FOLATE METABOLISM GENETIC VARIATION AND HEART DISEASE RISK IN HIV+ MEN UNDERGOING ANTIRETROVIRAL THERAPY

by

Anna Mamo

BS, Grove City College, 2003

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Anna Jean Mamo

It was defended on

April 22, 2010

and approved by

Thesis Advisor:

JeremyMartinson, DPhil

Assistant Professor

Department of Infectious Diseases and Microbiology and Human Genetics

Graduate School of Public Health

University of Pittsburgh

Committee Member:

Candace Kammerer, Ph.D.

Associate Professor

Department of Human Genetics

Graduate School of Public Health

University of Pittsburgh

Committee Member:

Todd Reinhart, Sc.D.

Professor and Assistant Chairman

Department of Infectious Diseases and Microbiology

Graduate School of Public Health

University of Pittsburgh

Jeremy Martinson, DPhil.

FOLATE METABOLISM GENETIC VARIATION AND HEART DISEASE RISK IN HIV+ MEN UNDERGOING ANTIRETROVIRAL THERAPY

Anna J. Mamo, M.S.

University of Pittsburgh, 2010

Background - The Martinson Lab is examining genetic characteristics related to cardiovascular disease (CVD) in men with HIV undergoing HAART therapy that are enrolled in the Multicenter Aids Cohort Study (MACS). CVD is a major side effect of HIV infection and HAART therapy. While the mechanism behind this remains unknown, the folic acid metabolic pathway may be involved. This study examines genes that encode enzymes involved in this pathway. Polymorphisms in these genes may lead to increased risk of CVD due to altered function of the enzymes they encode. The following polymorphisms in the MTHFR, MS, and MTRR genes have been found to affect enzymatic function of Methylenetetrahydrofolate reductase (MTHFR), Methionine Synthase (MS or MTR), and Methionine Synthase Reductase (MTRR) respectively: MTHFR C677T, MTHFR C1068T, MS A2756G and MTRR A66G. SNP genotypes for these loci were characterized for the MACS DNA samples. These results were compared with corresponding clinical data and statistics were used to determine how polymorphisms affect cardiovascular disease when influenced by HIV infection and HAART therapy. **Methods** – MACS DNA samples were amplified using PCR, and each SNP was characterized using a Fluorescence Polarization Assay (FP). These data and clinical data for LDL, HDL, triglyceride, BMI, HIV status, HAART status, age, gender, and family history were analyzed using box-and-whisker diagrams, Kruskal-Wallis test, odds ratio calculations, and logistic regression analysis. **Results** – Visual trends were seen between LDL levels and polymorphisms in MTHFR C1068T and MS A2756G. However, no significant associations were found statistically between SNP

genotypes and LDL levels. A protective association may exist between the MS A2756G GG genotype and not-high LDL levels, but the small sample size of this genotype means that statistical significance was not reached. We intend to obtain data for more samples and repeat statistical calculations. This may lead to statistically significant outcomes. This thesis contributes to public health by furthering knowledge of how an individual's genetics influences CVD risk while being HIV+ and undergoing HAART therapy. This knowledge enables the patient to be given the best care possible with regards to their individual situations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

"Small minds discuss people, average minds discuss events, and great minds discuss ideas."

   -H.G. Rickover

I would like to thank the following people for helping me with my most significant idea yet:

Dr. Jeremy Martinson

 He has been my advisor, committee member, and the originator of this idea, which I have spent the past 2 years exploring. His countless hours of help, patient teaching, and good humor have been appreciated, beyond words.

My committee members

 Dr. Todd Reinhart and Dr. Candace Kammerer have been dedicated committee members, whose academic insights have enabled this thesis to reach its full potential.

My lab

I would like to thank Matt Nicholaou, Becky Bosko, and Eden Huang for all of their help, insight, and between-experiment conversations that helped to pass the long hours in lab enjoyably.

All GSPH faculty and students

I would like to thank each of them for their unique contributions to my graduate school education.

My Parents

From the very beginning, they have always stressed ideas. Without them, I certainly would not have made it this far. I cannot begin to thank them for all they've done for me.

My sister and fiancé

They have been the best cheerleaders I could ask for. Their constant support has been very much appreciated.

1.0   INTRODUCTION

1.1   THE PROBLEM

For many years, an association has been observed between some HIV-infected individuals undergoing highly active anti-retroviral therapy (HAART) and the risk of cardiovascular disease [1]. This risk is characterized by increased prevalence of a high risk atherogenic phenotype. An atherogenic phenotype is illustrated by abnormally high levels of triglycerides and low density lipoproteins (LDLs), and abnormally low levels of high density lipoproteins (HDL). Premature atherosclerosis has been observed to result in a subset of individuals with the high risk atherogenic phenotype, as well as insulin resistance, central fat accumulation, and peripheral fat wasting [2, 3]. These associations generate concern for the long term ramifications of HAART therapy on various individuals. We think that genetic factors may be involved in the development of a high risk atherogenic phenotype. If genetics determines the amount of risk a certain individual has of developing cardiovascular disease, they can be tested when HAART is initiated and carefully monitored so that the disease does not develop [4].

1.2   THE IMPACT OF HIV ON LIPIDS

HIV is a virus that wreaks havoc on the lipid metabolism of its host. HIV positive individuals not undergoing therapy have a higher prevalence of metabolic diseases, such as hypertriglyceridaemia, impaired glucose tolerance, and high plasma triglyceride concentration, than those who are not infected with HIV [5]. Highly active anti-retroviral therapy (HAART) has been shown to further exacerbate these conditions in individuals where they are already present [6].

The complex interactions between the HIV virus, HAART, and environmental and genetic factors from the host may all play an integral part in the development of cardiovascular disease [6]. The HIV virus relies on the presence of cholesterol-dense structures known as lipid rafts in the host cell membrane in order to negotiate its entry from the extra cellular matrix into the cell [7]. The exact mechanisms are not known, but it is hypothesized that HIV may manipulate the body's lipids to enable cell entry. One proposed theory for HAART's contribution to cardiovascular disease is that long term effects of HAART therapy include endothelial oxidative stress, which causes attraction and activation of mononuclear cells. The congestion of these cells in the arteries contributes to atherosclerosis, and thus cardiovascular disease [8]. We think that genetic and environmental factors of the host, which influence development of cardiovascular disease in the non-HIV positive population, also contribute to cardiovascular disease in the HIV positive population.

1.3   HAART THERAPY

HAART typically involves prescribing a combination of three antiretroviral drugs that function in reducing plasma levels of HIV RNA. These drugs fall into the three main categories of nucleoside reverse transcriptase inhibitors (NRTIs), non nucleoside reverse transcriptase inhibitors (NNRTIs), and protease inhibitors (PIs). These drugs inhibit replication of HIV by competitive inhibition, non-competitive inhibition, and enzymatic inhibition, respectively. NRTIs are analogs of adenine, cytosine, guanine, and thymine that compete with normal nucleotides for access to binding sites on the HIV RNA template and thus cause defective RNA to be produced by HIV reverse transcriptase. NNRTIs are non-competitive inhibitors, and they are designed to bind to a site on the reverse transcriptase enzyme that is not directly involved in template binding, but which also inhibits HIV replication from occurring. Protease inhibitors deactivate the HIV enzyme protease, which functions in processing the new strands of HIV polyprotein so that they can form new virus particles and increase plasma RNA levels [9]. The overall effects of HAART therapy are to decrease viral load in the blood, semen, vaginal secretions, and to increase the number of CD4 T cells in the body. Since these effects contribute to maintaining a stronger immune system, HAART decreases the amount of opportunistic infections suffered by HIV positive individuals [10].

The Office of AIDS Research Advisory Council suggests that HAART therapy should be administered to any HIV+ individual with a CD4 T cell count of 350 cells/mm^3 or less [11]. Once an individual begins taking the therapy, it is essential that the medication is taken on time as prescribed. If timelines are disregarded or a dose is skipped, resistance to the drugs may develop. If this occurs, stronger medications must be prescribed. Some medications cause side effects such as myalgia, rash, nausea, vomiting, and diarrhea, among others. Resistance to

milder drugs should be avoided at all costs by taking medications as prescribed. It is not ideal to switch a patient to a stronger drug with more severe side effects if it is not necessary [11].

1.4   CARDIOVASCULAR DISEASE

As previously mentioned, Cardiovascular Disease (CVD) is a common side effect of HAART. In the HIV negative United States population, CVD is the most common cause of death independently in each state, and is also responsible for 40% of total deaths, nationwide [12]. It is characterized by heart attack, angina, heart failure and arrhythmias. These symptoms are caused by atherosclerosis, or hardening and narrowing of the blood vessels due to blockages of lipids and cellular debris that prevent sufficient amounts of blood from getting to the heart and brain [13]. Since CVD is such a major concern in the general HIV negative population, studies have been performed in attempts to identify genes and environmental factors that affect the pathogenesis of the disease. The folic acid metabolic pathway and specific genes may influence CVD susceptibility in this pathway [14].
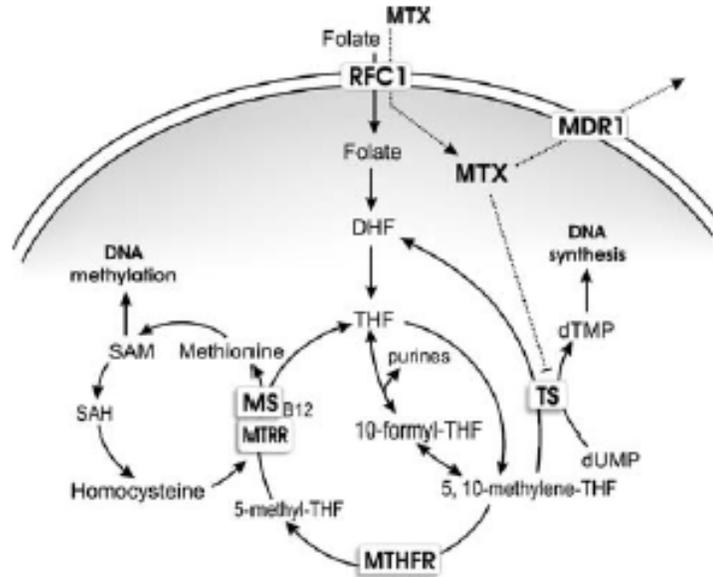
Figure 1. Folate Metabolic Pathway: essential genes and enzymes involved in the metabolic pathway

[15]

1.5   FOLIC ACID METABOLIC PATHWAY

The folic acid metabolic pathway results in two main processes: DNA methylation and purine production for DNA synthesis (Figure 1). The DNA methylation process has been our main focus due to the association that has been shown between inadequate DNA methylation and risk of cardiovascular disease [14], although the precise relationship between folate metabolism and cardiovascular disease remains unclear. As shown in Figure 1, methyl groups are produced through the enzymatic metabolism of the amino acid homocysteine. These methyl groups are used to methylate the cytosine-5 region of genomic DNA, which is essential for DNA stability, chromatin structure, and in controlling transcription, or the "turning on and off" of genes [14]. The following enzymes, named after the gene from which they are produced, catalyze the DNA

5

methylation function of the folic acid metabolic pathway and will be discussed later: Methylenetetrahydrofolate reductase (MTHFR), Methionine Synthase (MS), and Methionine Synthase Reductase (MTRR).

A second essential function of the methylation process of the folic acid metabolic pathway is the conversion of homocysteine into methionine. This step is very important, as excess levels of homocysteine are toxic to the body and have been associated with vascular complications [16, 17]. A possible explanation, called the homocysteine-thiolactone hypothesis, has been proposed to explain why this occurs. This proposal suggests that genetic polymorphisms or insufficient concentrations of folate, vitamin B6, or vitamin B12, can cause excess levels of homocysteine to be present in the body. These factors disrupt the folic acid metabolic pathway shown in Figure 1 and the homocysteine is, instead, directed towards an alternate biochemical pathway where it is converted into homocysteine-thiolactone, which is then converted into an N-homocysteine protein. This protein elicits an autoimmune response, cell death, and inflammation, which in turn, directly cause atherosclerosis and cardiovascular problems [17]. The folate metabolic pathway is very complex, however, and it is not known which of these variables directly affects cardiovascular disease. We plan to analyze genetic factors that are involved in different parts of the pathway, hoping that it will help reveal the precise factors that contribute to cardiovascular disease.

1.6   SINGLE NUCLEOTIDE POLYMORPHISMS

Single nucleotide polymorphisms (SNPs) are genetic variations in a gene involving a single nucleotide change at a particular location on the gene sequence. Typically there are two

variations. The more common variant is called the major allele, and the least common variant is called the minor allele [18]. SNP nomenclature involves a number encased by two uppercase letters. The number is typically the nucleotide's position within the gene. The first letter indicates the major allele in the population of the SNP discovery and the second letter indicates the minor allele. In order for a single-base mutation to be classified as a SNP, the minor allele must occur in at least 1% of the population [18]. SNPs in a gene can have negative ramifications. If a certain variation occurs during transcription that causes a protein's translation to be terminated early, the protein will be defective and will not be able to fulfill its biochemical purpose. For example, a mid-protein amino acid could undergo a SNP mutation in transcription that changes it into a stop codon, therefore terminating translation of the protein prematurely. It is thought that the SNPs in this study reduce the functional ability of the enzymes previously mentioned, therefore causing a higher risk of CVD. The following SNPs were genotyped: MTHFR C677T, MTHFR C1068T, MS A2756G, and MTRR A66G. These SNPs were chosen because the literature identifies them as being SNPs associated with high risk of diseases resulting from folate metabolism dysfunction [19].

1.6.1 MTHFR

The gene MTHFR, encoding the enzyme Methylenetetrahydrofolate reductase, is located on chromosome 1, and is 19,322 base pairs long. The enzyme product specifically reduces the molecule 5,10-methylenetetrahydrofolate (methylene-THF) to 5-methyltetrahydrofolate (methyl-THF) [15, 20]. The product of this reaction is a methyl group which is necessary for the conversion of homocysteine to methionine. It also contributes to the production of purines for DNA synthesis in conjunction with the enzymes MS and MTRR [20]. It has been hypothesized

that in the presence of certain SNPs, dysfunctional enzyme is produced and homocysteine is therefore inefficiently detoxified. Consequently, the risk of cardiovascular disease, neural tube defects, colon cancer and acute leukemia increases [21]. Specifically, the SNP C677T has been studied in linking elevated levels of homocysteine to coronary artery disease and is one of the SNPs that are the focus of our study [22]. In previous studies, the major allele was C, and was observed ~56% of the time in a control population [23]. The C→T mutation causes an alanine to valine change, which results in a thermolabile product with reduced activity [24]. This could be a factor in the reduced efficacy of homocysteine detoxification. Another commonly studied SNP on MTHFR which we will study is C1068T. In previous studies, the major allele was C, and was observed ~86% of the time in a control population [25]. The C→T mutation is a silent change in the third base of the codon for Ser353. This causes the loss of a Hha1 restriction site and could be related to the previously mentioned risks involved with SNPs in MTHFR [25]. A diagram of the MTHFR gene is shown in figure 2.
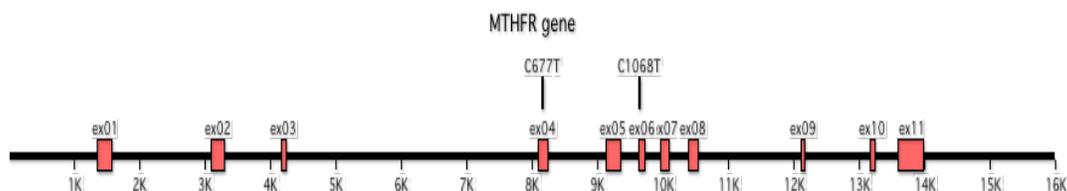
Figure 2. Diagram of the MTHFR gene: exon and SNP location (C677T and C1068T)

1.6.2 MS

The MS gene, which encodes the enzyme methionine synthase, is also on chromosome 1, and is 105,244 base pairs long. The enzyme is sometimes also referred to as 5-methyltetrahydrofolate-homocysteine methyltransferase, or MTR [26]. It catalyzes the final step

in methionine synthesis by producing enzyme that, in conjunction with the enzyme Methionine Synthase Reductase, or MTRR, catalyzes the methylene-THF to 5-methyl-THF reaction. Mutations in this gene have been shown to be associated with susceptibility to cleft lip/palate, Down syndrome, methylcobalamin deficiency, and neural tube defects [27]. A common SNP on this gene that has been studied, which we will also examine, is A2756G. In previous studies, the major allele was A, and was observed in ~81% of a control population [23]. The A→G mutation results in a glycine to aspartic acid substitution and could adversely affect the final step in methionine synthesis, leaving the body with higher levels of toxic homocysteine [28]. Figure 3 shows a diagram of the MS gene structure.
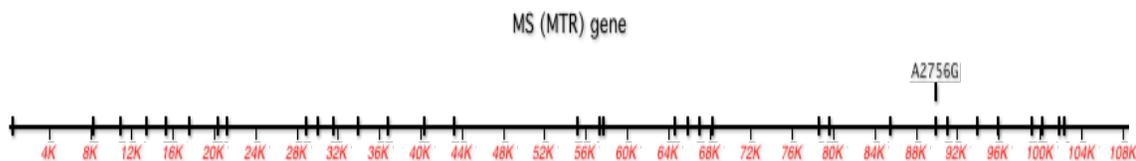
Figure 3. Diagram of the MS (MTR) gene: exon and SNP location (A2756G)

1.6.3   MTRR

The MTRR gene, which encodes the enzyme Methionine Synthase Reductase, is located on chromosome 5, and is 32,020 base pairs long. The MTRR enzyme is responsible for producing enzyme that helps to regenerate Methionine Synthase [29]. This allows the process of detoxification of homocysteine to occur. A common SNP on this gene that has been studied for its association with coronary artery disease is A66G [30]. In previous studies, the major allele was A, and was observed ~51% of the time in a control population [23]. The A→G mutation results in an isoleucine to methionine substitution, which could potentially cause the previously

mentioned conditions [31].   Figure 4 shows a diagram of the MTRR gene structure.
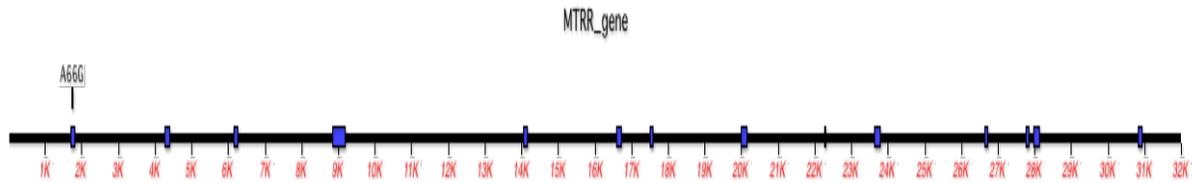
Figure 4. Diagram of the MTRR gene: exon and SNP location (A66G)

1.7   MACS COHORT STUDY

The Multi Center AIDS Cohort Study (MACS) is an ongoing multi-location study, initiated in 1984, of the natural and treated histories of HIV-1 infection in homosexual and bisexual men. The locations participating in the study include Baltimore, Chicago, Los Angeles, and Pittsburgh. The study includes men who are HIV-seronegative, HIV-infected men who are not on HAART, and HIV-infected men who are on HAART. The criteria for our investigation are that men must currently be enrolled in the MACS and they must have given a blood sample in the past year [1, 32]. These criteria are important because classifying men into high risk and low risk atherogenic phenotypes requires that clinical data and cholesterol data be collected from men enrolled in the study. The following clinical information was obtained from the MACS participants, and were considered as variables during analysis: age, family history of CVD, race, fasting glucose, HbA1c, total cholesterol (TC), low-density lipoprotein-cholesterol (LDL), high-density lipoprotein-cholesterol (HDL), triglycerides (TG), Apolipoprotein A, Apolipoprotein B, insulin, body mass index (BMI), HIV status, CDC-AIDS classification, date of AIDS onset, duration of HAART, and use of cholesterol-lowering drugs such as statins.

1.8   ATHEROGENIC PHENOTYPE

The term "atherogenic phenotype" refers to observable blood serum characteristics describing the likelihood of developing arterial plaques and susceptibility to cardiovascular disease. It is customary to categorize people as having a high risk or low risk atherogenic phenotype, which are sometimes also referred to as phenotype B and phenotype A, respectively [33]. A high risk atherogenic phenotype is illustrated through the presence of high levels of low density lipoprotein (LDL), high levels of triglycerides, and low levels of high density lipoprotein (HDL). This phenotype has been associated in past studies with cardiovascular disease [2]. The low risk atherogenic phenotype is just the opposite, and can be illustrated through the presence of low levels of low density LDL, lower levels of triglycerides, and higher levels of HDL [33]. LDL levels that are considered "high" are typically classified as above 160mg/dL and increase the risk of cardiovascular disease. LDL levels that are less than 100mg/dL are considered to have a protective effect against cardiovascular disease. HDL levels that are less than 50mg/dL are considered "low" and are a risk factor for cardiovascular disease. HDL levels that are above 60mg/dL are considered "high" and are protective against cardiovascular disease. Triglyceride levels are considered high when they are over 500mg/dL and are risk factors while levels under 150mg/dL are considered normal and are protective factors against cardiovascular disease [34].

Individuals infected with HIV have historically been reported to experience problems with lipid metabolism. Past studies have reported hypocholesterolemia and hypertriglyeridemia in HIV+ individuals, as well as peripheral fat wasting (lipodystrophy), central fat accumulation, and insulin resistance [3]. HAART therapy was developed to reduce viral load in the blood, but has appeared to take part in causing these side effects that lead to cardiovascular disease [4]. It is

thought that these side effects are more severe amongst those with high risk atherogenic phenotypes [35].

1.9   HYPOTHESIS AND AIMS

We hypothesize that Infection with HIV and HAART therapy exacerbates underlying genetic predispositions to cardiovascular disease. The folate metabolism pathway may be an important component of this. We intend to investigate this hypothesis by accomplishing the following two aims:

1) Genetic characterization of members enrolled in the MACS.

2) Statistical determination of whether a certain SNP characterization correlates with presence or absence of a high/low risk atherogenic phenotype.

2.0   METHODS AND MATERIALS

2.1   STUDY OVERVIEW

The object of this study was to generate and analyze folate metabolism genotype data for the MACS participants against their clinical atherogenic phenotype data to see if there was an association between having a high risk atherogenic phenotype and developing cardiovascular disease while undergoing HAART therapy. In order to carry this out, genotype data was generated for each of the previously mentioned SNPs on MACS DNA samples. After generating genotype data, analyses were performed through generation of box-and-whisker diagrams, the Kruskal-Wallis Test, the odds ratio test, and logistic regression analysis.

2.2   GENERATING GENOTYPE DATA

This was performed using a four step process that shall be referred to as a fluorescence polarization assay. The four steps included PCR amplification, agarose gel verification, primer/dNTP degradation, and fluorescence polarization. The DNA utilized in these steps had previously been extracted from blood samples donated by the MACS participants. After extraction, they were diluted to levels of 5ng/μL by members of the Martinson lab.

2.2.1 PCR AMPLIFICATION

Polymerase chain reaction amplification (PCR) is an *in vitro* process performed to amplify a desired segment of DNA. PCR is performed by heating a reaction mixture containing many reagents, including DNA, until the DNA denatures and the two strands separate. The mixture is then cooled and primers, which are small complementary DNA sequences to the segment to be amplified, anneal to the edges of the parent DNA strand. The mixture also contains deoxyribonucleoside triphosphates (dNTPs) in excess, which are added onto the primer strand by a thermostable DNA polymerase. We used Taq polymerase. The DNA polymerase functions in adding the dNTPs downstream of the primer in a complementary manner to the parent DNA strand. At the completion of this event, there are two brand new DNA strands annealed to the two parent DNA strands. These steps are repeated and the DNA segment is amplified exponentially every cycle. The reaction mixture also contains sterile deionized water (SDW), a buffer, and magnesium chloride to optimize reaction conditions. Table 1 shows the exact amounts of reagents added.

The 10X buffer, magnesium chloride, dNTPs, and Taq polymerase were all purchased from Sigma. The dNTPs came in individual tubes of 2'-deoxyguanosine 5'-triphosphate trisodium, thymidine 5'-triphosphate sodium salt solution, 2'-deoxyadenosine 5'-triphosphate sodium salt solution, and 2'-deoxycytidine 5'-triphosphate disodium solution, and were mixed together in a 1:1:1:1 ratio to make the dNTP mixture.

Table 1. PCR Reagents

Reagent	Mixture for 1 well	Mixture for 96-well plate
DNA	1μL	96 μL
10X Buffer	1 μL	100 μL
MgCl2	1 μL	100 μL
dNTPs	0.08 μL	8 μL
Primer Mixture	0.1 μL	10 μL
Taq Polymerase	0.04 μL	4 μL
SDW	8 μL	800 μL

The primers were ordered from Integrated DNA Technologies and were received at an original concentration of 5ng/μL. Their sequences can be found in Table 2. They were diluted from their original state by adding 8μL of TE buffer to 1μL of the forward primer and 1μL of the reverse primer for each 96-well plate PCR reaction. Eppendorf skirted twin.tec 96-well plates were used to carry out the PCR reaction. As described above, 10μL of reaction mixture was administered into each well using an Eppendorf Multipipette Plus. DNA, in 1μL aliquots, was added to each well using a Gene Mate 8-channel pipette from ICS Bio Express. The 96-well plate was then sealed with a rubber mat and placed into an Eppendorf Mastercycler.

The following program was run for all plates involving primers MTHFR C1068T, MS A2756G, and MTRR A66G:

1) 95.0° C for 2 minutes

2) 95.0° C for 30 seconds

3) 67.0° C for 30 seconds

4) Temperature increases at a rate of 3.0° C/s until 72.0° C is reached

5) Steps 2-4 are repeated 13 times

6) 95.0° C for 30 seconds

7) 60.0° C for 30 seconds

8) 72.0° C for 1 minute

9) Steps 6-8 are repeated 21 times

MTHFR C677T :

1) 95.0° C for 2 minutes

2) 95.0° C for 30 seconds

3) 64.0° C for 30 seconds

4) Temperature increases at a rate of 3.0° C/s until 72° C is reached

5) Steps 2-4 are repeated 13 times

6) 95.0° C for 30 seconds

7) 57.0° C for 30 seconds

8) 72.0° C for 1 minute

9) Steps 6-8 are repeated 21 times

A different procedure was used for MTHFR C677T because better PCR results were obtained when a lower temperature was used for primer annealing (57.0° C). The other primers performed best at 60.0° C.

<p align="center">**Table 2. Forward and Reverse PCR Primers, 5'→3'**</p>

Forward and Reverse PCR Primers, 5'→ 3'		
SNPS	Forward	Reverse
MTHFR C677T	CATCCCTATTGGCAGGTTACCC	AACTCAGCGAACTCAGCACTCC
MTHFR C1068T	GGCACATCGAAGGCTCCAGTC	TGACTCCTCCCAGGTCAACCAC
MS A2756G	GCTATCTTGCATTTTCAGTGTTCCC	AATGATCCAAAGCCTTTTACACTCC
MTRR A66G	GGATCTTTTTTCCCCCATTTTTC	CTTCCCAACCAAAATTCTTCAAAGC

Once the PCR Mastercycler completed the program, the plate was removed from the machine so that the next step, agarose gel verification, could be performed.

2.2.2 AGAROSE GEL VERIFICATION

Agarose gel verification involved running the PCR product on a gel to verify the presence of PCR product. It was a measure of success that the PCR reaction was completed successfully. In each 96-well plate, a column of positive controls and water blanks were included. The liquid in these wells was removed and placed in an 8 strip PCR Tube Zipperstrip from Phenix Research

Products. Dye composed of bromophenol blue, xylene cyanol, and ficoll-400 was added and the contents were loaded into a 2% agarose gel. The gel was made by adding 4 grams of GenePure LE agarose to 200mL of TBE buffer. This solution was heated until the agarose completely dissolved. After cooling to slightly above room temperature, 10μL of ethidium bromide, a DNA-binding fluorescent tag, was added. Then the agarose mixture was poured into a gel encasement and left to solidify. Once solidification occurred, the gel was loaded with the DNA product and placed into the GelXLPlus from LabNet International Inc. 100V of electricity was applied for 30 minutes and the gel was photographed under a UV light. Presence of fluorescently glowing bands verified the existence of PCR product and the experiment moved onto the primer/dNTP degradation step.

2.2.3   PRIMER/dNTP DEGRADATION STEP

This step is performed to rid the PCR product of excess dNTPs and primers that remain from the PCR Amplification step. These can interfere with the fluorescence polarization step and must therefore be removed. This is done by adding an additional reaction mixture to the PCR product including exonuclease to degrade free primers and shrimp alkaline phosphatase to degrade free dNTPs. Table 3 shows the exact amount of reagents added.

The 10X rApid alkaline phosphatase buffer and rApid alkaline phosphatase (1U/μL) were purchased from Roche and the exonuclease (20,000 U/mL) was purchased from New England BioLabs. This reaction mixture was added in 10 μL aliquots to the PCR product in the 96-well plate using the Eppendorf Multipipette Plus.

18

Table 3. Degradation Step Reagents

Reagent	Mixture for 1 well	Mixture for 96-well plate
10X rApid Alkaline Phosphatase Buffer	1 μL	100μL
rApid Alkaline Phosphatase	1 μL	100 μL
Exonuclease	0.05 μL	5 μL
SDW	8 μL	800 μL

The 96-well plate was then placed in the Mastercycler and the following program was run:

1) 37°C for 1 hour

2) 85°C for 15 minutes

During the first step, the exonuclease and rApid alkaline phosphatase were placed at a temperature at which they function optimally (37°C). They degraded the free primers and free dNTPs respectively. The second step was intended to raise the temperature high enough to denature the exonuclease and shrimp alkaline phosphatase (85°C) to ensure that they do not interfere with the following fluorescence polarization step. Once the program was completed, the 96-well plate was removed from the Mastercycler in order to continue with the final fluorescence polarization step.

2.2.4   FLUORESCENCE POLARIZATION

Fluorescence polarization is a single base primer extension assay that uses a fluorescently tagged dideoxynucleotide chain terminator to classify the genotype of the SNP. The primers for

this assay are designed to anneal immediately upstream of the base pair of interest for SNP classification. The fluorescently tagged base pair is a dideoxynucleotide and halts further annealing after it incorporates into the SNP position. The dye labeled terminator is then read and the allele present can be deduced. The reading of the dye labeled terminator is possible due to the principle that when a fluorescent molecule is excited by plane polarized light, it emits polarized light in response. The fluorescence polarization, or FP, recorded is proportional to the molecule's rotational relaxation time. This is the amount of time it takes the molecule to rotate 68.5°. When viscosity and absolute temperature of the solution are held constant, the FP is also directly proportional to molecular weight of the molecule occupying the SNP position. It is expressed in a fluorescence ratio of the horizontal and vertical axes. This allows differences in alleles at the SNP position to be identified. Two fluorescent dyes are used in the assay and each is tagged to a different one of the two possibilities of alleles for the SNP. When these alleles incorporate and the FP is recorded, the homozygous and heterozygous allele genotypes separate into different quadrants on the horizontal and vertical axis [36]. An example of this can be observed in Figure 5. Tamra and R110 were dyes used to tag the A and G dideoxynucleotides, respectively. When an FP was read that showed proportions of the molecular weight for two AA fluorescently tagged molecules, a data point was recorded in the upper left hand quadrant of the graph. When an FP was read showing a proportional molecular weight for an A, tagged with Tamra, and a G, tagged with R110 dye, a data point was recorded in the upper right hand quadrant of the graph. Similarly, a GG FP reading recorded a data point in the lower right hand quadrant of the graph. The lower left hand cluster was a negative control and failed sample quadrant, which showed samples that failed to incorporate any fluorescent signal. Failed samples were disregarded during analysis.

20
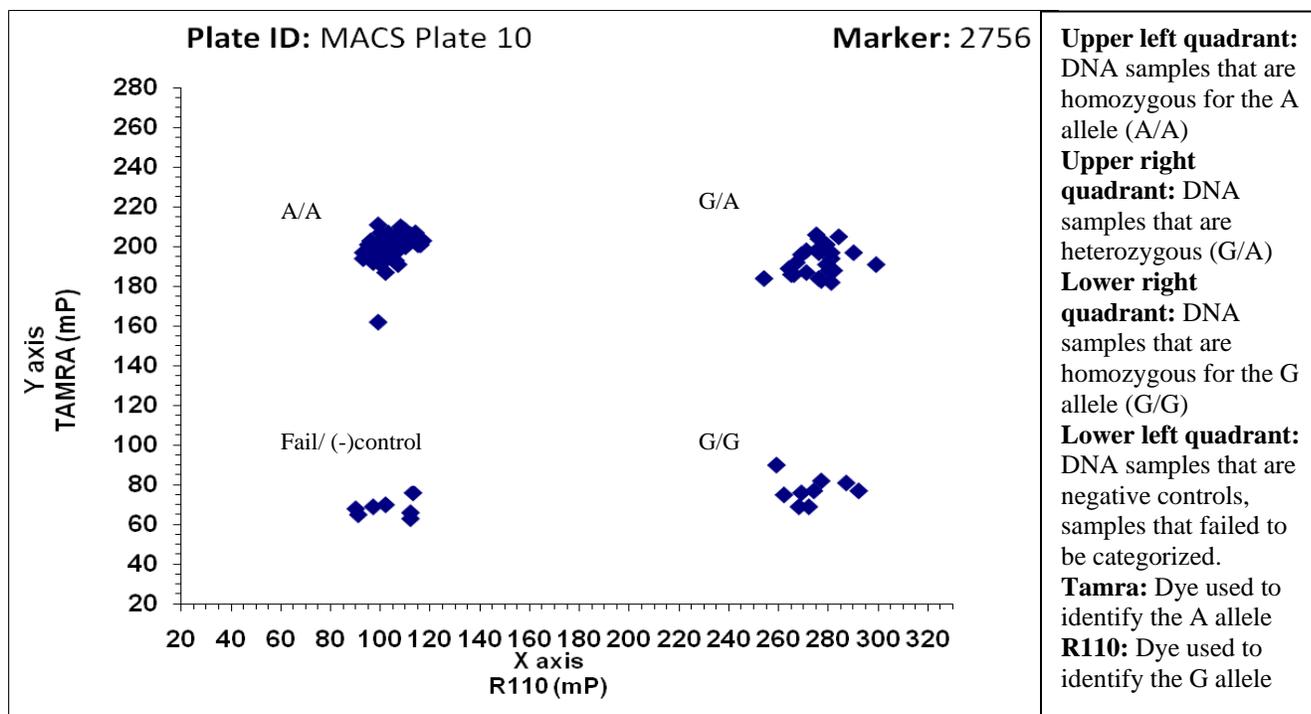
Figure 5. Data Results for Fluorescence Polarization Assay: Genotype Clusters

Table 4 shows the exact amount of reagents used for the Fluorescence Polarization step:

Table 4. Fluorescence Polarization Step Reagents

Reagents	Mixture for 1 well	Mixture for 96-well plate
Thermosequenase 10X Buffer	1μL	100μL
Primer	1μL	100μL
ddNTP mixture	0.05μL	5μL
Thermosequenase	0.05μL	5μL
SDW	8μL	800μL

The Thermosequenase 10X Buffer and Thermosequenase DNA polymerase w/pyrophosphatase (4Un/μL) were purchased from USB Corporation. The ddNTP mixture was composed of a 16:16:15:15:1:1 ratio of ddCTP:ddATP:ddGTP:ddTTP:R110G:tamraA. The

ddGTP Ultrapure (10mM pH7.5), ddTTP Ultrapure (10mM pH 7.5), ddATP Ultrapure (10mM pH 7.5), and ddCTP Ultrapure (10mM pH 7.5) were purchased from USB Corporation. Tamra ddATP (NEL 474 2.5Nmoles) and R110 ddGTP (NEL 495001 EA 2.5Nmoles) were purchased from Perkin Elmer Life Sciences Inc. The primer was composed of 10μL of internal primer added to 100μL of 1X TE Buffer. A table of internal primers is listed below. The internal primers were purchased from Integrated DNA Technologies. Forward and reverse internal primers were purchased for each SNP, however, we found that one primer typically gave higher quality results than the other. Therefore we used the highest performing primer for each SNP.

Table 5. Internal Primers for Fluorescence Polarization, 5'→ 3'

Internal Primers for Fluorescence Polarization, 5'→3'			
SNP	Primer Orientation	Internal Primer Sequence	Fluorescent Dye
MTHFR C677T	Reverse Internal	AAGCTGCGTGATGATGAAATCG	A/G
MTHFR 1068T	Reverse Internal	CCTCTCGGCGCTTGGGGTGGGC	A/G
MS A2756	Forward Internal	GAATATGAAGATATTAGACAGG	A/G
MTRR A66G	Forward Internal	AAAGGCCATCGCAGAAGAAAT	A/G

These reagents were mixed together and 10μL aliquots were placed into the 96-well plates using the Eppendorf Multipipette Plus. The plates were then placed into the Mastercycler and the following program was run for all plates:

1) 94.0°C for 1 minute

2) 94°C for 10 seconds

3) 52°C for 30 seconds

4) Steps 2-4 are repeated 40 times

5) 72°C for 10 seconds

6) Temperature is held at 10.0°C until plate is removed from cycler

When the plate was removed at the finish of the program, it was wrapped in foil to keep out any light that would prematurely excite the fluorescent dyes. The plate was then taken to the LJL Biosystems Analyst HT where the fluorescently tagged ddNTPs were read. The data collected were formatted into a graph as seen in Figure 5.

2.3   STATISTICAL ANALYSIS

The following methods were used in data analysis, and the statistical package, R (Lucent Technologies), was used to generate the plots and statistical calculations.

2.3.1   BOX-AND-WHISKER DIAGRAMS

Once all data were collected, box-and-whisker diagrams were generated to showcase the entire data set, and to enable observations of basic trends in the data. A box-and-whisker diagram is a descriptive, non-parametric, method to showcase data. It provides important information concerning the data set such as the median value and quartile locations of the data points. In a box-and-whisker diagram, the horizontal, bold line represents the median. This median line is surrounded by a box, which represents the middle 2 quartiles, or 50% of the data (Q2 and Q3). The vertical lines are called whiskers. The bottom whisker represents one half of Quartile 1 and the top whisker represents one half of Quartile 4. Together the box and whiskers encompass 75% of the data. Any data points residing outside of the box and whisker plot are extreme outliers and represent the data points that fall into the extreme high and low 12.5% of

the population. On the box-and-whisker diagram that we generated, we inserted thin horizontal and dotted lines at threshold levels. These were visual aids for identifying LDL, HDL, and triglyceride levels that were high or low. These plots helped us to observe any obvious associations and helped to narrow our focus for the next statistical test.

2.3.2   KRUSKAL-WALLIS TEST

The Kruskal-Wallis Test is a conservative non-parametric test performed to observe the general shape of a distribution. We used it so that we could observe the degree of equality amongst the population medians for SNPS C1068T and A2756G. The nominal variable was the SNP classification and the measurable variable was the LDL level. The test was performed to see if the LDL level difference was significant enough in each allele classification to result in a significant p-value. This would indicate that one grouping of data points had significantly different median values than another grouping. This would be strong evidence for an association between two variables.

2.3.3   ODDS RATIO CALCULATION

An odds ratio measures the size of the difference between populations by describing the strength of association between two binary values. For this test, we compared genetic classification and LDL levels. For the genetic classifications, AA and AG were grouped together into one category (the dominant inheritance group) and GG was the other category (the recessive inheritance group). The LDL levels were classified as high (above 159 mg/dL) or not high (159mg/dL or lower). If an odds ratio value of 1 is obtained, it describes no association between

the two variables. We set up the calculation so that if an odds ratio value of greater than 1 is obtained, it describes the genotype GG as being a predictive factor for high LDL levels. If a value of less than one is obtained for the odds ratio, GG is described as being a protective factor against high LDL levels, or a predictor of low LDL levels.

2.3.4 LOGISTIC REGRESSION

A logistic regression test was also performed. This test allows prediction of the probability of an event by fitting data to a calculated logistic curve. This is performed by quantifying the association between one or more independent variables and a binary response variable. For this study, we used BMI, HIV status, HAART, and allele classification (AA, AG, GG) for C677T, C1068T, A2756G, and A66G as independent variables and high or not high LDL levels for the binary response variable. LDL levels of above 159 were considered "high" and LDL levels of 159 and below were considered "not high." The statistics package, R was used for this [37].

3.0   RESULTS

A total number of 1,945 samples were typed for the four folate metabolism SNPs. Table 6 summarizes the results obtained from the FP assays.

Table 6. FP Assay Data Summary

Locus	Sample Size	Minor Allele Frequency (MAF)	Major Allele Homozygote	Heterozygotes	Minor Allele Homozygote
MTRR A66G	1573	(A) 49.1%	914 (21.9%)	344 (58.1%)	315 (20.0%)
MS A2756G	1730	(G) 19.8%	1127 (65.1%)	520 (30.1%)	83 (4.8%)
MTHFR C1068T	1565	(T) 6.90%	1358 (86.8%)	195 (12.5%)	12 (0.8%)
MTHFR C677T	1493	(T) 28.9%	806 (54.0%)	510 (34.2%)	177 (11.9%)

Before beginning analysis, the racial background of the MACS participants was considered. This is important because genetic differences among various racial groups can skew data. Table 7 was generated, listing the genotypic data for the following self reported categories: whites, blacks, and other. The other category was composed of various other races such as

Hispanic, Asian, and Native American. Since the majority of the data was for the white category, all analysis beyond this table will be performed on that group in order to ensure the most consistent results.

Table 7. Racial background summary of MACS participants

Locus	Race	Major Allele	Heterozygotes	Minor Allele	Blank/Fail	HWE with Bonferroni correction
MTHFR C677T	White	259	197	75	129	0.041
	Black	225	49	2	53	0.984
	Other	42	35	27	26	--
MTHFR C1068T	White	500	70	3	87	0.948
	Black	251	34	2	40	0.779
	Other	103	10	3	14	--
MTRR A66G	White	94	343	144	79	$*3\text{x}10^{-5}$
	Black	68	182	31	49	$*7 \text{ x } 10^{-6}$
	Other	24	67	16	23	--
MS A2756G	White	421	162	29	48	0.041
	Black	170	126	20	14	0.870
	Other	78	37	5	100	--

In addition to listing the genotype data for each race, a Hardy Weinberg Equilibrium calculation with a Bonferroni correction was also performed for the white and black groups. The Hardy Weinberg principle states that allele and genotype frequencies remain constant in a

homogenous population unless disturbed by non-random mating, mutations, selection, limited population size, "overlapping generations", random genetic drift and gene flow [38]. The number shown in the chart is the p-value. P-values that are not significant are considered within Hardy Weinberg Equilibrium. A Bonferroni correction was applied to the Hardy Weinberg Equilibrium to correct for Family wise Type I error that occurs when multiple independent comparisons are made [39]. This enables small amounts of error, involved in each comparison, to be corrected. The correction involves dividing the Hardy Weinberg Equilibrium p-value by the number of independent comparisons that were made [39]. Since four SNP comparisons were made for both black and white races, this meant that there were 8 comparisons to consider. The original p-value of 0.05 was therefore divided by 8. After application of the Bonferroni correction, the p-value for each category was 0.00625. All categories fell into Hardy Weinberg Equilibrium except for the MTRR A66G white and black categories, whose p-values remained more extreme than 0.00625. This is most likely due to a complication in the assignment of genotypes for that particular SNP. This will be further analyzed at a later date. The other SNPs fell into Hardy Weinberg Equilibrium and their data will be used for additional analysis. The Hardy Weinberg Equilibrium calculation was not performed for the "other" category because it was a heterogeneous group. The results would not be significant since the conditions for the Hardy Weinberg Equilibrium principle do not apply to heterogeneous groups [38].

Since the data involved several variables and the potential for their associations were very complex, we analyzed the data on several different levels. A box-and-whisker diagram was made to observe any obvious associations between LDL, HDL, and total triglyceride levels versus each SNP's genotype classifications.
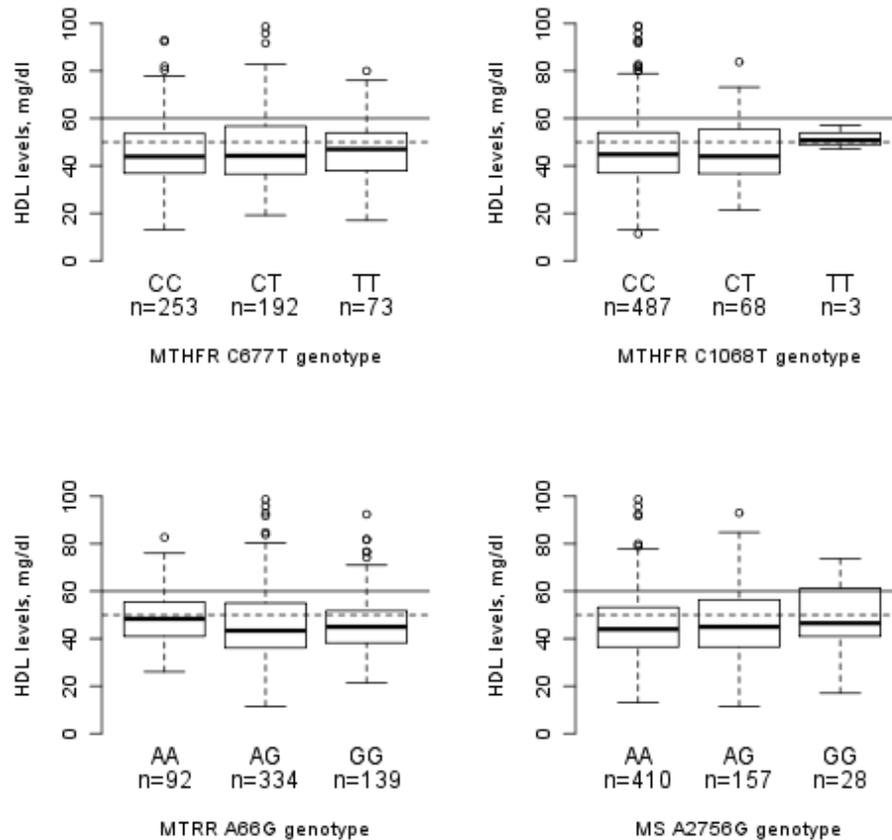
Figure 6. HDL levels versus SNP genotype classifications of white samples

As previously mentioned, high HDL levels (60mg/dL and above) are protective factors against cardiovascular disease and low HDL levels (50mg/dL or less) are risk factors. The horizontal, solid black line in Figure 6 represents the high HDL level cutoff and the dotted horizontal line represents the low HDL level cutoff. The thick horizontal line represents the median for the data set, the box represents 50% of the data set, and the whiskers encompass 75% of the data set. The pictured data points are extreme outliers and show the most extreme 12.5% of the data set. There are no obvious associations in Figure 6. However, due to the complexity of variables involved, it is not expected to find a simple association at this point.

29
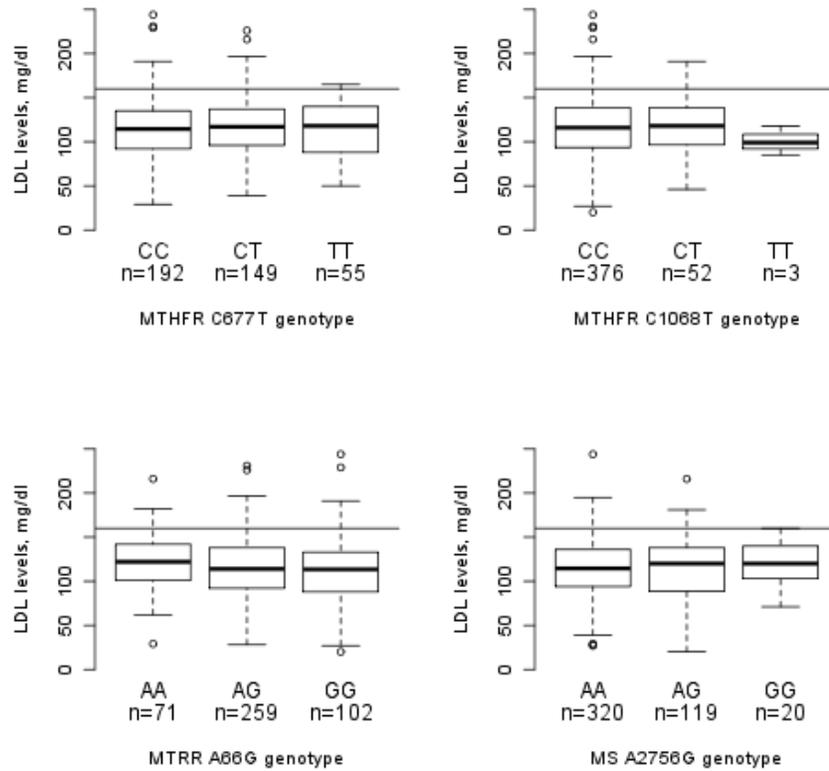
Figure 7. LDL levels versus genotype classification of white samples

As previously mentioned, high levels of LDL (160mg/dL and above) are risk factors for CVD and low levels of LDL (100mg/dL and below) are protective factors. The solid horizontal line in Figure 7 represents the high LDL level cutoff. There appears to be a trend in MTHFR C1068T that the samples genotyped as TT have lower LDL levels than those with CT or CC genotypes. There also appears to be a trend in MS A2756G that the samples genotyped as GG have lower LDL levels than those with AG or AA genotypes. This will be further investigated with other statistical tests.

30
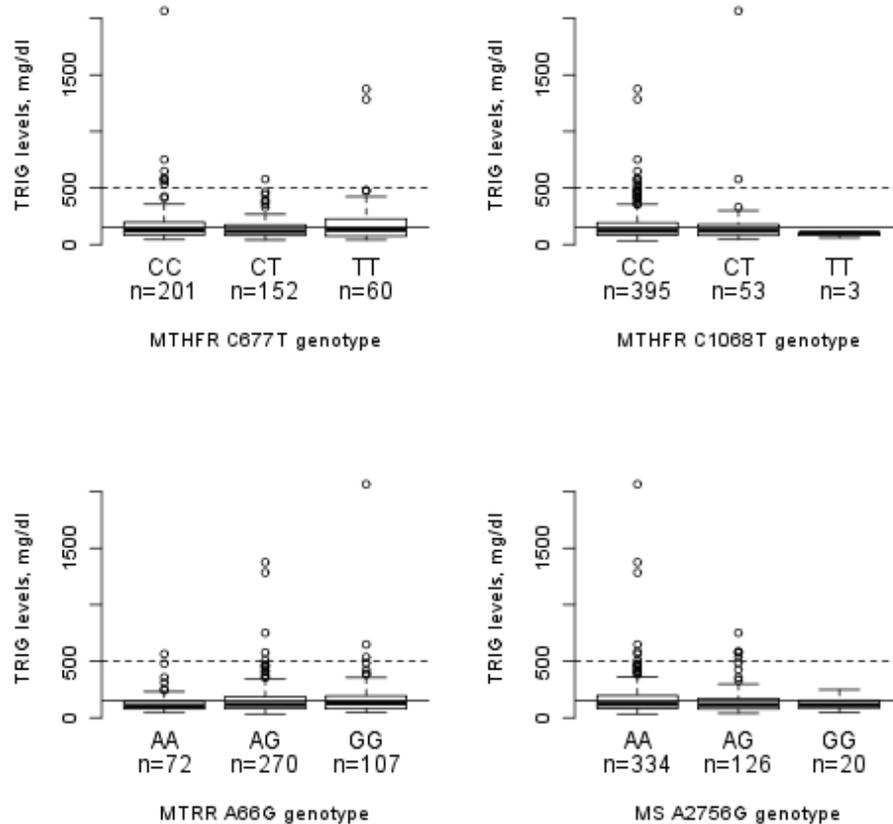
Figure 8. Total Triglyceride Levels versus genotype classification for white samples

As previously mentioned, high levels of triglycerides (500mg/dL and above) are risk factors for CVD and low levels of triglycerides (150mg/dL and below) are protective factors. The dotted horizontal line in Figure 8 shows the 500mg/dL cutoff and the solid horizontal line shows the 150mg/dL cutoff. There do not appear to be any obvious associations in this figure.

Since LDL levels for SNPs C1068T and A2756G (Figure 7) were the only ones that showed a possible association with genotype, more descriptive box-and-whisker diagrams were developed for them. The population was separated into individuals on HAART therapy and individuals not on HAART therapy. The population of individuals not on HAART is made up of primarily individuals (n=583) not infected with HIV and is therefore meant to be a control group.
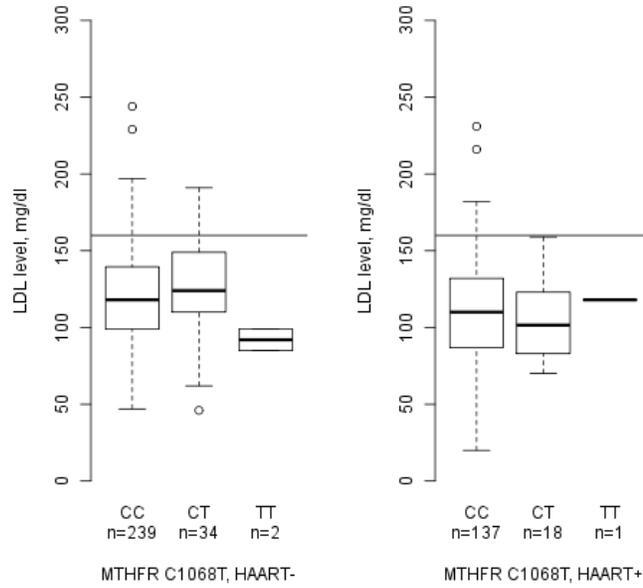
Figure 9. LDL levels versus MTHFR C1068T genotypes for white samples
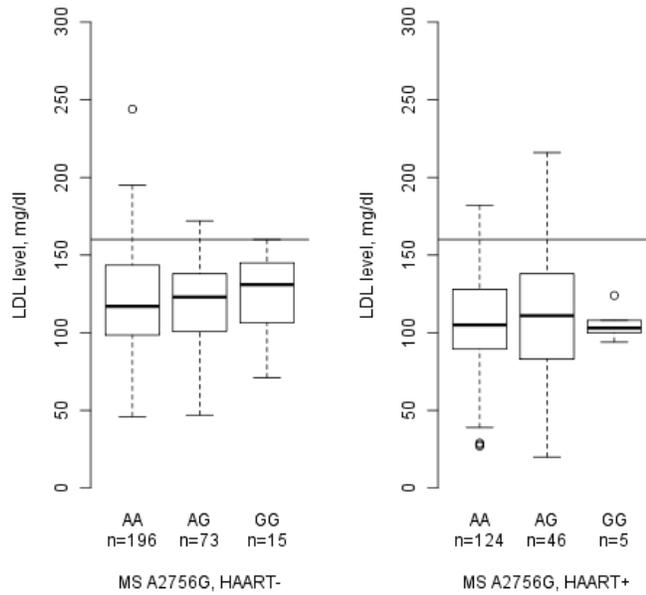
Figure 10. LDL levels versus MS A2756G genotypes for white samples

In Figure 9, it initially appears that the TT genotype in the C1068T SNP is a protective factor for LDL levels, since all of the data points are below the 159mg/dL threshold. However, since there was only one data point in that category, we cannot say if it is a significant result without conducting further tests. However, in figure 10 there may be a positive association between individuals undergoing HAART therapy with the genetic characterization of GG for the A2756G SNP and low LDL levels. There were 5 individuals in this category and we decided to investigate this possible association more thoroughly.

Next the Kruskal-Wallis Test was performed. Two degrees of freedom and a 95% confidence interval (CI) were used. We found that the p-values were not significant (data not shown) indicating that there is no significant difference in the distributions of LDL levels between the three genotype groups overall. This test does not, however, consider the 159 mg/dL threshold of LDL cholesterol levels that distinguishes high cholesterol individuals from low cholesterol ones.

We then performed an odds ratio calculation, as this enables us to consider the influence of genotype on the binary outcome of high LDL vs. non-high LDL. A 95% CI was used for this test as well. An odds ratio was calculated for the overall MS A2756G data, MS A2756G data for individuals on HAART therapy, and MS A2756G data for individuals not on HAART therapy. We expected to see a higher odds ratio in the HAART+ group than the HAART- group since we hypothesize that HAART in addition to HIV infection and genetic factors are related to high LDL levels. For the overall data set, we found an odds ratio of 0.24 with a CI of 0.03-1.75. For the HAART+ data set, we found an odds ratio of 0.49 with a CI of 0.06-3.82. For the HAART-data set, we found an odds ratio of 0.22 with a CI of 0.01-3.66. Although this suggests that the presence of the GG genotype at MS A2756G is associated with non-high LDL levels, in each

case the 95% CI included 1.0, so we cannot conclude that any of these results are statistically significant. We suspect that this was due to the abnormally large CI caused by a small population for some of the genotype classifications. The odds ratio was highest for the HAART+ data set (0.49) and, still being less than 1, shows a small association with low LDL levels.

To further investigate how variables are influenced by A2756G SNP classification, we performed a logistic regression analysis. None of the variables (BMI, HIV status, HAART status, or SNP classification) showed a significant p-value. This means that individually, these variables do not predict LDL level classification. However, the intercept p-value was significant (0.0139). This means the data set as a whole may predict LDL level classification.

Table 8. Summary of Logistic Regression p-values

Variable	z-value	p-value
Intercept	-2.166	0.0303 *
BMI	0.350	0.7261
HAART	-0.976	0.3293
C677T (C/T to C/C)	0.358	0.7201
C677T (T/T to C/C	-0.797	0.4252
C1068T (C/T to C/C)	-0.038	0.9694
C1068T (T/T to C/C)	-0.013	0.9898
A66G (A/G to A/A)	0.163	0.8703
A66G (G/G to A/A)	0.407	0.6844
A2756G (A/G to A/A)	0.975	0.3296
A2756G (G/G to A/A)	-0.164	0.8701

4.0   DISCUSSION

Although none of the statistical tests performed showed definitively significant associations, we still believe that the genetics of this metabolic pathway warrant further investigation.

The Hardy Weinberg Equilibrium calculation with the Bonferroni correction in Table 7 showed that the most normal data was for the MTHFR and MS SNPs. Therefore, that data was considered more reliable than the data for the MTRR SNP, A66G. For consistency, further calculations were performed on the white category, as it was the racial group of the highest prevalence in the data. This was performed so any genetic variation between races could be eliminated.

The box-and-whisker diagrams, shown in Figures 6, 7, and 8, show the entire set of genotype data generated versus HDL, LDL, and triglyceride levels. We were looking for a linear trend in data points, showing a certain genotype with extreme levels of LDL, HDL, or triglycerides (high or low) and the LDL levels decreasing or increasing across the X axis. If GG, for example, contained all data points with low LDL levels, AG contained slightly higher LDL levels, and AA contained high LDL levels, it would suggest that the GG genotype is protective against high levels of LDL. There were no trends seen in any of the graphs for HDL or triglycerides. However, we thought that there could potentially be an association between genotype classification and LDL levels for MTHFR C1068T and MS A2756G. The TT

genotype for MTHFR C1068T showed low levels of LDL which increased for CT, and was highest for CC. For MS A2756G, LDL levels were highest for the AA classification, decreased for AG, and were lowest for GG. We therefore continued our analysis focusing on these two SNPs and the LDL variable.

We then generated more descriptive box-and-whisker diagrams of MHFR C1068T and MS A2756G versus LDL levels in Figures 9 and 10, respectively. Individuals undergoing HAART therapy were separated from the individuals not undergoing HAART therapy. This was done so that we could observe any difference, to which HAART therapy contributed. We continued to look for linear trends in LDL levels. We also looked for a recessive inheritance trend. For a trend like this, we would notice that low or high levels of LDL would only be observed for a homozygote phenotype. For example, if GG was a recessive trait that predicted low LDL levels, the low levels would not be present in AG or AA. Taking this into consideration, we also looked for any results that showed tight box-and-whisker groupings of homozygote data in either high or low positions in comparison to the other two groupings. Conversely, a dominantly inherited trend would show high or low LDL level groupings in both the heterozygote and dominant homozygote groupings. In Figure 9, we observed a box and whisker grouping contained completely in the not high LDL level region for the genotype classification TT. Initially it seemed as though this might have been a recessive inheritance trend. However, the sample size for this classification was so small (n=1), it was likely that the trend was not significant. In Figure 10, there was a box and whisker grouping contained completely in the not high LDL level region for MS A2756G GG. The AA and AG samples had data points in the high LDL level region and the sample size for GG was 5. We considered this

to be a significant number of data points and a possible recessive inheritance trend. We then decided to pursue this possible association by performing more specific statistical tests.

A Kruskal-Wallis test was performed next to quantify the extent of the trends we saw in the scatter plots and box-and-whisker diagram. This test was performed for both C1068T and A2756G and no significant p-values were obtained. This shows that there were also no statistically significant overall differences in LDL levels between the genotype classifications. However, this test is a conservative non-parametric test and it is possible that an association exists, but is too small to be detected by this general test. This test was also performed to generally characterize the shape of the distribution without regard to LDL thresholds. We next decided to address this by performing an odds ratio calculation.

Since the box-and-whisker diagram showed A2756G as having a more promising data set for association, an odds ratio was calculated to see if GG was protective against high LDL levels. As mentioned in the results, the odds ratio showed a small association for this. The CI was very large and prohibited us from being able to make a strong assertion that GG protected against high LDL levels. The CI was very large due to the small sample size of individuals that were classified as having a GG genotype for the SNP A2756G. While we think that there might be a stronger association between GG and hot high LDL levels, we cannot confidently state this until we have data for a larger sample size of individuals homozygous for the G allele. This could easily be attained in a few weeks time. The Center of Analysis for the Multicenter AIDS Cohort Study (CAMACS) keeps a library of clinical data for all of the MACS participants. We collaborate with them by sending them genotypic data for these participants in exchange for the clinical data. They have been unable, so far, to send us data for a large number of samples that

we have already genotyped. Once these data, comprising ~800 additional samples, are received, our calculations can be revised and possibly statistically significant results will be obtained.

The last statistical calculation that we performed was logistic regression. This test enabled us to consider the possible effect of several different variables on high/not high LDL levels. Once again, none of the individual p-values were significant, but the intercept p-value was significant. The intercept p-value is the value of the logit score when all independent variables have a value of zero. This estimates the degree of non-randomness in the dependent variable. If data collected for the dependent variable differs from data expected when a random sampling is undertaken, then the p-value will be significant. The significant p-value that we found suggests that there truly is a trend in this data set that can be considered predictive for LDL level classification. Although none of the independent variables that we tested proved to be the cause of this significance, there could be other variables that were outside the scope of this study that caused the significant intercept p-value. From the other plots and statistics we have generated, we believe that homozygous G alleles for A2756G are protective against high levels of LDL. This, however, would not be directly observed in logistic regression. In order for the A2756G p-value to be significant, the entire A2756G data set would have to correlate with low LDL levels. Logistic Regression would probably not detect one subgroup that influences LDL levels. The lack of sensitivity of this statistical test is a disadvantage, but it is useful in enabling the observation of several variables and their predictive abilities for a binary factor.

The folic acid metabolic pathway is a very influential part of the body's biochemistry that has been associated, repeatedly, with CVD. Although the specific mechanism has not been discovered, it is clearly a complex pathway, and host genetics influence it in a way that is still not fully understood. When the variables of HIV infection and HAART are also considered, the

complexity increases. A recently published paper in the journal of Acquired Immune Deficiency Syndromes has made an association between elevation of plasma homocysteine levels and HAART, therefore reinforcing our idea that the folic acid metabolic pathway is associated with HAART and CVD in HIV+ individuals [40]. Despite the fact that our study has not resulted in statistically significant associations, this conclusion can still be considered an important result. The elimination of MTHFR C677T, MTHFR C1068T, MS A2756G, and MTRR A66G as strongly causative factors leads us in pursuit of other SNPs in the folic acid metabolic process that may be involved, and to find associations that have yet to be discovered.

BIBLIOGRAPHY

1.   Riddler, S.A., et al., *Antiretroviral therapy is associated with an atherogenic lipoprotein phenotype among HIV-1-infected men in the Multicenter AIDS Cohort Study.* J Acquir Immune Defic Syndr, 2008. **48**(3): p. 281-8.
2.   Pearson, T.A., et al., *The association of low levels of HDL cholesterol and arteriographically defined coronary artery disease.* Am J Epidemiol, 1979. **109**(3): p. 285-95.
3.   Martinez, E. and J.M. Gatell, *Metabolic abnormalities and body fat redistribution in HIV-1 infected patients: the lipodystrophy syndrome.* Curr Opin Infect Dis, 1999. **12**(1): p. 13-9.
4.   Riddler, S.A., et al., *Impact of HIV infection and HAART on serum lipids in men.* JAMA, 2003. **289**(22): p. 2978-82.
5.   Isezuo, S.A. and M.A. Makusidi, *Metabolic dysfunctions in non-antiretroviral treated HIV/AIDS patients.* Niger J Clin Pract, 2009. **12**(4): p. 375-8.
6.   Domingos, H., et al., *Metabolic effects associated to the highly active antiretroviral therapy (HAART) in AIDS patients.* Braz J Infect Dis, 2009. **13**(2): p. 130-6.
7.   Verma, S.P., *HIV: A Raft-Targeting Approach for Prevention and Therapy Using Plant-Derived Compounds* Current Drug Targets. **10**(1): p. 51-59.
8.   Mondal, D., et al., *HAART drugs induce oxidative stress in human endothelial cells and increase endothelial recruitment of mononuclear cells: exacerbation by inflammatory cytokines and amelioration by antioxidants.* Cardiovasc Toxicol, 2004. **4**(3): p. 287-302.
9.   Smith, D., *HIV treatment options.* AIDS Treat News, 1998(No 294): p. 4-5.
10.   Carter, C.C., et al., *HIV-1 infects multipotent progenitor cells causing cell death and establishing latent cellular reservoirs.* Nat Med, 2010.
11.   McNicholl, I. *Adverse Effects of Antiretroviral Drugs.* 2009 [cited 2010 12 April]; Available from: http://hivinsite.ucsf.edu/InSite?page=ar-05-01.
12.   Staff, M.C. *Heart Disease.* 2010 April 8 2010 [cited 2010 April 25]; Available from: http://www.mayoclinic.com/health/heart-disease/DS01120.
13.   *About Heart Disease.* 2009 November 16 [cited 2010 25 April]; Available from: http://www.cdc.gov/heartdisease/about.htm.
14.   Myungjin Kim, T.I.L., Kazuko Arakawa, Renwei Wang, Mimi C. Yu, and Peter W. Laird, *DNA Methylation as a Biomarker for Cardiovascular Disease Risk.* PLoS One, 2010. **5**(3): p. e9692.
15.   Grabar, P.B., Dusan Logar, Boris Lestan, Vita Dolzan, *Genetic determinants of methotrexate toxicity in rheumatoid arthritis patients: a study of polymorphisms affecting methotrexate transport and folate metabolism.* European Journal of Clinical Pharmacology, 2008. **64**: p. 1057-1068.
16.   Malinowska, A. and A. Chmurzynska, *Polymorphism of genes encoding homocysteine metabolism-related enzymes and risk for cardiovascular disease.* Nutr Res, 2009. **29**(10): p. 685-95.
17.   Jakubowski, H., *Pathophysiological consequences of homocysteine excess.* J Nutr, 2006. **136**(6 Suppl): p. 1741S-1749S.

18.	Vignal, A., Denis Milan, Magali SanCristobal and André Eggen, *A	review on SNP and other types of molecular markers and their use in animal genetics* Genetics Selection Evolution 2002: p. 275-305.

19.	Friso, S., et al., *A common mutation in the 5,10-methylenetetrahydrofolate reductase gene affects genomic DNA methylation through an interaction with folate status.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5606-11.

20.	*MTHFR*. Genetics Home Reference 2010; Available from: http://ghr.nlm.nih.gov/gene=mthfr.

21.	*MTHFR 5,10-methylenetetrahydrofolate reductase (NADPH)* 10 April 2010 [cited 2010 12 April]; Available from: http://www.ncbi.nlm.nih.gov/gene/4524.

22.	Tsai MY, L.C., Cao J, Kim Y, Siscovick DS, Schreiner PJ, Hanson NQ, *Polygenic association with total homocysteine in the post-folic acid fortification era: the CARDIA study.* Molecular Genetics and Metabolism, 2009. **98**(1-2): p. 181-186.

23.	Bosco, P., et al., *Methionine synthase (MTR) 2756 (A --> G) polymorphism, double heterozygosity methionine synthase 2756 AG/methionine synthase reductase (MTRR) 66 AG, and elevated homocysteinemia are three risk factors for having a child with Down syndrome.* Am J Med Genet A, 2003. **121A**(3): p. 219-24.

24.	Scazzone, C., et al., *Methionine Synthase Reductase (MTRR) A66G polymorphism is not related with plasma homocysteine concentration and risk for vascular disease.* Exp Mol Pathol, 2009.

25.	Morrison, K., et al., *Susceptibility to spina bifida; an association study of five candidate genes.* Ann Hum Genet, 1998. **62**(Pt 5): p. 379-96.

26.	*MTR*. Genetics Home Reference 2010 [cited 2010 13 April]; Available from: http://ghr.nlm.nih.gov/gene=mtr.

27.	*MTR 5-methyltetrahydrofolate-homocysteine methyltransferase*. 10 April 2010 [cited 2010 12 April]; Available from: http://www.ncbi.nlm.nih.gov/gene/4548.

28.	Ott, N., H. Geddert, and M. Sarbia, *Polymorphisms in methionine synthase (A2756G) and cystathionine beta-synthase (844ins68) and susceptibility to carcinomas of the upper gastrointestinal tract.* J Cancer Res Clin Oncol, 2008. **134**(3): p. 405-10.

29.	*MTRR*. Genetics Home Reference 2010; Available from: http://ghr.nlm.nih.gov/gene=mtrr.

30.	Tsai, M.Y., et al., *Polygenic association with total homocysteine in the post-folic acid fortification era: the CARDIA study.* Mol Genet Metab, 2009. **98**(1-2): p. 181-6.

31.	Gaughan, D.J., et al., *The methionine synthase reductase (MTRR) A66G polymorphism is a novel genetic determinant of plasma homocysteine concentrations.* Atherosclerosis, 2001. **157**(2): p. 451-6.

32.	*Multi Center AIDS Cohort Study*. July 2009; Available from: http://www.statepi.jhsph.edu/macs/macs.html.

33.	Austin, M.A., et al., *Atherogenic lipoprotein phenotype. A proposed genetic marker for coronary heart disease risk.* Circulation, 1990. **82**(2): p. 495-506.

34.	*What your Cholesterol Levels Mean*. Available from: http://www.americanheart.org/presenter.jhtml?identifier=183#HDL.

35.	Sudano, I., et al., *Cardiovascular disease in HIV infection.* Am Heart J, 2006. **151**(6): p. 1147-55.

36.	Pui-Yan Kwok, M., PhD, *Single Nucleotide Polymorphisms Methods and Protocols*. Methods in Molecular Biology, ed. J.M. Walker. Vol. 212. 2003, Totowa: Humana Press. 177-187.

37.	*The R Project for Statistical Computing*. 2010 [cited 2010 April 16]; Available from: http://www.r-project.org.

38.	Clark, D.L.H.a.A.G., ed. *Principles of Population Genetics*. Fourth ed. 2007, Sinauer Associates, Inc. Publishers: Sunderland.

39.	Quinn, G.P.K., M.J., ed. *Experimental Design and Data Analysis for Biologists*. 2002, Cambridge University Press: Cambridge. 48-50.

40.	Coria-Ramirez, E., et al., *Effect of Highly Active Antiretroviral Therapy on Homocysteine Plasma Concentrations in HIV-1-Infected Patients.* J Acquir Immune Defic Syndr, 2010.