PREDICTIVE BIOMARKERS FOR THE ASSESSMENT OF DEVELOPING HIV-ASSOCIATED NEUROCOGNITIVE DECLINE IN HIV-POSITIVE MEN

by

Gregory James Joseph

BS, BA, Case Western Reserve University, 2010

Submitted to the Graduate Faculty of

the Department of Infectious Diseases and Microbiology,

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Gregory James Joseph

It was defended on

August 8th, 2013

and approved by

Thesis Advisor: Assistant Professor, Infectious Diseases and Microbiology/Human Genetics, Graduate School of Public Health, University of Pittsburgh

Lawrence Kingsley, Professor, Infectious Diseases and Microbiology, Graduate School of Public Health, University of Pittsburgh

Eleanor Feingold, Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Jeremy Martinson, DPhil

PREDICTIVE BIOMARKERS FOR THE ASSESSMENT OF DEVELOPING HIV-ASSOCIATED NEUROCOGNITIVE DECLINE IN HIV-POSITIVE MEN

Gregory James Joseph, M.S.

University of Pittsburgh, 2013

ABSTRACT

The advent of Highly Active Antiretroviral Therapy (HAART) has greatly improved the control of HIV replication in infected individuals. Viral replication has become increasingly well managed and HAART has extended the time to the development of Acquired Immunodeficiency Syndrome (AIDS) for many patients. These now chronically HIV-infected individuals, while enjoying longer more healthy lives, have begun to encounter diseases that are not related to AIDS. These non-AIDS defining illnesses (NADIs) are becoming a prominent area of research and this project sought to study one such NADI, HIV-associated Neurocognitive Decline (HAND). This dementia, associated with HIV-infection, has become the focus of research seeking to understand the mechanism for the disease and also to identify risk factors for identifying patients at risk for HAND development. A pilot study was designed which sought to examine four biomarkers and assess their predictive ability for the development of HAND: single nucleotide polymorphisms (SNPs) in the RYR3 gene region and associated with Atherosclerosis, the Alzheimer's disease risk allele APOE*4, telomere length shortening over time, and DNA methylation change were each examined in a case-control study across three sample groups: HAND+/HIV+ (cases), HAND-/HIV+ (seropositive control), HAND-/HIV- (seronegative control). The genotypes for RYR3 and APOE were assessed through Taqman and Sanger Sequencing, respectively. Absolute telomere length was assessed through quantitative

PCR (qPCR) and normalized against the ribosomal single copy housekeeping gene 36B4. The methylation analysis was conducted by identifying 8 cases and matching 8 individuals to form the two control groups. Archival PBMC pellets collected at ten-year intervals were extracted from this group of 24 individuals, and their genomewide DNA methylation profile was determined via bisulfite conversion on the Methyl450 array from Illumina. It was observed that dementia status, rather than HIV serostatus, was the unifying trait that generated unexpected methylation patterns and requires further investigation. The relevance of this work in the public health field is to enable scientists to more accurately identify those at risk for developing HAND through the assessment of the four predictive biomarkers examined in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank my advisor, Dr. Jeremy Martinson, and my thesis committee, Dr. Eleanor Feingold and Dr. Larry Kingsley for their support, guidance, and advice. I would also like to thank my fellow classmates, the faculty, and staff of the department of Infectious Diseases and Microbiology for the knowledge and support they have shown me over my time here at IDM. I would also like to acknowledge the members of the Martinson Lab during my time at IDM: Dr Matthew Nicholau, Rebecca Bosko Marino, and Blair Gleeson for their support, friendship, and camaraderie. I would also like to specifically thank my parents, Ronald and Virginia Joseph, my sisters Deanna and Jocelyn, and my brother Ron, for all the love, support, and encouragement they showed me during this project's completion. It was not an easy road, but it was made easier by the love and support from all of you and I am forever grateful.

1.0 INTRODUCTION

1.1 HUMAN IMMUNODEFICIENCY VIRUS

1.1.1 Epidemiology

Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS). AIDS has become a pandemic and one of the most pervasive challenges currently facing the medical and scientific communities. It is estimated that 34 million people (Figure 1) currently live with HIV infection worldwide.[1] In 2011, 2.5 million people were estimated to have been newly infected with HIV (Figure 2) and 1.7 million people were estimated to have died from AIDS.

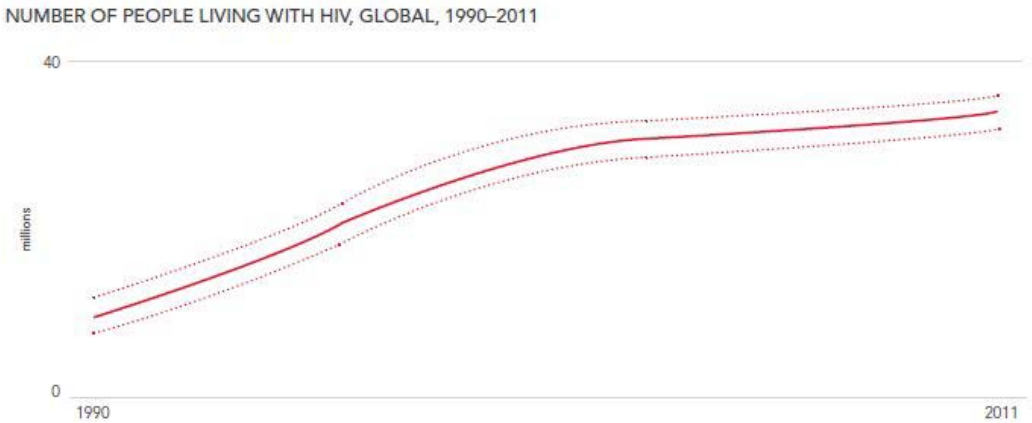

Figure 1. Graphical representation of the number of persons living with HIV globally, in millions (y-axis), between the years 1990 and 2011 (x-axis).[1]

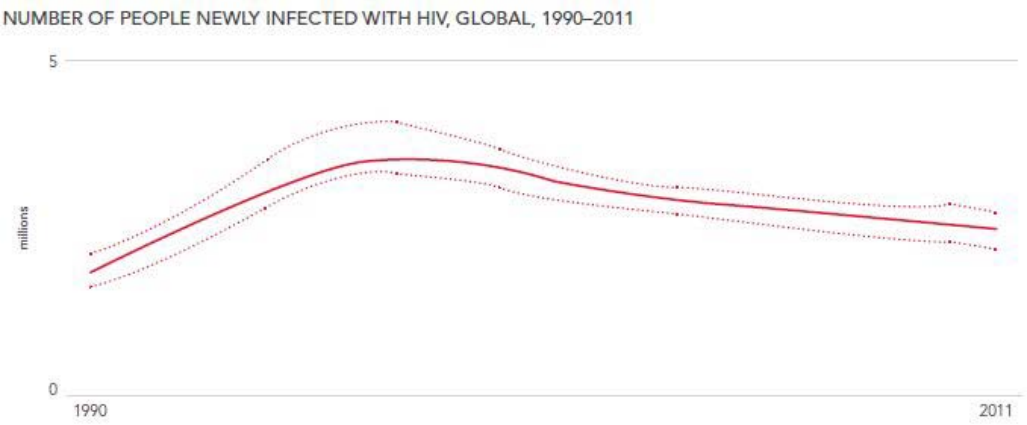

Figure 2. Graphical representation of the number of persons newly infected with HIV globally, in millions (y-axis), between the years 1990 and 2011 (x-axis).[1]

60 million people have contracted HIV and nearly 30 million have died of HIV-related causes, since the beginning of this epidemic in the late 1980s. Initially, HIV infection occurred predominantly within a subset of the population defined as men who have sex with men (MSM), but as the pandemic continued to progress through the years, the demographics affected evolved as well. In the late 1980's, MSM were the largest group at risk for HIV infection, but more recently heterosexual individuals and intravenous drug injection users have shown increased transmission rates. While several risk factors have been identified for HIV transmission, infection rates have begun to decline, due in large part to the action of public health interventions programs aimed at educating at risk individuals. However, much work remains with intervention development and implementation as it remains that over two thirds, approximately 69%, of all individuals infected with HIV, live in Sub-Saharan Africa.[2]

Two identified strains of HIV exist: HIV-1 and HIV-2. HIV strains are believed to have originated from Simian Immunodeficiency Virus (SIV). A cross-species transmission event most likely occurred to spawn both strains of HIV. HIV-1, the virulent HIV strain responsible for the pandemic, is most closely related to the SIV found in chimpanzees. Therefore, it is believed that

transmission most likely occurred through blood to blood contact during the butchering of SIV positive chimpanzees for bush meat. HIV-2, the rarer strain of HIV and concentrated in West Africa, is most closely related to the SIV found in sooty mangabeys. These small monkeys are kept as pets and so the transmission route is believed to have been through bites/scratches from an infected monkey. HIV-1 evolved quickly due to rapid replication errors, a quality discussed later in this introduction, and several HIV-1 subtypes arose. Currently, HIV-1 pandemic form, M for "main", is divided into subtypes: A, B, C, D, F, G, H, J, and K, as well as circulating recombinant forms (CRFs) and unique recombinant forms (URFs). Genetic variation within a subtype has been shown to approach 15 to 20%, while variation between subtypes is usually 25% to 35%. [3]

1.1.2 Viral Classification and Structure

HIV is an enveloped, single stranded, positive sense RNA virus with a genome approximately 9.8 kilobases in length. HIV is classified as a member of the *Retroviridae* family. The RNA genome of HIV encodes the retrovirus proteins Gag, Pol, and Env. The Gag protein is cleaved to form the matrix, capsid, and nucleocapsid proteins. Pol is cleaved into the protease, the reverse transcriptase, and integrase. Env is a 160-kiloDalton glycoprotein that is cleaved into the gp120 external subunit and the gp41 transmembrane subunit, which forms the spikes on the virion surface. The viral genome also encodes several regulatory and accessory proteins that aid in replication and virion packaging.[4]

All HIV strains possess a cone-shaped core, which is comprised of two constituents: the viral capsid protein and the Gag protein. Inside the capsid, two identical, positive sense RNA strands are closely associated with the viral RNA-dependent DNA polymerase (Pol) and the

nucleocapsid proteins. The inner lining of the virion membrane possesses myristoylated p17 core (Gag) proteins, which provide the matrix (MA) for, and aid in maintaining the integrity of, the virion. The virion membrane is comprised of the gp120 external surface envelope protein and the gp41 transmembrane protein. The gp120 envelope protein of the virus surface contains the binding sites for the cellular receptors and the major neutralizing domains.[5,6] The gp41 protein, meanwhile, has been shown to hold a crucial role in the depletion of CD4+ T cells by inducing the cell death.[3]

1.1.3 Infection Cycle

HIV is a blood-borne pathogen and can be transmitted via contact with contaminated blood or needles, sexual contact with infected individuals, or vertically from infected mother to child during birth.[7] Once HIV enters its host, it gains entry to a human cell through a coordinated binding cascade. First, the HIV virion gp120 surface protein component must bind the human cell surface receptor CD4, a 55 kiloDalton (kDa) molecule used by the T cell receptor for communication during antigen presentation. The CD4 surface receptor is present on T helper cells and macrophages. However, this interaction alone is not sufficient for viral entry. HIV requires the interaction of a coreceptor, a second cell surface receptor necessary for successful attachment and entry of the virion into the host's cell. HIV coreceptors are CCR5 and CXCL4, which are chemokine receptors expressed by CD4+ immune cells. Fusion of the viral envelope to the cellular membrane is achieved by the binding of gp41.[6–8] Once viral entry has occurred, reverse transcription of the viral RNA genome is initiated and a double stranded DNA version of the HIV genome is generated. The newly synthesized proviral DNA is then integrated into the host cell's genome by the virally encoded integrase (IT) enzyme.[8]

Initial viral infection presents with flu-like symptoms for a span of 1-3 weeks. During the initial infection, virus replication is rapid and brings with it a reduction in the CD4+ T cell population. Roughly 14-21 days after the initial infection, a strong cytotoxic T lymphocyte response is observed which drives down blood plasma viral titers. After approximately 6-8 weeks, the immune system has generated HIV specific antibodies which reduces, but cannot wholly abolish, the viral load in the plasma. This reduced, but consistent, level of virus present in the plasma is defined as the set point for the virus and is a useful predictor for an individual's rate of progression to AIDS, the endpoint condition of HIV infection. AIDS arises when the host's T cell population has been depleted to a point where an insufficient cell population is present for mounting an effective immune response. It has been observed that HIV rarely causes the host's death, but rather, opportunistic infections can overwhelm what remains of the host's immune system after the T cell population had been depleted by HIV infection. One factor, which is believed to contribute to the ability of HIV to evade the host immune system response and antiretroviral drugs, is the highly error-prone nature of the viral reverse transcriptase. The reverse transcriptase incorporates mutations within the replicating HIV genome and enables escape mutants to be formed, which can then proceed, unregulated by the immune system. [4,7]

1.1.4 Treatment

The advent of antiretroviral therapy can be broadly placed at a point in the mid- to late-1980s beginning with a study by Fischl *et al* that examined the effect of zidovudine (AZT), a nucleoside analog Reverse Transcriptase Inhibitor (NRTI), versus a placebo on HIV infection. Analysis of the study participants at six months post study initiation, showed a reduction in frequency of opportunistic infections and increased survivorship among the AZT treatment

group as compared to the placebo, which resulted in a halting of the study and placement of all participants on AZT regimes, since a clear benefit to treatment was observed. This initial experiment and several subsequent studies brought about the recommendation that therapy should be provided to any patient, regardless of the presence of symptoms, who presented a CD4+ cell count less than 500 cells/μL. Studies since have shown mixed, and in some instances conflicting, conclusions as to when is the most effective time point to initiate therapy, however the fact that therapeutic interventions are necessary for the treatment of HIV was undeniably confirmed in all studies since the advent of HIV/AIDS.[9] More recent studies have shown that combination therapy is the most effective means of reducing the capacity of HIV to replicate in a host's cells. Combination therapy is a treatment program in which several classes of drugs are used in concert with one another, as a means of having a synergistic effect on reducing the replicative ability of HIV. The formalized term for this combination treatment strategy is Highly Active Antiretroviral Therapy (HAART). HAART does not cure an individual infected with HIV, but rather attempts to limit viral replication and viral RNA production to the highest degree possible, while attempting to limit adverse affects on the health of the infected individual. HAART is able, in many cases, to reduce HIV levels below the limit of detection of most clinical assays, effectively preventing disease progression and greatly extending the life of the individual.

Recent studies have shown that a decrease in HIV viral RNA reduces the risk of disease progression and death independent of the baseline CD4 count or the subsequent increase in CD4 count once therapy has begun.[10] Therefore, treatment efficacy evaluations are based on how the viral RNA load responds to antiretroviral therapy regimen. If the viral load decreases, then therapy is deemed effective and utilized, however if viral load increases, then therapy is

considered to be ineffective and a new combinational drug therapy is devised once other possible contributing factors to viral load increases have been ruled out.[10]

1.1.4.1 Reverse Transcriptase Inhibitors

Initially, reverse transcriptase inhibitors were developed as a means of treating HIV infection by attempting to prevent viral replication. However, mono-therapy was soon recognized as an ineffective means of controlling HIV's replication within a host, as escape mutants arose which avoided the action of the drug, and so new drug classes were slowly identified and the combinational therapies were devised. The first anti-HIV drugs to be developed were the Nucleoside analog Reverse Transcriptase Inhibitors (NRTIs), and prodrugs that require metabolism to become NRTIs. This class of drug needs phosphorylation to become active, and this phosphorylation is necessary because this class of drugs inhibits viral replication by competing with natural deoxynucleotide triphosphates (dNTPs) at the active site of the viral RT for subsequent incorporation into DNA. Once the NRTIs become incorporated to the growing DNA strand, the NRTIs act as a chain terminator because they lack the hydroxyl group on the 3' end, necessary for elongation of the chain and polymerization is halted.[8]

Non-nucleoside Reverse Transcriptase Inhibitors (NNRTIs) have a different mechanism of action than NRTIs but retain the same goal of inhibiting the action of RT. NNRTIs bind to RT at a site that is distinct from the substrate binding site, and they do not require phosphorylation to become active as they do not act as competitive inhibitors of RT via DNA chain termination.[8] This ancillary action of NNRTIs is the reason why they are utilized in combination with a drug from the NRTI class.

While effective when utilized appropriately, the RT inhibitors are not without disadvantages. The drugs present issues such as toxicity concerns, an increase in viral resistance

when used as a mono-therapy, and the fact that once the HIV genome becomes integrated, the role of RT is no longer required for the replication of the virus.[11]

1.1.4.2 Protease Inhibitors

The viral genome of HIV-1 requires the action of a protease, an enzyme which cleaves a proprotein, to generate the biologically relevant forms of several of the viral proteins. For the specific protease of HIV-1, the protease cleaves the polyprotein of the gag precursor into the two subunits which are necessary for HIV infectivity: p24, the capsid protein (CA), and p17, the matrix protein (MA). Since these two proteins are needed for infectivity, a class of drugs which sought to inhibit the action of the HIV-1 protease was developed.[6,8]

The action of the protease is also pivotal to the maturation of the HIV virion and so interference with this action diminishes the capacity of the virus to complete its infection cycle and continue replication. The action of the protease inhibitor on the protease enzyme results in the release of structurally disorganized and non-infectious viral particles.[10]

The combination of agents that can inhibit de novo infection of cells (i.e., RT inhibitors) with agents that lead to the release of defective, noninfectious virions from cells already infected by HIV (i.e., protease inhibitors) results in potent antiviral activity. Protease inhibitors have rapidly become important components of combination anti-HIV drug regimens.

1.1.4.3 Consequences of HIV Infection and Treatment

Abnormalities in serum lipid concentrations have been associated with both HIV-1 infection itself and its treatment with highly active antiretroviral therapy (HAART). In addition to lipid changes, HAART may result in metabolic disturbances such as insulin resistance, peripheral lipoatrophy, and central adiposity. Large prospective, observational studies have

suggested that use of HAART, particularly HIV-1 protease inhibitors (PIs), was associated with an increased risk of myocardial infarction that is partially mediated by serum lipid abnormalities. The apparent risk to heart health and risk of heart disease is addressed in the experimental design of this thesis. However, there is another disease of HIV infection which has garnered increasing focus as HAART showed the capacity to increase the quality and length of life for HIV-positive patients. The condition in question is known as HIV-Associated Neurocognitive Decline.

1.1.5 HIV-Associated Neurocognitive Decline (HAND)

It has been well established that HIV is the causative agent for the manifestation of Acquired Immunodeficiency Syndrome (AIDS) in HIV-positive individuals, but what has become increasingly clear, in recent studies, is that HIV infection can bring with it a litany of other disorders. Recent research has focused on understanding the neurocognitive disorders that HIV has been attributed to causing in seropositive individuals on HAART. One of the first associations between HIV and neurocogntive decline was identified and published by Navia *et al* in 1986.[12,13] In one of the landmark works on the subject, Igor Grant and his research team published one of the first connections between HIV infection and the observed neurocognitive disorders which manifested in these individuals.[14] It has been shown since, that HIV has been associated with a wide array of neurocognitive complications and that these complications have been categorized under the abbreviation HAND, which stands for HIV-Associated Neurocognitive Disorders. These various complications, now associated with HAND, have been largely attributed to neuropathological changes, within the basal ganglia and the white matter of the brain, resulting in mental slowing, a loss of memory, motor disorders, and difficulties with

tasks which require executive functions.[15] While these disorders have been attributed to HIV, the exact mechanism by which the effects are generated, has yet to be elucidated.

1.1.5.1 Clinical Definition of HAND

To understand the importance of HAND to the objectives of this study, it is beneficial to first understand the nature and characteristics of HAND. In the early years of the HIV epidemic, there was only one category for the observed neurocognitive impairments that manifested during HIV infection and progression. This most severe presentation was defined by clinicians as HIV-associated dementia (HAD). HAD was a prominent outcome of HIV infection early in the AIDS epidemic, occurring in nearly 15% of individuals prior to treatment [16], but most usually presented in patients with low CD4+ cell counts and who were in the advanced stages of HIV progression. The advent of HAART in the mid-1990's improved the available treatment of HIV infection and subsequently altered the progression of HAD in these individuals as HIV disease progression was also altered by the combinational drug therapy. As the newly developed combinational therapies began improving, the morbidity and mortality of HIV-positive individuals and the incidence of the dementia as a hallmark symptom of AIDS also increased.[17] This increase has been so profound in fact, that HAD now stands as an independent risk factor for death due to AIDS.[18]

In 2007, a new classification system was designed to streamline the case definition of HIV-related Central Nervous System (CNS) impairments. This new system also sought to include the milder presentations of neurocognitive disturbances in HIV-positive patients, instead of simply the most severe presentation, HAD. The newly revised nomenclature, known as the *Frascati criteria*, created two additional categories for neurocognitive impairment to account for the more mild manifestations of HAND. The first category is termed asymptomatic

neurocognitive impairment (ANI) and the second is minor neurocognitive disorder (MND). The clinical definition for these two mild categories is impairment as determined by scores lower than the norms on neuropsychological tests. HAD is defined as an individual having a neuropsychological test result that is greater than 2 standard deviations below demographically-related means.[19] The only symptom which separates the classifications of MND and ANI, is that ANI has no overt interference in the daily functions of the individual, whereas MND individuals have a mild level of interference on their daily functioning.[19]

As described above, when HAND presents as a dementia, there are several hallmark features in which neurocognitive declines are predominant features. HIV-1 encephalitis (HIVE) is a histologically identifiable condition that presents in samples taken from HAND-positive individuals. HIVE has been shown to occur in most incidents of HAD, the most severe form of HAND, but the presence is not always 100%. Histologically, HIVE manifests at the neuropathological level, through the loss of a neuronal subpopulation in the neocortex, limbic system, and basal ganglia in association with synaptic and dendritic damage, astrogliosis, and formation of microglial nodules and multinucleated giant cells.[20]

1.1.5.2 HAND Pathogenesis

The need to understand the cause(s) of HAND is of immense importance to the field of HIV study because individuals who are HIV-positive are now living longer, healthier lives due in large part to the benefits seen from HAART. However, HAART's effect on controlling viral replication has brought with it increases in disease conditions outside of the traditional AIDS classification. This new category of diseases is known as non-AIDS defining illnesses or NADIs. Since these disease conditions are now becoming prevalent in the HIV-positive population, and are therefore becoming of increasing concern to the public health and clinical sectors, it is

important to consider what is known regarding the way in which HIV infection contributes to the pathogenesis of HAND in HIV-positive patients.

In a study conducted by Koenig *et al*, on HIV-positive samples collected during the early years of the AIDS epidemic, it was determined that mononucleated and multinucleated macrophages were the major cell types generating HIV in the brains of the HIV-positive patients with AIDS.[21] Once this characteristic was confirmed experimentally, it next became important to understand how the virus came to be present in the brain tissue of these HIV-positive AIDS patients.

Subsequent research identified HIV tropism to specific cells of the immune system, suggesting that HIV traveled to the brain by way of infected monocytes and CD4+ T-lymphocytes crossing the blood-brain barrier (BBB). As described previously, the immune cell tropism of the HIV virus is determined by the cellular surface receptor which the gp120 glycoprotein interacts with on the host's cells. Macrophage-specific HIV viruses use the CCR5 coreceptor, while those specific for T-cells use the CXCR4 coreceptor. HIV has also shown dual tropism, where both receptors are utilized, as well as a broad range of other receptors being employed for viral entry.[20]

Post-mortem sample studies have suggested that macrophages, as well as microglial cells, those cells response for the care, maintenance, and protection of the CNS cells, can contribute to the pathogenesis which is a hallmark of neurological disease associated with HIV infection.[20,22] While the explicit mechanism, by which this recruitment and activation leads to damage, has yet to be defined, studies have shown that chemokine and chemokine receptors experience upregulation in HIV-positive, HIVE-positive patients.

The presence of the chemokines and receptors was examined in the Cerebrospinal Fluid (CSF) of HIV infected, cognitively impaired patients, and compared to the chemokine levels of non-infected control samples. The work focused on determining the concentrations of MCP-1/CCL2, MIP-1alpha/CCL3, MIP-1beta/CCL4, RANTES/CCL5, IL-8/CXCL8 and fractalkine/CX3CL1. From these comparisons, it was determined that a positive correlation existed, most strongly for MCP-1, between the CSF concentration of the chemokine, the CSF viral load and the severity of the dementia presenting in the HIV-positive patients. MCP-1, known as a potent monocyte-attracting chemokine, was indicated to possess a correlation to pathogenesis of inflammatory diseases of the CNS characterized by macrophage infiltration. This correlation suggested that HIV may be a precipitating factor for brain damage which subsequently manifests in the form of HAND through increased macrophage infiltration of the CNS.[23,24] From this data, several studies were conducted in the years that followed to identify possible mechanisms or other contributing factors that could aid in determining the role these chemokines and/or the chemokine receptors play in the development or progression of HAND.

Cartier *et al* studied the roles and localizations of these chemokines and chemokine receptors, and determined that several of them, including the known HIV co-receptor CXCR4, were each upregulated in regions of the brain known to be associated with the Limbic system, a brain region possessing key roles in memory and other cognitive functions.[20] These authors also showed that CXCR4 may be directly involved in gp120-mediated neurotoxicity through the activation of neuronal receptors by gp120 recognition, or indirectly through the stimulation of glial cells to release neurotoxic factors.[20] This study is important to understanding the mechanism by which HIV infection and localization in the brain could be leading to the adverse effect on memory and cognitive functionality, hallmark symptoms of diseases such as HAND.

HAART drugs have shown considerable efficacy in reducing viral RNA load and inhibiting infection, however only a few of the HAART drugs consistently cross the Blood Brain Barrier. Therefore, the seemingly contradictory nature of how HIV-positive patients, undergoing HAART, could have viral titers in the brain above those found elsewhere in the body becomes clear. The drugs effectively inhibit the viral replication in the blood and immune cells but with limited penetrance to the neural tissue, localized viral replication increases.

1.1.5.3 Treatment Options for HAND

Grant *et al*'s work, on the presentation of neurocognitive impairments during HIV infection and progression to AIDS[14], presented a link between a disease state, currently defined as HAND, and the progression of HIV infection to AIDS. Since the work's publication in 1987 however, the majority of research has been focused on identifying mechanistic traits of the development of HAND and the other diseases HIV infection can beget. While the development of HAART and continued laboratory research have paved the way for many significant advances in the management of HIV infection, a similarly productive progression in the treatment and management of neurological symptoms of HIV infection has not been observed. Research has yielded several pharmacological therapies to aid in improving the neurocognitive outcomes of these patients; however study designs which critically evaluate the efficacy of neurocognitive rehabilitation for alleviating the HAND symptoms are still greatly lacking.[25]

Another pressing issue surrounding therapeutic interventions for HAND is that clinicians have identified a 16-21% false-positive rate for the MNI and ANI classification of HAND patients when using the current classification standards for neuropsychological examinations, the *Frascati criteria*.[19] In addition to this high false-positive rate it has also been suggested, by various studies, that HAART cannot completely protect an individual from HAD, the most

severe form of HAND, and that long-term exposure to HAART has also been linked to toxicological issues that may result in the neurological tissue damage that is the hallmark of HAND.[17]

This brief discussion of the history of HAND research has shown that the greatest focus has been placed on the understanding of the development and pathogenesis of HAND, but that little research has yielded positive therapeutic interventions. In addition to this trend, there is a high false-positive rate which can lead to a crippling lack of a definitive identification of the risk, an HIV-positive patient possesses for the development of HIV-associated neurocognitive decline. Therefore, it is the objective of this project to identify objective, biological traits, known as biomarkers, which can be experimentally utilized, to assess an individual's risk of HAND progression once HIV-positive status has been identified. The remainder of this introductory section will discuss the history of the four biomarkers this research sought to examine for use in assessing their correlation to HAND progression.

1.2 PREDICTIVE BIOMARKERS AND HIV-ASSOCIATED NEUROCOGNITIVE DECLINE

Research in the development of therapies and interventions to the underlying condition is critical, but it is also pragmatic to have objective measures, by which individuals suspected of being positive for a disease can be assessed. It has been stated previously, in the literature focused on this subject that a disparate amount of research focus has been allotted to the determination of therapeutics and mechanistic understandings of the infection which lead to HIV-associated Neurocognitive Decline. However, the troubling reality is that the measures by which a HIV-

positive patient is assessed and classified, in the spectrum of HAND severity, are too subjective and inconsistent, often resulting in misdiagnoses and inadequate determination of disease severity in the clinical setting. This fact highlights the necessity for the identification of biologically relevant measures by which an individual can be objectively examined and assessed for their risk(s) for the development of future conditions post infection, such as HAND. These measures, henceforth referred to as predictive biomarkers, are paramount to the successful identification of individuals most in need of treatment and can potentially aid in the determination of the patient's treatment strategy.

These predictive measures potentially possess the power to identify an individual's risk for a particular condition, prior to the clinical manifestation of that disease state. It is the objective of this research project, to assess the capacity of four biomarkers independently identified, outside the context of HIV infection and the development of HAND, which can be applied to assessing an individual's risk for the development of HAND.

1.3 HOST GENETIC BIOMARKERS

A brief discussion of the four predictive biomarkers selected for this experimental design will serve to summarize their importance, as they relate to each marker's association to the development of dementia, as well as the way in that association is relevant to a neurocognitive declining state within an HIV-positive patient, represented in the presence of HAND.

1.3.1 RYR3 Single Nucleotide Polymorphism

Therapeutic interventions for HIV infection have brought with them incredible advances in the quality of life and control of the viral load for HIV-positive individuals. However, research has shown that individuals have an increase in risk for developing detrimental health conditions, as a result of exposure to HAART regimens. Cardiovascular disease (CVD) has now become a leading cause of morbidity and mortality for persons with HIV. Risk factors, which have been attributed to setting the stage for the development of CVD within the HIV-positive population, include a greater prevalence of traditional coronary artery heart disease risk factors, such as smoking, hypertension, diabetes, fibrinogen, and low-density lipoprotein cholesterol (LDL-C). In addition to these traditional risk factors, research has also shown that HIV infection itself, and the combination drug therapies utilized to combat HIV disease progression are also driving the manifestation of CVD consequences in HIV-positive patients.[26,27] There are a growing number of reports on lipodystrophy, which is an abnormal accumulation of fat without body weight variations, in patients treated with HAART.[9] These traditional risk factors for CVD are also associated with preclinical atherosclerosis, generally measured as the intima-media thickness (IMT) of carotid arteries by B-mode ultrasound.[27] Armed with the understanding that individuals on HAART are enduring an increased risk for the development of CVD, as well an increased prevalence of traditional risk factors for the development of CVD, it became pertinent to seek out the existence of predictive biomarkers that could identify the risk in HIV-positive, HAART-positive individuals for the genetic predisposition for the development of CVD. Pre-existing CVD conditions or risk factors for CVD development early in life, have shown correlation to the development of dementia later in life. A study conducted by Wright *et al,* found that individuals with pre-existing CVD risk factors possess a 6.2-fold increase in the risk for developing forms of

HAND. This link to pre-existing CVD risk factors highlights the predictive power of CVD risk factors for dementia development. [16]

In 2010, a study conducted by Shrestha *et al* sought to identify, through use of a genome-wide association study (GWAS), any SNP biomarkers which could be used to assess atherosclerosis risk in HIV-positive individuals. Common and internal carotid intima-media thicknesses (cIMT) were measured by B-mode ultrasound, as a subclinical measure of atherosclerosis. From the patient samples taken, SNPs were assayed and assessed for correlation to the clinical manifestation of atherosclerosis. Copy Number Variation (CNV) was also assessed by statistical analysis. The GWAS conducted identified two SNPs, rs2229116 (a missense, ~~A~~ G substitution) and rs7177922, located in the Ryanodine receptor (RYR3) gene on chromosome 15. These two SNPs were also significantly associated with common cIMT (p-value$<1.61\times10-7$). [28,29]

RYR3 is a member of a family of genes known to encode calcium (Ca^{2+}) ion channels. RYR3 is known to be coexpressed within the cardiac skeletal muscle tissues of the body and has shown associations to endothelial vasodilatation, a process compromised in atherosclerosis. Since the explicit functioning of RYR3 is yet to be fully realized, the authors made the hypothesis that RYR3 could potentially share similar linkage to CVD outcomes to those observed with RYR1 and RYR2. [28,29]

A replication study of this association, using a sample cohort of 244 individuals enlisted in the Multi-Center AIDS Cohort Study (MACS) CVD sub-study was conducted at the University of Pittsburgh. The genotype of two RYR3 gene SNPs were determined by Taqman Assay. The results of the initial study were replicated in the study conducted using MACS CVD sub-study samples, suggesting that this gene is indeed associated with CVD risk. However, there

are several caveats to these results that warrant discussion. Firstly, the measurement methodologies for atherosclerosis risk between the two studies, the exploratory study was predicated on and the CVD sub-study of the MACS, were different and so this could impact the results which were obtained by the SNP assay. More diverse sample populations, ie various ethnic groups, age ranges, etc, would greatly aid in determining and differentiating the mechanism which drives this observed linkage between the SNP and cIMT. Also, there are a large number of splice site variants of the RYR3 gene and so this could further compound the association and further research was deemed warranted by our collaborators and is a future research aim for their study. [28,29]

The current study sought to examine the SNPs within the RYR3 gene as a predictive biomarker, to assess the existence of a correlation between RYR3 point mutations, rs2229116 and rs7177922, and the progression to HAND in HIV-positive, HAART-positive individuals. Previous research has shown that individual undergoing HIV treatment comprised of Anti-Retroviral Therapy (ART) can reduce their atherosclerosis risk, by reducing the inflammation caused by HIV viral activity. It has also been shown that infected individuals have subclinical atherosclerosis, associated with T-cell activation and elevated levels of other inflammatory markers, as compared to those without infection.[30–32] HAND, as explained previously, has shown that disease progression is the linked to the inflammation caused by HIV activity and the subsequent upregulation of pro-inflammatory cytokines and receptors in the brain of infected individuals. Therefore, it is the hypothesis of this study that RYR3 could serve a useful purpose in assessing both risks for CVD as well as HAND in the HIV-positive population. Since both conditions arise as a result of pro-inflammatory conditions, it is the hypothesis of this study that

the two SNPs of the RYR3 gene have the potential to correlate to HAND development and progression and could therefore potentially serve as predictive biomarkers for HAND.

1.3.2 Apolipoprotein E

As considerations were being made to identify markers which exist elsewhere, to aid in determining the link to neurocognitive decline in the HIV-positive population, it seemed logical to include previously associated biomarkers that correlated to an increased risk of dementia, regardless of infection status. One such predictive biomarker identified to have a link to a risk for dementia in the healthy population, which could then be applied to the HIV-positive population, as a potential predictive biomarker for dementia in the form of HAND, is Apolipoprotein E (ApoE).

A strong association between the ApoE e4 allele (APOE*4) and increased risk of Alzheimer's disease (AD) has been established, based on extensive clinical and basic studies. APOE*4 carriers have an increased risk of late onset AD approximately 3 to 15 times greater than non-carriers, in a gene dose-dependent manner. The APOE*4 carriers also encounter ages of AD onset 10 to 20 years earlier than non e4 carriers. [33] The two main brain functions affected by the progression of AD are thought and memory, but other presentations and symptoms have been identified as research on this disease has progressed. In 1984, the National Institutes of Communicative Diseases and Stroke and the Alzheimer's Disease and Related Disorders Association drafted criteria that would serve to identify a clinical presentation for AD. McKhann *et al* outlined various judgment criteria which would ultimately lead to the identification of three classifications of AD: Possible, Probable, and Definite. The published report sought to classify a patient's presentation of AD by examining medical history, clinical examination,

neuropsychological testing, and laboratory assessments. Once the group's determined potential criteria were settled, they were discussed among study participants and a finalized set of criteria were established for clinical diagnoses of AD. It is important to note that these criteria sought to exclude confounders in the diagnosis of AD and so other causes of memory loss and impaired cognitive function were excluded. [34]

The patient with AD experiences gradual increases in forgetfulness, diminishing attention span, and alterations in mood, often with frustration and agitation, resulting in increasing difficulties in meeting the demands of daily living. These problems progress until the patient ultimately cannot attend to his or her simplest needs and becomes bedridden, totally dependent on caregivers. The interval between initial diagnosis and death varies considerably, usually between 3 and 15 years. According to the defined diagnostic criteria, a diagnosis of definite AD can be made only by microscopic examination of brain tissue, either at biopsy or, more commonly, autopsy. The neuropathological criteria, also established by a consensus group (Khachaturian 1985); require the presence of neuritic plaques and neurofibrillary tangles (NFTs), at specified densities. The diagnosis of definite AD has therefore been defined by phenotypic neuropathological findings. The relationship between the plaques and tangles and the mechanism causing Alzheimer's dementia is unknown and somewhat controversial. Discovering the relationship between AD neuropathology and the pathogenesis of the disease is a central issue in developing and testing hypotheses to uncover the molecular and cellular mechanisms that ultimately result in the dementia.[35]

In 2011, The National Institute on Aging and the Alzheimer's Association tasked a workgroup to revise the criteria for successful identification of AD, which had been established in 1984. A key component to the new workgroup's task was to ensure that their new criteria were

flexible enough to be of use to general healthcare providers lacking access to neuropsychological testing, advanced imaging, and cerebrospinal fluid measures, but not too relaxed as to reduce the criteria's value to the investigators possessing these advanced metrics. The general framework of probably AD dementia, from the original 1984 criteria, was kept in the 2011 revisions. Improvements to clinical criteria for diagnosis were made and further refinements to the definition of the AD classifications were made in the new report. As new research has identified biomarkers, such as ApoE genotype, that can impact risk for and diagnosis of AD, these two were integrated into the new workgroup's criteria. At its core however, AD dementia will continue to be the foundation for clinical diagnosis, but biomarkers and other characteristics will further aid in the manifestation of symptoms and identifying the specific form of AD in patients.[36]

Given the nature of the association observed between the APOE*4 allele and the risk for AD, its inclusion in the study of neurological issues culminating in dementia within the HIV-positive population, appears practical. While it is true that individuals who possess the APOE*4 allele are at the highest risk for developing AD, the inclusion of the genetic predictor for this particular form of dementia serves a crucial objective in regards to HAND. HAND is a disease whose onset is observed in a manner disproportionate to age. The age is disproportionate because the neurological conditions, which are observed as HAND progresses, manifest in patients at earlier ages than are expected for neurodegenerative conditions, much the same as AD in individuals possessing APOE*4. Therefore, by assessing the genotype of ApoE present, one is assessing a host genetic factor's potential to augment the progression of HAND in HIV-positive individuals. Since HAND and AD resulting from APOE*4 expression share the analogous trait of having a disease of aging encountered in patients who are younger than when symptoms

normally manifest, a further examination of possible correlation between ApoE genotype and HAND progression was desired.

1.3.3 Telomere Shortening

Dementia is often referred to as a disease of aging, in so far as symptoms largely present in individuals considered biologically advanced in age. HAND, on the contrary, is a disease in which the symptoms of neurological decline, culminating in dementia, occur in individuals who are not biologically advanced in age, an unexpected trait and a focal point of current research. Therefore, this study sought to identify a potential measure, which could be employed within the context of HAND, which showed independent association to other disease states, relating to dementia and diseases of aging. With this goal in mind, it became important to identify a biological trait known to relate to senescence, which could serve as a potential predictive biomarker for the risk of age related conditions for this study design. Extensive research has been conducted on the field of senescence, or biological aging, and it has been shown that the ends of the chromosomes, the telomeres, have positive association with determining biological age and one's risk for developing diseases of aging.

Telomeres are concatemers of "TTAGGG" sequences at the ends of DNA molecules. These repeats are true, conserved, repeat sequences and not remnants of splicing or replication errors. Telomeres have been shown to possess associated proteins, which aid the telomere's function of maintaining genomic and cellular stability and replication. It has been shown that the telomeres shorten as repeated cellular divisions occur and that once a critical threshold of telomere length has been surpassed, the cell enters a stage of senescence and then the cell death, or apoptotic, pathway. Knowing that a critical telomere length signaled the death of a cell, the

telomere is often considered as a potential biological clock by which the lifespan of a cell can be determined.[37,38] Since telomere length is a critical signal to the cell for the initiation of senescence and cell death it seems logical that the cell would need to maintain telomere length or cells would die too quickly. It was this thought that lead to the identification of an enzyme known as telomerase which is responsible for maintaining the length of the telomere, thereby extending cell viability and function. It has been shown that telomere length is equivalent in different organs and blood leukocytes from the same individual at birth and later in life, thus confirming the role of telomerase in the maintenance of telomere length across cellular divisions.[39] To measure telomere length, it is most common to measure the Leukocyte Telomere Length (LTL) by Southern blotting or qPCR, but this measure can be misleading when considering tissues with low mitotic activity, such as the brain, because these tissues usually possess longer telomere lengths inherently. Southern blotting has been shown to be more accurate, but qPCR is more cost effective. [37]

As addressed previously, repeated cell divisions progressively shorten the telomere regions of the DNA in a cell. Therefore, aging is a leading cause of telomere shortening, a process formally referred to as attrition. As a cell divides, the telomeres become shorter, and as the telomere length diminishes, the cell becomes closer to entering cell death. The aging of the cell is a cause of the telomere attrition and thereby also can be associated with other causes of aging, such as oxidative stresses and inflammation. [37]

In a number of cross-sectional studies, short peripheral blood cell telomere length was associated with an increased prevalence of various age-related diseases including myocardial infarction, atherosclerosis, Alzheimer's dementia, dementia after stroke, and other cardiovascular comorbidities.[37,40,41] To have associations with such a breadth of disease states, one begins to

question whether shortening of the telomere is linked, directly or indirectly, to the causal pathway of these disorders. These associations also raised into question whether an acceleration of the rate of telomere shortening could expedite the onset of these age-related diseases? Martin-Ruiz *et al,* in a study of telomere length as a predictor for poststroke mortality, observed that telomere length measured within 3 months of stroke was a significant predictor of risk for dementia within 2 years after stroke. [42] This study showed that telomere length measurement had predictive power as a biomarker of neurocognitive decline in poststroke patients.

1.3.4 Cytosine Residue Methylation

The naturally occurring, chemical methylation of DNA has been theorized to play an integral role in gene expression and regulation for some time; however research has shown that DNA methylation is not an explicitly conserved process, as some organisms completely lack the capacity to methylate DNA. However, DNA methylation has been shown to play a vital role in several DNA regulatory mechanisms, such as chromatin packaging and transcription, in vertebrates. [43] Since DNA methylation has been shown to hold a role in the pathway for gene expression, as well as regulation, why is it important for vertebrate gene regulation, and is it possible that methylation contributes to the disease condition of HAND?

1.3.4.1 DNA Methylation

DNA methylation is an enzyme-catalyzed process in which a methyl group is biochemically transferred from S-adenosyl-L-methionine to specific DNA residues within the genome of an organism. On the biochemical level, methyl group additions occur through the action of enzymes known as DNA methyltransferases. The process of DNA methylation is

carried out when a methyl group is taken from the universal methyl donor, S-adenosyl-L-methionine, and added to one of the DNA nucleotides Adenine (A) or Cytosine (C). In eukaryotes, DNA methylation only occurs in the context of cytosine resides in DNA and more specifically, occurs only in the context of Cytosine-Guanine residue tandems, known as CpG islands.[44]

DNA methylation is defined as an epigenetic change. This term literally means "above the gene" and is fitting, since the biochemical addition of methyl groups to DNA residues serves to augment the existing base, without causing a heritable base pair change in the way that a genetic mutation does. DNA methylation also exhibits functionality dependent on the context in which it occurs. The location within the gene where methylation occurs has different outcomes upon gene regulation. Methylation of gene promoters represses the promoter's corresponding gene. Conversely, methylation of DNA within the gene's coding region has been shown to be increased in active transcription.[45,46] Therefore, methylation of DNA has both positive and negative regulatory power over gene expression and it is the location of the CpG methylation site that conveys the nature of the control over the gene in question. Knowing that methylation has regulatory power over gene expression, it next became necessary for scientists to be able to examine the methylation of DNA as a means of understanding its influence in a biological setting.

Scientists devised three main biochemical methodologies to enable the study of DNA methylation: Endonuclease Digestion, Affinity Enrichment, and Bisulphite Conversion. Endonuclease digestion is a process that combines two important components of a biological system, endonucleases and methylation. Endonucleases cleave DNA at sequence-specific restriction sites, and over time, organisms have evolved a means of protection against undesired

restriction site digestion. That defense is to methylate the restriction site at the Cytosine residue. Some endonucleases are sensitive to this modification and are therefore no longer effective at cleaving the DNA. This sensitivity was observed and endonucleases were further identified to be either, methylation-sensitive or methylation-indiscriminant. Therefore, in the process of endonuclease digestion, scientists can examine DNA for methylation patterns by using methylation-sensitive and methylation-indiscriminant endonucleases and observing how the restriction digestion patterns change differentially between the different endonucleases.

Affinity enrichment is a process in which antibodies, specific for the methylated carbons of CpG islands, recognize their specific methylation sites and bind to them, forming complexes that can then pulled down using Chromatin Immunoprecipitation (ChIP) for further investigation by array hybridization of next-generation sequencing. Affinity enrichment can also use methyl-binding proteins with affinity for methylated native genomic DNA.[45] These techniques are useful and informative means of examining methylation, however the technique that was selected for analysis of the samples in this experimental design was process known as Bisulphite conversion.

Bisulphite conversion is an experimental procedure in which DNA is treated with sodium bisulphite ($NaHSO_3$), as a means of detecting the Cytosine specific methylation of a genetic sample. This chemical treatment deaminates non-methylated Cytosine residues much more rapidly than it does methylated Cytosines. This unique chemical disparity is important, because as a Cytosine is deaminated, an epigenetic difference can be manipulated into a genetic difference that can be detected using standard genetic analysis techniques. Unmethylated Cytosines become deaminated in the presence of sodium bisulphite, resulting in the replacement of the deaminated Cytosine with Uracil. The newly substituted Uracil, not conventionally present in DNA, is then replaced with Thymine as the DNA polymerase copies the DNA during PCR of

the treated samples. Therefore, the unmethylated Cytosines in the DNA will be converted to Thymines, while the methylated Cytosine residues will remain unaffected. This pretreatment is very important when utilizing any biomolecular process of analyzing DNA methylation because the methylation information of a cell's genetic material is effectively erased by standard molecular biology techniques. Therefore, it was necessary to devise methylation-dependent pretreatments of DNA which could maintain the methylation for analysis. [45] The fact that the methylated Cytosines remain in the newly amplified PCR product is what gives this technique its analytic power. The amplified product is then examined by hybridization or next-generation sequencing platforms to determine the location of CpG island(s) that is/are methylated.

1.3.4.2 DNA Methylation and HAND

DNA methylation patterns must be properly established and maintained for mammalian development and for normal functionality of DNA in the adult organism. As was discussed previously, DNA methylation is a powerful tool for the regulation of gene expression and maintenance of genome stability. It is when this regulation of DNA methylation is not tightly regulated, that the propensity for disadvantageous genetic events can occur. Incorrect and/or detrimental recombination events can arise, leading to transcriptional deregulation of genes. [46] This discovery is embodied in 1983 discovery of a link between DNA methylation and cancer. It was first described by Feinberg *et al*, that the genomes of cancer cells are hypomethylated relative to their normal counterpart. [47] This reduction in methylation across genes, most notably in the repetitive DNA sequences, sets the stage for the aberrant gene expression that is indicative of cancer.

Hypermethylation possesses a similar effect. Unregulated and excessive methylation, or aberrant hypermethylation, has been reported in cancer to occur at CpG islands which are

normally unmethylated in non cancerous cells. Since methylation has the capacity to upregulate or downregulate gene expression depending on context, the seemingly counterintuitive fact that hyper- and hypo-methylation can increase gene expression becomes clear.[46] Depending on the location of the aberrant methylation of DNA, the gene expression profiles can be increased or decreased, and so the context of these methylation changes become vital to determining what downstream effect, if any, these changes will impart. Methylation changes, like those observed in cancers, have also been shown to hold association with a variety of biologically relevant situations associated with aging and other genetic events, both aberrant and essential. [48]

Methylation control of gene silencing has been shown to enforce its effect through the use of noncoding RNA (ncRNA) molecules. Therefore, when the regulation of these ncRNA molecules becomes dysfunctional, it can lead to the genesis of pathologies in cells and the tissues they comprise. These pathologies can manifest in the forms of cancer, various neurological and cardiovascular diseases, and also bring about the initiation of senescence. [49] The knowledge that methylation can enact both genetic control, and also serve as the source for the manifestation of a variety of pathological situations associated with or related to neurological and cardiovascular maladies, is evidence for the consideration of a potential link between methylation pattern changes and to a condition such as HAND, which has shown association with a variety of risk factors for neurological and cardiovascular diseases.

1.4 HYPOTHESIS AND STUDY DESIGN

The information provided above forms a sound case for the consideration of how these four independently associated biomarkers could enact a level of predictive power in the assessment of

HAND in HIV+ individuals on HAART, and could well be informative about the development of cognitive decline in the HIV- population also. Therefore, we hypothesized that host genetic and epigenetic factors can influence the development of HAND in individuals who are HIV-positive and also currently receiving HAART. Furthermore, we also hypothesized that the measurement of these biomarkers could assess an individual's susceptibility to HAND while still in the early stages of HIV infection. Our initial goal was to restrict this study to HIV+ MSM who developed HAND, and to compare them to control samples. Later re-evaluation of the serostatus of some of the samples in our study population enabled us to broaden these aims, to consider the similarities and differences between seropositive and seronegative individuals who develop signs of cognitive decline. To evaluate this hypothesis, research was conducted to examine the following specific aims:

I. To determine the genotype frequencies of several candidate SNPs in a cohort of men with neurocognitive decline, both HIV+ and HIV-, and to compare them to the two control groups consisting of one population of HIV+ MSM who do not present with symptoms of HAND, and another population of seronegative individuals with no signs of dementia

II. To measure telomere length in the same three cohorts and to assess whether MSM with dementia show greater telomere shortening compared with seropositive and/or seronegative controls

III. To determine whole-genome measures of DNA methylation in these same cohorts, evaluate whether MSM with cognitive decline have differing patterns of methylation overall compared with seropositive or seronegative controls, and also to identify any specific CpG islands that show consistently different methylation patterns between the three sample populations.

2.0 MATERIALS AND METHODS

2.1 SAMPLES

Peripheral Mononuclear Blood Cell (PBMC) pellets were obtained from the Multi-Center AIDS Cohort Study (MACS) laboratory, located at the Graduate School of Public Health in the University of Pittsburgh. PBMC pellets were isolated from blood draws of participants enrolled in the MACS's longitudinal HIV study. Participants of the MACS study from which samples were collected were also members of a Cardiovascular Disease (CVD) sub-study. Individuals were initially classified into one of three groups: HAND positive and HIV positive (cases), HAND negative and HIV positive (seropositive controls), and HAND negative and HIV negative (seronegative controls) (Table 1). Later re-analysis of our samples showed that several of our case group samples were in fact seronegative, although all showed clear signs of cognitive decline. Cases and controls were matched on various selection criteria. Cognitive decline was identified through the observation of deviations in cognitive ability survey results administered during each visit to the MACS clinic for checkups and sample draws. Some cases also were confirmed through the use of Functional Magnetic Resonance Imaging (FMRI) of MACS study participants. Controls were matched to the identified cases on the conditions of age and duration of HIV infection. Matching of cases with controls was necessary for the differential methylation analysis, which seeks to examine how methylation changes over time within the three sample

groups. Samples in the cognitive decline group were considered together, regardless of serostatus, for the SNP and telomere length analyses in Aims #1 and #2, but were considered separately for the methylation analysis of Aim #3.

A pair of samples, for each individual, was identified as close to ten years apart as possible given the limitations of sample availability. Once all samples were identified, the pellets were pulled from the -80^{o}C freezers and prepared for DNA extraction.

Table 1. Counts of PBMC pellets identified for MACS samples

Cognitive Decline (Cases)	HAND-/HIV- (Seropositive Controls)	HAND-/HIV- (Seronegative Controls)
33 Pellets	9 Pellets	10 Pellets

The MACS IDs are withheld to blind sample identification. Controls are less abundant than cases because the matching was only a requirement for the differential methylation analysis

2.2 DNA EXTRACTIONS

DNA was extracted from these PBMC pellets (101 in total) using the Qiagen QIAmp DNA Mini and Blood Mini Kit. Tubes containing the PBMC cells of the experimental samples and all experimental buffers and solutions were equilibrated to room temperature. The initial steps of the extraction procedure were performed under BSL2+ safety protocol due to the HIV positive nature of the several samples. Gloves and a laboratory gown were worn to guard against accidental spills and all work, from initial temperature equilibration to ethanol addition, was performed inside a laminar flow hood.

The extraction procedure was initiated by adding 20µL of QIAGEN protease and 200µL of PBS to each sample tube. Next 200µL of Buffer AL was added to the sample tube and mixed, by pulse-vortexing, for fifteen seconds. Next the samples were incubated in a water bath at 56^{o}C for 10 minutes. 200µL of 96-100% ethanol was then added to each sample and pulse-vortexed for 15

seconds. Samples were then relocated to a bench top lab station from the laminar flow hood for the spin column work to be completed. The mixture was carefully added to the QIAamp Mini spin column, and centrifuged at full speed (14.5 x 1000 rpm) for 1 minute to avoid clogging of the column. The spin column was then placed in a clean 2 mL collection tube and the filtrate was discarded into an empty bottle with lid to be gathered for proper disposal after all extractions were performed. The QIAamp Mini spin column was opened and 500µL of Buffer AW1 was added to the column. The cap was closed and the column was centrifuged at 6000 x g (8000 rpm) for 1 minute. The column was placed in a new 2mL collection tube and the filtrate was discarded. The QIAamp Mini spin column was opened and 500µL of Buffer AW2 was added and the column was spun at full speed (14.5 x 1000 rpm) for 3 minutes. The column was placed in a new 2 mL collection tube, and centrifuged a final time at full speed for 1 minute, to remove any possible carryover from the preceding spin. Finally, the QIAamp Mini spin column was placed in a clean 1.5 microcentrifuge tube. 200µL of the Buffer AE was added for the elution step. The tube and buffer was incubated at room temperature for 1 minute and then centrifuged at 6000 x g (8000 rpm) for 1 minute. This final elution step was repeated to yield a final resuspension volume of 400µL for each of the 101 extracted PBMC DNA samples. This second elution step was recommended because it will increase the DNA yield by up to 15% for each sample.

2.3 SINGLE NUCLEOTIDE POLYMORPHISM DETECTION

Single Nucleotide Polymorphism (SNP) detection was conducted in this experiment using two methodologies: TaqMan assay and dideoxy chain termination sequencing analysis. TaqMan assay was utilized to examine the RYR3 SNP genotypes while PCR and dideoxy sequencing was employed to examine Apolipoprotein E genotypes. In addition, PCR and dideoxy sequencing were employed to

determine the genotypes of samples that failed the TaqMan assay. Primers for PCR, sequencing, and TaqMan primer probes were ordered through Integrated DNA Technologies (IDT).

Table 2. Primer sequences for experimental PCR, qPCR, and Sanger Sequencing

Experimental Standards and Primer Sequences	
RYR3 Single Nucleotide Polymorphism	
rs2229116-F	5’ GGCCACAGTCTGAACACTCA 3’
rs2229116-R	5’ CTGAAGGAGACTGCTGGCTT 3’
rs7177922-F	5’ GCATTTGCATGTGTAGCTGG 3’
rs7177922-R	5’ TGGCACTGCTTTTTAGCCTT 3’
Apolipoprotein E Genotype	
Forward	5' CCCAAAGTGCTGGGATTAGA 3'
Reverse	5' CGTTCAGTGATTGTCGCT 3'
Absolute Telomere Length	
Telomere Standard	$(TTAGGG)_{14}$
36B4 Housekeeping Gene Standard	CAGCAAGTGGGAAGGTGTAATCCG TCTCCACAGACAAGGCCAGGACTCGTTT GTACCCGTTGATGATAGAATGGG
Telomere PCR Forward Primer	5’ CGGTTTGTTTGGGTTTGGGTTTGGGTTTGGG TTTGGGTT 3’
Telomere PCR Reverse Primer	5’ GGCTTGCCTTACCCTTACCCTTACCC TTACCCTTACCCT 3’
36B4 PCR Forward Primer	5’ CAGCAAGTGGGAAGGTGTAATCC 3’
36B4 PCR Reverse Primer	5’ CCCATTCTATCATCAACGGGTACAA 3’

2.3.1 Ryanodine Receptor 3 (RYR3) SNP Detection

Carotid Intima Media Thickness (cIMT), the biomarker for atherosclerosis risk in this experimental design, was assessed for Single Nucleotide Polymorphisms (SNPs) by TaqMan

assay. A previous genome wide association study [28] had identified two SNPs of interest, rs2229116 and rs7177922, within the gene encoding the Ryanodine Receptor 3 (RYR3), calcium ion channel. A TaqMan assay was conducted using Applied Biosystems TaqMan Genotyping Master Mix (Part No. 4371355) and TaqMan probes (Life Technologies) designed for the specific SNP in the gene region of RYR3 (Table 2).

Table 3. Master Mix reaction conditions for the TaqMan Genotyping assay

TaqMan Reaction Conditions	
Master Mix	300 µL
Probe (20x)	15 µL
Water	192.5 µL
DNA	1.0 µL

5µL of Master Mix was pipetted using the repeat dispenser and 1µL of DNA was then added to each sample well in the 96-well Phenix plate.

The Master Mix was assembled (Table 3) and then dispensed into a 96-well Phenix white well plate using the repeat dispenser at a volume of 5µL per well. Once the Master Mix was dispensed, a 1µL volume of sample DNA was dispensed by pipette into individual wells of the plate. A clear plastic, adhesive qPCR film was then placed over the 96-well plate and the plate was centrifuged at approximately 1300 x g to assure no portions of the Master Mix remained on the individual well's walls. The plate was then inserted into an Eppendorf MasterCycler RealPlex machine and qPCR was performed using the defined cycler program.

The TaqMan assay designed for this experimental approach utilizes a molecular fluorescent dye bound to the Quantitative Polymerase Chain Reaction (qPCR) probe. When bound, the dye on the probe is effectively quenched and no luminescence is recorded. Upon amplification, the 5'-3' exonucleolytic activity of the TaqMan DNA polymerase liberates the

dye, which emits light and this fluorescence is registered by the sensor. Two dyes were employed in this experiment and were individually linked to a particular primer for the qPCR. The probe for the "G" genotype was tagged with the FAM dye and the probe for the "A" genotype was tagged with the VIC dye to allow for the distinction of the two SNPs within the sample population during experimentation and subsequent data analysis. The fluorescence values of each dye were dot plotted via Microsoft Excel (see Figure 3) with FAM on the X-axis and VIC on the Y-axis. The resulting dot plot showed clusters of values and each cluster corresponded to a specific genotype (GG upper left, AG upper right, AA lower right, and "fails" lower left). The data plotted in Figure 3 were obtained from a larger study of the relationship between RYR3 genetic polymorphism and cIMT[28], which includes all of the samples analyzed for dementia later in this thesis.

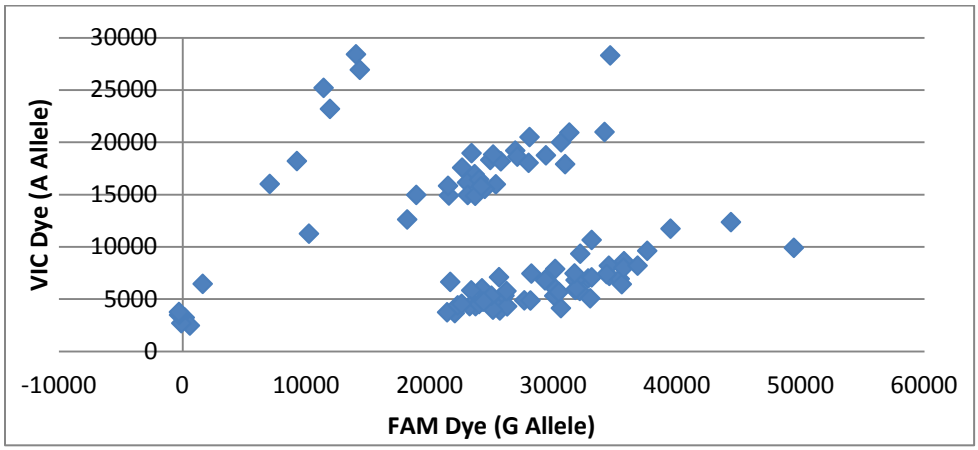

Figure 3. TaqMan allele clustering analysis

The samples which failed TaqMan assay, defined as either not exceeding threshold fluorescence for the sensor and/or recording values which were unable to be definitively assigned to a single genotype cluster, were examined using dideoxy sequencing.

2.3.2 Apolipoprotein E (APOE) Genotyping

The sequencing approach utilized for the genotyping of the Apolipoprotein E (APOE) gene region of all study samples was identical to the protocol outlined below for the Sanger Dideoxy Chain Terminator approach for gene sequencing. The only augmentation was that as part of the initial PCR reaction mixture, 0.5µL of Dimethyl Sulfoxide (DMSO) was added to the Master Mix reaction volume. DMSO was added to aid in linearizing the DNA template. The APOE gene region is AT-rich and tends to self anneal, limiting amplification efficiency, so the DMSO was added to the reaction volume to counteract this characteristic. This additional volume was accounted for in the total reaction volume by reducing the water volume in the Master Mix by 0.5 µL per reaction volume.

2.3.3 Polymerase Chain Reaction and Sanger Sequencing

Primer pairs were designed for each gene region (Table 2) containing the SNP of interest (RYR3 and APOE) and a temperature gradient and magnesium titration were performed to determine the optimal reaction conditions for each primer set. Four PCR master mixes were assembled using four different magnesium concentrations (1.0mM, 1.5mM, 2.0mM, and 2.5mM). The polymerase utilized for template amplification was AmpliTaq Gold (Applied Biosystems).

Four strips of 8-tubes were assembled with each tube having identical sample mixes. The PCR sample mixes were then placed into the Eppendorf Mastercycler and a temperature gradient program was employed so that each tube in the 8-strip tube ran at a different annealing temperature.

2.3.3.1 Gel Electrophoresis

PCR products were examined using gel electrophoresis to confirm PCR amplification. A 2% agarose gel was made by adding 4 grams of GenePure LE agarose powder (ISC BioExpress) into 200mL of 0.5X TBE (ISC BioExpress) buffer and heating in a microwave until agarose was successfully dissolved. 10µL of Ethidium Bromide was then added to the dissolved agarose and mixed. The liquid gel was then poured into a forming tray; a comb was inserted, and allowed to cool until the agarose had set. Once set, 1µL of the PCR sample was combined with 6µL of 6X loading dye and samples were loaded into the gel wells formed by removing the comb. A running ladder was also loaded into the gel, Phi-X174 RF DNA *Hae III* digest (Thermo Scientific). The fully loaded gel was run at 100 Volts for 30 minutes in a running buffer composed of 300 mL 0.5X TBE buffer with 15µL of Ethidium Bromide. After the gel electrophoresis was completed, the gel was taken to a gel examination system and the Ethidium Bromide-stained DNA was visualized using a UV transilluminator. Because Ethidium Bromide intercalates double stranded DNA and fluoresces under UV light, this characteristic was exploited for PCR amplification product visualization.

It was determined that the RYR3 primers required a magnesium concentration of 2.5mM and that both the RYR3 and the APOE samples required a PCR cycle with an annealing temperature of 60^{o}C. PCR amplifications were carried out using the newly determined conditions for all RYR3 and APOE samples.

2.3.3.2 Big Dye Sequencing

Once PCR samples were confirmed to have successfully amplified, 10 µL of the PCR product was taken for the sequencing reaction. The next step to begin Big Dye Sequencing is to conduct an ExoSAP procedure. The PCR samples are treated with Exonuclease I enzyme (New

England BioLabs), to remove any remaining primers, and alkaline phosphatase (Roche), to remove the phosphates from any remaining dNTPs and prevent them from being polymerized in the sequencing reaction.

The ExoSAP master mix for this reaction was constructed (Table 8). For each sample to be sequenced, 10 µL of PCR product was pipetted into the well of a 96-well PCR plate. 10 µL of the ExoSAP master mix was subsequently added to each well, by use of a repeat dispenser, for a total reaction volume of 20 µL. The plate was then placed into the Eppendorf Mastercycler and the "ExoSAP" program was run which has the samples at 37^{o}C for 60 minutes and 85^{o}C for 15 minutes.

Table 4. Master Mix conditions for the ExoSAP treatment of PCR samples prior to Sanger Dideoxy sequencing experiment

ExoSAP Master Mix	
10X Buffer	1.0 µL
Shrimp Alkaline Phosphatase	1.0 µL
EXO Nuclease I	0.05 µL
H_2O	7.95 µL
TOTAL	10 µL

The next step after the ExoSAP treatment is the "Big Dye" Dideoxy chain terminator sequencing reaction. 5 µL of the ExoSAP reaction described previously is taken and placed into a new 96-well skirted optical plate. The 5 µL aliquots of the ExoSAP product were pipetted into two rows such that the same sample was in two side by side wells. This is because the Big Dye reaction is run in tandem with each run for a "forward" and "reverse" reaction using that specific

direction primer exclusively. The Big Dye Master Mix was created for each individual sample by using the Big Dye mix (Applied Biosystems) and then a primer dilution of 1:500 for each direction (forward and reverse) which makes a 10mM concentration for the primer. 5 µL of the Big Dye Master Mix (table 4) is then added together into the well with the ExoSAP product and the sequencing PCR reaction is run.

Table 5. Big Dye Dideoxy chain terminator reaction Master Mix

Big Dye Master Mix	
Primers (1:500)	2.5 µL
5x Dilution Buffer	2.0 µL
Big Dye	0.5 µL
TOTAL	5 µL

Big Dye mix (Applied Biosystems) comes as a pre-mixed solution and is added to the reaction mixture in a volume of 0.5 µL

Once the Big Dye cycle was finished, the 10 µL of the sequencing reaction was placed into a clean 96-well optical plate. To each well of the plate, 125 mM of ethylenediaminetetraacetic acid (EDTA) and 60 µL of 100% Ethanol were added. The plate was sealed, inverted several times to mix contents, and incubated for 15 minutes in the dark to avoid degradation of the light sensitive Big Dye reagents. The plate was then centrifuged at 2500 x g for 30 minutes at 4^{o}C to pellet the DNA contents of the wells. The supernatant was then removed by inverting the plate over the sink and a flick of the wrist to expel the liquid supernatant from the plate wells. 60 µL of 70% Ethanol was then added to the plate wells. The plate was sealed once more and spun at 1650 x g for 15 minutes at 4^{o}C. The supernatant was removed a final time and the plates were inverted on a stack of paper towels and spun in the centrifuge for several seconds to pull out any remaining ethanol from the wells. The plate was then allowed to air dry

for approximately 20 minutes in the dark. The plate was then sealed with a non-transparent film to guard against degradation from light and then wrapped in aluminum foil and stored in the freezer. Once the sample preparation was completed, plates were taken to the University of Pittsburgh Genomics and Proteomics CORE laboratories (GPCL) for capillary electrophoresis. The sequence data generated by the CORE facility was downloaded from their website and the sequences files were examined using the Sequencher program.

2.4 ABSOLUTE TELOMERE LENGTH MEASUREMENT

The protocol for a quantitative PCR method to determine absolute telomere length was presented by O'Callaghan and French. [50] This experimental protocol was followed using the experimental design presented in the work, with primer sequences for the telomere repeat region and the 36B4 housekeeping gene being ordered from IDT (Table 2). The reference standards for both the telomere repeat region as well as the 36B4 housekeeping gene were ordered from IDT as well. The only deviation utilized from the methodology presented by O'Callaghan and French was that the dilutions of the standard curves were both calculated to be at a starting concentration of 500 picograms/µL and then followed a five-step, ten-fold serial dilution for each standard with three water blanks after the standard curve to provide baseline fluorescence values. Samples were designed to have 10µL/well on a 96-well Phenix white well plate.

The primers were diluted to a 10µM working concentration for the experimental procedure. The RT^2 SYBR Green ROX ^{TM} qPCR Mastermix kit (Qiagen) was used for this experiment. SYBR Green was used for detection of PCR amplification and a reference dye, ROX, was used for signal optimization. The kit contains a real–time PCR buffer, a high-

performance, HotStart DNA Taq polymerase, nucleotides, SYBR Green dye, and ROX dye for optimization of the instrument optics. The use of HotStart enzyme improves qPCR results by avoiding the amplification of primer dimers and other non-specific and unintended products. Sample and standard reaction mixes were constructed as illustrated below. (Tables 6 and 7)

Table 6. Sample Master Mix for the SYBR Green ROX

Sample SYBR Green ROX ™ Master Mix	
SYBR Green Buffer	5µL
Primer	0.2 µL
H_2O	3.8 µL
DNA	1.0 µL
TOTAL	10 µL

The Master Mix was generated by combining Primers, Buffer, and Water. Once generated, 9 µL of the mix was pipetted into the Phenix White well 96-well plate. 1 µL of sample DNA was then added to wells

Table 7. Standard Master Mix for the SYBR Green ROX

Standard SYBR Green ROX ™ Master Mix	
SYBER Green Buffer	5 µL
Primer	0.2 µL
H_2O	2.8 µL
DNA (pGEM)	1.0 µL
Standard (TELO/36B4)	1.0 µL
TOTAL	10 µL

The Master Mix was generated by combining Primers, Buffer, pGEM plasmid DNA, and Water. 10 µL of the mix was pipetted into the Phenix White well 96-well plate. 1 µL of synthesized Standard (TELO/36B4) was added

Sample plates were placed into the Eppendorf RealPlex Cycler and a qPCR program was run. A 50 cycle program was completed and data was then analyzed using the Eppendorf RealPlex software. A line was fit to the standard curve on each plate. An r^2, the linear correlation coefficient, value between 0.98 – 1.0 was used to determine acceptable results. The efficiency value reported by the software, calculated by the formula $E= 10^{[-1/slope]}-1$, was also expected to be approximately 1.0 and indicates an efficient doubling of the PCR product with each cycle of the qPCR program. Noise filtering for the data was completed by the data collection software.

2.5 **DNA METHYLATION ARRAY**

Extracted DNA, from PBMC pellets, was sent to the Genomics and Proteomics CORE Lab of the University of Pittsburgh where the analysis for methylation was conducted. 500 nanograms (ng) of DNA was necessary for methylation analysis. Sample DNA concentration was determined through use of PicoGreen assay, conducted by the CORE labs, and samples below this necessary DNA concentration were supplemented by additional volumes as determined by the CORE lab. Bisulphite conversion of DNA was carried out using the EZ DNA Methylation™ Kit (Zymo Research Corp., CA) to convert unmethylated cytosine nucleotides to uracil. Converted DNA samples are amplified, fragmented, denatured, and hybridized to prepared Infinium BeadChips (Illumina). Single base extension is performed with labeled nucleotides, which are then stained with florescent dyes and scanned using an Illumina BeadArray Reader and the data analyzed using Bead Studio 2.0. This provides methylation data on over 450,000 different CpG dinucleotide sites throughout the genome. This data was then downloaded from the CORE facility and analyzed using GenomeStudio and other software packages.

Table 8. DNA methylation sample collection dates

Cognitive Decline (Cases)		HAND-/HIV+ (Seropositive Controls)		HAND-/HIV- (Seronegative Controls)	
Case 1-1 5/25/03	Case 1-2 10/27/11	Pos 1-1 4/19/01	Pos 1-2 11/10/11	Neg 1-1 7/14/04	Neg 1-2 7/28/12
Case 2-1 10/6/01	Case 2-2 10/22/11	Pos 2-1 8/11/02	Pos 2-2 6/23/12	Neg 2-1 11/16/02	Neg 2-2 6/8/05
Case 3-1 11/8/01	Case 3-2 10/13/11	Pos 3-1 10/8/01	Pos 3-2 4/10/12	Neg 3-1 12/17/02	Neg 3-2 7/24/12
Case 4-1 10/4/01	Case 4-2 6/7/08	Pos 4-1 10/18/01	Pos 4-2 5/15/12	Neg 4-1 1/5/02	Neg 4-2 2/2/12
Case 5-1 12/15/07	Case 5-2 12/17/11	Pos 5-1 11/21/02	Pos 5-2 7/12/12	Neg 5-1 10/8/02	Neg 5-2 5/9/12
Case 6-1 10/23/01	Case 6-2 4/19/11	Pos 6-1 3/3/01	Pos 6-2 4/21/12	Neg 6-1 6/19/02	Neg 6-2 4/21/12
Case 7-1 2/27/03	Case 7-2 4/27/11	Pos 7-1 11/22/00	Pos 7-2 4/6/12	Neg 7-1 7/17/97	Neg 7-2 10/23/99
Case 8-1 5/9/91	Case 8-2 1/9/03	Pos 8-1 4/30/03	Pos 8-2 4/10/12	Neg 8-1 4/6/01	Neg 8-2 1/22/11

The above are the sample collection dates for samples sent to CORE lab for methylation array analysis. The samples are listed in their corresponding sample category and oriented horizontally according to their matched triad. Samples have been de-identified for this analysis

2.6 **DATA ANALYSIS**

The RYR3 TaqMan genotype data was analyzed by scatter plotting in Excel, the TaqMan call data which was exported from the Thermocycler. The data was plotted with the fluorescence

reading from the FAM probe dye along the X-axis and the fluorescence reading from the VIC dye along the Y-axis. Sample clustering was observed visually and the genotypes were determined such that homozygotes (AA and GG) were the upper left and bottom right clusters and the heterozygotes comprised the upper right cluster. (Figure 3) The samples that failed the experimental run or had fluorescence values below the machine's threshold were closest to the origin. A computer algorithm was also utilized to verify visual "calls" of genotypes. Once all sample genotypes were determined through this method, the allele frequencies of these data were tabulated and analyzed for distinguishing or significant patterns of the "risk" allele being present within and between the three experimental groups.

The APOE samples were sequenced and the genotypes were determined through the use of the Sequencher program. Sequences were aligned to a reference of the APOE gene to identify which of the APOE alleles were present. APOE genotypes were recorded for each patient sample and allele frequencies were calculated within the three experimental groups.

Absolute telomere length calculations were carried as outlined in the work of O'Callaghan and Fenech with a minor modification to the final value reported. [50] Best fit lines were applied to the standard (Telomere and 36B4 Housekeeping) dilution curves for each plate of the SYBR Green assay conducted. Line equations were deemed acceptable that possessed a linear fit coefficient, r^2, between 0.98 and 1.0. The line equation was utilized to calculate the telomere length in kilobases (KB). Using the formula $y = mx + b$, where m is the slope of the best fit line for each plate, b is the y-intercept, and y is the C_T value from the plate sample, the variable x, the telomere length for each sample in KB, was determined. An identical analysis process was carried out for the housekeeping gene where x was the diploid genome copies since there are two copies of 36B4 in a diploid genome. These calculations were completed for each

triplicate run of the sample data. The newly calculated values for the length of telomere were divided by the calculated diploid genome copies of the housekeeping gene to normalize the data and calculate the absolute length of telomere per diploid genome. Each sample was run in triplicate, and three values for each sample of the absolute telomere length per diploid genome were computed. This process was carried out for each of the two sample time points and values were recorded separately for each. The telomere length per diploid genome of the second time point was then subtracted from the first time point's telomere length per diploid genome value to give a net difference of the two sample points. This was carried out for each of the three triplicate values and these three net difference values were then averaged for each sample to give the mean net difference. Telomere length analysis was completed by dividing each returned mean net difference by the amount of years between the pellet sample dates for each MACS sample. This last data step generates a data value that is the rate of telomere length change per year for each of the samples analyzed in this study.

The methylation data was first generically visualized and examined using the GenomeStudio program (Illumina). This program aided the identification of any samples that failed the array analysis and was used to generate the files necessary for downstream processing. The downstream processing and analysis of the methylation was completed using the caGEDA program, developed by the University of Pittsburgh Cancer Institute.[51] The normalization procedure selected was global mean adjustment and this test was determined by running several normalization procedures and identifying the method which gave the most accurate data normalization. The test selected for differential methylation data analysis was the J5 test. Patel *et al*, explain that the J5 test is best for pilot studies because the high variance in pilot study samples can lead to t-tests with unacceptable specificity.[51] The J5 and global mean adjustment

methodologies were selected as the most appropriate through use of an efficiency analysis program created by Jordan *et* al, from the biomedical informatics department of the University of Pittsburgh.[52] The caGEDA program returns a large collection of data for analysis but the two data returned that are most informative for this analysis were the top 50 differentially expressed genes and the differential gene expression patterns for each sample group run on the array.

3.0 RESULTS

3.1 RYR3 SINGLE NUCLEOTIDE POLYMORPHISM

The point mutation associated with the rs2229116 SNP is classified as a nonsynonymous polymorphism because it results in an amino acid residue change. In this instance, the rs2229116 SNP results in the interchanging of an Isoleucine (Ile) with a Valine (Val), by way of the Adenine to Guanine substitution in the RYR3 gene sequence. Research by Shrestha *et al* showed that individuals possessing the GG genotype at the location of the rs2229116 SNP showed significantly higher carotid intima media thickness (cIMT) than those with either the AA or AG genotypes.[28] Allele and genotype frequencies were calculated for the study population within the respective sample groups (Table 9 and 10). Four individuals in the HIV-/HAND- sample group failed TaqMan analysis and still require SNP confirmation through Sanger sequencing.

Table 9. Genotype frequency for rs2229116 SNP of the RYR3 gene

	rs2229116		
	AA	**AG**	**GG**
Cases	25/33 (0.7576)	5/33 (0.1515)	3/33 (0.0909)
Seropositive controls	4/9 (0.4444)	3/9 (0.3333)	2/9 (0.2222)
Seronegative controls	2/6 (0.3333)	1/6 (0.1667)	3/6 (0.5000)

The individuals possessing the GG genotype have shown the largest cIMT, a genetic trait that is associated most frequently with the development of atherosclerosis.

Table 10. Allele frequencies for rs2229116 of the cIMT SNP

rs2229116 Allele Frequency			
Sample Group	**Allele**	**Sample Size**	**Allele Frequency**
Cases	$Prob_G$	N = 66 (2*33)	11/66 (0.167)
	$Prob_A$	N= 66 (2*33)	55/66 (0.833)
Seropositive Controls	$Prob_G$	N = 18 (2*9)	7/18 (0.389)
	$Prob_A$	N = 18 (2*9)	11/18 (0.611)
Seronegative Controls	$Prob_G$	N = 12 (2*6)	7/12 (0.583)
	$Prob_A$	N = 12 (2*6)	5/12 (0.417)

Three individuals in the dementia cases, two individuals in the seropositive group, and three individuals in the seronegative group possess the "GG" genotype putting them at increased risk of Atherosclerosis. Only two of these individuals in the seronegative group, "Negative 3" and "Negative 5", also appeared within the triads for the methylation analysis, limiting the capacity to assess this biomarker's predictive ability for the development of HAND.

The point mutation for rs7177922 is similar to the SNP of rs2229116 with the difference that instead of Guanine it is an Adenine substitution, and the "AA" genotype is the risk allele. Shrestha *et al*, found that rs7177922 was not significantly associated with atherosclerosis, like the rs2229116 SNP, yet maintained a similar trend in the population allele frequency.[28] Genotype and allele frequency analyses were conducted for the rs7177922 SNP in the three study sample populations (Table 11 and 12). Detailed statistical analysis was not conducted due to sample size limitations.

Table 11. Genotype frequency for rs7177922 SNP of the RYR3 gene

rs7177922			
	AA	**AG**	**GG**
Cases	4/33 (0.1212)	4/33 (0.1212)	25/33 (0.7576)
Seropositive controls	2/6 (0.3333)	1/6 (0.1667)	3/6 (0.5000)
Seronegative controls	0/8 (0.0000)	5/8 (0.625)	3/8 (0.375)

The individuals possessing the AA genotype have shown a similar trend as rs2229116 but the allele is not associated with the development of atherosclerosis

Table 12. Allele frequency for rs7177922 of the RYR3 gene

rs717922 Allele Frequency			
Sample Group	**Allele**	**Chromosome Count**	**Allele Frequency**
Cases	$Prob_G$	N = 66 (2*33)	54/66 (0.818)
	$Prob_A$	N = 66 (2*33)	12/66 (0.182)
Seropositive Controls	$Prob_G$	N = 12 (2*6)	7/12 (0.583)
	$Prob_A$	N = 12 (2*6)	5/12 (0.417)
Seronegative Controls	$Prob_G$	N = 16 (2*8)	11/16 (0.6875)
	$Prob_A$	N = 16 (2*8)	5/16 (0.3125)

3.2 **APOE GENOTYPE**

As reported previously, the APOE*4 allele has shown gene-dependent linkage to increased risk of Alzheimer's disease (AD) based on extensive clinical and basic studies. Increased risk of late onset AD for carriers is between 3 and 15 times greater than non-carriers. The APOE*4 carriers

also encounter ages of AD onset 10 to 20 years earlier than non E4 carriers.[33] The E4 allele therefore represents the greatest risk factor among the possible genotypes of the APOE gene, E2, E3, and E4. The E3 allele is considered the "neutral" gene form, being the most present in the at-large population.

Table 13. APOE genotype frequencies for the cognitive decline (case) samples

HAND+/HIV+ (Cases)	
Genotype	**Frequency**
E2/E3	**2/8**
E3/E3	**2/8**
E2/E4 or E1/E3*	**1/8**
Missing	**3/8**

Individuals who failed PCR and sequencing were deemed as missing until future analysis could be completed

Once the APOE sequencing data was examined it was determined that one case sample possessed the allele for E4 (Table 13). This individual is a heterozygote meaning that the risk for AD is lower than the homozygote E4/E4, yet remains higher than the risk the other case samples experience. This heterozygote E3/E4 sample was one of the cases examined on the methylation array and is identified by the sample name "Case 1" (Table 16). The three missing samples are individuals who failed PCR amplification and could not be sequenced and will be examined in future work. The discrepancy observed for the sample between the E2/E4 and the E1/E3 genotypes is a currently disputed concept, as some scientists argue that the allele combination does not exist in the human population. Therefore, for the purpose of this data analysis the "accepted" call was selected to be the E2/E4 genotype, however this particular sample requires further analysis to truly identify and resolve the discrepancy.

Table 14. APOE genotype frequencies for the Seropositive controls samples

HIV+/HAND- (Seropositive Controls)	
Genotype	**Frequency**
E3/E3	**7/9**
E3/E4	**1/9**
E4/E4	**1/9**

One seropositive control sample was homozygous for the APOE*4 (Table 14). This genotype corresponds to the highest risk of AD; however this individual was not one of the cases examined on the methylation array. The heterozygote E3/E4, which has a diminished risk of AD compared to the homozygote E4/E4, was included on the methylation array and is identified as "Positive 8" in the methylation data tables (Table 17).

Table 15. APOE allele frequencies for the Seronegative controls samples

HIV-/HAND- (Seronegative Controls)	
Genotype	**Frequency**
E2/E3	**1/8**
E3/E3	**4/8**
E3/E4	**3/8**

Three of the seronegative control samples were heterozygous for the APOE*3/APOE*4 genotype (Table 15). These same three individuals were also examined on the methylation array and are identified by the sample names "Negative 2", "Negative 6", and "Negative 7" in the methylation data table (Table 18). The seronegative control group contained the highest frequency of heterozygotes, a population percentage of 37.5% (3 samples of 8 total). Detailed statistical analysis was not conducted, however, due to sample size limitations.

3.3 ABSOLUTE TELOMERE LENGTH

The data analysis procedure for telomere length was carried out as indicated in the work of O'Callaghan and Fenech.[50] The only augmentation to the procedure was that the telomere length per diploid genome value of the second time point was subtracted from the first time point's telomere length per diploid genome value to give a net difference of the two sample points. This was carried out for each of the three triplicate values and these three net difference values were then averaged for each sample to give the mean net difference. Telomere length analysis was

completed by dividing each returned mean net difference by the amount of years between the pellet sample dates for each MACS sample. This last data step generates a data value that is the rate of telomere length change, in kilobases, per year for each of the samples analyzed in this study (Table 16 – 18).

Table 16. Case sample absolute telomere length

Cognitive decline (Cases) Telomere Data				
Sample	Time Point 1 (KB Telomere/Diploid Genome)	Time Point 2 (KB Telomere/Diploid Genome)	Years Between Time Points	Yearly Length Change (KB Telomere/Diploid Genome)
Case 1	**0.3538**	**0.4112**	**8**	**-0.0162**
Case 2	**0.3952**	**0.2191**	**10**	**0.0176**
Case 3	**0.4056**	**0.4073**	**10**	**-0.0002**
Case 4	**0.4846**	**0.4834**	**7**	**0.0002**
Case 5	**0.3952**	**0.3328**	**4**	**0.0329**
Case 6	**0.3582**	**0.4304**	**10**	**-0.0072**
Case 7	**0.4047**	**0.3948**	**8**	**0.0012**
Case 8	**0.3589**	**0.3524**	**12**	**0.0005**
Case 9	0.3472	0.2905	1	0.0566
Case 10	0.3894	0.3698	5	0.0039
Case 11	0.3908	0.4013	11	-0.0010
Case 12	-0.4106	-3.320	8	0.3637
Case 13	0.1891	0.1034	10	0.0086
Case 14	0.3127	0.5963	4	-0.0709
Case 15	0.3720	0.4975	7	-0.0179
Case 16	0.4123	0.4220	10	-0.0010

Table 16 continued

Case 17	0.5360	0.4422	8	0.0117
Case 18	0.5531	0.5335	9	0.0022
Case 19	0.5184	0.5302	8	-0.0015
Case 20	0.4266	0.3508	5	0.0152
Case 21	0.4371	0.4021	10	0.0035
Case 22	0.3266	0.3181	10	0.0008
Case 23	0.2918	0.4474	1	-0.1556
Case 24	0.4617	0.5076	4	-0.0115
Case 25	0.3966	0.4131	10	-0.0016
Case 26	0.4672	0.4100	11	0.0052
Case 27	0.3887	0.3177	1	0.0709
Case 28	0.4197	0.4206	10	0.0566

Telomere lengths are reported in kilobases (KB) of telomere per diploid genome. Samples in bold font are those which were analyzed on the methylation array. All samples are listed in order of their matched triad

Table 17. Seropositive controls absolute telomere length

HIV+/HAND- (Seropositive Controls) Telomere Data				
Sample	Time Point 1 (KB Telomere/Diploid Genome)	Time Point 2 (KB Telomere/Diploid Genome)	Years Between Time Points	Yearly Length Change (KB Telomere/Diploid Genome)
Positive 1	**0.3355**	**0.4112**	**10**	**0.0009**
Positive 2	**0.4724**	**0.3765**	**10**	**0.0120**
Positive 3	**0.4089**	**0.3594**	**11**	**0.0045**
Positive 4	**0.4624**	**0.3742**	**11**	**0.0080**
Positive 5	**0.3493**	**0.2618**	**10**	**0.0087**
Positive 6	**0.3485**	**0.2487**	**11**	**0.0091**

Table 17 continued

Positive 7	**0.5411**	**0.4506**	**12**	**0.0075**
Positive 8	**0.4465**	**0.3801**	**10**	**0.0066**
Positive 9	0.2838	0.4137	8	-0.0162

Telomere lengths are reported in kilobases (KB) of telomere per diploid genome. Samples in bold font are those which were analyzed on the methylation array. All samples are listed in order of their matched triad

Table 18. Seronegative controls absolute telomere length

HIV-/HAND- (Seronegative Controls) Telomere Data				
Sample	Time Point 1 (KB Telomere/Diploid Genome)	Time Point 2 (KB Telomere/Diploid Genome)	Years Between Time Points	Yearly Length Change (KB Telomere/Diploid Genome)
Negative 1	**0.4121**	**0.4622**	**8**	**0.0060**
Negative 2	**0.4657**	**0.4388**	**3**	**0.0183**
Negative 3	**0.4296**	**0.4517**	**10**	**0.0098**
Negative 4	**0.3962**	**0.3370**	**10**	**-0.0066**
Negative 5	**0.3897**	**0.4911**	**10**	**0.0055**
Negative 6	**0.4482**	**0.4460**	**10**	**0.0102**
Negative 7	**0.4641**	**0.4294**	**10**	**0.0873**
Negative 8	**0.5090**	**0.5437**	**2**	**0.0069**
Negative 9	0.4791	0.5254	10	0.0184
Negative 10	0.3715	0.4295	10	-0.0025

Telomere lengths are reported in kilobases (KB) of telomere per diploid genome. Samples in bold font are those which were analyzed on the methylation array. All samples are listed in order of their matched triad

The above data (Table 16-18) represents the calculated absolute telomere length per diploid genome for each sample at their two pellet time points, the years between when those samples were collected, and the calculated yearly absolute telomere length change. The expected

result was to observe negative values for the yearly rate of change in absolute telomere length. This expectation is rooted in the fact that cell differentiation establishes an initial telomere lengths and then this length shortens over the course of subsequent cellular divisions.[38,40] Observing a positive value for the calculated rates of change in telomere length, as occurs above for a majority of samples, implies that the absolute telomere length is increasing, not shortening. This does not appear to be a real result and it is suspected to be an artifact of the assay. This fact will be addressed in the discussion section.

3.4 **DNA METHYLATION**

DNA methylation data was analyzed through the use of the GenomeStudio program and comparisons were made between all sample cohorts. The first comparisons of the data sought to assess the existence of general trends in the DNA methylation patterns of the samples in the three sample cohorts. These comparisons were achieved through the generation of dendrograms which compared the two time points for each sample within a specific cohort.

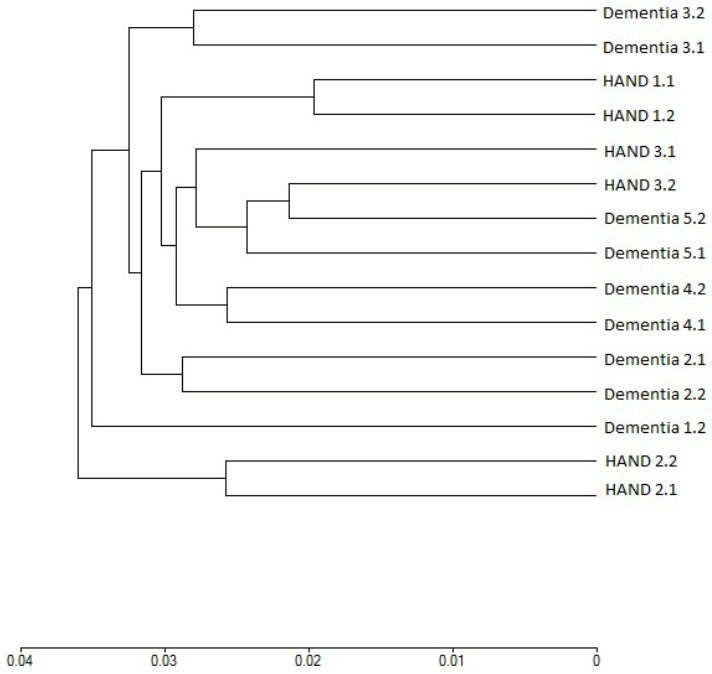

Figure 4. Dendrogram of case DNA methylation data

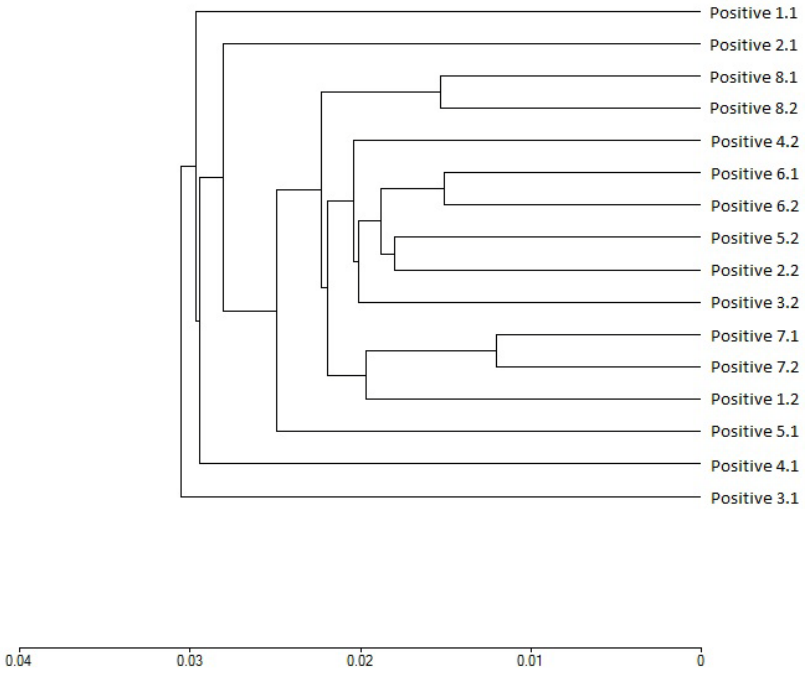

Figure 5. Dendrogram of seropositive control DNA methylation data

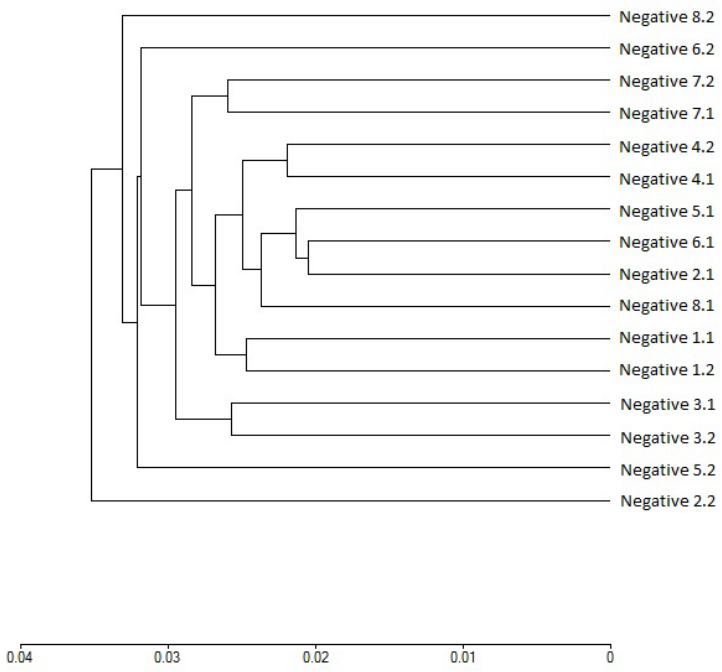

Figure 6. Dendrogram of seronegative control DNA methylation data

The above dendrograms compare the similarities of the DNA methylation profiles for each sample time point of the cases (Figure 4), seropositive controls (Figure 5), and seronegative controls (Figure 6). It was observed that the methylation profiles obtained at the two time points for each case sample were consistently clustering together, indicating that their methylation patterns were much more like each other than the profiles of any other case sample. Conversely, the methylation profiles for the samples in the two control groups showed much less intra-individual clustering; indicating methylation patterns were changing randomly between the collection dates for the samples. The finding that the cases' methylation profiles were remaining similar lead us to consider a comparison of the samples grouped on dementia status, instead of serostatus as before, and that this would aid in generating a more complete understanding of how the methylation patterns were changing for the DNA methylation data.

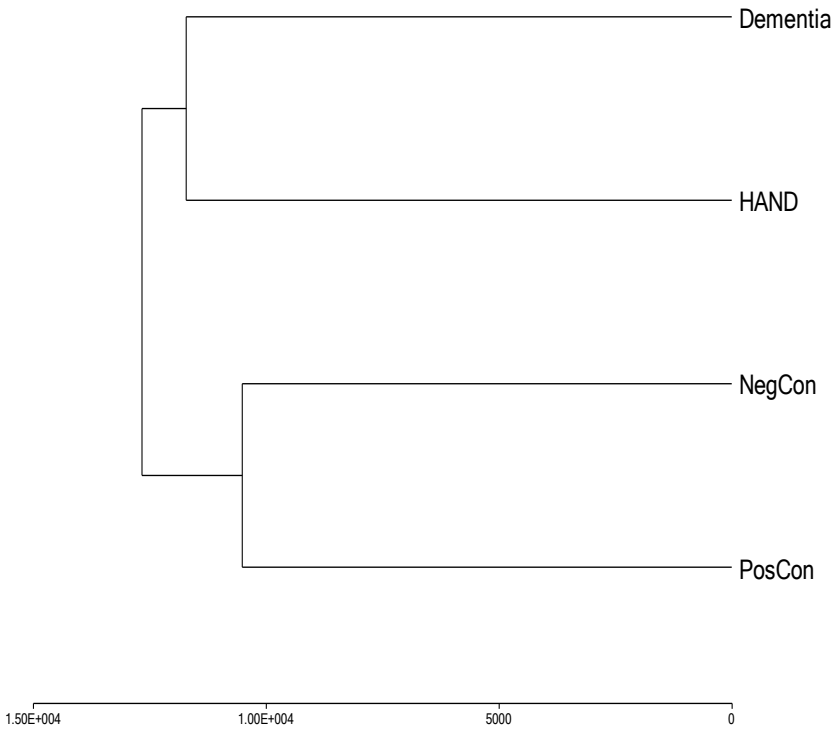

Figure 7. Dendrogram matching sample groups on dementia status

The above dendrogram of the sample methylation data, based on dementia status instead of serostatus, indicates that the samples with dementia (both HAND and non-HIV-related) were more similar to each other than to the two control populations (Figure 7). This observation then led to a comparative analysis of differential methylation patterns. This was achieved by comparing the changes in percent methylation between an individual sample's two time points and then comparing the differential methylation change across two unique individuals.

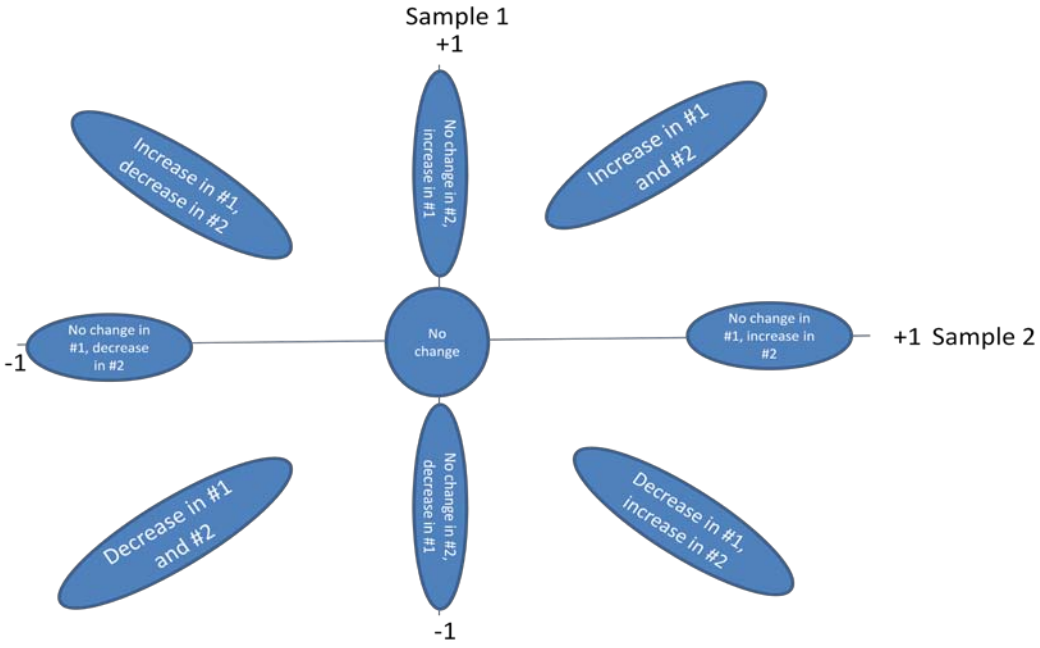

Figure 8. Quadrant key for diffrential methylation analysis

The above figure represents how the comparative differential methylation data for two different samples was analyzed. The various quadrants are read like a Cartesian coordinate plane and the corresponding change in methylation pattern for the two samples being compared is indicated (Figure 8).

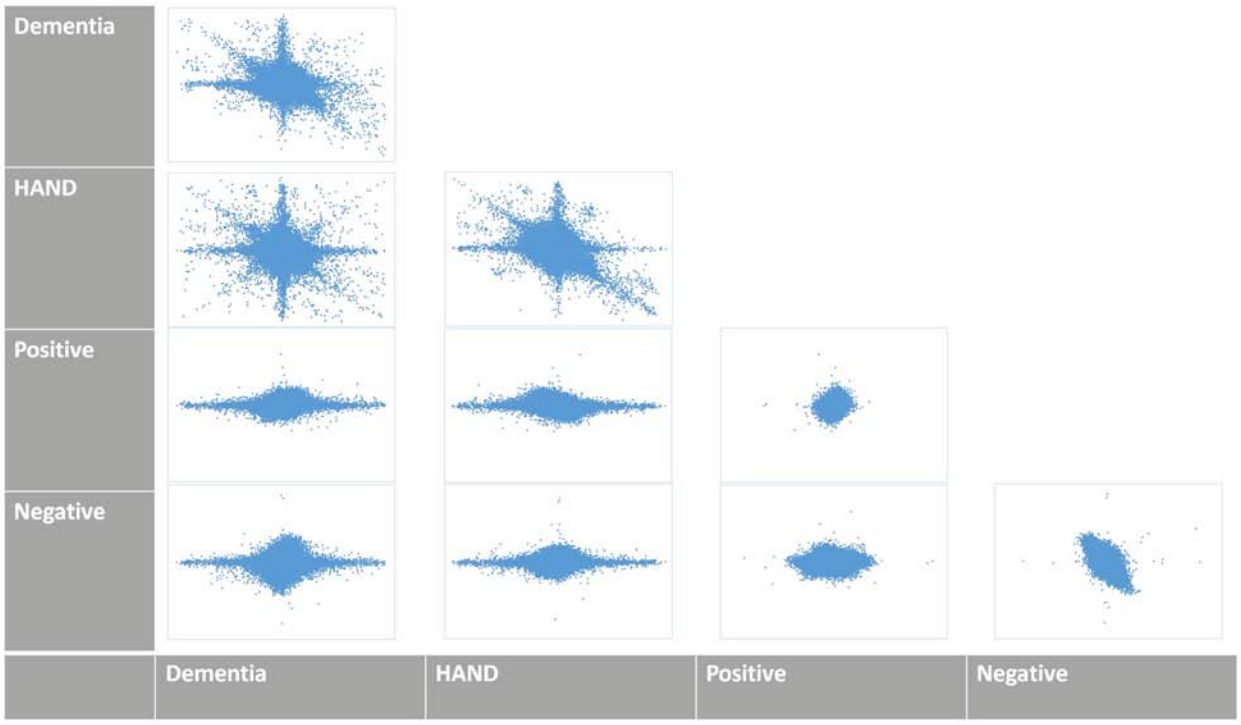

Figure 9. Differential DNA methylation analysis for two samples

The above figure (Figure 9) shows the differential DNA methylation for samples spanning each cohort group. Each dot represents a single CpG island. All axes have been normalized to a scale ranging from -1 to +1, with the origin (0,0) in the center. The sample type that is plotted on the x-axis of the differential plot is listed along the bottom of the figure and the sample type along the left side of the figure is the one plotted along the y-axis. This figure shows a remarkable difference in methylation profile between the dementia and control groups. As was suggested by the dendrograms of the cohort groups (Figures 4-6) the samples of the control groups are showing little to no significant directional change in methylation when compared to one another (all figures in the two rightmost columns). Comparison of a dementia sample (either HAND or seronegative dementia) with either a seropositive or a seronegative control sample shows that the dementia samples have a large number of CpG islands that have undergone dramatic increases or decreases in methylation during the time interval between sample

collection (bottom two rows of the two leftmost columns). The extent of these methylation changes becomes even clearer when two dementia samples are compared (top two rows of the two leftmost columns). The dots that appear furthest along the axes from the origin are those CpG sites experiencing the greatest change in methylation relative to one another when the two samples are compared. The dots that appear in the intermediate spaces are those CpG sites showing differential methylation as shown in the interpretation guide above (Figure 8). These intermediate sites are those which can experience positive association (upper right and lower left quadrants) or inverse association (upper right and lower left quadrants) to one another. Figure 10 shows the plots of two members of each group chosen at random, but these plots are representative of all plots obtained in a systematic pairwise comparison of all samples (data not shown). This result was unexpected and will be the basis for future investigation.

DNA methylation data was also analyzed through use of the caGEDA program created by Patel *et al* of the University of Pittsburgh Cancer Center.[51] The differential methylation analysis conducted in this experiment utilized the J5 test to do gene-specific comparisons and determine how the gene regions are expressed relative to one another. The figure below (Figure 10) shows those CpG islands with the greatest difference in methylation in a comparison of the case and seropositive control groups. This figure shows two separate differential methylation comparisons per figure. The image on the left is the comparison containing all the methylation samples per group without excluding any samples, based on the data pairs listed in bold in Tables 16-18. The image on the right contains the methylation samples excluding the samples with without a positive serostatus. In the figure below, a green bar corresponds to a gene that has decreased methylation in group 1 as compared to group 2 and a red bar corresponds to a gene that has increased methylation in group 1 as compared to group 2. The “group 1” and “group 2”

designations are defined by how the user groups the data during the loading of the methylation data into caGEDA. This differential analysis, performed in and generated by caGEDA will be utilized in the future to examine the various sites of differential methylation that are highlighted from the GenomeStudio output. This future work will determine the location of the CpG site in the genome, the corresponding gene, and downstream functionality of the gene and how the specific context of the methylation will impact gene functioning and regulation.

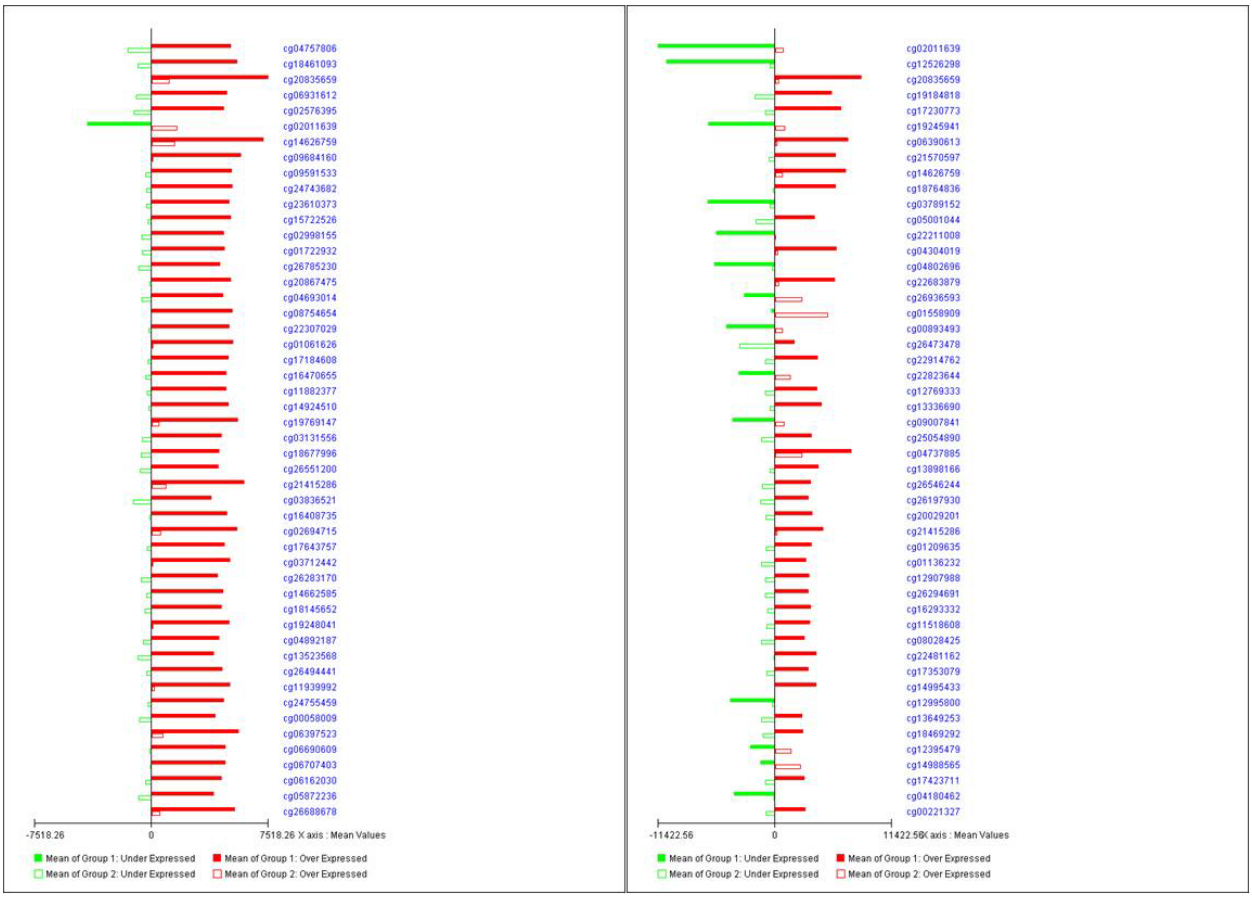

Figure 10. Differential Methylation in case vs seropositive control comparisons

4.0 DISCUSSION

The work presented in this thesis is a preliminary investigation of the molecular basis of the development of cognitive decline in the MSM population, and has already identified some exciting features that warrant further investigation. We had originally intended to focus on the development of HAND in HIV+ MSM, but a later re-evaluation of the demographic and serostatus information available on our samples indicated that we are in fact able to compare HIV+/HAND+ and HIV-/dementia+ MSM with both seronegative and seropositive controls. Five of the eight dementia case samples were seronegative, and not seropositive as originally thought (Table 16), and two of the eight seronegative samples that were examined on the methylation array were in fact seropositive (Table 16 and 18). This latter change did not materially affect our analysis, but the former revealed some fascinating features common to MSM with dementia, regardless of HIV serostatus.

Our initial goal was to identify a series of triad matches for the study, comprising HIV+/HAND+ samples and the corresponding HIV+/HAND- and HIV-/HAND- controls that would facilitate the comparison of our chosen biomarkers, but the distribution of serostatus values in our study set meant that this was not after all possible. We were, however, able to conduct several analyses and detect patterns in our data that suggest new directions for the analysis of aging and cognitive decline in MSM, and potentially in the broader population also.

Our first aim sought to determine the RYR3 genotype frequencies for the three sample groups examined (Table 9 and 10). The genotype frequencies observed were consistent with those described by Shrestha *et al*.[28] The TaqMan genotyping assay utilized for the RYR3 genotype data set of the regions of interest were consistent, indicating the assay is both specific and sensitive. Any samples failing the TaqMan were sequenced using PCR primers (Table 2) and the returned sequence data was analyzed using Sequencher. This strategy to capture failed samples through sequencing is viable and shall serve the future work of this pilot study well. The association of RYR3 with Atherosclerosis requires further examination to truly define the extent to which these factors can be attributed to the predisposition and the development of Neurocognitive Decline conditions such as HAND. However, previous research having indicated links between BMI and cardiovascular risk factors to dementia and other neurocognitive decline, it is recommended that RYR3 be examined in the future to truly characterize the existence of any association between RYR3, Atherosclerosis, and HAND.

The second component of our first aim sought to determine the APOE genotype frequencies for the three sample groups examined. The genotype frequencies were calculated and recorded (Table 11 – 13) but no strong associations or conclusions could be made from these allele frequencies since the sample size for the individuals possessing the risk allele was very small and lacked sufficient power when compared to those who do not have it. Also, a small proportion of study samples failed PCR amplification and this was believed to be attributable to template self-annealing due to the AT-rich nature of the APOE sequence. DMSO was used in the PCR reaction conditions to aid in linearizing the template because AT-rich primers require lower annealing temperatures; however some samples still failed amplification in spite of this DMSO addition. Bhaga *et al* determined that PCR conducted on GC-rich targets was optimized when

concentrations of a second chemical treatment, Betaine, were 1.5 M or less. It was also observed that the DMSO and Betaine strategy could be visualized on an agarose gel using Ethidium Bromide, whereas addition of 7-deaza-dGTP to PCRs prevent the application of this visualization strategy. [53]Therefore, it is recommended that Betaine be used in addition to DMSO and that reaction conditions be further specialized for these failed samples and those which could potentially fail amplification in the future.

The second aim of this project examined the change in absolute telomere length between two time points for each sample. As described above, synthetic standards were designed as amplification targets for the SYBR Green qPCR and a series of tenfold dilutions of these standards was prepared, to create standard curves from which calculations of the absolute telomere length would be rooted. It was observed through several attempts of this procedure that the standards were not amplifying in a consistent and reproducible manner. Normally, the Ct value for the standard would increase by roughly 3.3 cycles across each step of a tenfold serial dilution. This increase in the number of cycles is observed due to the increase in the number of cycles required for a perfect doubling of the PCR product when the sample is diluted tenfold. As the serial dilution makes the template more dilute, more time is required for a doubling of the reaction product to be observed. In this experiment, however, the increase in the amount of cycles across each step was not consistent and amplification of the standards (Telomere and 36B4) was variable. As a means of attempting to correct this issue, fresh sets of primers, SYBR Green reagent, and each of the synthetic standards were ordered based on the hypothesis that the variability was due in large part to a possible error during manufacturing or reagent degradation. However, this solution did not correct the inconsistency of the standards and so the values tabulated for this experimental aim are viewed as inaccurate. The fact that samples were

amplified but those results were inconsistent lead to the observation that the sensitivity of the assay, and not the specificity, was the issue. The extracted DNA samples ran consistently and their Ct values did not vary outside of the 0.5 cycles recommended by the SYBR Green troubleshooting guide. Although the data analysis was conducted as highlighted in the work by O'Callaghan and Fenech, alternative procedures are necessary for future work as we feel that this methodological approach is not a reliable one.[50]

One possible explanation for why the telomere data is inconsistent and recorded positive length changes was recently presented in several publications about methodologies for analysis of telomere length. It was observed that methods utilizing real time quantitative PCR, such as the SYBR Green methodology selected for this experiment, yielded inconsistent results which lead to the observation of unexpected data, such as positive telomere length changes. It has been suggested in these recent articles that methods utilizing hybridization (Southern Blotting) are considerably more accurate and reproducible in calculating the telomere length of extracted DNA samples.[54,55] Steenstrup *et al* concluded that the lengthening of telomeres, observed in longitudinal studies, is predominantly attributable to artifacts of measurement error and that short follow-up periods can make this issue worse. This group specifically determined that long follow-up periods and studies utilizing highly precise Southern blot methods had more accurate and reproducible telomere length results as compared to qPCR methods.[54] A second study, conducted by Aviv *et al* [55], showed similar results as the Steenstrup study and this validated the need to use alternative methods to qPCR.

Literature review of new experimental procedures for telomere length analysis, lead to the discovery of a technique in which DNA dot-blotting is employed to analyze telomere length as a function of and normalized against DNA load amount. Kimura *et al*, developed a DNA dot-

blot assay which uses a telomere probe to determine telomere length as a function of the ratio of the telomere probe signal divided by the amount of DNA present (Telomere/DNA).[56] The advantage observed in this paper is that only ~20ng of DNA is required for accurate results. Southern blotting can require comparatively larger quantities of DNA and so reducing the input DNA amount best serves future work. Since DNA was extracted from PBMCs fractionated out of blood draw samples, it is feasible that DNA loads could be too dilute from these extractions and undermine the capacity to employ Southern blotting. Therefore, it is the recommendation of this analysis to employ the DNA blot technique presented by Kimura *et al* in the future work of this study.

The third and final aim of this experiment sought to examine the differential methylation patterns through comparisons of the three experimental groups. By making comparisons between the Dementia, Seropositive, and Seronegative groups of the same cohort, it was hypothesized that genes of interest could be identified to guide future work, which will be rooted in this pilot study. These differential methylation comparisons were conducted using the GenomeStudio software provided by Illumina, and the caGEDA program developed at the University of Pittsburgh.[51]

The DNA methylation analysis for this pilot study was modified from what was initially intended, due to the earlier reported observation that the dementia group contained both seropositive and seronegative individuals. The analysis that was conducted, matching samples on dementia status instead of serostatus, yielded novel and unexpected results (Figures 4-9). The differential methylation analysis we conducted indicates that serostatus is not significantly predictive of the methylation pattern changes observed in the samples. This conclusion is based on the dendrogram of the dementia samples versus the controls (Figure 7) as well as the

differential methylation comparison between two samples from the various cohort groups (Figure 9). These comparisons indicate that it is the dementia status that is in fact associated with the aberrant methylation profile instead of the HIV serostatus, contrary to our initial hypothesis. This observation, that serostatus is not significantly linked to changes in DNA methylation pattern for dementia positive samples, is the landmark finding of our study. Our findings suggest that the phenomena seen in HIV+ MSM as HAND, and in HIV- MSM as cognitive decline, in fact share many molecular features, and may have similar underlying causes. Investigation of these claims will require a redefinition of the approach to the study of cognitive decline in MSM, and potentially in the broader population also. The differential methylation analysis conducted in GenomeStudio (Figure 9) shows that many CpG islands in the dementia positive samples are experiencing identical, as well as inversely related, degrees of methylation change when two samples are compared. Therefore, future analysis will strive to analyze the specific details of these CpG islands including: the context and extent to which CpG island methylation occurs within the genome, the corresponding gene in which the methylation change occurs, and also theorize about the downstream effects this methylation change will have on gene regulation and expression.

To achieve this goal for future work, the caGEDA program and packages developed in R will be utilized to examine the top differentially methylated sites (Figure 11). The returned differential methylation data from the caGEDA and other R packages will then guide the analysis of gene identification and downstream impact of the methylation change(s). For example, Figure 11 identifies the top differentially methylated CpG site for the Case vs seropositive group as cg04757806. Using GenomeStudio, it was determined that this CpG island is located in the FUT4 gene. It was determined that FUT4 codes for Fucosyltransferase 4, a protein which

transfers fucose to N-acetyllactosamine polysaccharides to generate fucosylated carbohydrate structures. Determining the importance of this methylation change and others within these samples will shed light onto their contributions to the development of neurocognitive decline related to HIV infection and potentially dementia conditions that are not related to HIV infection.

In conclusion, this pilot study sought to examine the predictive capacity of four unique biomarkers associated with conditions that relate to the development of neurocognitive decline. Cardiovascular health and risk factors such as BMI and Atherosclerosis have shown association with the development of neurocognitive decline and dementia. HIV-positive individuals on HAART therapy have experienced lipodystrophy and also have triglyceride levels that are elevated due to therapy. Therefore, a predictive biomarker related to heart disease which could then be applied to assessing the risk of a person on HAART developing HAND is still an important requirement, and our data warrants the inclusion of the RYR3 rs2229116 SNP in future work. This SNP is specifically associated with Atherosclerosis risk and warrants further assessment within the context of this study design to evaluate the predictive power the RYR3 SNP possesses as a biomarker for HAND.

The APOE*4 allele is well established as a risk factor for Alzheimer's Disease.[33,34,57–59] Therefore, its inclusion in future studies for the assessment of the risk of developing a neurocognitive decline in the HIV-positive population is clear. No evidence was generated in this pilot study to support the exclusion of APOE*4 from future work. The allele frequency for E4, compared to the E3 and E2 allele, was too small to draw definite conclusions. However, it is recommended that the supplement to DMSO treatment of the PCR, Betaine, be use on DNA samples that failed PCR amplification, as was observed in this study.

The telomere analysis was undermined by an inconsistent and non-reproducible assay design. This issue was addressed by the identification of the DNA dot-blot procedure defined by Kimura *et al* in a recent publication.[56] It is recommended that this DNA dot-blot technique be applied for the assessment of telomere length shortening in future study designs.

The methylation data analysis provided a novel, and unexpected, relationship between methylation profile and dementia status, rather than serostatus, in our samples. Therefore, future work will seek to more fully understand the similarities and differences between the development of cognitive decline in individuals with differing HIV serostatus. This research has identified a new means of characterizing cognitive decline at the molecular level, and is of great public health interest given the increase both in the number of aging individuals and in the prevalence of aging-related cognitive decline seen in the USA in recent decades.

BIBLIOGRAPHY

1. (UNAIDS), J. U. N. P. on H. *UNAIDS Global Report on the Global AIDS Epidemic 2012*. 110 (2012). at <http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_en.pdf>

2. AmfAR. Statistics: Worldwide. (2012). at <http://www.amfar.org/about_hiv_and_aids/facts_and_stats/statistics__worldwide/>

3. Araújo, L. A. L. & Almeida, S. E. M. HIV-1 diversity in the envelope glycoproteins: implications for viral entry inhibition. *Viruses* **5,** 595–604 (2013).

4. Girard, M. P., Osmanov, S., Assossou, O. M. & Kieny, M.-P. Human immunodeficiency virus (HIV) immunopathogenesis and vaccine development: a review. *Vaccine* **29,** 6191–218 (2011).

5. Duncan, C. J. a & Sattentau, Q. J. Viral determinants of HIV-1 macrophage tropism. *Viruses* **3,** 2255–79 (2011).

6. Flint, S. J., Enquist, L. W., Racaniello, V. R. & Skalka, A. M. *Principles of Virology: Pathogenesis and Control*. 1032 (ASM Press, 2009).

7. Levy, J. Pathogenesis of human immunodeficiency virus infection. *Microbiological reviews* **57,** 183–289 (1993).

8. Cohen, O. J. & Fauci, A. S. Current strategies in the treatment of HIV infection. *Disease-a-Month* **48,** 145–184 (2002).

9. Bonfanti, P., Capetti, A. & Rizzardini, G. HIV disease treatment in the era of HAART. *Biomedicine & pharmacotherapy = Biomédecine & pharmacothérapie* **53,** 93–105 (1999).

10. Temesgen, Z. Overview of HIV infection. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology* **83,** 1–5; quiz 6–7 (1999).

11. Stratigos, J. D., Johnson, R. a & Stratigos, a. HIV disease in the year 2000. *Clinics in dermatology* **18,** 375–9 (2000).

12. Navia, B. A., Jordan, B. D. & Price, R. W. The AIDS dementia complex: I. Clinical features. *Annals of neurology* **19,** 517–24 (1986).

13. Navia, B. A., Cho, E. S., Petito, C. K. & Price, R. W. The AIDS dementia complex: II. Neuropathology. *Annals of neurology* **19,** 525–35 (1986).

14. Grant, I. *et al.* Evidence for early central nervous system involvement in the acquired immunodeficiency syndrome (AIDS) and other human immunodeficiency virus (HIV) infections. Studies with neuropsychologic testing and magnetic resonance imaging. *Annals of internal medicine* **107,** 828–36 (1987).

15. Simioni, S. *et al.* Cognitive dysfunction in HIV patients despite long-standing suppression of viremia. *AIDS (London, England)* **24,** 1243–50 (2010).

16. Wright, E. J. *et al.* Cardiovascular risk factors associated with lower baseline cognitive performance in HIV-positive persons. *Neurology* **75,** 864–73 (2010).

17. Gras, G. & Kaul, M. Molecular mechanisms of neuroinvasion by monocytes-macrophages in HIV-1 infection. *Retrovirology* **7,** 30 (2010).

18. Ellis, R. J. *et al.* Neurocognitive Impairment Is an Independent Risk Factor for Death in HIV Infection. *Archives of Neurology* **54,** 416–424 (1997).

19. Gisslén, M., Price, R. W. & Nilsson, S. The definition of HIV-associated neurocognitive disorders: are we overestimating the real prevalence? *BMC infectious diseases* **11,** 356 (2011).

20. Cartier, L., Hartley, O., Dubois-Dauphin, M. & Krause, K.-H. Chemokine receptors in the central nervous system: role in brain inflammation and neurodegenerative diseases. *Brain research. Brain research reviews* **48,** 16–42 (2005).

21. Koenig, S. *et al.* Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science (New York, N.Y.)* **233,** 1089–93 (1986).

22. Glass, J. D., Fedor, H., Wesselingh, S. L. & McArthur, J. C. Immunocytochemical quantitation of human immunodeficiency virus in the brain: correlations with dementia. *Annals of neurology* **38,** 755–62 (1995).

23. Kelder, W., McArthur, J. C., Nance-Sproson, T., McClernon, D. & Griffin, D. E. Beta-chemokines MCP-1 and RANTES are selectively increased in cerebrospinal fluid of patients with human immunodeficiency virus-associated dementia. *Annals of neurology* **44,** 831–5 (1998).

24. Letendre, S. L., Lanier, E. R., Mccutchan, J. A. & Carolina, N. Cerebrospinal Fluid b Chemokine Concentrations in Neurocognitively Impaired Individuals Infected with Human Immunodeficiency Virus Type 1. 310–319 (1986).

25. Weber, E., Blackstone, K. & Woods, S. P. Cognitive neurorehabilitation of HIV-associated neurocognitive disorders: a qualitative review and call to action. *Neuropsychology review* **23,** 81–98 (2013).

26. Chambless, L. E. *et al.* Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: the Atherosclerosis Risk in Communities (ARIC) Study, 1987-1993. *American journal of epidemiology* **146,** 483–94 (1997).

27. Baker, J. V *et al.* Progression of carotid intima-media thickness in a contemporary human immunodeficiency virus cohort. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **53,** 826–35 (2011).

28. Shrestha, S. *et al.* Replication of RYR3 gene polymorphism association with cIMT among HIV-infected whites. *AIDS (London, England)* **26,** 1571–3 (2012).

29. Shrestha, S. *et al.* NIH Public Access. **24,** 583–592 (2011).

30. Kaplan, R. C. *et al.* T cell activation and senescence predict subclinical carotid artery disease in HIV-infected women. *The Journal of infectious diseases* **203,** 452–63 (2011).

31. Kuller, L. H. *et al.* Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. *PLoS medicine* **5,** e203 (2008).

32. Neuhaus, J. *et al.* Markers of inflammation, coagulation, and renal function are elevated in adults with HIV infection. *The Journal of infectious diseases* **201,** 1788–95 (2010).

33. Sun, X., Nicholas, J., Walker, A., Wagner, M. T. & Bachman, D. APOE genotype in the diagnosis of Alzheimer's disease in patients with cognitive impairment. *American journal of Alzheimer's disease and other dementias* **27,** 315–20 (2012).

34. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34,** 939–44 (1984).

35. Strittmatter, W. J. & Roses, a D. Apolipoprotein E and Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 4725–7 (1995).

36. McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **7,** 263–9 (2011).

37. Fyhrquist, F. & Saijonmaa, O. Telomere length and cardiovascular aging. *Annals of medicine* **44 Suppl 1,** S138–42 (2012).

38. Harley, C. B., Futcher, a B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345,** 458–60 (1990).

39. Kimura, M. *et al.* Synchrony of telomere length among hematopoietic cells. *Experimental hematology* **38,** 854–9 (2010).

40. Kong, C. M., Lee, X. W. & Wang, X. Telomere Shortening in Human Diseases. *The FEBS journal* (2013). doi:10.1111/febs.12326

41. Sanders, J. L. & Newman, A. B. Telomere Length in Epidemiology: A Biomarker of Aging, Age-Related Disease, Both, or Neither? *Epidemiologic reviews* (2013). doi:10.1093/epirev/mxs008

42. Martin-Ruiz, C. *et al.* Telomere length predicts poststroke mortality, dementia, and cognitive decline. *Annals of neurology* **60,** 174–80 (2006).

43. Pomraning, K. R., Smith, K. M. & Freitag, M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods (San Diego, Calif.)* **47,** 142–50 (2009).

44. Hornby, D. P. DNA Methylation. 1–4 (2002).

45. Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nature reviews. Genetics* **11,** 191–203 (2010).

46. Robertson, K. D. DNA methylation and human disease. *Nature reviews. Genetics* **6,** 597–610 (2005).

47. Feinberg, A. P. & Tycko, B. The history of cancer epigenetics. *Nature reviews. Cancer* **4,** 143–53 (2004).

48. Sinsheimer, J. S. *et al.* Epigenetic Predictor of Age. **6,** 1–6 (2011).

49. Aquila, P. D., Rose, G., Bellizzi, D. & Passarino, G. Maturitas Epigenetics and aging. **74,** 130–136 (2013).

50. O'Callaghan, N. J. & Fenech, M. A quantitative PCR method for measuring absolute telomere length. *Biological procedures online* **13,** 3 (2011).

51. Patel, S. & Lyons-Weiler, J. caGEDA: a web application for the integrated analysis of global gene expression patterns in cancer. *Applied bioinformatics* **3,** 49–62 (2004).

52. Jordan, R., Patel, S., Hu, H. & Lyons-Weiler, J. Efficiency analysis of competing tests for finding differentially expressed genes in lung adenocarcinoma. *Cancer informatics* **6,** 389–421 (2008).

53. Bhagya, C. H. W. M. R. C., Wijesundera Sulochana, W. S. & Hemamali, N. P. Polymerase chain reaction optimization for amplification of Guanine-Cytosine rich templates using buccal cell DNA. *Indian journal of human genetics* **19,** 78–83 (2013).

54. Steenstrup, T., Hjelmborg, J. V. B., Kark, J. D., Christensen, K. & Aviv, A. The telomere lengthening conundrum--artifact or biology? *Nucleic acids research* **41,** 1–7 (2013).

55. Aviv, A. *et al.* Impartial comparative analysis of measurement of leukocyte telomere length/DNA content by Southern blots and qPCR. *Nucleic acids research* **39,** e134 (2011).

56. Kimura, M. & Aviv, A. Measurement of telomere DNA content by dot blot analysis. *Nucleic acids research* **39,** e84 (2011).

57. Liu, C.-C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature reviews. Neurology* **9,** 106–18 (2013).

58. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* **90,** 1977–81 (1993).

59. Eriksson, U. K., Bennet, A. M., Gatz, M., Dickman, P. W. & Pedersen, N. L. Nonstroke cardiovascular disease and risk of Alzheimer disease and dementia. *Alzheimer disease and associated disorders* **24,** 213–9 (2010).