1.02,4.64)] + Assets [2.14 (1.06,4.32)] + Smoking [2.04 (1.06,3.90)]	− Satfat‡ [0.43 (0.22,0.86)] − Hygiene [0.74 (0.60,0.90)]
Healthy/Low BP		+ High WC† [1.92 (1.32,2.78)] + OW*[3.46 (1.95,6.16)] − Smoking [0.63 (0.44,0.89)]	+ High WC† [3.78 (1.77,8.06)] + OW* [6.80 (3.21,14.42)] + Assets [2.72 (1.42,5.24)]	+ OW* [5.12 (2.13,12.33)] − Satfat‡ [0.51 (0.27,0.98)] + Assets [1.94 (1.10,3.42)] − Smoking [0.56 (0.33,0.97)] + Alcohol [2.83 (1.19,6.72)] − Hygiene [0.83 (0.70,0.99)]
High BP			+ OW* [1.96 (1.02,3.77)] + Assets [2.42 (1.27,4.60)] + Smoking [2.28 (1.25,4.14)]	− High WC† [0.34 (0.18,0.64)] − Hygiene [0.82 (0.69,0.98)]
IR/High TG				− High WC† [0.17 (0.07,0.43)] − Smoking [0.39 (0.19,0.82)]
	Predicted female cluster			
Referent female cluster	Healthy/Low BP	High BP	IR/High TG	High CRP
Healthy/High HDL-C/High LDL-C		− Satfat‡ [0.46 (0.28,0.78)] + Energy§ [1.40 (1.00,1.96)] − HS Grad [0.51 (0.29,0.92)]	+ OW* [4.57 (1.90,10.95)]	− Satfat‡ [0.22 (0.08,0.61)] + Energy§ [1.73 (1.02,2.91)] + Urban [2.88 (1.30,6.39)]
Healthy/Low BP		+ High WC† [1.86 (1.24,2.77)] + OW* [2.24 (1.17,4.29)]	+ High WC† [2.94 (1.24,6.95)] + OW* [8.26 (3.50,19.50)]	+ OW* [4.12 (1.49,11.40)] − Satfat‡ [0.35 (0.13,0.92)] + Urban [2.81 (1.31,6.04)]
High BP			+ OW* [3.69 (1.72,7.92)]	+ Urban [2.82 (1.32,6.04)]
IR/High TG				

Cells display +/- association of predictors with cluster membership. Data are OR (95% CI). *Overweight; †Waist Circumference; ‡Percentage of total energy intake from saturated fat; this covariate was scaled (divided by 10) when imputed in the multinomial logistic regression to ease interpretation; §Energy intake was also scaled; units were *kJ/1000*.

Figure 3.1 Mean Z-scores of fasting biomarkers by cardiometabolic cluster



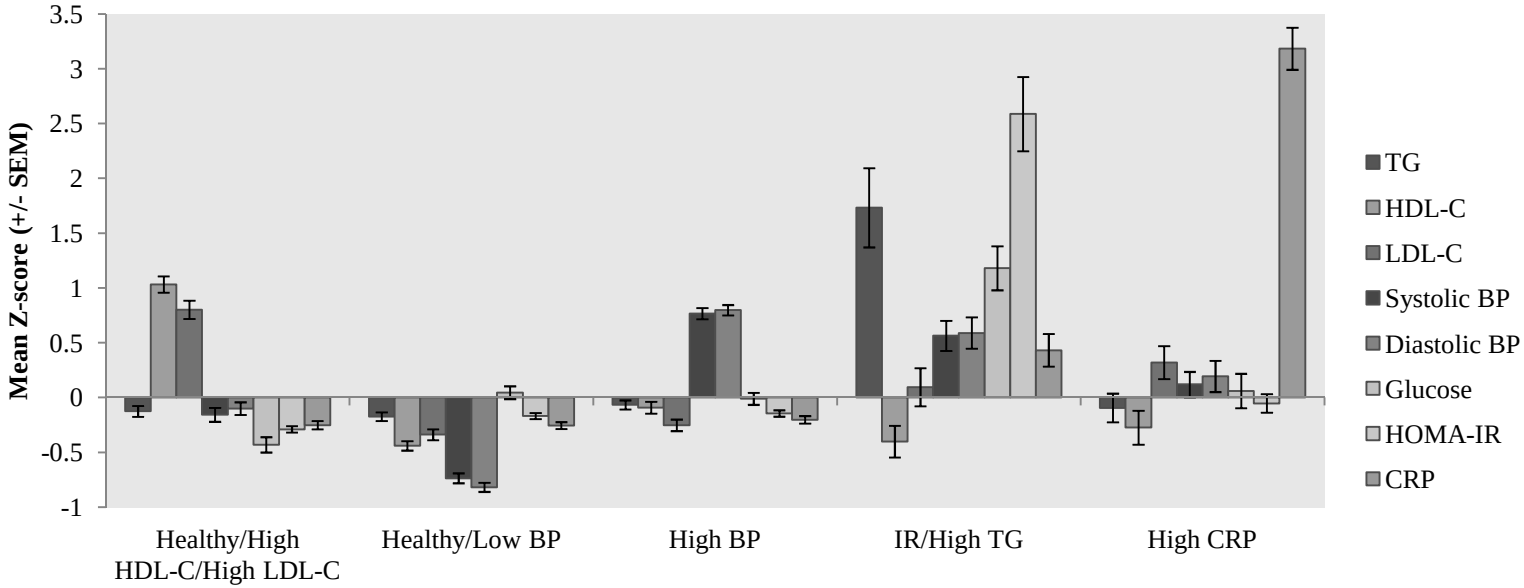


Figure 3.1: Mean Z-scores of fasting biomarkers by cardiometabolic cluster

Mean Z-scores by cardiometabolic cluster for the eight fasting biomarkers used as input variables in the cluster analysis. A: Mean Z-scores in young adult men. B: Mean Z-scores in young adult women.

Chapter 4. GENETIC RISK SCORE AND ADIPOSITY INTERACT TO INFLUENCE TRIGLYCERIDE LEVELS IN A COHORT OF FILIPINO WOMEN

OVERVIEW

Individually, genetic variants only moderately influence cardiometabolic (CM) traits, such as lipid and inflammatory markers. In this study we used genetic risk scores to combine the relatively small additive effects of individual variants to better capture the complex relationship between genetics, adiposity, and CM traits.

Participants included 1,649 women from the 2005 Cebu Longitudinal Health and Nutrition Survey. Three genetic risk scores were constructed for, C-reactive protein (CRP), high-density lipoprotein (HDL-C), and triglycerides (TG). We used linear regression models to assess the association between each genetic risk score and its related trait. We also tested for interactions between each score and measures of adiposity.

Each genetic risk score explained a greater proportion of variance in trait levels than any individual genetic variant. We found an interaction between the TG genetic risk score and waist circumference (WC) ($P_{interaction} = 1.66 \times 10^{-2}$). Based on model predictions, for individuals with a higher TG genetic risk score (75^{th} percentile = 12), having an elevated WC ($\geq$ 80cm) increased TG levels from 117 to 151 mg/dL. While for individuals with a lower score (25^{th} percentile = 7), having an elevated WC made no significant impact on TG levels (93 vs. 104 mg/dL).

In summary, combinations of genetic loci better explained the variation in CM traits and the TG genetic risk score interacted with adiposity to influence TG levels. Larger studies

are needed to support the potential clinical and public health utility of targeted prevention efforts using genetic profiling.

BACKGROUND

Recent studies in both European and Asian cohorts have found multiple genetic variants relating to cardiometabolic (CM) traits such as lipid and inflammatory levels.[5, 7, 9, 40, 122] Individually, the identified genetic variants only moderately influence these trait levels and are thought to provide only limited information in clinically assessing a person's risk. However, the combination of genetic variants, each with a relatively small effect, may better explain the variability of these complex traits.[79] Thus, the use of a genetic risk score has been proposed to better capture genetic variation.[80-83]

Genetic variants may interact with diet, environmental, and anthropometric factors to influence CM phenotypes; accounting for these synergistic effects may also help explain some of variability of these traits. [79 79] Excess adiposity is one of the strongest predictors of CM disease and risk.[54-57] Previous work suggests that measures of adiposity interact with specific genetic variants and predict CM traits.[124-126]

However, these synergistic effects are not well understood, especially in populations undergoing rapid nutritional and lifestyle changes. These lifestyle changes contribute to a growing burden of overweight, visceral adiposity, and thus associated CM diseases.[58-61] These concerns are especially pertinent for Asians; compared with Caucasians, Asians have increased visceral adiposity and greater insulin resistance at similar levels of BMI.[62-65] In addition, the World Health Organization concluded that the risk of CM based diseases is elevated for Asians with a BMI greater than 23 kg/m^2, suggesting the use of a lower cutpoint for overweight (OW).[105]

In this study we sought to: create genetic risk scores relating to inflammatory and lipid traits, examine the ability of these scores to explain the variation in these traits, and test whether these genetic risk scores interact with measures of adiposity to influence trait levels. We chose to specifically look at C-reactive protein (CRP), high-density lipoprotein (HDL-C), and triglycerides (TG) because previous research suggests that these traits interact with various measures of adiposity.[127-129] To accomplish this, we used an at-risk Asian population, from the 2005 Cebu Longitudinal and Health and Nutrition Survey (CLHNS), undergoing a nutrition and lifestyle transition. The identification of individuals with a genetic predisposition to elevated CM risk could help in creating targeted and thus more efficient prevention strategies.

METHODS

Survey design

The women in this study are participants in the CLHNS, which is described in detail elsewhere.[89] Briefly, the CLHNS is a community-based cohort of women and their index children followed since 1983. The original participants included all pregnant women in 33 randomly selected communities of Metro Cebu, who gave birth between May 1, 1983, and April 30, 1984. A baseline interview was conducted among 3,327 women in their 6th to 7th month of pregnancy. Subsequent surveys took place immediately after birth, bimonthly for 2 years, in 1991, 1994-5, 1998-99, 2002, and 2005. In 2005, fasting blood was drawn for CVD biomarkers and genetics. Here we use data from the mother cohort participating in the 2005 CLHNS. All data were collected under conditions of informed consent with institutional review board approval from the University of North Carolina, Chapel Hill, USA.

We excluded women who were pregnant at the time of blood draw, not fasting at the time of blood draw, and with CRP levels >10 mg/L (a level representing current/recent illness rather than low-level basal inflammation).[88] Even though we removed individuals with CRP levels >10 mg/L, we adjusted for the presence of any infectious symptoms at the time of blood collection to help control for any residual confounding in our analysis.[1] 1,649 women had complete biomarker, genetic, diet, socioeconomic, and anthropometric data. Medication use in this population was low: 0.1% took statins, 1.75% took diabetes medication, and 4% took anti-hypertensive medications. A sensitivity analysis showed that exclusion of these individuals did not impact results; therefore we did not exclude anyone taking medication. All data were collected with informed consent, using protocols approved by the institutional review board of the University of North Carolina, Chapel Hill.

Cardiometabolic biomarkers

Fasting plasma CM biomarkers used in the analyses included TG, HDL-C, and C- CRP. Blood samples were collected in participants' homes in the morning after an overnight fast. Venous blood was collected in EDTA anti-coagulant vacutainer tubes. Blood samples were stored on ice for no more than 2 hours and were then centrifuged to separate plasma. After separation, samples were frozen and remained frozen at -80 °C until ready for analysis. Total lipid concentrations were measured at the Emory Lipid Research Laboratory using enzymatic methods with reagents from Beckman Diagnostics on the Beckman Diagnostics CX5 chemistry analyzer (Fullerton, CA). HDL-C was determined using the homogeneous assay for direct determination (Genzyme Corporation, Exton, PA). The Emory Lipid Research Laboratory is a participant in the CDC/NHLBI Lipid Standardization Program to ensure accuracy and precision of the determinations. CRP concentrations were determined using a

high sensitivity immunoturbidimetric method (Synchron LX20, lower detection limit: 0.1 mg/L).

The cutpoints used to define elevated risk for each trait were: >3.0 mg/ L for CRP, < 50 mg/dL for HDL-C, and > 150 mg/dL for TG levels. These were based on recommendations from the International Diabetes Federation (IDF) and the American Heart Association.[88, 96, 102, 103]

Anthropometry

Body weight, height, and waist circumference (WC) were measured using standard techniques.[104] BMI was calculated as the ratio of weight (kg) to height (m^2). We used cutpoints for Asians to define normal weight as BMI < $23kg/m^2$, overweight (OW) as $23kg/m^2 \leq$ BMI < $27.5kg/m^2$, obese as BMI $\geq 27.5kg/m^2$, and central adiposity as WC $\geq$ 80 cm.[105 96]

Dietary data

Dietary data were derived from two 24-hour dietary recalls and the mean intake was used in the analyses. Data were collected during in-home interviews performed by highly trained local field staff. A nutritionist reviewed all dietary recalls immediately after collection. When implausible values were found, interviewers revisited respondents for verification. Energy and saturated fat intakes were calculated using the Philippines Food Composition Tables.[19, 106]

Sociodemographic and lifestyle characteristics

Highly trained interviewers collected reproductive history data; this included menopausal status beginning in the 1991 survey.

Socioeconomic status (SES) was measured by a factor score based on a principal

components analysis of household ownership of key assets such as television, vehicles, and furniture.[69]

Infectious illness was measured by asking participants if they were currently experiencing any symptoms of infection, consistent with prior research on CRP.[52] Symptoms included runny nose, cough, fever, diarrhea, and sore throat, as well as the more general categories of flu, cold, and sinusitis. Responses were used to construct a summary variable indicating the presence or absence of any infectious symptoms at the time of blood collection.

Environmental cleanliness and household hygiene was measured by a hygiene score based on data on the interviewer's rating of cooking area, immediate area around the house, toilet type, and water source. The score ranges from 0 to 9 with larger values indicating greater cleanliness.[52]

Genotyping, quality controls, and imputation

The complete methods for direct SNP genotyping, quality control, and SNP imputation have been described previously.[130] Briefly, genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 5.0. Quality control procedures excluded: samples with <97% genotyping call rate; members of estimated first-degree relative pairs; SNPs with a call rate < 90%; SNPs with a deviation from Hardy-Weinberg equilibrium ($P < 10^{-6}$); SNPs with ≥ 3 discrepancies among duplicate pairs; SNPs with Mendelian inheritance errors among five CEPH trios and/or CEPH sample genotype discrepancies with HapMap. Genotype imputation was conducted with MACH using phased haplotypes from the 1000 Genomes Project in both CEU and CHB+JPT samples (June 2010 Release).[131] In addition, we excluded any SNPs with poor imputation quality (MACH $r^2 < 0.3$) or estimated minor allele frequency

(MAF) ≤ 0.01.

Genetic marker selection

The process of choosing SNPs is depicted in Figure 1. The SNPs used to create the genetic risk scores were selected by finding SNPs associated with the individual CM traits of interest: CRP, HDL-C, and TG. We selected these SNPs from (1) genome-wide association studies (GWAS) conducted with our own study population,[1, 7] (2) published GWAS of East and South East Asian cohorts,[2, 9] and (3) published GWAS of European descent cohorts,[5] if the specific trait lacked studies conducted in populations of Asian descent. We limited our selection of studies to cohort-based studies and meta-analyses; case-control studies were not considered because we wanted to choose SNPs associated with the individual CM trait rather than disease state. From the studies identified, we selected SNPs with a $P < 5\times10^{-8}$ in the original study population for further analysis; we increased this threshold to a $P < 5\times10^{-5}$ for those studies conducted in our own study population due to the smaller sample size in the CLHNS.

The original studies were used to identify the risk allele. We designated the risk allele as the allele associated with an increased level of the specified trait, except in the case of HDL-C, for which the allele associated with lower levels was designated. For each individual, we coded imputed SNPs according to the dosage value or the expected number of copies of the risk allele (a continuous number between 0 and 2). This coding reflects the uncertainty in the imputation of the SNPs, for example a 1.5 suggests more uncertainty whereas a 1.9 suggests less uncertainty (an individual likely has two risk alleles).

Three genetic risk scores were constructed, one for each CM trait, CRP, HDL-C, and TG. Before creating each genetic risk score, we chose a subset of SNPs with nominal

significance ($P < 0.1$) and directional consistency of the effect estimate in our study population. This was based on adjusted linear regression models of the natural log-transformed CM trait on the individual SNP (see Model 2 in Methods). Then for each CM trait, we pruned SNPs for redundancy due to linkage disequilibrium ($r^2 > 0.2$). To do this we used the --clump procedure in PLINK to create "clumps" of correlated SNPs.[132] Each clump was represented by the top index SNP, designated as the SNP with the lowest *P* value (see Model 2 in Methods). Using these index SNPs, we calculated each genetic risk score by simply summing the risk alleles associated with the specific trait. We created an un-weighted score instead of weighting by the effect of each SNP because: (1) the current literature does not provide stable effect estimates of each SNP for each trait; (2) the outcomes (and thus effects) across studies were non-comparable (e.g. log-transformed trait vs. non-transformed trait); (3) studies used populations of various sample sizes and ethnicities; (4) using weights from the CLHNS data itself would have introduced bias.

Statistical Analysis

Linear regression models, with each of the three CM traits as a continuous outcome, were used. All traits were natural log-transformed to satisfy model assumptions of normally distributed residuals. Given the markedly skewed distribution of CRP concentrations and the presence of many values below the detectable level (0.1 mg/L), CRP values were natural log-transformed after adding the constant 0.10.

We constructed principal components (PCs) using the software EIGENSOFT to capture population substructure among CLHNS subjects.[133] We assessed the association between each of the first 10 PCs and each log-transformed CM trait to identify any potential ancestry explanatory PC; the 7th PC was significantly associated with CRP and HDL-C (no

PCs were significantly associated with log TG levels), thus the first 7 PCs were included as covariates in the linear regression models.

Two different models were examined. Model 1 was a linear regression model adjusted for age (categorical: ≤44 y, 45-49 y, 50-54 y, and ≥55 y) and population substructure. Model 2 included covariates adjusted for in Model 1 plus additional adjustment for postmenopausal status (yes/no), OW/obese status (BMI ≥ 23 kg/m^2), high WC (WC ≥ 80cm), % energy intake from saturated fat, energy intake, environmental hygiene, reported infectious illness (yes/no), and SES. The covariates chosen for adjustment in Model 2 were based on prior published studies in the CLHNS on these lipid and inflammatory traits.[52, 69, 115] We categorized age, BMI, and WC to account for their non-linear relationship with the log CM trait levels.

Models 1 and 2 were applied to test for the association between each candidate SNP and its related log-transformed CM trait (assuming an additive model). Then both Models were applied to test for the association between each genetic risk score (continuous) and its related log-transformed CM trait. In addition, Model 2 without a genetic component was estimated to examine the "non-genetic" factors associated with each log-transformed CM trait.

Lastly, for each CM trait we looked at interactions between the genetic risk score and measures of adiposity. We examined a genetic risk score × elevated WC interaction, both unadjusted and adjusted for BMI (using a 3-categorical dummy variable for normal weight, OW, and obese). Then we examined a genetic risk score × OW/obese status interaction, both unadjusted and adjusted for elevated WC. Each interaction was looked at separately while adjusting for the same covariates as Model 2.

For regression analyses we used a statistical significance criteria of P <0.05 (2-sided). For interaction terms, we considered P <0.1 as nominally significant. All regression analyses were performed with Stata 12.0 (Stata Corporation, College Station, TX, 2006).

RESULTS

The characteristics of 1,649 women participants in the 2005 CLHNS are presented in Table 1. In 2005, participants had a mean (SD) age of 48.41 (6.03) years. About 39% of women were postmenopausal, 52% had elevated WC, 60% were OW, 20% had elevated CRP, 82% had low HDL-C, and 29% had elevated TG.

Our selection strategy for candidate SNPs relating to CRP, HDL-C, and TG resulted in 46, 19, and 13 usable variants (Figure 1). After pruning to eliminate correlated SNPs in linkage disequilibrium (by trait), 6 CRP, 9 HDL-C, and 9 TG SNPs were used in the construction of the genetic risk scores (Figure 1; Table 2). Among participants, each genetic risk score was normally distributed (Figure 2). The mean score (SD) and range of number of risk alleles for CRP was 3.32 (1.37) with a range from 0.12-8.51; for HDL-C was 5.95 (1.58) with a range from 1.61-11.66; and for TG 9.42 (1.85) with a range from 2.29-14.34.

The regression results from Model 2 for each candidate SNP and its respective CM trait are shown in Table 2 (results from Model 1 were similar and thus not shown). Using Model 2, the individual SNP most strongly associated with CRP was rs876537 at the *CRP* loci (β= 0.33, 95% CI [0.24, 0.42], $P = 2.27 \times 10^{-12}$), with HDL-C was rs12708980 at the *CETP* loci (β= -0.05, 95% CI [-0.08, -0.03], $P = 6.61 \times 10^{-7}$), and with TG was rs964184 at the *APOC3* loci (β= 0.15, 95% CI [0.11, 0.19], $P = 3.37 \times 10^{-15}$). The same SNPs were found to be the most strongly associated with each trait in Model 1 as well.

As expected, each of the three genetic risk scores was associated with its respective log-transformed trait (Table 3). Specifically in Model 2, each additional CRP risk allele resulted in an estimated 18% increase in CRP levels (β= 0.18, 95% CI [0.14, 0.23]); each additional HDL-C risk allele resulted in an estimated 4% decrease in HDL-C levels (β= -0.04, 95% CI [-0.05, -0.04]); each additional TG risk allele resulted in an estimated 7% increase in TG levels (β= 0.07, 95% CI [0.06, 0.08]).

We compared the proportion of variance explained in the log-transformed CM trait by the genetic risk score vs. the most strongly associated individual SNP in Model 2 (Figure 3). To do this we first ran Model 2 without any genetic component (Table 4). The adjusted R-square obtained from this model represents the proportion of variance explained by specified "environmental" components (Rsq E). Running Model 2 with the individual SNP yielded an adjusted R-square representing the proportion of variance explained by the individual SNP and environmental components (Rsq SNP+E). Running Model 2 with the genetic risk score yielded an adjusted R-square representing the proportion of variance explained by the genetic risk score and environmental components (Rsq GRS+E). To obtain the proportion of variance explained by just the individual SNP (Rsq SNP) = (Rsq SNP+E)-(Rsq E). To obtain the proportion of variance explained by just the genetic risk score (Rsq GRS) = (Rsq GRS+E)-(Rsq E). For each CM trait, we plotted Rsq SNP vs. Rsq GRS (Figure 3). For all three traits, Rsq GRS > Rsq SNP; about 4% of log CRP levels, 7% of log HDL-C levels, and 6% of log TG levels were explained by the genetic risk score alone (Table 3).

We found significant interactions between measures of adiposity and the TG genetic risk score on log TG levels, while we found no evidence of such interactions on log CRP or log HDL-C levels.

Stratifying by normal WC (< 80cm) and elevated WC (≥ 80cm), the estimated % increase in TG levels for each additional TG risk allele was 5% (β= 0.05, 95% CI [0.03, 0.07]) in normal WC individuals, but increased to 8% (β= 0.08, 95% CI [0.06, 0.10]) in elevated WC individuals ($P_{interaction} = 1.66 \times 10^{-2}$) (Table 5). Here we present BMI adjusted results; we found no difference between the effect estimate and *P* value of the unadjusted model (results not shown).

Similarly, stratifying by normal weight (< 23kg/m^2) and OW/obese (≥23kg/m^2), the estimated % increase in TG levels for each additional TG risk allele was 5% (β= 0.05, 95% CI [0.03, 0.07],) in normal weight individuals, but increased to 8% (β= 0.08, 95% CI [0.06, 0.09]) in OW/obese individuals ($P_{interaction} = 2.73 \times 10^{-2}$) (Table 5). Here we present WC adjusted results; we found no difference between the effect estimate and *P* value of the unadjusted model (results not shown).

To better visualize these interactions (Figure 4), we predicted TG levels at the 25th and 75th percentile values of the genetic risk score (7 and 12 respectively) at varying levels of adiposity, while holding all other covariates in Model 2 at the mean. Based on model predictions, for individuals with a higher TG genetic risk score (= 12), having an elevated WC (≥ 80cm) increased TG levels from 117 to 151 mg/dL. While for individuals with a lower score (= 7), having an elevated WC made no significant impact on TG levels (93 vs. 104 mg/dL). Similar results were seen at varying levels of BMI.

We also examined all the above interactions with each individual SNP included in the TG genetic risk score, however none of these interactions were significant (results not shown).

DISCUSSION

In this study we used a genetic risk score to combine the relatively small additive effects of individual SNPs to better capture the complex relationship between genetics, adiposity, and CM risk. We found that for all three traits, the genetic risk score more strongly predicted biomarker levels than any individual SNP. In addition, the genetic risk score explained a greater proportion of variance in the specified trait than any given individual SNP. Lastly, we found that for individuals with a higher TG genetic risk score, having either an elevated WC or being OW/obese amplified the genetic risk score's effect by further increasing TG levels. While for individuals with a lower TG genetic risk score, measures of adiposity made almost no difference in TG levels. Interestingly for those women with a low TG genetic risk score and elevated levels adiposity, their predicted levels of TG equaled those of women with a high genetic risk score without any adverse levels of adiposity. Overall, these results demonstrate that combinations of multiple genetic loci better explain the variation in CRP, HDL-C, and TG levels and that the TG genetic risk score seemed interact with measures of adiposity to influence TG levels in this study population.

In support of our results, recent work using the same study population found that central obesity might accentuate the effect of the TG-increasing allele of an *APOA5* variant.[7] In addition, previous research has implicated several variants in the *LPL* gene (a gene included in our genetic risk score) as having an interactive effect with central adiposity on TG levels and the ratio of TG to HDL-C.[127 128] However, we did not find significant interactions between these individual loci and adiposity on TG levels, perhaps indicating that the interactive effect is driven by a collective result of all SNPs in the TG genetic risk score.

While the interaction between WC and the TG genetic risk score (adjusting for BMI) was stronger than that between OW/obese and the TG genetic risk score, we cannot conclusively say whether visceral adiposity, as proxied by WC, drives this interaction. However, it is interesting to note that the residuals of WC regressed on BMI also significantly interacted with the TG genetic risk score (results not shown). Previous studied have implicated visceral adiposity as a stronger predictor of TG levels and hypertriglyceridemia compared to subcutaneous adipose tissue.[134, 135]

From a clinical perspective, individuals with both a high TG genetic risk score and elevated WC, had predicted TG levels that meet the American Heart Association's level of "borderline high risk" (150 to 199 mg/dL).[102] This combination of elevated WC along with increased TG levels has been previously described as the "hypertriglyceridemic waist" phenotype. Individuals with this phenotype have a higher risk of increased visceral adiposity, CVD, insulin resistance, and other related outcomes.[56, 136-138] This is of particular concern for Asian populations, for whom increased levels of visceral adiposity are present at normal BMIs.[62-65] Work from Pollin *et al.* reinforces this concept by finding that an intensive lifestyle intervention appeared to partially mitigate the effect of the rs1260326 risk allele in the *GCKR* gene (a loci included in our genetic risk score) on higher TG levels.[139] Further research, especially clinical trials in larger populations, is needed to know whether such interventions could be useful, especially across different ethnicities.

Limitations of our study merit consideration. We used cross-sectional data since biomarker levels were only measured in 2005, thus no causal relationships can be inferred. In our literature search we found differing numbers of candidate SNPs for each trait. Although we used the same criteria in our search regardless of the CM trait, the variation in the number

of candidate SNPs could reflect the current state of the literature. In addition, there is concern with choosing SNPs from a European sample and applying them to an Asian sample, especially in terms of tagging the appropriate functional variant. We tried to mitigate this by choosing SNPs with nominal significance and directional consistent effect estimates in our study population, however due to the limited sample size in the CLHNS we may have lacked the power to detect the SNPs true effect. Also, using a threshold of r^2 <0.2 for linkage disequilibrium, still allows SNPs to partially tag the same underlying signal, potentially including some redundancy in the genetic risk score. While we used an un-weighted approach to create our genetic risk scores, it may be possible in the future to obtain stable and accurate estimates of genetic variants for use in a weighted risk score, which could improve predictability of CM risk.

In conclusion, using a study population of middle-aged Filipino women undergoing a nutrition and lifestyle transition, we found that CRP, HDL-C, and TG genetic risk scores explained a greater variance of the associated trait compared to a single SNP. We also found that the TG genetic risk score interacted with adiposity to synergistically influence TG levels. For individuals with a high genetic predisposition to elevated TG levels, our results suggest reducing adiposity could possibly prevent increases in TG levels and thereby reduce the likelihood of adverse health outcomes such CM diseases. Replication of these results in larger study populations is needed to support the potential clinical and public health utility of targeted prevention efforts using genetic profiling.

TABLES AND FIGURES

Figure 4.1: Choosing of SNPs to include in the genetic risk scores

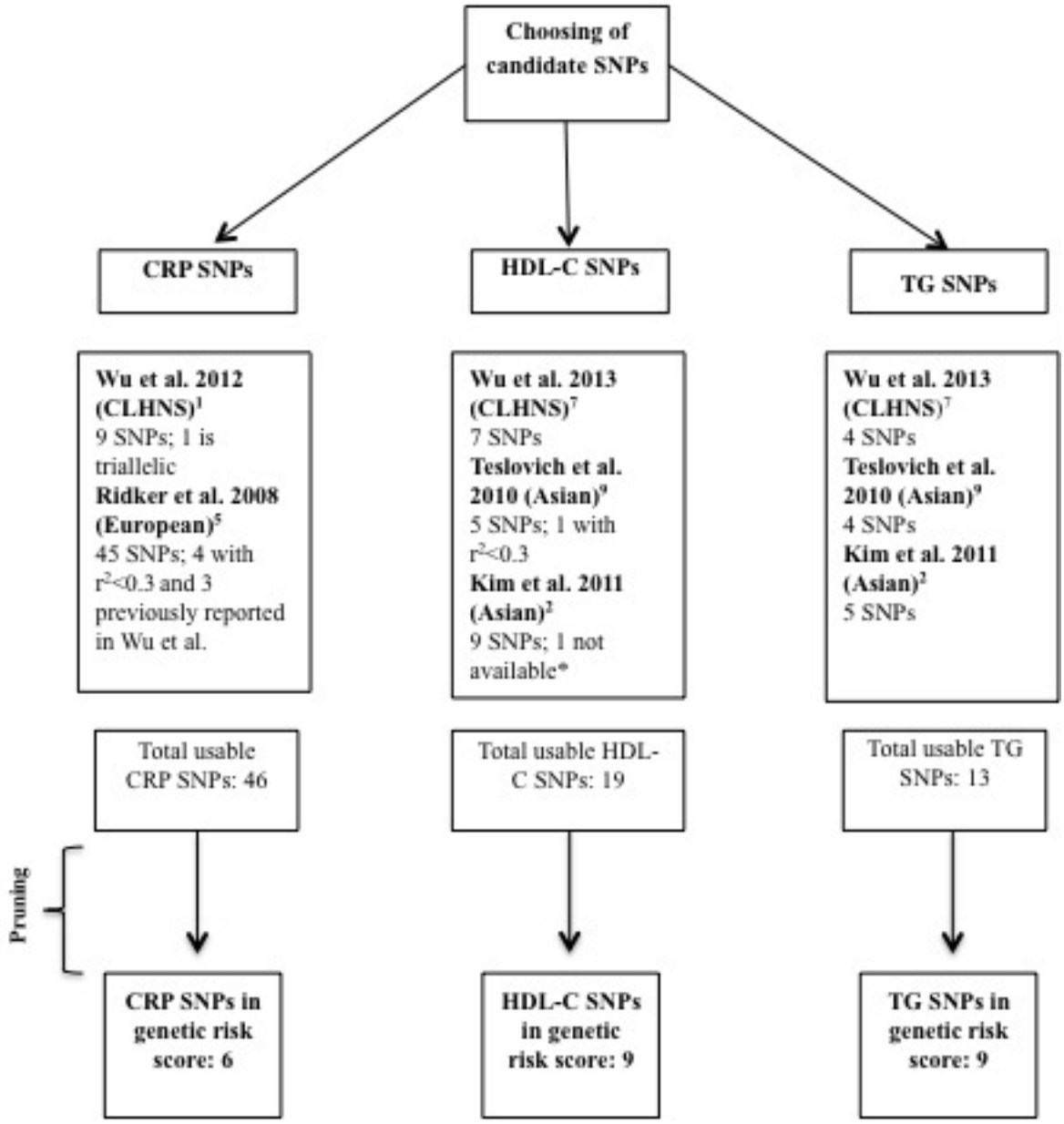


Figure 4.1: Choosing of SNPs to include in the genetic risk scores

A schematic representation describing the process of choosing SNPs associated with CRP, HDL-C, and TG. Parentheses indicate the specific study population in which analyses were conducted. *rs1268004 was not genotyped and no HapMap or 1000 Genomes imputed data was available. See Methods for further details.

Table 4.1: Characteristics of 1,649 women participants in the 2005 CLHNS

Age (%)	
≤44 y	32.4
45-49 y	31.5
50-54 y	20.9
≥55 y	15.1
Postmenopausal (%)	38.5
Illness* (%)	27.5
Energy intake (kcal)	1,128.8 ± 491.8
% Energy intake from saturated fat	5.4 ± 4.1
Waist circumference (WC; cm)	81.0 ± 10.9
Elevated WC† (%)	52.4
BMI (kg/m^2)	24.3 ± 4.3
Overweight† (%)	38.7
Obese† (%)	21.2
CRP (mg/L)	1.7 ± 2.1
Elevated CRP† (%)	19.6
HDL-C (mg/dL)	41.0 ± 10.3
Low HDL-C† (%)	82.4
TG (mg/dL)	131.0 ± 84.7
Elevated TG† (%)	28.8

Data are means ± SD or percentages. *Percentage of individuals reporting illness at time of blood draw †See Methods for cutpoint values

Table 4.2: SNPs selected from a literature search to generate cardiometabolic trait specific genetic risk scores

SNP	Chr	Nearest gene	Non-risk allele†	Risk allele†	Risk allele frequency	β	95% CI		*P* value	R square‡
CRP SNPs										
rs12093699	1	CRP	G	A	0.08	0.25	0.07	0.42	6.0E-3	0.19
rs876537	1	CRP	T	C	0.43	0.33	0.24	0.42	2.3E-12	0.21
rs1892534	1	LEPR	T	C	0.15	0.11	0.00	0.23	5.3E-2	0.18
rs1408282	6	6q16.1	G	A	0.09	0.42	0.24	0.59	2.3E-6	0.19
rs1169288	12	HNF1A	C	A	0.63	0.33	0.23	0.43	1.9E-10	0.2
rs1169302	12	HNF1A	G	T	0.29	0.09	-0.01	0.19	8.2E-2	0.18
HDL-C SNPs										
rs1544857	2	SLC4A10	G	C	0.17	-0.05	-0.08	-0.03	6.6E-6	0.08
rs17548357	2	BIRC6	G	A	0.02	-0.19	-0.28	-0.11	1.4E-5	0.08
rs3739440	9	PAX5	C	T	0.17	-0.07	-0.10	-0.04	2.3E-6	0.08
rs11227643	11	11q13.1	C	G	0.73	-0.05	-0.08	-0.02	5.8E-4	0.07
rs964184	11	APOC3	C	G	0.24	-0.02	-0.04	0.00	1.7E-2	0.07
rs1532085	15	LIPC	A	G	0.43	-0.04	-0.07	-0.02	4.3E-5	0.08
rs2070895	15	LIPC	A	G	0.62	-0.06	-0.09	-0.03	3.5E-5	0.08
rs12708980	16	CETP	T	G	0.19	-0.05	-0.08	-0.03	6.6E-7	0.08
rs138779	22	TOM1	T	C	0.39	-0.05	-0.06	-0.03	2.4E-6	0.08
TG SNPs										
rs780092	2	GCKR	G	A	0.68	0.09	0.05	0.12	3.2E-7	0.13
rs17023681	3	CNTN4	T	G	0.29	0.12	0.07	0.17	2.6E-6	0.13
rs7644509	3	Chr3q26.1	C	G	0.19	0.08	0.04	0.13	3.5E-4	0.13
rs2286276	7	TBL2-MLXIPL	T	C	0.9	0.05	-0.01	0.10	8.6E-2	0.12
rs12678919	8	LPL	G	A	0.95	0.09	0.02	0.17	1.8E-2	0.12
rs2001945	8	LPL	C	G	0.43	0.03	0.00	0.06	4.6E-2	0.12
rs603446	11	ZNF259	T	C	0.68	0.08	0.04	0.11	1.9E-5	0.13
rs964184	11	APOC3	C	G	0.24	0.15	0.11	0.19	3.4E-15	0.15
rs1893838	18	ZBTB7C	T	C	0.35	0.07	0.03	0.10	1.6E-4	0.13

*Association results from Model 2: the covariates were age (categorical: ≤44 y, 45-49 y, 50-54 y, and ≥55 y), PCs, postmenopausal status, overweight/obese status, elevated waist circumference, % energy intake from saturated fat, energy intake, environmental hygiene, illness, and SES (outcome was the log-transformed CM trait); †Risk allele as defined by the study from which it was chosen; ‡Adjusted R-square for Model 2

Figure 4.2: Distribution of the genetic risk score among 1,649 Filipino women

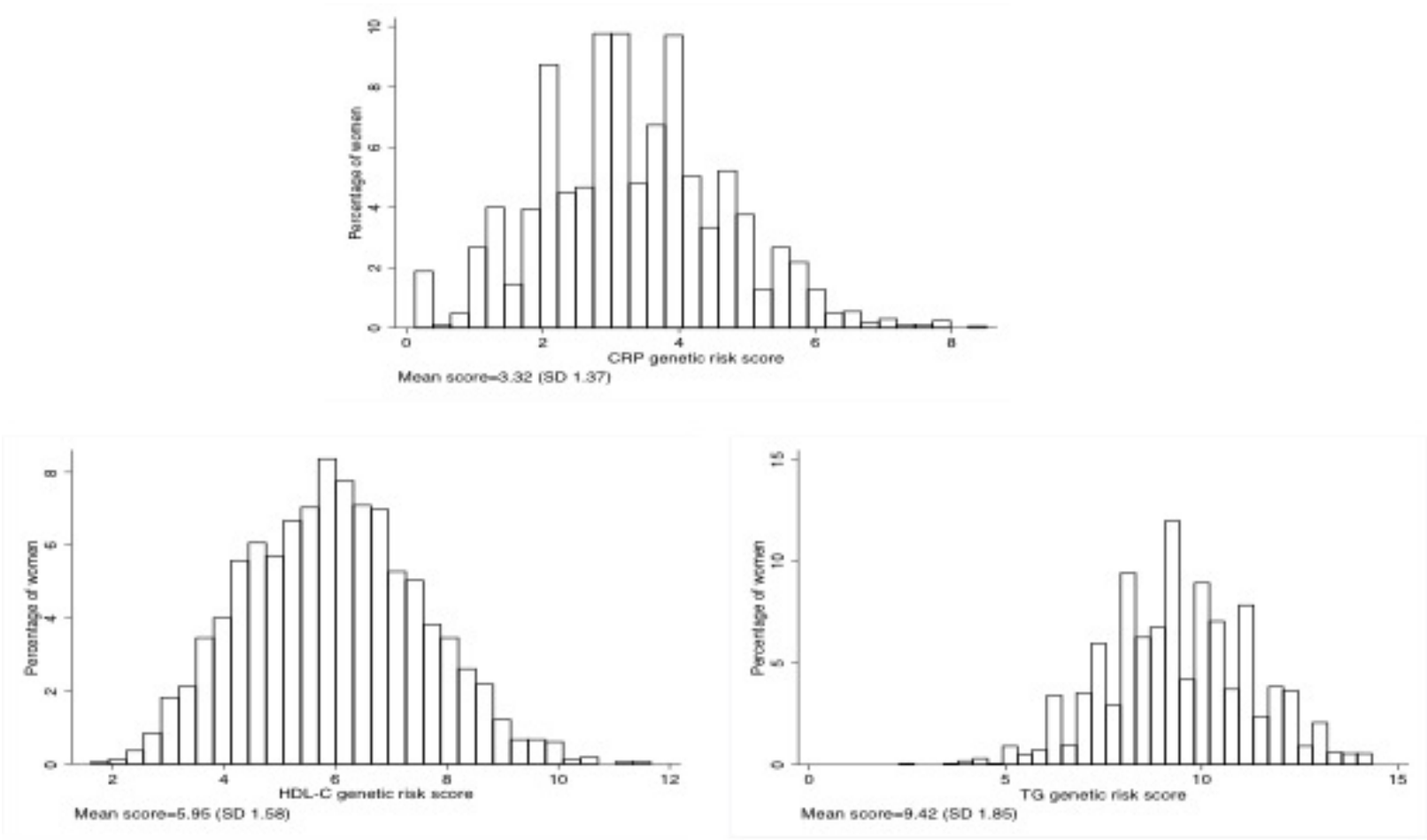


Figure 4.2: Distribution of the genetic risk score among 1,649 Filipino women

The distribution of the CRP, HDL-C, and TG genetic risk scores among participants

Table 4.3: Association of genetic risk scores with log-transformed CM trait levels in 1,649 women

Genetic risk score	β	95% CI		*P* value	R-square †	R square (genetic risk score alone)‡
CRP genetic risk score	0.19	0.15	0.23	4.81E-20	0.22	0.04
HDL-C genetic risk score	-0.04	-0.05	-0.04	1.81E-29	0.14	0.07
TG genetic risk score	0.07	0.06	0.08	3.38E-28	0.18	0.06

*Association results from Model 2; covariates were age (categorical: ≤44 y, 45-49 y, 50-54 y, and ≥55 y), PCs, postmenopausal status, overweight/obese status, elevated waist circumference, % energy intake from saturated fat, energy intake, environmental hygiene, illness, and SES (outcome was the log-transformed CM trait); †Adjusted R-square for Model 2; ‡Proportion of variance explained by the genetic risk score alone (see Methods)

Table 4.4: Association of "non-genetic" covariates with CM trait levels

Model 2 without genetic component	log CRP					log HDL-C					log TG				
Covariates	β	95% CI		*P* value	R-square*	β	95% CI		*P* value	R square*	β	95% CI		*P* value	R square*
					0.18					0.07					0.12
Age															
≤44 y									Reference						
45-49 y	0.03	-0.12	0.18	7.06E-01		0.00	-0.03	0.03	9.51E-01		0.08	0.02	0.14	8.17E-03	
50-54 y	0.02	-0.17	0.22	8.13E-01		-0.02	-0.06	0.03	4.56E-01		0.17	0.09	0.25	1.93E-05	
≥55 y	0.23	0.00	0.47	5.28E-02		0.01	-0.03	0.06	5.52E-01		0.15	0.06	0.24	1.70E-03	
Postmenopausal status	0.23	0.07	0.40	6.55E-03		-0.02	-0.06	0.01	1.94E-01		0.09	0.02	0.16	7.00E-03	
Elevated WC†	0.56	0.39	0.72	2.99E-11		-0.06	-0.10	-0.03	3.58E-04		0.18	0.12	0.25	2.28E-08	
OW/obese	0.43	0.26	0.60	9.30E-07		0.00	-0.04	0.03	8.39E-01		0.11	0.04	0.18	1.48E-03	
Energy intake	0.00	0.00	0.00	1.62E-02		0.00	0.00	0.00	3.04E-01		0.00	0.00	0.00	3.28E-01	
Saturated fat intake§	0.00	-0.01	0.02	8.35E-01		0.00	0.00	0.01	5.85E-03		0.00	0.00	0.01	6.28E-01	
Hygiene	-0.04	-0.09	0.01	1.03E-01		-0.01	-0.02	0.00	2.00E-01		0.00	-0.02	0.02	8.13E-01	
Illness	0.40	0.27	0.53	1.57E-09		-0.04	-0.07	-0.02	1.37E-03		0.04	-0.01	0.09	1.32E-01	
SES	0.09	0.03	0.15	2.11E-03		0.03	0.02	0.04	2.45E-06		0.02	0.00	0.04	6.42E-02	

*Adjusted R-square; †Waist circumference ≥ 80cm; ‡Overweight/obese, BMI ≥ 23kg/m^2; §Percentage of energy intake from saturated fat

Figure 4.3: The proportion of variance explained by genetic risk score vs. individual SNP most strongly associated with trait

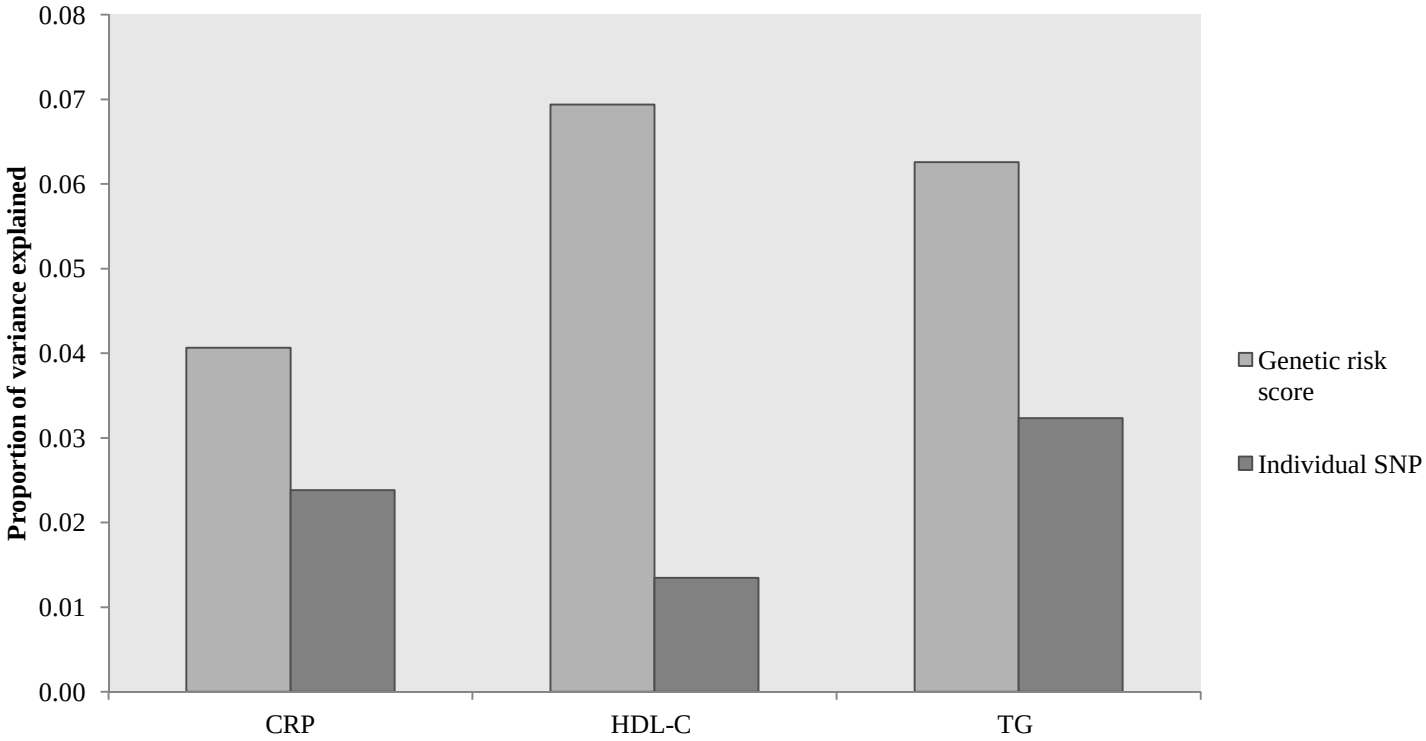


Figure 4.3: The proportion of variance explained by genetic risk score vs. individual SNP most strongly associated with trait

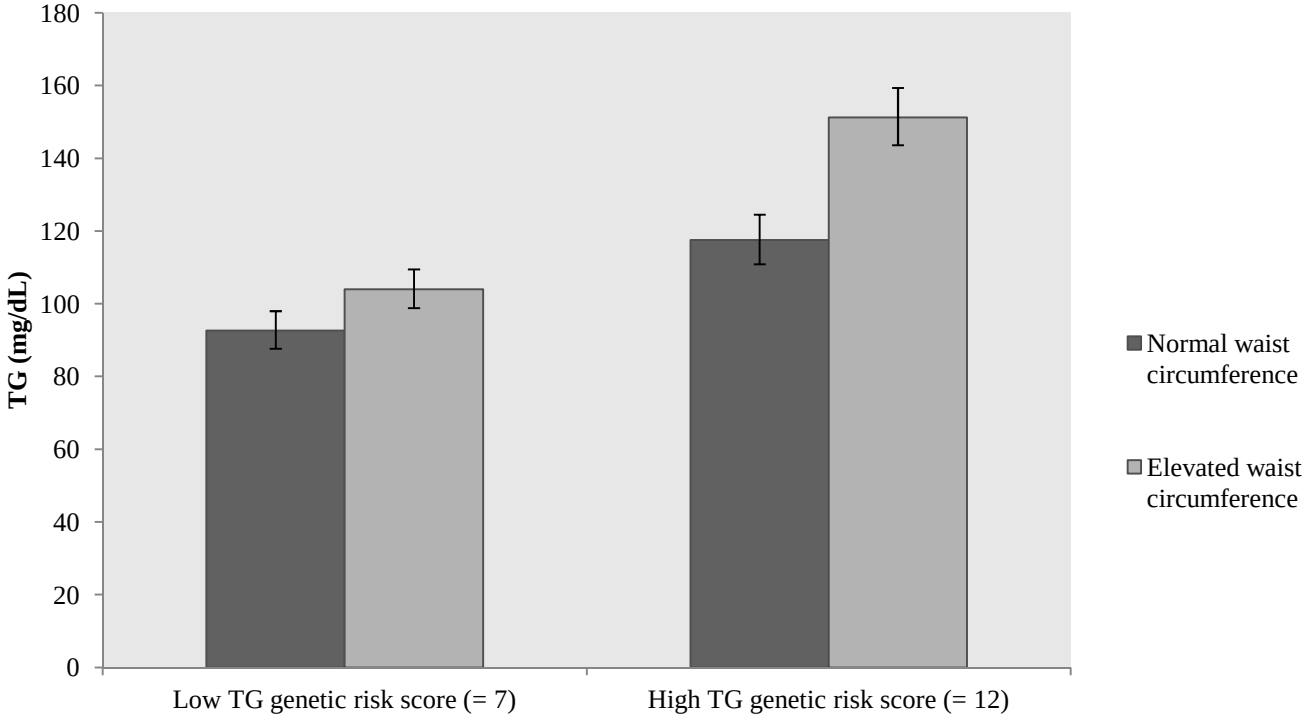
The proportion of variance in the CM trait explained by the genetic risk score vs. the individual SNP most strongly associated with the trait. The individual SNP most strongly associated with CRP was rs876537, with HDL-C was rs12708980, and with TG was rs964184. See Methods for further details.

Table 4.5: Evidence of interaction between TG genetic risk score and levels of adiposity on log-transformed TG levels

	Normal waist circumference† *n* = 785				Elevated waist circumference† *n* = 864				
	β	95% CI		*P* value	β	95% CI		*P* value	*P* interaction
TG genetic risk score	0.05	0.03	0.07	1.02E-08	0.08	0.06	0.1	4.96E-22	1.66E-02
	Normal weight‡ *n* = 661				Overweight and obese‡ *n* = 988				
	β	95% CI		*P* value	β	95% CI		*P* value	*P* interaction
TG genetic risk score	0.05	0.03	0.07	3.21E-07	0.08	0.06	0.09	5.21E-24	2.73E-02

*The covariates were age (categorical: ≤44 y, 45-49 y, 50-54 y, and ≥55 y), PCs, postmenopausal status, % energy intake from saturated fat, energy intake, environmental hygiene, illness, and SES (outcome was log-transformed TG levels); †Stratified by waist circumference ≥ 80cm; additional adjustment for normal weight, overweight, and obese BMI categories; ‡Stratified by normal weight: BMI < 23kg/m^2, overweight: 23kg/m^2 ≤ BMI < 27.5kg/m^2, obese: BMI ≥ 27.5kg/m^2; additional adjustment for elevated waist circumference

Figure 4.4: Predicted triglyceride levels at the 25th and 75th percentile values of the genetic risk score stratified by levels of adiposity



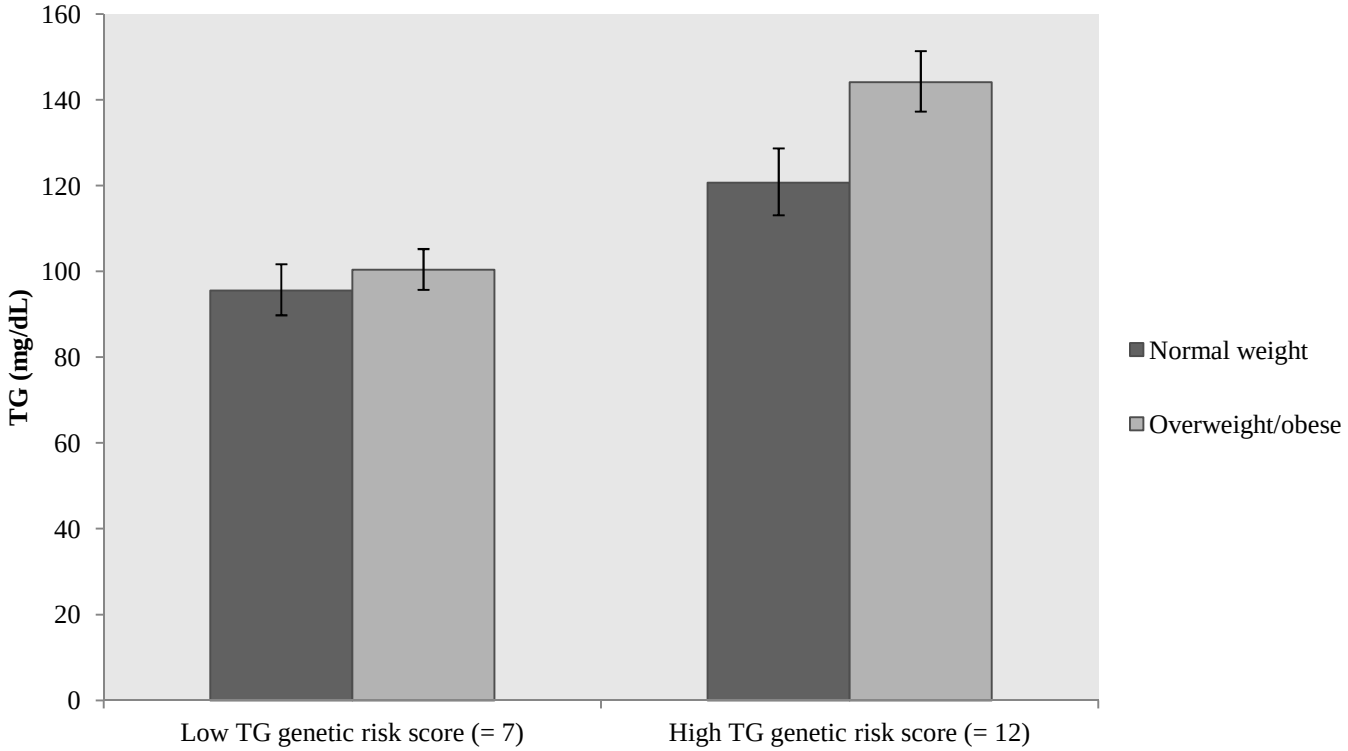


Figure 4.4: Predicted triglyceride levels at the 25th and 75th percentile values of the genetic risk score, stratified by levels of adiposity

Showing predicted geometric means (95% CI) of triglycerides (TG) at the 25th and 75th percentile values of the genetic risk score (7 and 12 respectively) at varying levels of waist circumference (WC) and BMI: A. Predicted levels of TG stratified by WC ≥ 80cm; B. Predicted levels of TG stratified by overweight/obese (BMI ≥ 23kg/m^2)

Chapter 5. IDENTIFICATION OF GENETIC RISK PROFILES ACROSS MULTIPLE CARDIOMETABOLIC PHENOTYPES IN FILIPINO WOMEN

OVERVIEW

The patterning of genetic risk across multiple phenotypes may better explain the underlying genetic susceptibility for cardiometabolic (CM) disease, especially since these diseases themselves display a specific clustering of risk factors including insulin resistance (IR), dyslipidemia, hypertension, and inflammation. Here we used a novel application of cluster analysis to identify groups of women who share similar patterns of genetic risk across multiple CM phenotypes.

Participants included 1,584 women from the 2005 Cebu Longitudinal Health and Nutrition Survey. Using cluster analysis, we grouped individuals by the following six CM genetic risk scores: triglycerides (TG), high-density lipoprotein (HDL-C), Low-density lipoprotein (LDL-C), blood pressure (BP), IR, and C-reactive protein (CRP). Using linear regression and multinomial logistic regression (mlogit) models we assessed how these genetic risk clusters, along with anthropometric, dietary, and other environmental factors predicted CM trait levels and CM clusters.

We identified 5 distinct genetic risk clusters: (1) Low CM risk variants, (2) Increased TG/BP variants, (3) Increased HDL-C variants, (4) Increased IR/BP variants, and (5) Increased LDL-C variants. Belonging to any one of the genetic risk clusters (as compared to the Low CM risk variants cluster) predicted increased levels in at least two CM traits (or decreased levels in terms of HDL-C).

Previous research found five CM risk clusters: (1) Healthy, (2) Low HDL-C, (3) Elevated BP, (4) Insulin resistant, and (5) Elevated CRP. In the mlogit, we found being in the Increased TG/BP variants cluster (vs. the Low CM risk variants cluster) increased the likelihood of being in the Elevated BP cluster (odds ratio [OR]= 1.71, 95% CI [1.03, 2.84]) and the Elevated CRP cluster (OR= 1.90, 95% CI [1.10, 3.27]). We also found that elevated WC increased the likelihood of being in all the CM risk clusters, whereas OW status only increased the likelihood of being in the Elevated BP and Elevated CRP clusters. In addition, a decrease in the percentage of energy intake from saturated fat uniquely increased the likelihood of being in the Low HDL-C cluster (OR= 0.94, 95% CI [0.90, 0.97]).

Genetic risk clusters, along with anthropometric and dietary factors, predicted CM trait levels and patterns in this population. By capturing the intricate relationship of these modifiable and non-modifiable factors with common complex traits we can further understand how to effectively reduce and prevent CM risk and its associated diseases.

BACKGROUND

Individually, common genetic variants only minimally explain common complex diseases such as cardiovascular disease (CVD), type 2 diabetes, and other related cardiometabolic (CM) conditions.[79] Jointly considering the relatively small effects of these individual SNPs may better capture underlying genetic risk associated with these diseases. Recently, genetic risk scores have been implemented to interrogate the joint impact of multiple SNPs of CM disease.[80-84]

Some studies create a genetic risk score by summing up the number of risk alleles pertaining to a single CM phenotype, while other studies similarly construct a genetic risk score, but choose risk alleles associated with a CM disease of interest. The latter method

combines risk variants pertaining to multiple phenotypes with the intention of better capturing the intricate relationship between genetics and common complex disease. Still, the majority of these studies find slight to no improvement in classifying at-risk individuals. [82, 83] This approach masks the actual patterning of genetic risk across phenotypes. Perhaps, understanding this heterogeneity in genetic risk clustering may aid in predicting and preventing CM disease,[79, 85, 86] especially since these diseases themselves display a specific patterning of risk factors including insulin resistance (IR), dyslipidemia, hypertension, and inflammation.[87, 88]

Driven by the downfalls of using a single genetic risk score approach, we used a novel application of cluster analysis to identify groups of women from the 2005 Cebu Longitudinal Health and Nutrition Survey (CLHNS), who share similar patterns of genetic risk across multiple CM phenotypes. Since CM risk factors tend to co-occur together, we hypothesized that genetic risk across CM phenotypes would also do the same. Cluster analysis is a valuable approach because it allows for the heterogeneous combinations of risk factors, which likely better reflect the underlying susceptibility for disease. To accomplish this goal, we first created six genetic risk scores, each score representing a summation of the genetic risk variants associated with a single CM trait. Next, we identified groups of women with similar profiles of genetic risk by using cluster analysis on the six genetic risk scores. We then examined how these genetic risk clusters related with CM risk in this population, while accounting for other factors such as age, diet, and anthropometry.

METHODS

Survey design

The women in this study are participants in the CLHNS, which is described in detail elsewhere.[89] Briefly, the CLHNS is a community-based cohort of women and their index children followed since 1983. The original participants included all pregnant women in 33 randomly selected communities of Metro Cebu, who gave birth between May 1, 1983, and April 30, 1984. A baseline interview was conducted among women in their 6th to 7th month of pregnancy. Subsequent surveys took place immediately after birth, bimonthly for 2 years, in 1991, 1994-5, 1998-99, 2002, and 2005. In 2005, fasting blood was drawn for CVD biomarkers and genetics. Here we use data from the mother cohort still participating in the 2005 CLHNS. All data were collected under conditions of informed consent with institutional review board approval from the University of North Carolina, Chapel Hill, USA.

We excluded women who were pregnant at the time of blood draw, not fasting at the time of blood draw, and with C-reactive protein (CRP) levels >10 mg/L (a level representing current/recent illness rather than low-level basal inflammation).[88] Even though we removed individuals with CRP levels >10 mg/L, we adjusted for the presence of any infectious symptoms at the time of blood collection to help control for any residual confounding in our analysis.[1] 1,584 women had complete biomarker, genetic, diet, socioeconomic, and anthropometric data. Medication use in this population was low: 0.1% took statins, 1.75% took diabetes medication, and 3% took anti-hypertensive medications. We adjusted for anti-hypertensive medication use in our analysis. All data were collected with informed consent, using protocols approved by the institutional review board of the University of North Carolina, Chapel Hill.

Cardiometabolic biomarkers

Fasting plasma CM biomarkers included triglycerides (TG), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), glucose, insulin, and CRP. Blood samples were collected in participants' homes in the morning after an overnight fast. Venous blood was collected in EDTA anti-coagulant vacutainer tubes. After mixing to inhibit clotting, glucose was measured immediately using the glucose dehydrogenase method (One Touch Ultra Blood Glucose Monitoring System, Lifescan Johnson and Johnson). Blood samples were stored on ice for no more than 2 hours and were then centrifuged to separate plasma.

After separation, samples were frozen and remained frozen at -80 °C until ready for analysis. Total lipid concentrations were measured at the Emory Lipid Research Laboratory using enzymatic methods with reagents from Beckman Diagnostics on the Beckman Diagnostics CX5 chemistry analyzer (Fullerton, CA). HDL-C was determined using the homogeneous assay for direct determination (Genzyme, Exton, PA). LDL-C was determined using the Friedewald formula, except if triglycerides exceeded 400 mg/dl then LDL-C was directly determined using a homogenous assay (Genzyme, Exton, PA). The Emory Lipid Research Laboratory is a participant in the CDC/NHLBI Lipid Standardization Program to ensure accuracy and precision of the determinations. Plasma insulin was measured using automated Bayer® ADVIA Centaur chemiluminescent methods.[101] CRP concentrations were determined using a high sensitivity immunoturbidimetric method (Synchron LX20, lower detection limit: 0.1 mg/L).

Other cardiometabolic biomarkers included homeostatic model assessment insulin resistance (HOMA-IR), and systolic and diastolic blood pressure (BP). HOMA-IR was

calculated as 22.5/ (insulin × glucose). Systolic BP and diastolic BP were measured in triplicate after a 10 minute seated rest using a mercury sphygmomanometer. The mean of the three measurements was used.

We used cutpoints for these biomarkers based on recommendations from the International Diabetes Federation, the American Heart Association, and other previously recognized and accepted cutpoints (Table 1).[88, 96, 102, 103] Before using cutpoints to identify participants with impaired fasting glucose, we applied a glucose correction factor to all fasting glucose levels. Glucometers overestimate glucose concentrations in whole venous blood as compared with standard laboratory methods.[140, 141] Therefore we subtracted 0.97 mmol/l from fasting glucose values to obtain the best equivalent to venous plasma as analysed by a laboratory autoanalyser.[140] The corrected fasting glucose values are reported in the analyses and tables.

Anthropometry

Body weight, height, and waist circumference (WC) were measured using standard techniques.[104] BMI was calculated as the ratio of weight (kg) to height (m^2). We used cutpoints for Asians to define overweight (OW) as BMI $\geq$ 23 kg/m^2 and central adiposity as WC $\geq$ 80 cm.[105 96]

Dietary data

Dietary data were derived from two 24-hour dietary recalls and the mean intake was used in the analyses. Data were collected during in-home interviews performed by highly trained local field staff. A nutritionist reviewed all dietary recalls immediately after collection. When implausible values were found, interviewers revisited respondents for verification.

Energy and saturated fat intakes were calculated using the Philippines Food Composition Tables. [19, 106]

Sociodemographic and lifestyle characteristics

Highly trained interviewers collected reproductive history data; this included menopausal status beginning in the 1991 survey.

Socioeconomic status (SES) was measured by a factor score based on a principal components analysis of household ownership of key assets such as television, vehicles, and furniture.[69]

Infectious illness was measured by asking participants if they were currently experiencing any symptoms of infection, consistent with prior research on CRP.[52] Symptoms included runny nose, cough, fever, diarrhea, and sore throat, as well as the more general categories of flu, cold, and sinusitis. Responses were used to construct a summary variable indicating the presence or absence of any infectious symptoms at the time of blood collection.

Environmental cleanliness and household hygiene was measured by a hygiene score based on data on the interviewer's rating of cooking area, immediate area around the house, toilet type, and water source. The score ranges from 0 to 9 with larger values indicating greater cleanliness.[52]

Genotyping, quality controls, and imputation

The complete methods for direct SNP genotyping, quality control, and SNP imputation have been described previously.[130] Briefly, genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 5.0. Quality control procedures excluded: samples with <97% genotyping call rate; members of estimated first-degree relative pairs; SNPs with a call

rate < 90%; SNPs with a deviation from Hardy-Weinberg equilibrium ($P < 10^{-6}$); SNPs with ≥ 3 discrepancies among duplicate pairs; SNPs with Mendelian inheritance errors among five CEPH trios and/or CEPH sample genotype discrepancies with HapMap. Genotype imputation was conducted with MACH using phased haplotypes from the 1000 Genomes Project in both CEU and CHB+JPT samples (June 2010 Release).[131] In addition, we excluded any SNPs with poor imputation quality (MACH $r^2 < 0.3$) or estimated minor allele frequency (MAF) ≤ 0.01.

G*enetic marker selection*

The process of choosing SNPs is depicted in Figure 1. The SNPs used to create the genetic risk scores were selected by finding SNPs associated with the individual CM traits of interest: TG, HDL-C, LDL-C, systolic BP, diastolic BP, glucose, HOMA-IR, and CRP. We selected these SNPs from (1) genome-wide association studies (GWAS) conducted with our own study population,[1, 7] (2) published GWAS of East and South East Asian cohorts,[2, 8, 9] and (3) published GWAS of European descent cohorts, [5, 71, 72] if the specific trait lacked studies conducted in populations of Asian descent. We limited our selection of studies to cohort-based studies and meta-analyses; case-control studies were not considered because we wanted to choose SNPs associated with the individual CM trait rather than the disease state. From the studies identified, we selected SNPs with a $P < 5\times10^{-8}$ in the original study population for further analysis; we increased this threshold to a $P < 5\times10^{-5}$ for those studies conducted in our own study population due to the smaller sample size in the CLHNS.

The original studies were used to identify the risk allele. We designated the risk allele as the allele associated with an increased level of the specified trait, except in the case of HDL-C, for which the allele associated with lower levels was designated. For each

individual, we coded imputed SNPs according to the dosage value or the expected number of copies of the risk allele (a continuous number between 0 and 2). This coding reflects the uncertainty in the imputation of the SNPs, for example a 1.5 suggests more uncertainty whereas a 1.9 suggests less uncertainty (an individual likely has two risk alleles).

Six genetic risk scores were constructed, one for each of the following traits: (1) TG, (2) HDL-C, (3) LDL-C, (4) BP (containing systolic and diastolic BP risk variants), (5) IR (containing glucose and HOMA-IR risk variants), and (6) CRP.

Before creating each genetic risk score, we chose a subset of SNPs with nominal significance (P <0.1) and directional consistency of the effect estimate in our study population. This was based on adjusted linear regression models of the natural log-transformed CM trait on the individual SNP (see Model 2 in Methods). Then for each CM trait, we pruned SNPs for redundancy due to linkage disequilibrium (r^2>0.2). To do this we used the --clump procedure in PLINK to create "clumps" of correlated SNPs.[132] Each clump was represented by the top index SNP, designated as the SNP with the lowest P value (see Model 2 in Methods). Using these index SNPs, we calculated each genetic risk score by summing all of the risk alleles associated with the specific trait. We used simple counts of the total number of risk alleles rather than weighting by the effect of each SNP. We created an un-weighted score instead of weighting by the effect of each SNP because: 1) the current literature does not provide stable effect estimates of each SNP for each trait; 2) the outcomes (and thus effects) across studies were non-comparable (e.g. log-transformed trait vs. non-transformed trait); 3) studies used populations of various sample sizes and ethnicities; 4) using weights from the CLHNS data itself would have introduced bias.

Cluster analysis

We performed cluster analysis to identify groups of women with similar genetic risk score patterns using SAS PROC FASTCLUS (SAS version 9.2, SAS Institute, Cary, NC). This procedure implements the K-means clustering algorithm (least squares method). K-means clustering uses the Euclidean distance, computed from input variables, to assign cluster membership by minimizing the distance among subjects in a cluster while maximizing the distance between clusters. The procedure first selects cluster seeds, a set of points calculated as a first guess of the cluster means. Next it calculates the Euclidean distance from each subject to each cluster seed; each subject is assigned to the nearest seed to form temporary clusters. The means of each of the temporary clusters are calculated and replace the seed values. Distance calculation and member assignment progress in an iterative fashion until no further changes occur.[108, 109]

Final cluster solutions are sensitive to initial seed values. To remedy this problem and to use a more objective approach to picking a cluster solution we created an algorithm modified from a previous clustering algorithm.[110] This algorithm performed 1,000 iterations of each cluster procedure using randomly generated initial cluster seeds. For each of the 1,000 cluster solutions it calculated the ratio of between-cluster variance to within-cluster variance or $R^2/(1-R^2)$, where R^2, pooled across all variables, representing the ability to predict each input variable from the cluster.[109] We wanted to maximize the ratio of between-cluster variance to within-cluster variance and therefore wanted to find the largest R^2. The algorithm identified the iteration/cluster solution with the largest R^2.[110]

The variables entered into the cluster analysis were sample-specific Z-scores of the six genetic risk scores: TG, HDL-C, LDL-C, BP, IR, and CRP. These genetic risk score variables were standardized because they did not have equal variance.

Statistical Analysis

We conducted 2 sets of analyses: (1) linear regression models to examine the association of each genetic risk cluster with each of the CM traits; (2) a multinomial logistic regression model (mlogit) to examine the association of the genetic risk clusters with CM risk patterns.

In the first analysis we used linear regression models to examine the association of the genetic risk cluster (coded as 5 dummy variables) with each of the CM traits. All traits were continuous and natural log-transformed to satisfy model assumptions of normally distributed residuals. Given the markedly skewed distribution of CRP concentrations and the presence of many values below the detectable level (0.1 mg/L) CRP values were natural log-transformed after adding the constant 0.10.[1]

Two linear regression models were examined. Model 1 was a linear regression model adjusted for age (categorical: ≤44 y, 45-49 y, 50-54 y, and ≥55 y) and principal components (PCs) representing population substructure among CLHNS subjects.[133] PCs were constructed using the software EIGENSOFT. We assessed the association between each of the first 10 PCs and each log-transformed CM trait to identify any potential ancestry explanatory PCs. The 7th PC was significantly associated with HDL-C, LDL-C, systolic BP, diastolic BP, and CRP (no PCs were significantly associated with log TG levels). Thus we included the first 7 PCs as covariates in all analyses. Model 2 included covariates adjusted for in Model 1 plus additional adjustment for postmenopausal status (yes/no), OW status (BMI ≥ 23 kg/m^2), high

WC (WC ≥ 80cm), % energy intake from saturated fat, energy intake, environmental hygiene, reported infectious illness (yes/no), anti-hypertensive medication use (yes/no), and SES. The covariates chosen for adjustment in Model 2 were based on prior published studies in the CLHNS on these traits.[52, 69, 115] We categorized age, BMI, and WC to account for their non-linear relationship with the log CM trait levels.

In the second analysis we used an mlogit to examine how genetic risk clusters, along with anthropometric, dietary, and other environmental factors predicted CM patterns. Here CM patterns represent the results from a cluster analysis previously published in this population on the following biomarkers: TG, HDL-C, LDL-C, BP, glucose, HOMA-IR, and CRP; this cluster analysis used the same approach as described for the cluster analysis on genetic risk scores (see Methods). We found five biologically relevant groups, which we named according to their predominant CM characteristics: (1) Healthy, (2) Low HDL-C, (3) Elevated BP, (4) Insulin resistant, and (5) Elevated CRP (Figure 1).[69] The outcome in this mlogit was one of the CM clusters, where the "Healthy" cluster served as the referent group. The mlogit included the same covariates as in Model 2.

For regression analyses we used a statistical significance criteria of $P < 0.05$ (2-sided). All regression analyses were performed with Stata 12.0 (Stata Corporation, College Station, TX, 2006).

RESULTS

The characteristics of all 1,584 women participants in the 2005 CLHNS are presented in Table 2. In 2005, participants had a mean (SD) age of 48.4 (6.0) years. About 38% of women were postmenopausal, 53% had elevated WC, 60% were OW, 29% had elevated TG, 82% had low HDL-C, 35% had elevated LDL-C, 36% had hypertension, 24% had elevated

fasting glucose, 16% had elevated HOMA-IR, and 20% had elevated CRP. Based on fasting glucose levels (≥ 7mmol/L), 8% of women were diabetic (although only 1.75% of women were taking medication). Results did not differ when we excluded these women in sensitivity analysis, thus we retained these women throughout our analysis.

Our selection strategy for candidate SNPs resulted in 13 TG, 19 HDL-C, 8 LDL-C, 9 BP, 22 IR, and 46 CRP usable SNPs (Figure 1). After pruning to eliminate correlated SNPs in linkage disequilibrium (by trait), 9 TG, 9 HDL-C, 4 LDL-C, 2 BP, 3 IR, and 6 CRP SNPs were used in the construction of the genetic risk scores (Figure 1; Table 3). The means and distributions for the genetic risk scores are presented in Table 4.

We conducted a series of cluster analyses with 3 to 6 clusters specified, and chose a 5-cluster solution because it yielded distinct genetic risk score patterns and each cluster had sufficient numbers (each approximately >17% of the sample).[110] The 5-cluster solution had an $R^2 = 0.39$, indicating the clusters explained about 40% of the variance in genetic risk scores.

We identified the five genetic risk clusters as: (1) Low CM risk variants, (2) Increased TG/BP variants, (3) Increased HDL-C variants, (4) Increased IR/BP variants, and (5) Increased LDL-C variants. We named the clusters according to what genetic risk score(s) had the highest/lowest mean relative to other clusters (Figure 2 and Table 4). The "Low CM risk variants" group represents low Z-scores for all of the genetic risk scores. Other characteristics of these clusters are highlighted in Table 2.

The results from regressions from each log-transformed CM trait (except log diastolic BP) on the genetic risk clusters are shown in Tables 5A and 5B. The results for log diastolic

BP were similar to that of log systolic BP and were therefore not shown. Here we report the results from Model 2 since similar associations were found in Model 1.

Belonging to the Increased TG/BP variants cluster (vs. the Low CM risk variants cluster) resulted in a 14% increase in TG levels (β= 0.14, 95% CI [0.07, 0.21]), a 7% decrease in HDL-C levels (β= -0.07, 95% CI [-0.10, -0.03]), and a 24% increase in CRP levels (β= 0.24, 95% CI [0.06, 0.43]).

Belonging to the Increased HDL-C variants cluster (vs. the Low CM risk variants cluster) resulted in a 9% increase in TG levels (β= 0.09, 95% CI [0.02, 0.17]), a 9% decrease in HDL-C levels (β= -0.09, 95% CI [-0.12, -0.05]), and a 39% increase in CRP levels (β= 0.39, 95% CI [0.20, 0.58]).

Belonging to the Increased IR/BP variants cluster (vs. the Low CM risk variants cluster) resulted in a 2% increase in systolic BP levels (β= 0.02, 95% CI [0.00, 0.05]) and a 31% increase in CRP levels (β= 0.31, 95% CI [0.13, 0.48]).

Belonging to the Increased LDL-C variants cluster (vs. the Low CM risk variants cluster) resulted in a 4% increase in LDL-C levels (β= 0.04, 95% CI [0.00, 0.09]) and a 22% increase in CRP levels (β= 0.22, 95% CI [0.04, 0.40]).

The results of the mlogit, with the outcome being one of the CM risk clusters (as compared to the Healthy cluster), are presented in Table 6. First looking at the genetic predictors, we found being in the Increased TG/BP variants cluster (vs. the Low CM risk variants cluster) increased the likelihood of being in the Elevated BP cluster (odds ratio [OR]= 1.71, 95% CI [1.03, 2.84]) and the Elevated CRP cluster (OR= 1.90, 95% CI [1.10, 3.27]). Next looking at non-genetic predictors, we found elevated WC increased the likelihood of being in all the CM risk clusters (while controlling for OW status), whereas

OW status only increased the likelihood of being in the Elevated BP and Elevated CRP clusters (while controlling for WC). In addition, a decrease in the percentage of energy intake from saturated fat uniquely increased the likelihood of being in the Low HDL-C cluster (OR= 0.94, 95% CI [0.90, 0.97]).

DISCUSSION

In this study we sought out to identify groups of individuals with similar profiles of genetic risk across multiple CM phenotypes. Cluster analysis was a useful tool for identifying groups of individuals who share similar patterns of genetic risk scores. By using cluster analysis, we were able to capture the heterogeneity in patterns of genetic risk across various phenotypes. To our knowledge, cluster analysis has never been used before to create genetic risk patterns for CM associated traits. From this we were able to identify which genetic risk patterns most strongly predicted CM trait levels. We also found that these genetic risk clusters, as well as anthropometric and dietary factors, independently predicted CM risk patterns in this population.

Belonging to any one of the genetic risk clusters (as compared to the Low CM risk variants cluster) predicted increased levels in at least two CM traits (or decreased levels in terms of HDL-C). Interestingly, each genetic risk cluster most strongly predicted log CRP levels as compared to all other CM traits. The properties of the referent cluster, the Low CM risk variants cluster, likely drove these findings: this cluster had the lowest relative mean CRP genetic risk score compared to all other clusters.

Among all the CM traits, we did not find an association between the genetic risk clusters and fasting glucose or HOMA-IR levels. This could simply reflect the lack of variation in these traits due to the low prevalence of IR/diabetes in this population, also noted by the

small size of the Insulin resistant CM cluster (n = 80) we previously found.[69] Another possibility is that the effects of adiposity overwhelmed the effects of the genetic risk cluster on glucose and HOMA-IR levels. Levels of adiposity highly influence glucose and HOMA-IR levels in our population as well as in other studies.[62, 69, 87, 120] It is also important to note that the SNPs chosen for the IR genetic risk score came largely from studies conducted in European populations. Therefore these SNPs may not be tagging the appropriate functional variant in our population, which could also explain why we saw non-significant associations between the genetic risk clusters and glucose/HOMA-IR.

Belonging to a specific CM cluster likely reflects variations in genetic risk and other modifiable and non-modifiable characteristics. Because CM risk factors tend to co-occur together, it seems likely that genetic risk across phenotypes would also do the same. In the mlogit, we found that being in the Increased TG/BP variants cluster (vs. the Low CM risk variants cluster) increased the likelihood of being in the Elevated BP and Elevated CRP clusters (vs. the Healthy CM cluster). It's unclear why this genetic risk cluster predicted the Elevated CRP group. Perhaps using the mlogit decreased our power to detect genetic effects (due to the categorical nature of the predictor and the outcome). Further research is needed to understand how/whether genetic risk profiles translate to phenotypic risk profiles.

In this same mlogit model (while adjusting for OW status) we found that WC was the most pervasive predictor of CM cluster membership, underscoring the adverse health effects of excess visceral fat deposition, assuming WC is an indicator of visceral fat. [142, 143] WC is among the best-established predictors of CM risk and past work in the CLHNS and studies in other Asian populations support this notion. [45, 52, 68, 70] We found that OW status only predicted membership to the Elevated BP and Elevated CRP clusters (when controlling for

WC). Research demonstrates that increased WC predicts CM abnormalities in both normal weight and OW individuals, highlighting the potential for visceral fat to influence development of CM risk factors independent of overall BMI status.[44]

In relation to dietary intake, we found that a decrease in the percentage of energy intake coming from saturated fat uniquely predicted membership in the Low HDL-C cluster. Most dietary recommendations suggest limiting saturated fat intake, since it elevates total and LDL cholesterol. However, recent studies have shown that lauric acid has a more favorable effect on the total cholesterol to HDL cholesterol ratio than any other fatty acid, either saturated or unsaturated, primarily by increasing HDL-C levels.[49] The most common cooking oil in Cebu is coconut oil, which is rich in lauric acid.[116] Our results suggest that decreased saturated fat intake, perhaps from coconut oil, increase the likelihood of membership into the Low HDL-C cluster. This is supported by recent findings by Feranil et al. that dietary coconut oil intake was positively associated with HDL-C levels in pre-menopausal CLHNS women.[144]

Limitations of our study merit consideration. Our sample size is relatively small therefore replication of our results in other Asian populations would reinforce our findings. In our literature search we found differing numbers of candidate SNPs for each trait. Although we used the same criteria in our search regardless of the CM trait, the variation in the number of candidate SNPs could reflect the current state of the literature. In addition, there is concern with choosing SNPs from a European sample and applying them to an Asian sample, especially in terms of tagging the appropriate functional variant. We tried to mitigate this by choosing SNPs with nominal significance and directional consistent effect estimates in our study population, however due to the limited sample size in the CLHNS we may have lacked the power to detect a SNPs true effect. Also, using a threshold of $r^2<0.2$ for linkage

disequilibrium, still allows SNPs to partially tag the same underlying signal, potentially including some redundancy in the genetic risk score.

A limitation of assigning names to the clusters is that not all individuals within a certain cluster necessarily share the ascribed characteristics. For example, in our "Low CM risk variants" cluster we found the average Z-scores for genetic risk scores were relatively low, but we cannot ascribe these low values to each individual in the cluster.

In conclusion, by using cluster analysis we were able to find distinct patterns of genetic risk. This method made fewer assumptions and allowed for relationships among CM genetic risk scores to emerge from the data themselves.[85] By finding combinations of genetic risk across multiple phenotypes, we can hopefully better explain the underlying genetic susceptibility for CM disease in this population, especially since these diseases themselves display a specific patterning of risk factors including IR, dyslipidemia, hypertension, and inflammation.[87, 88] These genetic risk clusters, along with anthropometric and dietary factors, predicted both CM trait levels and patterns in this population. By capturing this intricate relationship of these modifiable and non-modifiable factors with common complex traits we can further understand how to effectively reduce and prevent CM risk and its associated diseases.

Figure 5.1: Choosing of SNPs to include in the genetic risk scores

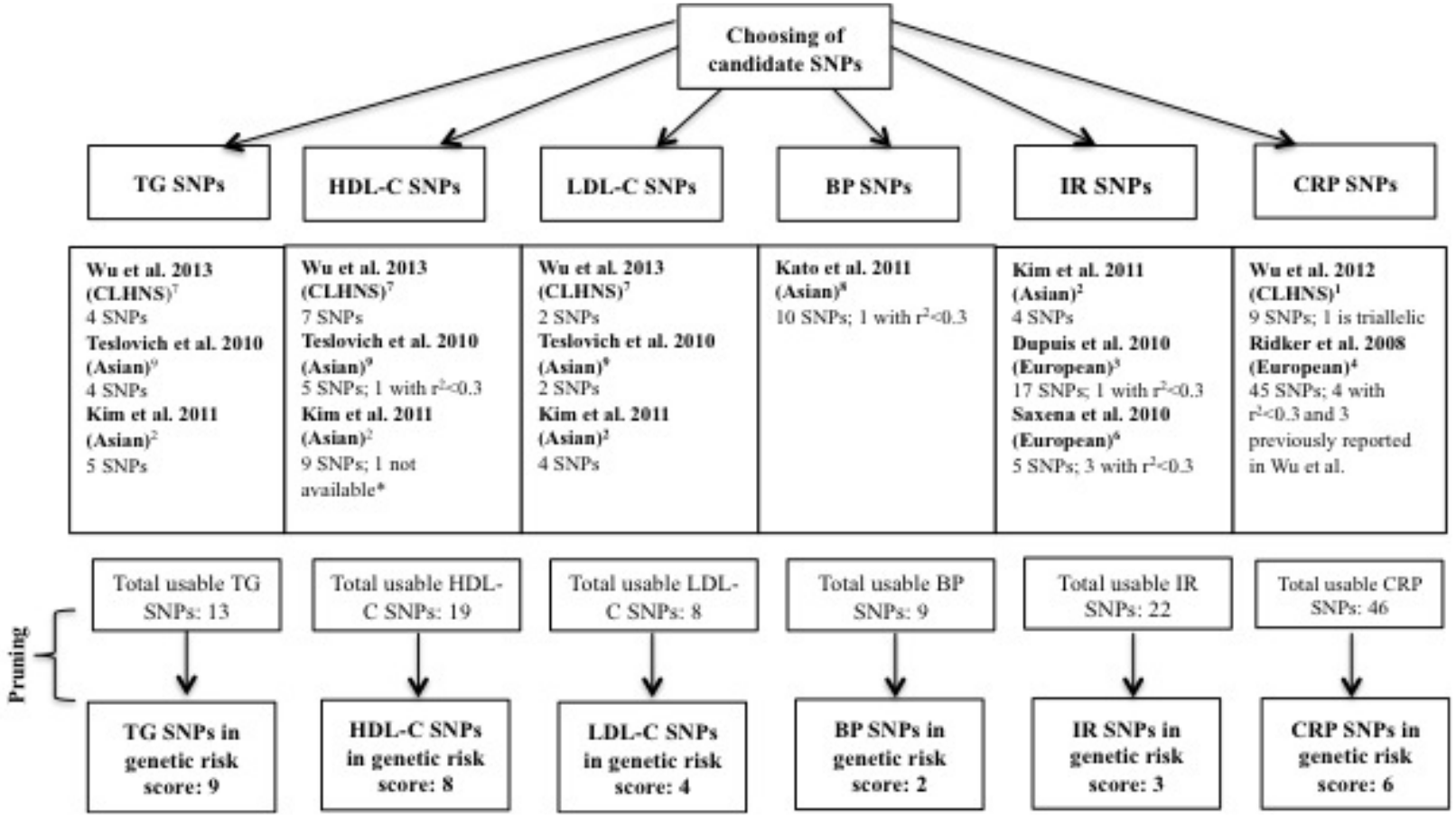


Figure 5.1: Choosing of SNPs to include in the genetic risk scores

A schematic representation describing the process of choosing SNPs associated with the following CM traits. Parentheses indicate the specific study population in which analyses were conducted. *rs1268004 was not genotyped and no HapMap or 1000 Genomes imputed data was available. See Methods for further details.

Table 5.1. Criteria for defining elevated cardiometabolic risk

Risk factors	Cutpoint
Triglycerides*	≥ 150 mg/dL
HDL cholesterol*	< 50 mg/dL
LDL cholesterol†	≥ 130 mg/dL
Systolic BP*	≥ 130 mm Hg
Diastolic BP*	≥ 85 mm Hg
Glucose*	≥ 100 mg/dL
HOMA-IR‡	≥ 4.65 mg/dL x μg/mL
Diabetes*	Fasting glucose ≥ 126 mg/dL
CRP§	> 3.0 mg/dL

*Cutpoints are defined by the IDF.[96] †Cutpoint is defined by the National Cholesterol Education Program.[102] ‡Cutpoint is defined by Stern et al.[103] §Cutpoint is defined by the American Heart Association.[88]

Figure 5.2: Mean Z-scores of fasting biomarkers by cardiometabolic cluster in Filipino women

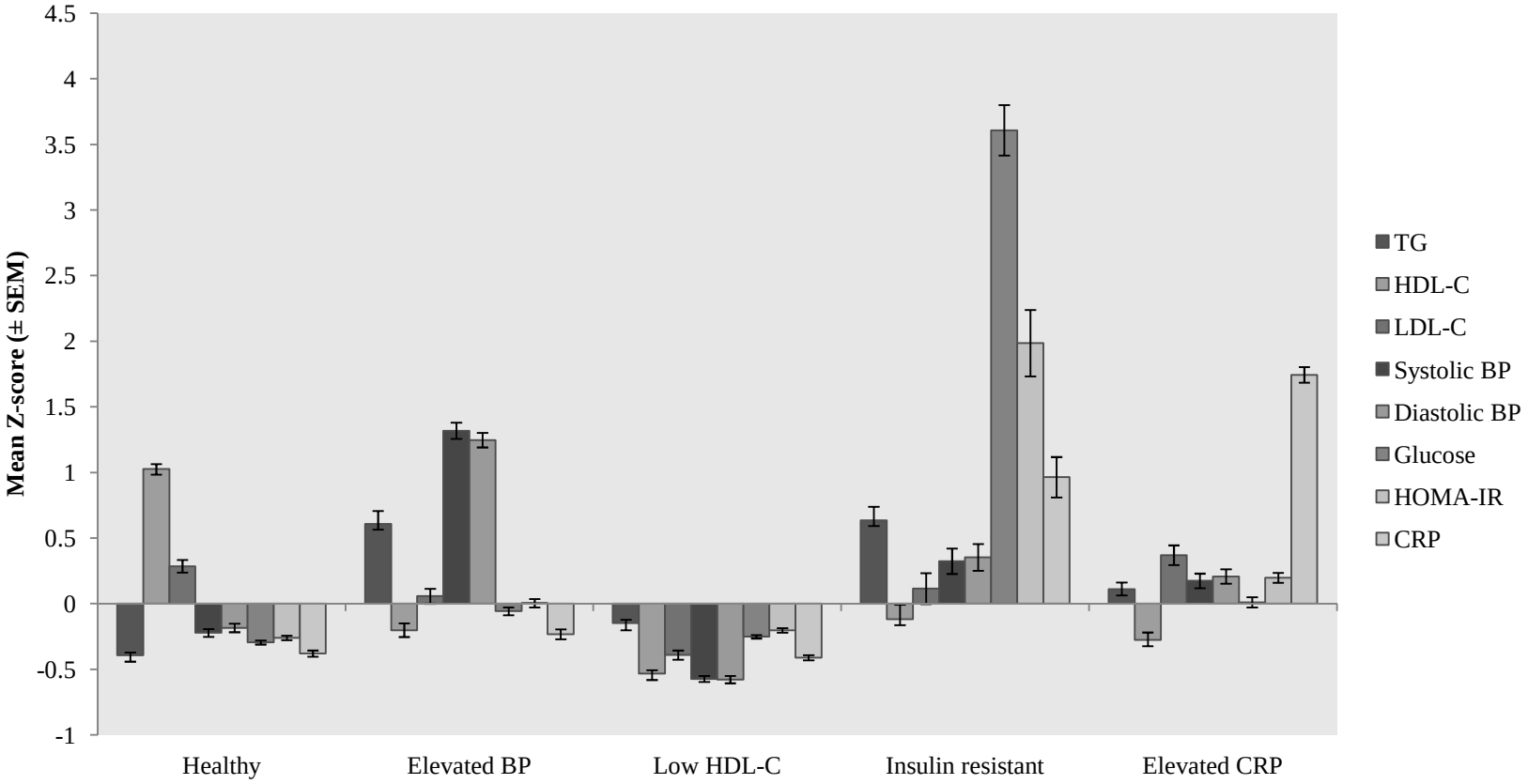


Figure 5.2: Mean Z-scores of fasting biomarkers by cardiometabolic cluster in Filipino women

Mean Z-scores by cardiometabolic cluster for the eight fasting biomarkers used as input variables in the cluster analysis for 1,584 Filipino women.

Table 5.2: Characteristics of 1,584 women participants in the 2005 CLHNS

	All women (n= 1,584)	Low CM risk variants (n= 335)	Increased TG/BP variants (n= 296)	Increased HDL-C variants (n= 268)	Increased IR/BP variants (n= 357)	Increased LDL-C variants (n= 328)
Age (%)						
≤44 y	32.7	31	29.4	31	36.7	34.5
45-49 y	31.3	32.2	32.1	34.3	27.5	31.4
50-54 y	20.8	19.1	23.6	23.1	20.7	18
≥55 y	15.2	17.6	14.9	11.6	15.1	16.2
Postmenopausal (%)	38.3	39.4	41.2	38.4	38.4	34.1
Illness* (%)	27.4	28.1	31.4	26.1	27.2	24.4
Anti-hypertensive medication (%)	3.5	4.5	5.4	2.2	2.5	2.7
Energy intake (kcal)	1,131.8 ± 496.8	1,105.4 ± 500.9	1,110.6 ± 431.1	1,148.3 ± 491.9	1,155.6 ± 527.3	1,138.7 ± 518.0
% Energy intake from saturated fat	5.4 ± 4.1	5.4 ± 4.4	5.3 ± 4.0	5.4 ± 4.0	5.6 ± 4.1	5.3 ± 3.9
Waist circumference (WC; cm)	81.1 ± 10.9	81.6 ± 11.2	81.3 ± 10.8	80.7 ± 11.4	80.8 ± 10.6	81.1 ± 11.0
Elevated WC‡ (%)	52.6	55.8	53.7	50	49.6	53.7
BMI (kg/m^2)	24.4 ± 4.4	24.6 ± 4.3	24.4 ± 4.5	24.0 ± 4.3	24.3 ± 4.3	24.4 ± 4.4
Overweight§ (%)	60	62.7	60.8	56.7	59.7	59.8
TG (mg/dL)	131.1 ± 85.0	127.0 ± 78.0	151.6 ± 120.9	141.6 ± 91.7	114.2 ± 57.0	126.8 ± 66.2
Elevated TG† (%)	28.7	27.8	36.5	34	19.9	27.7
HDL-C (mg/dL)	41.0 ± 10.3	41.7 ± 9.7	39.3 ± 9.5	39.1 ± 10.6	41.2 ± 10.5	43.1 ± 10.6
Low HDL-C† (%)	82.4	81.2	88.2	84.3	82.9	76.2
LDL-C (mg/dL)	119.7 ± 33.6	119.2 ± 35.6	122.1 ± 33.8	115.0 ± 29.8	118.3 ± 31.8	123.5 ± 35.5
Elevated LDL-C† (%)	35.3	34.9	37.8	30.2	35	37.8
Systolic BP (mm Hg)	119.8 ± 20.2	119.0 ± 18.4	121.3 ± 21.9	118.5 ± 18.9	121.0 ± 21.1	119.1 ± 20.4
Diastolic BP (mm Hg)	79.9 ± 12.5	79.4 ± 11.8	80.6 ± 12.9	78.8 ± 11.6	80.7 ± 13.1	79.5 ± 13.0
Hypertension (%)	36.2	34.9	39.9	32.5	38.7	34.5
Glucose (mmol/L)	5.5 ± 2.0	5.5 ± 1.9	5.3 ± 1.5	5.6 ± 2.3	5.7 ± 2.4	5.4 ± 1.6

Elevated glucose (%)	23.5	23	21.6	25.4	25.5	22.3
HOMA-IR (mmol/L x μIU/mL)	3.0 ± 3.2	3.1 ± 3.0	2.8 ± 2.3	3.0 ± 3.4	3.1 ± 3.9	3.1 ± 3.2
Elevated HOMA-IR (%)	16	17.6	14.5	16.4	14.3	17.4
Diabetecs† (%)	8.2	7.2	6.4	9.7	8.4	9.5
CRP (mg/L)	1.7 ± 2.1	1.5 ± 2.0	1.9 ± 2.3	1.9 ± 2.2	1.8 ± 2.1	1.7 ± 2.1
Elevated CRP† (%)	20	15.5	23.3	20.5	23	18

Data are means ± SD or percentages. *Percentage of individuals reporting illness at time of blood draw †See Table 1 for cutpoint values; ‡Waist circumference ≥ 80cm; §Overweight, BMI ≥ 23kg/m^2

Table 5.3: SNPs used to create trait specific genetic risk scores

Trait	Chr	SNPs	Nearest gene	Non-risk allele†	Risk allele	Risk allele frequency	β‡	95% CI		*P* value
TG	2	rs780092	GCKR	G	A	0.68	0.09	0.05	0.12	3.2E-7
	3	rs17023681	CNTN4	T	G	0.29	0.12	0.07	0.17	2.6E-6
	3	rs7644509	Chr3q26.1	C	G	0.19	0.08	0.04	0.13	3.5E-4
	7	rs2286276	TBL2-MLXIPL	T	C	0.90	0.05	-0.01	0.10	8.6E-2
	8	rs12678919	LPL	G	A	0.95	0.09	0.02	0.17	1.8E-2
	8	rs2001945	LPL	C	G	0.43	0.03	0.00	0.06	4.6E-2
	11	rs603446	ZNF259	T	C	0.68	0.08	0.04	0.11	1.9E-5
	11	rs964184	APOC3	C	G	0.24	0.15	0.11	0.19	3.4E-15
	18	rs1893838	ZBTB7C	T	C	0.35	0.07	0.03	0.10	1.6E-4
HDL-C	2	rs1544857	SLC4A10	G	C	0.17	-0.05	-0.08	-0.03	6.6E-6
	2	rs17548357	BIRC6	G	A	0.02	-0.19	-0.28	-0.11	1.4E-5
	9	rs3739440	PAX5	C	T	0.17	-0.07	-0.10	-0.04	2.3E-6
	11	rs11227643	11q13.1	C	G	0.73	-0.05	-0.08	-0.02	5.8E-4
	11	rs964184	APOC3	C	G	0.24	-0.02	-0.04	0.00	1.7E-2
	15	rs1532085	LIPC	A	G	0.43	-0.04	-0.07	-0.02	4.3E-5
	15	rs2070895	LIPC	A	G	0.62	-0.06	-0.09	-0.03	3.5E-5
	16	rs12708980	CETP	T	G	0.19	-0.05	-0.08	-0.03	6.6E-7
	22	rs138779	TOM1	T	C	0.39	-0.05	-0.06	-0.03	2.4E-6
LDL-C	1	rs629301	SORT1	G	T	0.95	0.04	0.00	0.09	7.1E-2
	8	rs4570159	TNKS	A	G	0.69	0.04	0.02	0.06	1.0E-4
	16	rs4787103	A2BP1	G	A	0.37	0.04	0.02	0.06	1.1E-3
	19	rs2738446	LDLR	C	G	0.24	0.02	0.00	0.05	3.5E-2
BP	1	rs17030613	ST7L, CAPZA1	A	C	0.42	0.01	0.00	0.02	4.7E-2
	4	rs16998073	FGF5	A	T	0.48	0.01	0.00	0.02	6.1E-2
IR	2	rs780092	GCKR	T	C	0.44	0.02	0.00	0.03	4.4E-2
	2	rs16856247	ABCB11	C	T	0.35	0.02	0.00	0.03	7.3E-2
	8	rs11558471	SLC30A8	G	A	0.59	0.03	0.01	0.05	1.2E-2
CRP	1	rs12093699	CRP	G	A	0.08	0.25	0.07	0.42	6.0E-3
	1	rs876537	CRP	T	C	0.43	0.33	0.24	0.42	2.3E-12
	1	rs1892534	LEPR	T	C	0.15	0.11	0.00	0.23	5.3E-2
	6	rs1408282	6q16.1	G	A	0.09	0.42	0.24	0.59	2.3E-6
	12	rs1169288	HNF1A	C	A	0.63	0.33	0.23	0.43	1.1E-10
	12	rs1169302	HNF1A	G	T	0.29	0.09	-0.01	0.19	8.2E-2

*SNPs used to create genetic risk score after pruning (see Methods); †Risk allele as defined by the study from which it was chosen; ‡Coefficient represents % change in CM trait level per each risk allele; results from Model 2 (see Methods) where the outcome was the natural-log transformed CM trait

Table 5.4: Genetic risk score distribution for all 1,584 women participants and by genetic risk cluster

Genetic risk score	All women (n = 1,584)		Low CM risk variants (n = 335)		Increased TG/BP variants (n = 296)		Increased HDL-C variants (n = 268)		Increased IR/BP variants (n = 357)		Increased LDL-C variants (n = 328)	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range
TG	9.4 ± 1.8	2.3 - 14.3	8.8 ± 1.6	4.3 - 14.2	11.1 ± 1.3	6.1 - 14.3	10.2 ± 1.6	5.1 - 14.1	8.1 ± 1.4	2.3 - 11.3	9.3 ± 1.6	3.6 - 13.9
HDL-C	5.9 ± 1.6	1.6 - 11.7	5.0 ± 1.2	1.6 - 8.3	6.8 ± 1.4	2.9 - 11.2	7.4 ± 1.2	4.1 - 11.7	5.6 ± 1.4	2.5 - 9.3	5.3 ± 1.4	2.2 - 9.8
LDL-C	4.5 ± 1.1	1.1 - 8.0	3.8 ± 0.8	1.3 - 5.8	4.3 ± 0.9	1.1 - 7.0	4.3 ± 0.9	1.2 - 6.9	4.4 ± 1.0	2.1 - 8.0	5.7 ± 0.7	4.0 - 8.0
BP	1.8 ± 0.9	0.0 - 3.9	1.3 ± 0.7	0.0 - 3.8	2.6 ± 0.7	0.9 - 3.9	0.9 ± 0.6	0.0 - 2.3	2.3 ± 0.7	0.3 - 3.9	1.6 ± 0.7	0.0 - 3.0
IR	3.0 ± 1.1	0.2 - 6.0	3.1 ± 1.0	0.4 - 6.0	2.8 ± 0.9	0.4 - 6.0	2.7 ± 1.1	0.3 - 6.0	3.8 ± 0.9	1.7 - 6.0	2.4 ± 1.0	0.2 - 5.0
CRP	3.3 ± 1.4	0.1 - 8.5	2.3 ± 1.0	0.1 - 5.9	3.1 ± 1.3	0.2 - 7.5	3.9 ± 1.3	0.2 - 7.9	4.0 ± 1.3	0.9 - 8.5	3.3 ± 1.2	0.2 - 6.7

Displayed means are mean ± SD

Figure 5.3: Mean Z-scores of genetic risk scores by genetic risk cluster in Filipino women

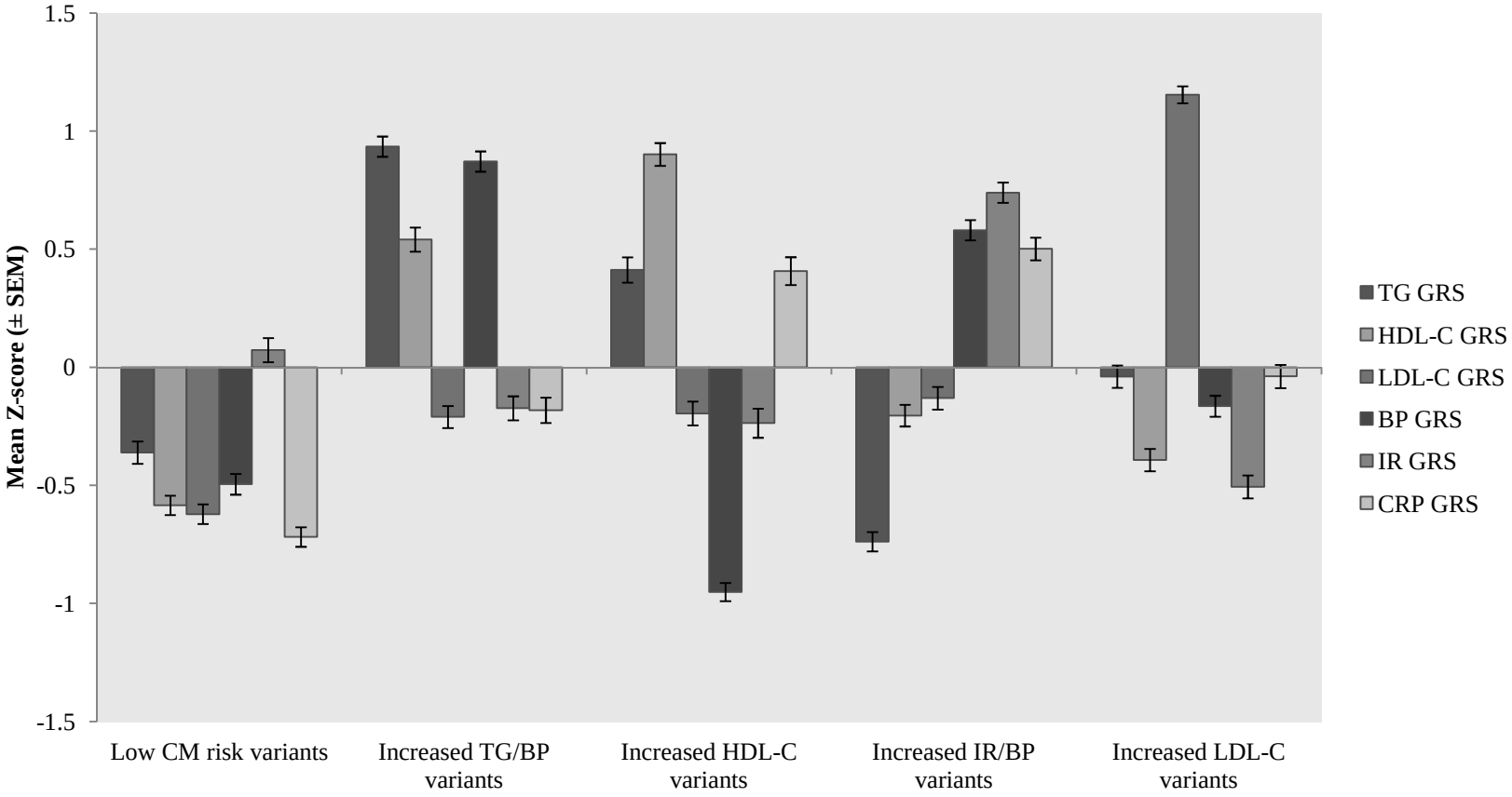


Figure 5.3: Mean Z-scores of genetic risk scores by genetic risk cluster in Filipino women

Mean Z-scores by genetic risk cluster (GRS) for the six genetic risk scores used as input variables in the cluster analysis for 1,584 Filipino women.

Table 5.5A: Association of genetic risk clusters with log-transformed lipid and blood pressure levels in 1,584 women

	Log TG				Log HDL-C				Log LDL-C				Log systolic BP			
Covariates	β	95% CI		*P* value	β	95% CI		*P* value	β	95% CI		*P* value	β	95% CI		*P* value
Genetic risk score cluster																
Low CM risk variants	Referent				Referent				Referent				Referent			
Increased TG/BP variants	0.14	0.07	0.21	0.00	-0.07	-0.10	-0.03	0.00	0.03	-0.01	0.07	0.20	0.02	-0.01	0.04	0.19
Increased HDL-C variants	0.09	0.02	0.17	0.01	-0.09	-0.12	-0.05	0.00	-0.03	-0.07	0.02	0.24	0.01	-0.02	0.03	0.67
Increased IR/BP variants	-0.06	-0.13	0.00	0.07	-0.03	-0.06	0.01	0.17	0.00	-0.04	0.05	0.83	0.02	0.00	0.05	0.03
Increased LDL-C variants	0.03	-0.05	0.10	0.49	0.02	-0.02	0.05	0.39	0.04	0.00	0.09	0.04	0.01	-0.02	0.03	0.69
Age																
≤44 y	Referent				Referent				Referent				Referent			
45-49 y	0.07	0.01	0.13	0.02	0.00	-0.03	0.04	0.81	0.05	0.01	0.08	0.01	0.03	0.01	0.05	0.00
50-54 y	0.16	0.08	0.24	0.00	-0.01	-0.05	0.03	0.70	0.09	0.04	0.14	0.00	0.04	0.01	0.06	0.01
≥55 y	0.13	0.04	0.23	0.01	0.01	-0.04	0.06	0.60	0.08	0.03	0.14	0.00	0.07	0.04	0.10	0.00
Postmenopausal status	0.09	0.02	0.15	0.01	-0.02	-0.06	0.02	0.29	0.05	0.01	0.09	0.02	0.02	0.00	0.04	0.04
Elevated WC†	0.18	0.11	0.24	0.00	-0.06	-0.10	-0.03	0.00	0.00	-0.04	0.04	0.95	0.05	0.03	0.07	0.00
OW status‡	0.12	0.05	0.19	0.00	-0.01	-0.04	0.03	0.67	0.08	0.04	0.12	0.00	0.04	0.02	0.06	0.00
Energy intake	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.21	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.65
Saturated fat intake§	0.00	0.00	0.01	0.30	0.00	0.00	0.01	0.02	0.00	0.00	0.01	0.08	0.00	0.00	0.00	0.80
Hygiene	0.00	-0.02	0.02	1.00	-0.01	-0.02	0.00	0.18	0.00	-0.01	0.01	0.99	-0.01	-0.01	0.00	0.03
Illness	0.03	-0.03	0.08	0.33	-0.04	-0.07	-0.01	0.00	-0.01	-0.04	0.03	0.72	0.01	-0.01	0.03	0.31
SES	0.02	-0.01	0.04	0.16	0.03	0.02	0.04	0.00	0.03	0.02	0.04	0.00	0.01	0.00	0.01	0.13
Anti-hypertensive medication	0.12	-0.01	0.25	0.08	-0.04	-0.11	0.03	0.31	-0.04	-0.12	0.04	0.31	0.13	0.09	0.18	0.00

*Coefficient represents % change in CM trait level per unit change in predictor; results from Model 2 (see Methods) where the outcome was the natural-log transformed CM trait; †Waist circumference ≥ 80cm; ‡Overweight, BMI ≥ 23kg/m^2; §Percentage of energy intake from saturated fat

Table 5.5B: Association of genetic risk clusters with log-transformed glucose, HOMA-IR, and CRP levels in 1,584 women

Covariates	Log glucose β	Log glucose 95% CI		Log glucose *P* value	Log HOMA-IR β	Log HOMA-IR 95% CI		Log HOMA-IR *P* value	Log CRP β	Log CRP 95% CI		Log CRP *P* value
Genetic risk score cluster												
Low CM risk variants		Referent				Referent				Referent		
Increased TG/BP variants	-0.02	-0.06	0.02	0.31	-0.08	-0.17	0.02	0.11	0.24	0.06	0.43	0.01
Increased HDL-C variants	0.03	-0.01	0.07	0.13	-0.03	-0.13	0.07	0.54	0.39	0.20	0.58	0.00
Increased IR/BP variants	0.03	0.00	0.07	0.07	-0.04	-0.13	0.05	0.41	0.31	0.13	0.48	0.00
Increased LDL-C variants	0.00	-0.03	0.04	0.87	0.00	-0.09	0.09	0.99	0.22	0.04	0.40	0.02
Age												
≤44 y		Referent				Referent				Referent		
45-49 y	0.01	-0.02	0.04	0.37	-0.01	-0.09	0.06	0.74	0.03	-0.12	0.18	0.69
50-54 y	0.05	0.01	0.10	0.01	0.00	-0.10	0.11	0.97	0.04	-0.16	0.24	0.71
≥55 y	0.07	0.02	0.11	0.01	-0.02	-0.14	0.10	0.75	0.27	0.03	0.51	0.03
Postmenopausal status	0.01	-0.03	0.04	0.77	0.02	-0.07	0.10	0.73	0.22	0.04	0.39	0.01
Elevated WC†	0.08	0.05	0.11	0.00	0.39	0.31	0.48	0.00	0.56	0.39	0.72	0.00
OW status‡	0.03	-0.01	0.06	0.16	0.38	0.29	0.47	0.00	0.46	0.29	0.64	0.00
Energy intake	0.00	0.00	0.00	0.77	0.00	0.00	0.00	0.97	0.00	0.00	0.00	0.03
Saturated fat intake§	0.00	0.00	0.00	0.85	0.01	0.00	0.02	0.13	0.01	-0.01	0.02	0.55
Hygiene	-0.01	-0.02	0.00	0.12	-0.02	-0.05	0.00	0.07	-0.04	-0.09	0.01	0.16
Illness	0.05	0.02	0.08	0.00	0.13	0.06	0.20	0.00	0.43	0.29	0.57	0.00
SES	0.01	0.00	0.02	0.12	0.09	0.06	0.12	0.00	0.09	0.03	0.15	0.00
Anti-hypertensive medication	0.01	-0.06	0.08	0.75	0.17	-0.01	0.34	0.06	-0.07	-0.40	0.27	0.69

*Coefficient represents % change in CM trait level per unit change in predictor; results from Model 2 (see Methods) where the outcome was the natural-log transformed CM trait; †Waist circumference ≥ 80cm; ‡Overweight, BMI ≥ 23kg/m^2; §Percentage of energy intake from saturated fat

Table 5.6: Association of genetic risk cluster with cardiometabolic risk cluster membership

	Elevated BP (n = 278)				Low HDL-C (n = 582)				Insulin resistant (n = 80)				Elevated CRP (n = 221)			
Covariates	OR	95% CI		*P* value	OR	95% CI		*P* value	OR	95% CI		*P* value	OR	95% CI		*P* value
Genetic risk score cluster																
Low CM risk variants		Referent				Referent				Referent				Referent		
Increased TG/BP variants	1.71	1.03	2.84	0.04	1.27	0.83	1.93	0.27	0.81	0.33	1.99	0.64	1.90	1.10	3.27	0.02
Increased HDL-C variants	1.43	0.84	2.43	0.19	1.28	0.84	1.96	0.26	1.55	0.68	3.51	0.29	1.73	0.98	3.04	0.06
Increased IR/BP variants	1.26	0.77	2.05	0.36	1.10	0.74	1.62	0.64	1.76	0.85	3.61	0.13	1.23	0.72	2.10	0.45
Increased LDL-C variants	1.15	0.71	1.88	0.57	0.87	0.59	1.30	0.50	1.02	0.47	2.22	0.96	1.07	0.62	1.83	0.81
Age																
≤44 y		Referent				Referent				Referent				Referent		
45-49 y	2.02	1.30	3.14	0.00	1.04	0.75	1.44	0.82	1.01	0.51	1.99	0.98	1.06	0.68	1.65	0.78
50-54 y	3.41	1.94	6.01	0.00	1.19	0.74	1.91	0.47	1.69	0.74	3.88	0.22	1.45	0.80	2.64	0.22
≥55 y	3.91	2.03	7.56	0.00	0.89	0.51	1.56	0.69	1.85	0.69	4.94	0.22	1.56	0.77	3.16	0.22
Postmenopausal status	0.80	0.50	1.27	0.34	0.68	0.46	1.02	0.06	0.99	0.49	2.00	0.97	1.03	0.62	1.70	0.91
Elevated WC†	2.83	1.79	4.48	0.00	1.44	1.00	2.08	0.05	4.04	1.91	8.56	0.00	2.56	1.58	4.14	0.00
OW status‡	1.61	0.99	2.62	0.05	0.88	0.61	1.27	0.49	1.48	0.67	3.28	0.33	3.23	1.86	5.62	0.00
Energy intake	1.00	1.00	1.00	0.68	1.00	1.00	1.00	0.29	1.00	1.00	1.00	0.31	1.00	1.00	1.00	0.94
Saturated fat intake§	0.96	0.92	1.01	0.10	0.94	0.90	0.97	0.00	0.93	0.87	1.00	0.06	0.97	0.93	1.02	0.27
Hygiene	0.97	0.84	1.11	0.66	1.10	0.98	1.23	0.10	0.93	0.75	1.15	0.50	0.91	0.78	1.05	0.19
Illness	1.42	0.97	2.08	0.07	1.06	0.77	1.47	0.71	4.14	2.44	7.04	0.00	2.25	1.52	3.32	0.00
SES	0.87	0.74	1.02	0.08	0.74	0.65	0.85	0.00	0.98	0.77	1.26	0.89	0.97	0.82	1.15	0.71
Anti-hypertensive medication	5.70	1.99	16.30	0.00	2.18	0.71	6.63	0.17	1.96	0.47	8.15	0.36	2.20	0.69	7.04	0.18

*Odds ratio represents likelihood of membership to specified cardiometabolic risk cluster (as compared to the Healthy cluster, n = 423) per unit change in predictor; results from mlogit adjusted for covariates in Model 2 (see Methods); †Waist circumference ≥ 80cm; ‡Overweight, BMI ≥ 23kg/m^2; §Percentage of energy intake from saturated fat

Chapter 6: SYNTHESIS

OVERVIEW OF FINDINGS

This research examined the interactive influence of dietary, lifestyle, and genetic factors on cardiometabolic (CM) traits and patterns in Filipino young adults and women. We used cross-sectional data from the 2005 Cebu Longitudinal Health and Nutrition Survey (CLHNS) of Filipino middle-aged women and their young adult offspring.

First we examined the patterns and determinants of CM risk factors among individuals in this study population. Our previous research identified five profiles of Filipino middle-aged women with similar CM characteristics and found modifiable risk factors for these CM patterns, including measures of adiposity and saturated fat intake. We extended this analysis to their young adult offspring, for whom prevention is still an important goal.

In order to further understand the etiology of CM risk in these older women, we used genetic risk scores, which combined the relatively small additive effects of individual single nucleotide polymorphisms (SNPs), to better capture the complex relationship between genetics, adiposity, and CM risk.

Lastly, we used cluster analysis to identify groups of these women who share similar patterns of genetic risk scores across multiple CM phenotypes. We then examined how these genetic risk clusters related with CM traits and patterns in this population, while accounting for other factors such as age, diet, and anthropometry.

The following section provides a summary and synthesis of our primary findings.

CLUSTERING AND DETERMINANTS OF CARDIOMETABOLIC RISK FACTORS AMONG FILIPINO YOUNG ADULTS

With modernization, CM disease risk has increased in low and middle-income countries. To better understand CM disease etiology, we evaluated the patterning of CM risk factors in a young adult population. This population displayed a low prevalence of overweight (18%). Despite leanness, we identified 5 distinct sex-specific clusters: (1) Healthy/high high-density lipoprotein cholesterol (HDL-C) (with the addition of high low-density lipoprotein cholesterol, LDL-C, in women), (2) Healthy/low blood pressure (BP), (3) High BP, (4) Insulin resistant (IR)/high triglycerides (TG), (5) High C-reactive protein (CRP). Though we did not identify a specific cluster primarily defined by low HDL-C, over 65% of men and 70% women had this trait, making low HDL-C the most pervasive CM risk factor. While our analysis found relatively similar CM risk clusters among men and women, the predictors of these clusters varied by sex. Perhaps as these young adults age more distinct CM patterns between men and women will emerge.

In both men and women, decreased % energy intake from saturated fat predicted membership in the High CRP group when compared to the two Healthy clusters. In addition, a decrease in % saturated fat intake predicted membership in the High BP group in women, compared to the Healthy/high HDL-C/high LDL-C group. The association of saturated fat intake with healthy CM profiles could reflect the types of saturated fats consumed in this population. Coconut oil, the most common and traditional cooking oil in Cebu, is rich in lauric acid.[116] Lauric acid improves the total cholesterol to HDL-C ratio, more than any other saturated or unsaturated fatty acid, primarily by increasing HDL-C levels.[49] Other studies have found diets rich in coconut oil or in saturated fat do not alter markers of inflammation,

fasting glucose, fasting insulin, homeostatic model assessment insulin resistance (HOMA-IR), or incident diabetes.[117, 118]

Men with poorer environmental hygiene (increased pathogenicity) were more likely to be in the High CRP cluster, compared to the two Healthy clusters. These results support previous research conducted in the CLHNS and reinforce the involvement of pathogen exposure in activating pro-inflammatory pathways.[50-52] But why do we fail to observe this hygiene effect in women? Adiposity relates more strongly with inflammation in women than in men, thus it is possible the effects of adiposity overwhelmed the effects of the hygiene score in women.[100, 119]

As expected, waist circumference (WC) and overweight (OW) status were the strongest predictors of membership in the IR/high TG cluster, underscoring the adverse health effects of excess visceral adipose tissue, for which WC serves as a proxy.[120] WC is among the best-established predictors of CM risk and past work in the CLHNS and other populations support this notion. [45, 50, 52, 68] Research has also demonstrated that increased WC predicts CM abnormalities in both normal weight and OW individuals, highlighting the potential for visceral fat to influence the development of CM risk factors, independent of body mass index (BMI). [121]

In conclusion, despite leanness, cluster analysis found patterns of CM risk. While measures of adiposity predicted some of these patterns, modifiable factors such as dietary intake and pathogen exposure also independently predicted cluster membership. This emphasizes the importance monitor and screen lean individuals for CM risk and future CM diseases, especially in Asian populations where the risk of CM diseases is elevated at a lower BMI (likely due to increased visceral fat at lower BMIs).[92]

GENETIC RISK SCORE AND ADIPOSITY INTERACT TO INFLUENCE TRIGLYCERIDE LEVELS IN A COHORT OF FILIPINO WOMEN

In this study we sought to: create three genetic risk scores relating to CRP, HDL-C, and TG traits, examine the ability of these scores to explain the variation in these traits, and test whether these genetic risk scores interact with adiposity to influence trait levels. Here participants included middle-aged Filipino women, for whom CM risk is more established.

The genetic risk score explained a greater proportion of variance in the specified trait than any given individual SNP. In addition, we found that for individuals with a higher TG genetic risk score, having either an elevated WC or being OW/obese amplified the genetic risk score's effect by further increasing TG levels. While for individuals with a lower TG genetic risk score, measures of adiposity made almost no difference in TG levels. Interestingly for those women with a low TG genetic risk score and elevated levels adiposity, their predicted levels of TG equaled those of women with a high genetic risk score without any adverse levels of adiposity.

In support of our results, recent work using the same study population found that central obesity might accentuate the effect of the TG-increasing allele of an *APOA5* variant.[7] In addition, previous research has implicated several variants in the *LPL* gene (a gene included in our genetic risk score) as having an interactive effect with central adiposity on TG levels and the ratio of TG to HDL-C.[127 128] However, we did not find significant interactions between these individual loci and adiposity on TG levels, perhaps indicating that the interactive effect is driven by a collective result of all SNPs in the TG genetic risk score.

In conclusion, our results suggest for individuals with a high genetic predisposition to elevated TG levels, reducing adiposity could possibly prevent increases in TG levels and thereby reduce the likelihood of CM diseases. Replication of these results in larger study

populations is needed to support the potential clinical and public health utility of targeted prevention efforts using genetic profiling.

IDENTIFICATION OF GENETIC RISK PROFILES ACROSS MULTIPLE CARDIOMETABOLIC PHENOTYPES IN FILIPINO WOMEN

In this study we sought out to identify groups of middle-aged Filipino women with similar profiles of genetic risk across multiple CM phenotypes. We then examined how these genetic risk scores, along with anthropometric and dietary factors, predicted CM trait levels and patterns in this population.

By using cluster analysis, we were able to capture the heterogeneity in patterns of genetic risk across various phenotypes. We identified five genetic risk clusters as: (1) Low CM risk variants, (2) Increased TG/BP variants, (3) Increased HDL-C variants, (4) Increased IR/BP variants, and (5) Increased LDL-C variants.

Belonging to any one of the genetic risk clusters (as compared to the Low CM risk variants cluster) predicted increased levels in at least two CM traits (or decreased levels in terms of HDL-C). Each genetic risk cluster most strongly predicted log CRP levels as compared to all other CM traits. The properties of the referent cluster, the Low CM risk variants cluster, likely drove these findings: this cluster had the lowest relative mean CRP genetic risk score compared to all other clusters.

Among all the CM traits, we did not find an association between the genetic risk clusters and fasting glucose or HOMA-IR levels. This could simply reflect the lack of variation in these traits due to the low prevalence of IR/diabetes in this population.[69] Another possibility is that the effects of adiposity, one of the strongest predictors of glucose and HOMA-IR levels, overwhelmed the effects of the genetic risk cluster.[62, 69, 87, 120] It is also important to note that the SNPs chosen for the IR genetic risk score came largely from

studies conducted in European populations, therefore these SNPs may not be tagging the appropriate functional variant in our population, which could also explain why we saw non-significant associations between the genetic risk clusters and glucose/HOMA-IR.

In a multinomial logistic regression (mlogit), we found that being in the Increased TG/BP variants cluster (vs. the Low CM risk variants cluster) increased the likelihood of being in the Elevated BP and Elevated CRP clusters (vs. the Healthy CM cluster). It's unclear why this genetic risk cluster, as opposed to the Increased IR/BP variants cluster, predicted the Elevated CRP group. Further research is needed to understand how/whether genetic risk profiles translate to phenotypic risk profiles.

In this same mlogit model (while adjusting for OW status) we found that WC was the most pervasive predictor of CM cluster membership. WC, a proxy for visceral fat, is among the best-established predictors of CM risk and past work in the CLHNS and studies in other Asian populations support this notion. [45, 52, 68, 70] We found that OW status only predicted membership to the Elevated BP and Elevated CRP clusters (while adjusting for WC). Research demonstrates the potential for visceral fat to influence development of CM risk factors independent of overall BMI status.[44]

In relation to dietary intake, we found that a decrease in the percentage of energy intake coming from saturated fat uniquely predicted membership in the Low HDL-C cluster. The most common cooking oil in Cebu is coconut oil, which is rich in lauric acid.[116] Earlier studies have shown that lauric acid has a more favorable effect on the total cholesterol to HDL cholesterol ratio than any other fatty acid, either saturated or unsaturated, primarily by increasing HDL-C levels.[49] This result is supported by recent findings by Feranil et al. that

dietary coconut oil intake was positively associated with HDL-C levels in pre-menopausal CLHNS women.[144]

In conclusion, genetic risk clusters, along with anthropometric and dietary factors, predicted CM trait levels and patterns in this population. By understanding how these modifiable and non-modifiable factors predict common complex traits we can further recognize how to effectively reduce and prevent CM risk and its associated diseases.

LIMITATIONS

Several limitations warrant mention. Migration of the more educated, urban segment of the original cohort has left us with a sample that is no longer representative of the population from which it was drawn.[89] Compared with those lost to follow-up, individuals who participated in the 2005 survey were less educated and came disproportionately from rural and poorer households. Given that permanent migrants from the Metro Cebu area were not followed, the remaining sample is therefore selective of households with more residential stability and lower SES.

Our study is cross-sectional since CM biomarkers were only measured in the 2005 CLHNS survey. Due to the nature of this study we cannot determine when CM risk first developed. This limits our ability to infer causality. Since the CLHNS was not originally designed with the study of CM risk in mind, some of the measures are adequate, but not ideal, for our research aims. For example, dietary intake was measured using two 24-hour recalls, which may not represent usual intake at the time of each survey. However, such dietary recalls can appropriately rank an individual's intake and thus accurately predict a variety of biological and health outcomes. Prior peer-reviewed publications utilized this diet recall data.[55, 116, 145]

Regarding cluster analysis, a limitation of assigning names to the clusters is that not all individuals within a certain cluster necessarily share the ascribed characteristics. For example, in our "Low CM risk variants" cluster we found the average Z-scores for genetic risk scores were relatively low, but we cannot ascribe these low values to each individual in the cluster.

In our literature search we found differing numbers of candidate SNPs for each trait. Although we used the same criteria in our search regardless of the CM trait, the variation in the number of candidate SNPs could reflect the current state of the literature. In addition, there is concern with choosing SNPs from a European sample and applying them to an Asian sample, especially in terms of tagging the appropriate functional variant. We tried to mitigate this by choosing SNPs with nominal significance and directional consistent effect estimates in our study population, however due to the limited sample size in the CLHNS we may have lacked the power to detect the SNPs true effect.

Using a threshold of $r^2<0.2$ for linkage disequilibrium, still allows SNPs to partially tag the same underlying signal, potentially including some redundancy in the genetic risk score. While we used an un-weighted approach to create our genetic risk scores, it may be possible in the future to obtain stable and accurate estimates of genetic variants for use in a weighted risk score, which could improve predictability of CM risk.

Lastly, our sample size is relatively small, especially in the scope of genetic epidemiology studies; therefore replication of our results in larger populations would reinforce our findings.

SIGNIFICANCE AND PUBLIC HEALTH IMPACT

This study addresses an important international public health issue: understanding the multifactorial etiology of CM diseases. Research demonstrates that CM risk factors tend to co-occur and may be causally interrelated.[23, 87] Furthermore, differences in the prevalence and patterns of co-occurrence of CM risk factors likely reflect variation in diet, lifestyle, and genetics. However, there is insufficient research on the interplay between these modifiable and non-modifiable factors and how they relate to CM risk patterns. In this study we used innovative methods to account for potential synergistic effects.

Here we utilize data from the CLHNS, which contains detailed diet, lifestyle, and genetic data; this unique data along with the rapid nutrition and lifestyle transition make the CLHNS an ideal dataset for our study aims**.** These findings could apply not only to Asians but to more modernized countries such as the U.S.

Despite the population's young age, lack of clinical disease, and relative leanness, cluster analysis identified distinct patterns of CM risk factors in Filipino young adults. By using cluster analysis we made fewer assumptions regarding the underlying etiology and allowed relationships among CM risk factors to emerge from the data themselves. As expected, measures of adiposity predicted specific CM risk patterns. However, diet and environmental factors also independently predicted risk factor clustering. This emphasizes the importance of screening both lean and overweight individuals for CM risk, especially in Asian populations where the risk of CM diseases is elevated at lower BMI.[92] Finding modifiable and non-modifiable predictors of CM risk in early adulthood could help inform targeted prevention efforts for future CM disease.

To further understand the etiology of CM risk in middle-aged Filipino women, for whom CM risk is more established, we used genetic risk scores, which combined the relatively small additive effects of individual single nucleotide polymorphisms (SNPs), to better capture the complex relationship between genetics, adiposity, and CM risk. Our results suggest that for individuals with a high genetic predisposition to elevated TG levels, reducing adiposity could possibly prevent increases in TG levels and thereby reduce the likelihood of adverse health outcomes such as CM disease. Replication of these results in larger study populations is needed to support the potential clinical and public health utility of targeted prevention efforts using genetic profiling.

Lastly, we used a novel application of cluster analysis to identify groups of these Filipino women who share similar patterns of genetic risk scores across multiple CM phenotypes. By finding combinations of genetic risk across multiple phenotypes, we can hopefully better explain the underlying genetic susceptibility for CM disease in this population, especially since these diseases themselves display a specific patterning of risk factors including insulin resistance, dyslipidemia, hypertension, and inflammation.[87, 88] These genetic risk clusters, along with anthropometric and dietary factors, predicted both CM trait levels and patterns in this population.

In summary, by capturing the intricate relationship of these modifiable and non-modifiable factors with common complex traits we can further understand how to effectively reduce and prevent CM risk and its associated diseases.

DIRECTIONS FOR FUTURE RESEARCH

The CLHNS provided us with the unique opportunity to examine how modifiable and non-modifiable factors predicted CM risk. However, the cross-sectional nature of this study limits us in our ability to infer causality. With the collection of CM biomarkers in future follow-up surveys, we could examine how CM risk patterns change longitudinally and whether predictors (both genetic and environmental) for these patterns change longitudinally as well. This could provide insight to how CM risk evolves across the life course. In addition, collecting medical records or detailed mortality data, could help us further understand whether these CM traits and patterns actually predict the development of disease.

We found that dietary factors, specifically decreased saturated fat intake, predicted CM risk; however research demonstrates that the quality of foods consumed may better predict CM risk than the quantity.[146, 147] Therefore in future work we could look at how dietary patterns or food groups predict CM risk in this population.

We used an un-weighted approach to create our genetic risk scores. However, it may be possible in the future, as more genetic studies in Asian populations get published, to obtain stable and accurate estimates of genetic variants for use in a weighted risk score. This weighted risk score could potentially improve predictability of CM risk.

If genetic expression data becomes available in the CLHNS, we can observe how expression changes over time and with environmental stimuli. By studying these epigenetic modifications, we can perhaps understand the triggers for CM disease progression. With this expression data we could also examine the effects of epistatic interactions on CM risk. Epistatic interactions can occur when two genes are mutated (genetic–genetic interaction), when one gene is mutated and the other gene varies in expression (genetic–epigenetic

interaction), or when two genes simultaneously vary in expression (epigenetic–epigenetic interaction).[148]

We found that for individuals with a higher TG genetic risk score, having elevated levels of adiposity amplified the genetic risk score's effect by further increasing TG levels, while for individuals with a lower TG genetic risk score, measures of adiposity made almost no difference in TG levels. Earlier research suggests that an intensive lifestyle intervention may mitigate the effect of the rs1260326 risk allele in the *GCKR* gene on higher TG levels.[139] However, replication of our results in larger studies is needed before further research (e.g. clinical trials) can examine whether weight-loss or lifestyle interventions could be useful for those with a genetic predisposition to elevated TG levels.

REFERENCES

1. Wu Y, McDade TW, Kuzawa CW, Borja J, Li Y, Adair LS, et al. Genome-wide association with C-reactive protein levels in CLHNS: evidence for the CRP and HNF1A loci and their interaction with exposure to a pathogenic environment. *Inflammation*. 2012;35:574-583

2. Kim YJ, Go MJ, Hu C, Hong CB, Kim YK, Lee JY, et al. Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits. *Nat Genet*. 2011;43:990-995

3. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42:105-116

4. Ridker PM, Wilson PW, Grundy SM. Should C-reactive protein be added to metabolic syndrome and to assessment of global cardiovascular risk? *Circulation*. 2004;109:2818-2825

5. Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, et al. Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am J Hum Genet*. 2008;82:1185-1192

6. Saxena R, Hivert M-F, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet*. 2010;42:142-148

7. Wu Y, Marvelle AF, Li J, Croteau-Chonka DC, Feranil AB, Le NA, et al. Genetic association of lipoprotein cholesterol and triglycerides in Filipinos: Waist circumference modifies the effects of APOA5 variants on triglyceride levels *In Preparation; See Supplemental Reference*. 2013

8. Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, Sim X, et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature genetics*. 2011;43:531-538

9. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713

10. Popkin BM. Urbanization, lifestyle changes and the nutrition transition. *World Development*. 1999;27:1905-1916

11. Drewnowski A, Popkin BM. The nutrition transition: new trends in the global diet. *Nutr Rev*. 1997;55:31-43

12. Janus ED, Postiglione A, Singh RB, Lewis B. The modernization of Asia: implications for coronary heart disease. *Circulation*. 1996;94:2671

13. Deurenberg Yap M, Chew SK, Deurenberg P. Elevated body fat percentage and cardiovascular risks at low body mass index levels among Singaporean Chinese, Malays and Indians. *Obesity Reviews*. 2002;3:209-215

14. Gu D, Reynolds K, Wu X, Chen J, Duan X, Reynolds RF, et al. Prevalence of the metabolic syndrome and overweight among adults in China. *The Lancet*. 2005;365:1398-1405

15. Li JKY, Ng MCY, So WY, Chiu CKP, Ozaki R, Tong PCY, et al. Phenotypic and genetic clustering of diabetes and metabolic syndrome in Chinese families with type 2 diabetes mellitus. *Diabetes/metabolism research and reviews*. 2006;22:46-52

16. Patel A, Huang KC, Janus ED, Gill T, Neal B, Suriyawongpaisal P, et al. Is a single definition of the metabolic syndrome appropriate?--A comparative study of the USA and Asia. *Atherosclerosis*. 2006;184:225-232

17. Nogueira A, Marcopito L, Lanas F, Galdames D, Jialiang W, Jing F, et al. Risk factors for cardiovascular disease in the developing world. A multicentre collaborative study in the International Clinical Epidemiology Network (INCLEN). *Journal of Clinical Epidemiology*. 1992;45:841-847

18. Nogueira A, Marcopito L, Lanas F, Galdames D, Jialiang W, Jing F, et al. Socio-economic status and risk factors for cardiovascular disease: A multicentre collaborative study in the International Clinical Epidemiology Network (INCLEN). *Journal of Clinical Epidemiology*. 1994;47:1401-1409

19. Philippine Nutrition Facts & Figures. 2001

20. Mortality Country Fact Sheet, Philippines. 2006

21. Mackay J, Mensah GA, Mendis S, Greenlund K. *The atlas of heart disease and stroke*. World Health Organization; 2004.

22. Morales DD, Punzalan FER, Paz-Pacheco E, Sy RG, Duante CA. Metabolic syndrome in the Philippine general population: prevalence and risk for atherosclerotic cardiovascular disease and diabetes mellitus. *Diabetes and Vascular Disease Research*. 2008;5:36

23. Reaven GM. Role of Insulin Resistance in Human Disease (Syndrome X): An Expanded Definition. *Annual Review of Medicine*. 1993;44:121-131

24. Das M, Pal S, Ghosh A. Factor analysis of risk variables associated with metabolic syndrome in adult Asian Indians. *J Cardiovasc Dis Res*. 2010;1:86-91

25. de Ferranti SD, Gauvreau K, Ludwig DS, Neufeld EJ, Newburger JW, Rifai N. Prevalence of the metabolic syndrome in American adolescents: findings from the Third National Health and Nutrition Examination Survey. *Circulation*. 2004;110:2494

26. McLaughlin T, Abbasi F, Cheal K, Chu J, Lamendola C, Reaven G. Use of Metabolic Markers To Identify Overweight Individuals Who Are Insulin Resistant. *Annals of internal medicine*. 2003;139:802-809

27. Bray GA, Bellanger T. Epidemiology, trends, and morbidities of obesity and the metabolic syndrome. *Endocrine*. 2006;29:109-117

28. Bray GA. Medical consequences of obesity. *Journal of Clinical Endocrinology & Metabolism*. 2004;89:2583

29. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. *JAMA: the journal of the American Medical Association*. 1999;282:1523

30. Kahn R, Buse J, Ferrannini E, Stern M, American Diabetes A, European Association for the Study of D. The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*. 2005;28:2289-2304

31. Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, et al. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA: the journal of the American Medical Association*. 2002;288:2709

32. Lehto S, Rˆnnemaa T, Pyˆr‰l‰ K, Laakso M. Cardiovascular risk factors clustering with endogenous hyperinsulinaemia predict death from coronary heart disease in patients with type II diabetes. *Diabetologia*. 2000;43:148-155

33. Wang J, Qiao Q, Miettinen ME, Lappalainen J, Hu G, Tuomilehto J. The metabolic syndrome defined by factor analysis and incident type 2 diabetes in a Chinese population with high postprandial glucose. *Diabetes Care*. 2004;27:2429

34. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure. *Hypertension*. 2003;42:1206-1252

35. Smith SC, Jackson R, Pearson TA, Fuster V, Yusuf S, Faergeman O, et al. Principles for national and regional guidelines on cardiovascular disease prevention. *Circulation*. 2004;109:3112-3121

36. Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, III, Criqui M, et al. Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: A Statement for Healthcare Professionals From the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*. 2003;107:499-511

37. Alberti KG, Zimmet P, Shaw J. Metabolic syndrome--a new world-wide definition. A Consensus Statement from the International Diabetes Federation. *Diabet Med*. 2006;23:469-480

38. Kuk JL, Ardern CI. Age and sex differences in the clustering of metabolic syndrome factors: association with mortality risk. *Diabetes Care*. 2010;33:2457-2461

39. Duncan GE, Li SM, Zhou XH. Prevalence and trends of a metabolic syndrome phenotype among US adolescents, 1999–2000. *Diabetes Care*. 2004;27:2438

40. Kim MH, Kim MK, Choi BY, Shin YJ. Prevalence of the metabolic syndrome and its association with cardiovascular diseases in Korea. *Journal of Korean Medical Science*. 2004;19:195

41. Goldman N, Weinstein M, Cornman J, Singer B, Seeman T, Chang MC. Sex differentials in biological risk factors for chronic disease: estimates from population-based surveys. *Journal of Women's Health*. 2004;13:393-403

42. Ridker PM, Wilson PWF, Grundy SM. Should C-Reactive Protein Be Added to Metabolic Syndrome and to Assessment of Global Cardiovascular Risk? *Circulation*. 2004;109:2818-2825

43. Ford ES. The metabolic syndrome and C-reactive protein, fibrinogen, and leukocyte count: findings from the Third National Health and Nutrition Examination Survey. *Atherosclerosis*. 2003;168:351-358

44. Wildman RP, Muntner P, Reynolds K, McGinn AP, Rajpathak S, Wylie-Rosett J, et al. The Obese Without Cardiometabolic Risk Factor Clustering and the Normal Weight With Cardiometabolic Risk Factor Clustering: Prevalence and Correlates of 2 Phenotypes Among the US Population (NHANES 1999-2004). *Archives of Internal Medicine*. 2008;168:1617-1624

45. Ramachandran A, Snehalatha C, Yamuna A, Murugesan N, Narayan KM. Insulin resistance and clustering of cardiometabolic risk factors in urban teenagers in southern India. *Diabetes Care*. 2007;30:1828-1833

46. Barker DJ. The fetal and infant origins of adult disease. *British Medical Journal*. 1990;301:1111

47. Yao M, Lichtenstein AH, Roberts SB, Ma G, Gao S, Tucker KL, et al. Relative influence of diet and physical activity on cardiovascular risk factors in urban Chinese adults. *International journal of obesity*. 2003;27:920-932

48. Glew RH, Williams M, Conn CA, Cadena SM, Crossey M, Okolo SN, et al. Cardiovascular disease risk factors and diet of Fulani pastoralists of northern Nigeria. *The American Journal of Clinical Nutrition*. 2001;74:730

49. Mensink RP, Zock PL, Kester AD, Katan MB. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am J Clin Nutr*. 2003;77:1146-1155

50. Rexrode KM, Pradhan A, Manson JE, Buring JE, Ridker PM. Relationship of total and abdominal adiposity with CRP and IL-6 in women. *Ann Epidemiol*. 2003;13:674-682

51. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451:990-993

52. McDade TW, Rutherford JN, Adair L, Kuzawa C. Adiposity and pathogen exposure predict C-reactive protein in Filipino women. *J Nutr*. 2008;138:2442-2447

53. WHO. The World Health Report 2004 --changing histroy. 2004

54. Rexrode KM, Carey VJ, Hennekens CH, Walters EE, Colditz GA, Stampfer MJ, et al. Abdominal adiposity and coronary heart disease in women. *JAMA*. 1998;280:1843-1848

55. Adair LS. Dramatic Rise in Overweight and Obesity in Adult Filipino Women and Risk of Hypertension. *Obesity*. 2004;12:1335-1341

56. Tanko LB, Bagger YZ, Qin G, Alexandersen P, Larsen PJ, Christiansen C. Enlarged waist combined with elevated triglycerides is a strong predictor of accelerated atherogenesis and related cardiovascular mortality in postmenopausal women. *Circulation*. 2005;111:1883-1890

57. Yusuf S, Hawken S, Ounpuu S, Bautista L, Franzosi MG, Commerford P, et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. *Lancet*. 2005;366:1640-1649

58. Popkin BM, Adair LS, Ng SW. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev*. 2012;70:3-21

59. Tudor-Locke C, Ainsworth BE, Adair LS, Popkin BM. Physical activity in Filipino youth: the Cebu Longitudinal Health and Nutrition Survey. *International Journal of Obesity*. 2003;27:181-190

60. World Health Organization. Global Strategy on Diet, Physical Activity and Health: Obesity and Overweight. 2004

61. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2224-2260

62. McKeigue PM, Shah B, Marmot MG. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. *Lancet*. 1991;337:382-386

63. Deurenberg-Yap M, Chew SK, Deurenberg P. Elevated body fat percentage and cardiovascular risks at low body mass index levels among Singaporean Chinese, Malays and Indians. *Obes Rev*. 2002;3:209-215

64. Sung RY, Yu CC, Choi KC, McManus A, Li AM, Xu SL, et al. Waist circumference and body mass index in Chinese children: cutoff values for predicting cardiovascular risk factors. *International journal of obesity (2005)*. 2007;31:550-558

65. Wang J, Thornton JC, Russell M, Burastero S, Heymsfield S, Pierson RN, Jr. Asians have lower body mass index (BMI) but higher percent body fat than do whites: comparisons of anthropometric measurements. *Am J Clin Nutr*. 1994;60:23-28

66. Adair LS, Gultiano S, Suchindran C. 20-Year Trends in Filipino Women's Weight Reflect Substantial Secular and Age Effects. *The Journal of Nutrition*. 2011

67. Rutherford JN, McDade TW, Feranil AB, Adair LS, Kuzawa CW. High prevalence of low HDL-c in the Philippines compared to the US: population differences in associations with diet and BMI. *Asia Pac J Clin Nutr*. 2010;19:57-67

68. Rutherford JN, McDade TW, Lee NR, Adair LS, Kuzawa C. Change in waist circumference over 11 years and current waist circumference independently predict elevated CRP in Filipino women. *Am J Hum Biol*. 2010;22:310-315

69. Zubair N, Kuzawa CW, McDade TW, Adair LS. Cluster analysis reveals important determinants of cardiometabolic risk patterns in Filipino women. *Asia Pac J Clin Nutr*. 2012;21:271-281

70. Satoh H, Fujii S, Furumoto T, Kishi R, Tsutsui H. Waist Circumference can Predict the Occurrence of Multiple Metabolic Risk Factors in Middle-aged Japanese Subjects. *Industrial health*. 2010;48:447-451

71. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42:105-116

72. Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nature genetics*. 2010;42:142-148

73. Sjögren M, Lyssenko V, Jonsson A, Berglund G, Nilsson P, Groop L, et al. The search for putative unifying genetic factors for components of the metabolic syndrome. *Diabetologia*. 2008;51:2242-2251-2251

74. Zabaneh D, Balding DJ. A Genome-Wide Association Study of the Metabolic Syndrome in Indian Asian Men. *PLoS ONE*. 2010;5:e11961

75. Benyamin B, Sørensen T, Schousboe K, Fenger M, Visscher P, Kyvik K. Are there common genetic and environmental factors behind the endophenotypes associated with the metabolic syndrome? *Diabetologia*. 2007;50:1880-1888-1888

76. Pollex R, Hegele R. Genetic determinants of the metabolic syndrome. *Nat Clin Pract Cardiovasc Med*. 2006;3:482 - 489

77. Tam CHT, Ma RCW, So WY, Wang Y, Lam VKL, Germer S, et al. Interaction Effect of Genetic Polymorphisms in Glucokinase (GCK) and Glucokinase Regulatory Protein (GCKR) on Metabolic Traits in Healthy Chinese Adults and Adolescents. *Diabetes*. 2009;58:765-769

78. Miyaki K, Oo T, Song Y, Lwin H, Tomita Y, Hoshino H, et al. Association of a Cyclin-Dependent Kinase 5 Regulatory Subunit-Associated Protein 1-Like 1 (CDKAL1) Polymorphism With Elevated Hemoglobin A1c Levels and the Prevalence of Metabolic Syndrome in Japanese Men: Interaction With Dietary Energy Intake. *American Journal of Epidemiology*. 2010;172:985-991

79. Manolio TA. Cohort studies and the genetics of complex disease. *Nat Genet*. 2009;41:5-6

80. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet*. 2009;41:47-55

81. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation*. 2011;123:731-738

82. Paynter NP, Chasman DI, Pare G, Buring JE, Cook NR, Miletich JP, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA*. 2010;303:631-637

83. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet*. 2010;376:1393-1400

84. Qi Q, Forman JP, Jensen MK, Flint A, Curhan GC, Rimm EB, et al. Genetic predisposition to high blood pressure associates with cardiovascular complications among patients with type 2 diabetes: two independent studies. *Diabetes*. 2012;61:3026-3032

85. Dasgupta A, Sun YV, Konig IR, Bailey-Wilson JE, Malley JD. Brief review of regression-based and machine learning methods in genetic epidemiology: the Genetic Analysis Workshop 17 experience. *Genet Epidemiol*. 2011;35 Suppl 1:S5-11

86. Sun YV. Multigenic modeling of complex disease by random forests. *Advances in genetics*. 2010;72:73-99

87. Kahn R, Buse J, Ferrannini E, Stern M, American Diabetes Association, European Association for the Study of Diabetes. The metabolic syndrome: time for a critical appraisal: joint statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care*. 2005;28:2289-2304

88. Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon RO, Criqui M, et al. Markers of Inflammation and Cardiovascular Disease: Application to Clinical and Public Health Practice: A Statement for Healthcare Professionals From the Centers for Disease Control and Prevention and the American Heart Association *Circulation*. 2003;107:499-511

89. Adair LS, Popkin BM, Akin JS, Guilkey DK, Gultiano S, Borja J, et al. Cohort profile: the Cebu longitudinal health and nutrition survey. *Int J Epidemiol*. 2011;40:619-625

90. Yoon KH, Lee JH, Kim JW, Cho JH, Choi YH, Ko SH, et al. Epidemic obesity and type 2 diabetes in Asia. *Lancet*. 2006;368:1681-1688

91. Popkin BM. The Nutrition Transition and Obesity in the Developing World. *J Nutr*. 2001;131:871S-873S

92. Choo V. WHO reassesses appropriate body-mass index for Asian populations. *Lancet*. 2002;360:235

93. Singh A, Mulder C, Twisk JWR, Van Mechelen W, Chinapaw MJM. Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obes Rev*. 2008;9:474-488

94. Guo SS, Huang C, Maynard LM, Demerath E, Towne B, Chumlea WC, et al. Body mass index during childhood, adolescence and young adulthood in relation to adult overweight and adiposity: the Fels Longitudinal Study. *Int J Obes Relat Metab Disord*. 2000;24:1628-1635

95. Zimmet P, Alberti G, Kaufman F, Tajima N, Silink M, Arslanian S, et al. The metabolic syndrome in children and adolescents. *Lancet*. 2007;369:2059

96. Alberti KG, Zimmet P, Shaw J, IDF Epidemiology Task Force Consensus Group. The metabolic syndrome--a new worldwide definition. *Lancet*. 2005;366:1059-1062

97. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. *JAMA*. 1999;282:1523

98. Lakka HM, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, et al. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA*. 2002;288:2709

99. Lehto S, Laakso M. Cardiovascular risk factors clustering with endogenous hyperinsulinaemia predict death from coronary heart disease in patients with type II diabetes. *Diabetologia*. 2000;43:148-155

100. de Roos NM, Bots ML, Katan MB. Replacement of dietary saturated fatty acids by trans fatty acids lowers serum HDL cholesterol and impairs endothelial function in healthy men and women. *Arterioscler Thromb Vasc Biol*. 2001;21:1233-1237

101. Manley SE, Stratton IM, Clark PM, Luzio SD. Comparison of 11 Human Insulin Assays: Implications for Clinical Investigation and Research. *Clinical Chemistry*. 2007;53:922-932

102. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*. 2002;106:3143-3421

103. Stern SE, Williams K, Ferrannini E, DeFronzo RA, Bogardus C, Stern MP. Identification of individuals with insulin resistance using routine clinical measurements. *Diabetes*. 2005;54:333-339

104. Lohman TG, Roche AF, Martorell R. *Anthropometric Standardization Reference Manual*. Human Kinetics Books Champaign; 1988.

105. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet*. 2004;363:157-163

106. FNRI. Food Composition Tables Recommended for Use in the Philippines. 1997

107. Dahly DL, Adair LS. Quantifying the urban environment: a scale measure of urbanicity outperforms the urban-rural dichotomy. *Soc Sci Med*. 2007;64:1407-1419

108. Aldenderfer M, Blashfield R. *Cluster Analysis*. Newbury Park, CA: SAGE Publications, Inc.; 1984.

109. SAS 9.2 Documentation. 2010

110. Boone-Heinonen J, Gordon-Larsen P, Adair L. Obesogenic Clusters: Multidimensional Adolescent Obesity-related Behaviors in the U.S. *Annals of Behavioral Medicine*. 2008;36:217-230

111. Willett W. *Nutritional Epidemiology*. Oxford University Press, USA; 1998.

112. Boone-Heinonen J, Gordon-Larsen P, Adair LS. Obesogenic clusters: multidimensional adolescent obesity-related behaviors in the U.S. *Ann Behav Med*. 2008;36:217-230

113. Morales DD, Punzalan FE, Paz-Pacheco E, Sy RG, Duante CA. Metabolic syndrome in the Philippine general population: prevalence and risk for atherosclerotic cardiovascular disease and diabetes mellitus. *Diab Vasc Dis Res*. 2008;5:36-43

114. Huxley RR, Barzi F, Lam TH, Czernichow S, Fang X, Welborn T, et al. Isolated low levels of high-density lipoprotein cholesterol are associated with an increased risk of coronary heart disease: an individual participant data meta-analysis of 23 studies in the Asia-Pacific region. *Circulation*. 2011;124:2056-2064

115. Feranil AB, Duazo PL, Kuzawa CW, Adair LS. Coconut oil is associated with a beneficial lipid profile in pre-menopausal women in the Philippines. *Asia Pac J Clin Nutr*. 2011;20:190-195

116. Kuzawa CW, Adair LS, Avila JL, Cadungog JH, Le NA. Atherogenic lipid profiles in Filipino adolescents with low body mass index and low dietary fat intake. *Am J Hum Biol*. 2003;15:688-696

117. Voon PT, Ng TK, Lee VK, Nesaretnam K. Diets high in palmitic acid (16:0), lauric and myristic acids (12:0 + 14:0), or oleic acid (18:1) do not alter postprandial or fasting plasma homocysteine and inflammatory markers in healthy Malaysian adults. *Am J Clin Nutr*. 2011;94:1451-1457

118. Tinker LF, Bonds DE, Margolis KL, Manson JE, Howard BV, Larson J, et al. Low-fat dietary pattern and risk of treated diabetes mellitus in postmenopausal women: the Women's Health Initiative randomized controlled dietary modification trial. *Arch Intern Med*. 2008;168:1500-1511

119. Micha R, Mozaffarian D. Saturated fat and cardiometabolic risk factors, coronary heart disease, stroke, and diabetes: a fresh look at the evidence. *Lipids*. 2010;45:893-905

120. Olefsky JM, Glass CK. Macrophages, inflammation, and insulin resistance. *Annu Rev Physiol*. 2010;72:219-246

121. Araneta MRG, Barrett-Connor E. Ethnic Differences in Visceral Adipose Tissue and Type 2 Diabetes: Filipino, African-American, and White Women. *Obesity*. 2005;13:1458-1465

122. Curocichin G, Wu Y, McDade TW, Kuzawa CW, Borja JB, Qin L, et al. Single-nucleotide polymorphisms at five loci are associated with C-reactive protein levels in a cohort of Filipino young adults. *Journal of Human Genetics*. 2011

123. Dumitrescu L, Goodloe R, Brown-Gentry K, Mayo P, Allen M, Jin H, et al. Serum vitamins A and E as modifiers of lipid trait genetics in the National Health and Nutrition Examination Surveys as part of the Population Architecture using Genomics and Epidemiology (PAGE) study. *Hum Genet*. 2012

124. Wang X, Ding X, Su S, Spector TD, Mangino M, Iliadou A, et al. Heritability of insulin sensitivity and lipid profile depend on BMI: evidence for gene-obesity interaction. *Diabetologia*. 2009;52:2578-2584

125. Surakka I, Isaacs A, Karssen LC, Laurila PP, Middelberg RP, Tikkanen E, et al. A Genome-Wide Screen for Interactions Reveals a New Locus on 4p15 Modifying the Effect of Waist-to-Hip Ratio on Total Cholesterol. *PLoS Genet*. 2011;7:e1002333

126. Lamina C, Forer L, Schonherr S, Kollerits B, Ried JS, Gieger C, et al. Evaluation of gene-obesity interaction effects on cholesterol levels: a genetic predisposition score on HDL-cholesterol is modified by obesity. *Atherosclerosis*. 2012;225:363-369

127. Li J, Huang A, Hu Y, Chen D. Association of the lipoprotein lipase gene T+495G polymorphism with central obesity and serum lipids in a twin study. *Ann Epidemiol*. 2008;18:760-767

128. Senti M, Bosch M, Aubo C, Elosua R, Masia R, Marrugat J. Relationship of abdominal adiposity and dyslipemic status in women with a common mutation in the lipoprotein lipase gene. The REGICOR investigators. *Atherosclerosis*. 2000;150:135-141

129. Eiriksdottir G, Smith AV, Aspelund T, Hafsteinsdottir SH, Olafsdottir E, Launer LJ, et al. The interaction of adiposity with the CRP gene affects CRP levels: age, gene/environment susceptibilty-Reykjavik study. *International journal of obesity (2005)*. 2009;33:267-272

130. Lange LA, Croteau-Chonka DC, Marvelle AF, Qin L, Gaulton KJ, Kuzawa CW, et al. Genome-wide association study of homocysteine levels in Filipinos provides evidence for CPS1 in women and a stronger MTHFR effect in young adults. *Hum Mol Genet*. 2010;19:2050-2058

131. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34:816-834

132. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007;81:559-575

133. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature*. 2006;38:904-909

134. Nazare JA, Smith JD, Borel AL, Haffner SM, Balkau B, Ross R, et al. Ethnic influences on the relations between abdominal subcutaneous and visceral adiposity, liver

fat, and cardiometabolic risk profile: the International Study of Prediction of Intra-Abdominal Adiposity and Its Relationship With Cardiometabolic Risk/Intra-Abdominal Adiposity. *Am J Clin Nutr*. 2012;96:714-726

135. Fox CS, Massaro JM, Hoffmann U, Pou KM, Maurovich-Horvat P, Liu CY, et al. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. *Circulation*. 2007;116:39-48

136. Sam S, Haffner S, Davidson MH, D'Agostino RB, Sr., Feinstein S, Kondos G, et al. Hypertriglyceridemic waist phenotype predicts increased visceral fat in subjects with type 2 diabetes. *Diabetes care*. 2009;32:1916-1920

137. Lemieux I, Pascot A, Couillard C, Lamarche B, Tchernof A, Almeras N, et al. Hypertriglyceridemic waist: A marker of the atherogenic metabolic triad (hyperinsulinemia; hyperapolipoprotein B; small, dense LDL) in men? *Circulation*. 2000;102:179-184

138. Bos G, Dekker JM, Heine RJ, Hoom s. Non-HDL cholesterol contributes to the "hypertriglyceridemic waist" as a cardiovascular risk factor: the Hoorn study. *Diabetes care*. 2004;27:283-284

139. Pollin TI, Jablonski KA, McAteer JB, Saxena R, Kathiresan S, Kahn SE, et al. Triglyceride response to an intensive lifestyle intervention is enhanced in carriers of the GCKR Pro446Leu polymorphism. *J Clin Endocrinol Metab*. 2011;96:E1142-1147

140. Kumar CPTG, Sng CPTBL, Kumar COLS. Correlation of capillary and venous blood glucometry with laboratory determination *Prehospital Emergency Care*. 2004;8:378-383

141. Petersen JR, Graves DF, Tacker DH, Okorodudu AO, Mohammad AA, Cardenas Jr VJ. Comparison of POCT and central laboratory blood glucose results using arterial, capillary, and venous samples from MICU patients on a tight glycemic protocol. *Clinica Chimica Acta*. 2008;396:10-13

142. Montague CT, O'Rahilly S. The perils of portliness: causes and consequences of visceral adiposity. *Diabetes*. 2000;49:883-888

143. Janssen I, Heymsfield SB, Allison DB, Kotler DP, Ross R. Body mass index and waist circumference independently contribute to the prediction of nonabdominal, abdominal subcutaneous, and visceral fat. *The American Journal of Clinical Nutrition*. 2002;75:683-688

144. Feranil A, Duazo PL, Kuzawa CW, Adair LS. Coconut oil predicts a beneficial lipid profile in pre-menopausal women in the Philippines. *Asia Pac J Clin Nutr*. *In press*

145. Adair LS, Popkin BM. Are Child Eating Patterns Being Transformed Globally? . *Obesity*. 2005;13:1281-1299

146. Mendez MA, Popkin BM, Jakszyn P, Berenguer A, Tormo MJ, Sanchez MJ, et al. Adherence to a Mediterranean diet is associated with reduced 3-year incidence of obesity. *J Nutr*. 2006;136:2934-2938

147. Mozaffarian D, Hao T, Rimm EB, Willett WC, Hu FB. Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men. *New England Journal of Medicine*. 2011;364:2392-2404

148. Park S, Lehner B. Epigenetic epistatic interactions constrain the evolution of gene expression. *Molecular systems biology*. 2013;9:645