2013-03-19

Mitochondrial Genetics of Alzheimer's Disease and Aging

Perry Gene Ridge
Brigham Young University - Provo

Mitochondrial Genetics of Alzheimer's Disease and Aging

Perry G. Ridge

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

John S. K. Kauwe, Chair
Karl V. Voelkerding
Mark J. Clement
Dennis L. Eggett
Nathan E. Lewis

Department of Biology

Brigham Young University

March 2013

ABSTRACT

Mitochondrial Genetics of Alzheimer's Disease and Aging

Perry G. Ridge
Department of Biology, BYU
Doctor of Philosophy

Mitochondria are essential cellular organelles and the location of the electron transport chain, the site of the majority of energy production in the cell. Mitochondria contain their own circular genome approximately 16,000 base pairs in length. The mitochondrial genome (mtDNA) encodes 11 protein-coding genes essential for the electron transport chain, 22 tRNA genes, and two rRNA genes. Mitochondrial malfunction occurs in many diseases, and changes in the mitochondrial genome lead to numerous disorders. Multiple mitochondrial haplotypes and sequence features are associated with Alzheimer's disease. In this dissertation we utilized TreeScanning, an evolutionary-based haplotype approach to identify haplotypes and sequence variation associated with specific phenotypes: Alzheimer's disease case-control status, mitochondrial copy number, and 16 neuroimaging phenotypes related to Alzheimer's disease neurodegeneration. In the first two studies we utilized 1007 complete mitochondrial genomes from participants in the Cache County Study on Memory Health and Aging. First, individuals with mitochondrial haplotypes H6A1A and H6A1B showed a reduced risk of AD. Our study is the largest to date and the only study with complete mtDNA genome sequence data. Next, each cell contains multiple mitochondria, and each mitochondrion contains multiple copies of its own circular genome. The ratio of mitochondrial genomes to nuclear genomes is referred to as mitochondrial copy number. Decreases in mitochondrial copy number are known to occur in many tissues as people age, and in certain diseases. Three variants belonging to mitochondrial haplogroups U5A1 and T2 were significantly associated with higher mitochondrial copy number in our dataset. Each of these three variants was associated with higher mitochondrial copy number and we suggest several hypotheses for how these variants influence mitochondrial copy number by interacting with known regulators of mitochondrial copy number. Our results are the first to report sequence variation in the mitochondrial genome that lead to changes in mitochondrial copy number. The identification of these variants that increase mtDNA copy number has important implications in understanding the pathological processes that underlie these phenotypes. Lastly, we used an endophenotype-based approach to further characterize mitochondrial genetic variation and its relationship to risk markers for Alzheimer's disease. We analyzed longitudinal data from non-demented, mild cognitive impairment, and late onset Alzheimer's disease participants in the Alzheimer's Disease Neuroimaging Initiative with genetic, brain imaging, and behavioral data. Four clades were associated with three different endophenotypes: whole brain volume, percent change in temporal pole thickness, and left hippocampal atrophy over two years. This was the first study of its kind to identify mitochondrial variation associated with brain imaging endophenotypes of Alzheimer's disease. Together, these projects provide evidence of mtDNA involvement in the risk and physiological changes of Alzheimer's disease.

Keywords: mitochondrial genetics, haplotypes, mitochondrial copy number, Alzheimer's disease, whole genomes, endophenotypes

ACKNOWLEDGEMENTS

I have received tremendous assistance as I have pursued my doctoral degree and I cannot hope to name all the individuals that contributed to my success; however, I would like to name a few. My path to a PhD has been very untraditional. I started my PhD at the University of Nebraska-Lincoln in 2006, nearly seven years ago. At the time my wife and I, Kristen, had been married about a year and had just been blessed with our first child, Shaelynn. After three years we moved from Lincoln to Meridian, ID, unsure of what our future plans were and how continued education would be possible with our growing family (Grace joined our family in 2008). While we were living in Meridian I found employment with ARUP Laboratories in Salt Lake City and we moved two more times, first to North Salt Lake and then a few months later to Bountiful. A short time later Caroline, our third daughter, was born. After I had been at ARUP Laboratories for about two years, I sought the advice of Keith Crandall, a previous mentor, and determined that I would finish my PhD at BYU. Subsequently, we moved two more times (Bountiful to Orem and then Orem to Provo) and had our first son (Nathaniel). I share this so that all will know what Kristen has endured so that I can pursue my dreams: six moves in three states in seven years, four children at two year intervals, and four different jobs. She has never complained about my educational pursuits and has always been supportive when I asked for more time. A lesser woman would not have sacrificed many of her own dreams so that I can have mine, nor stood by me and sacrificed the way Kristen has. I love Kristen and am eternally grateful for her support and loyalty.

At many times in this journey it appeared that the trail ended, that there was no way I could continue my education. At each of these points the Lord has opened doors and presented a path, often better than what I had anticipated or hoped for. I am grateful to my Father in Heaven

for providing a way for me to become what I want to be. Looking back I recognize that there are no coincidences in my journey, rather they are many times that the Lord opened doors that I did not know even existed.

One of the reasons I chose to accept employment with ARUP Laboratories was their commitment to me to support me while I pursued my PhD and allow me to continue my full time employment. I did not feel that with a growing family I could leave steady employment, benefits, and a salary. I am grateful for Karl Voelkerding and ARUP Laboratories for supporting me in my education. When it became clear that good choices for graduate study for someone working did not exist at the University of Utah, Karl helped me find suitable options and even though I know my productivity at ARUP has been less while I have worked on my PhD and split time between BYU and ARUP, Karl has always been supportive. Even now Karl and I regularly brain storm about different ways (mostly his own ideas) that I can get to where I want to be professionally, even if it means leaving ARUP.

Finally, I am grateful for Keoni, the Biology Department, and BYU. When it became clear that the University of Utah was not going to be an option for graduate study and full time employment, I went on a crusade to find a graduate program that would allow me to work while studying. With one exception, I found no schools willing to allow me to work, in spite of my willingness to do whatever it took to be successful. When I sought Keith's advice, without hesitation he told me that BYU could make it work and he suggested Keoni's lab. Keoni has been the type of mentor every faculty member should strive to be. He has supported my educational pursuits, helped me maintain a healthy balance between work and family (and has walked the walk), like Karl he has constantly fought for me and my personal and professional successes (and I am confident he will as long as I want his support), and has become a close

friend. I never felt like his indentured servant, even if he was entitled to that type of effort and time from me, and in fact I was treated more like his professional colleague than student.

There are many other people that have helped me and my family along the way, but there is not space here to acknowledge every person that has blessed my life. I would however, like to acknowledge Taylor Maxwell who helped me (and is a co-author) with each of these projects, the faculty and staff at Utah State University who are involved with the Cache County Study on Memory Health and Aging that provided much of the data I analyzed, and the staff at the Fulton Supercomputing Lab. Nearly every analysis I have run since arriving at BYU has been on Marylou. My education experience has been unique, but I feel that because of the people I have mentioned here, my committee, and BYU I have been able to receive training that is competitive with what is available at any university in the world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Genetics of Alzheimer's disease

Perry G. Ridge

Introduction

Alzheimer's disease (AD) is the most common form of dementia. Worldwide estimates of incidence vary, with estimates of 24 to 35 million people affected (Brookmeyer et al. 2007; Querfurth, LaFerla 2010; Ballard et al. 2011) and with an aging population, incidence is expected to increase to 1 in 85 people by 2050 (Brookmeyer et al. 2007). Typically, persons diagnosed with AD survive 3 to 9 years after diagnosis (Querfurth, LaFerla 2010), with a median of 7 years (Molsa, Marttila, Rinne 1986), and fewer than 3% survive more than 14 years (Molsa, Marttila, Rinne 1995). The disease is progressive and people with late stage AD typically require full time care.

Given the need for full time care in late stages of the disease, early diagnosis is important. Effective AD diagnostics remain challenging given its overlap with other dementias. The National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's disease and Related Disorders Association have jointly established criteria for the diagnosis of AD (Dubois et al. 2007). A diagnosis of probable AD is made based on meeting criteria in two areas: core diagnostic criteria and supportive features. To receive a diagnosis of probable AD a person must meet all criteria for core diagnostic criteria and one of four possible supportive features. Certain exclusion criteria exist, which if present, prevent diagnosis of probable AD. Diagnosis based on the core criteria is challenging because the criteria rely primarily on clinical

observations/history. A full description of AD diagnosis can be found in Dubois et al. (Dubois et al. 2007).

One of the obstacles to effective diagnosis is that, while a number of hypotheses exist, the exact cause of AD is unknown. The most widely accepted hypothesis is the amyloid cascade hypothesis (Hardy, Higgins 1992). In the amyloid cascade hypothesis, amyloid precursor protein (APP) is cleaved by various secretases (α, β, and γ) into peptides, one of which is $A\beta_{42}$. $A\beta_{42}$ self-aggregates and can grow into extracellular fibrils arranged into β-pleated sheets which are the insoluble fibers of amyloid or senile plaques (SP) (Querfurth, LaFerla 2010). This is thought to be the first step in the development of AD (Armstrong 2011). Subsequently, intracellular neurofibrillary tangles (NFT) form, largely composed of hyperphosphorylated tau proteins. The formation of NFTs is largely thought to be driven by the accumulation of SPs (Querfurth, LaFerla 2010). The presence of SPs and NFTs are the hallmark pathologies of AD (Newell et al. 1999).

Another hypothesis of AD involves the mitochondria. It is widely accepted that mitochondrial function is disrupted in the brains of AD patients (Swerdlow, Khan 2004; Onyango et al. 2006; Mancuso et al. 2008; Ankarcrona, Mangialasche, Winblad 2010; Querfurth, LaFerla 2010; Swerdlow, Burns, Khan 2010). It is also known that SPs aggregate within mitochondria (Devi et al. 2006; Anandatheerthavarada, Devi 2007). However, it is not known whether mitochondrial dysfunction is a cause or effect of SP aggregation (Mancuso et al. 2008). These questions led to the proposal of the mitochondrial cascade hypothesis (Swerdlow, Khan 2004). Briefly, mitochondrial function and morphology change and decline with age (Chan 2006; Swerdlow,

Burns, Khan 2010). As function begins to decline, mitochondria try to compensate. During this phase, the compensation causes alterations in the mitochondria. Finally, as the mitochondria begin to fail there are additional compensatory changes. Changes such as Aβ aggregation and tau phosphorylation are some of the transformations that occur as a result of compensating and failing mitochondria. In contrast, in the amyloid cascade hypothesis, changes such as Aβ aggregation and tau phosphorylation happen first and lead to the dysfunction of mitochondria (Swerdlow, Khan 2004; Swerdlow, Khan 2009; Swerdlow, Burns, Khan 2010). Each of these hypotheses is likely to be affected by both genetic and non-genetic factors.

Various non-genetic factors have an impact on both risk for and protection from AD. The greatest non-genetic risk factor is age (Herrup 2010; Querfurth, LaFerla 2010). Other risk factors include hypertension, use of estrogen (Patterson et al. 2008), smoking (Almeida et al. 2002; Cataldo, Prochaska, Glantz 2010), stroke, heart disease, depression, arthritis, and diabetes (Lindsay et al. 2002). On the other hand, certain lifestyle choices appear to decrease the risk of AD: exercise (Podewils et al. 2005), intellectual stimulation (Wang et al. 2002), and maintaining a Mediterranean diet (including fish) (Scarmeas et al. 2006; Patterson et al. 2007).

The genetics of AD are complicated. AD is a highly heterogeneous disorder and in this review we will discuss the genetics of AD, including a discussion of causative genes as well as genes with replicable association with AD. Additionally, we will review cerebrospinal fluid biomarkers that can be used to assess both risk for and rate of progression of AD.

Genetics

Several genes, or haplotypes, are known to harbor either causative or risk variants for AD. There are two primary types of AD as defined by age. The first is early onset, or familial AD, and the second type is late onset AD, or sporadic AD. Each has a set of causative or risk genetic factors.

Early onset Alzheimer's disease

Early onset AD (EOAD) typically begins before age 65, and estimates for prevalence vary depending on the study, but range from 0-1% (Blennow, de Leon, Zetterberg 2006) to 6%-7% (Patterson et al. 2008) of total AD cases. While it is believed that familial AD is dominantly inherited, it is not completely penetrant, and in fact fewer than 13% of EOAD cases demonstrate a fully penetrant autosomal dominant inheritance for multiple generations (Patterson et al. 2008). Mutations in three different genes are known to cause EOAD: amyloid beta (A4) precursor protein (APP) (Goate et al. 1991), presenilin 1 (PSEN1) (Sherrington et al. 1995), and presenilin 2 (PSEN2) (Levy-Lahad et al. 1995).

APP

APP was one of the first genes identified to cause AD. APP is located on chromosome 21 (21q21.2-21q21.3). There are at least 10 different APP isoforms. The primary transcript (NM_000484, NP_000475) is also the longest transcript with 18 exons. The exact function of APP is not certain, but several possible functions have been suggested such as synaptic development (Priller et al. 2006), neuronal migration (Young-Pearse et al. 2007), or as a receptor, although there have been arguments against this (Thinakaran, Koo 2008). It is clear, however, that APP is cleaved into Aβ molecules, including $A\beta_{42}$, which are secreted and can then

accumulate in the brain forming the hallmark SPs of AD (Querfurth, LaFerla 2010). At least 25 pathogenic mutations have been identified in APP, the majority located within or adjacent to the Aβ domain (http://www.molgen.ua.ac.be/ADMutations) (Cruts, Van Broeckhoven 1998a; Thinakaran, Koo 2008). Mutations in APP account for 13-16% of all EOAD cases (Janssen et al. 2003; Raux et al. 2005).

PSEN1

PSEN1 is located on chromosome 14 (14q24.3) and there are at least two known isoforms. Of the three genes known to cause EOAD, mutations in PSEN1 account for a greater percentage of EOAD cases (18-50%) than either of the other genes (Campion et al. 1995; Hutton et al. 1996; Cruts et al. 1998). To date, there are at least 185 known AD causing mutations in PSEN1 (http://www.molgen.ua.ac.be/ADMutations) (Cruts, Van Broeckhoven 1998b; Cruts, Van Broeckhoven 1998a). PSEN1 EOAD is autosomal dominant, however it is incompletely penetrant. Furthermore, there can be substantial variation in age at onset, rate of progression, and severity of disease. Some of the variation is attributable to specific mutations in PSEN1 (Moehlmann et al. 2002; Heckmann et al. 2004; Rudzinski et al. 2008).

PSEN1 is a component of γ-secretase, which is one of the secretases responsible for APP cleavage (Steiner et al. 2002). Mutations in PSEN1 can change the secretase activity of γ-secretase and increase the ratio of $A\beta_{42}$ to $A\beta_{40}$-and $A\beta_{42}$ more readily forms SPs (Schellenberg et al. 1992; Citron et al. 1997).

PSEN2

PSEN2 is located on chromosome 1 (1q31-q42) and has two known isoforms. EOAD causing mutations in PSEN2 are relatively rare compared to PSEN1 and appear to have a more variable penetrance (Sherrington et al. 1996). To date, there are at least 12 known pathogenic mutations in PSEN2, in contrast to the 185 and 25 known pathogenic mutations of PSEN1 and APP, respectively (Cruts, Van Broeckhoven 1998a; Cruts, Van Broeckhoven 1998b). While the exact function of PSEN2 is unknown, it is believed to have very a similar function to PSEN1 (as described above) (Kovacs et al. 1996), and to cause AD pathology by increasing levels of $A\beta_{42}$ (Steiner et al. 2002).

Late onset Alzheimer's disease

The second type of AD is late onset AD (LOAD) or sporadic AD. Even though numerous genetic risk factors and biomarkers have been identified for LOAD, no causative gene has been identified. Ten different loci (Table 1) meet all the criteria to be included on the "Top Results" list of the Alzheimer Research Forum or ALZGENE (accessed October 2011, for details about construction of the list see www.alzgene.org) for associations with AD. In this section we briefly introduce each of these loci in the following groups (grouped by common function, pathway, or family): apolipoproteins and lipid homeostasis, genes involved in endocytosis, MS4 family proteins, and others.

Apolipoproteins and Lipid Homeostasis

Apolipoproteins are a family of proteins involved in lipid homeostasis. These proteins bind and transport lipids through the lymphatic and circulatory systems. Two different apolipoproteins and

an ABC transporter have been shown to associate with AD. The first is apolipoprotein E (APOE), which is located on chromosome 19 (19q13.2) and consists of four total exons (three coding). There is only one major isoform (NM_000041, NP_000032), which encodes a protein 317 amino acids in length. APOE is a component of the chylomicron and plays a pivotal role in very low density lipoprotein clearance from circulation (Mahley 1988). Impaired function of APOE results in increased plasma levels of cholesterol and triglycerides (Mahley 1988).

There are three primary APOE alleles: ε2 (rs429358), ε3, and ε4 (rs7412). These alleles differ by substitutions at positions 112 and 158 (protein positions correspond to the processed protein). ε3, the wild type allele, is Cys112 and Arg158, ε2 is Cys112 and Arg158Cys, and ε4 is Cys112Arg and Arg158. ε3, on average, has a population frequency of 78.3% (8.5%-98% depending on population), whereas ε2 has a population frequency of 6.4% (0%-37.5%, depending on population) and ε4 14.5% (0%-49%, depending on population) (Eisenberg, Kuzawa, Hayes 2010). The ε4 allele is the risk allele, and is the most significant known genetic risk factor for LOAD. This allele was first identified as a genetic risk factor for LOAD in 1993 by Corder et al (Corder et al. 1993). The association for this allele has been replicated numerous times in various ethnic groups including: Caucasians (Corder et al. 1993), African Americans (Hendrie et al. 1995; Maestre et al. 1995), Asians (Noguchi, Murakami, Yamada 1993; Ueki et al. 1993), and Hispanics (Maestre et al. 1995). The ε4 allele is the only widely accepted genetic risk factor for LOAD (Harold et al. 2009).

AD risk increases with increasing dosage of ε4, whereas ε2 appears to be protective (Corder et al. 1994). Possible genotypes (listed in increasing risk of AD) are: ε2/ε2, ε2/ε3, ε3/ε3 or ε2/ε4, ε3/ε4,

and ε4/ε4 (Corder et al. 1994). Although risk for AD is much higher in persons with one or more ε4 alleles, there exist ε4/ε4 individuals who never develop AD (Corder et al. 1993).

Despite APOE's importance in AD genetics its exact role in AD is unknown; however, levels of $A\beta_{42}$ deposition in the brain are correlated with the number of ε4 alleles (Reiman et al. 2009) and it is hypothesized that APOE is involved in the clearance of $A\beta_{42}$ from the brain, proteolytic degradation of $A\beta_{42}$, and astrocyte mediated degradation of $A\beta_{42}$ (Bales et al. 2002; Koistinaho et al. 2004; Jiang et al. 2008).

The second apolipoprotein associated with AD is clusterin (CLU). A single variant, rs11136000, in CLU has been associated with AD in multiple different ethnic groups as a protective allele (Tycko et al. 1996; Giedraitis et al. 2009; Harold et al. 2009; Lambert et al. 2009; Biffi et al. 2010; Carrasquillo et al. 2010; Corneveaux et al. 2010; Guerreiro et al. 2010; Jun et al. 2010; Kamboh et al. 2010; Seshadri et al. 2010; Yu et al. 2010; Hu et al. 2011; Naj et al. 2011; Wijsman et al. 2011). CLU, also known as apolipoprotein J, is located on chromosome 8 (8p21-p12). It has been suggested that CLU may increase the toxicity of $A\beta_{42}$ (DeMattos et al. 2002), but that it might also be involved in $A\beta_{42}$ clearance (DeMattos et al. 2004; Bell et al. 2007). Additionally, it has been shown that AD affected people have increased CLU in circulation, and that increased CLU levels are suggestive of a higher rate of cognitive decline (Thambisetty et al. 2010; Schrijvers et al. 2011; Thambisetty et al. 2012). Lastly, there is evidence that Aβ increases expression of CLU (LaDu et al. 2000) and of a direct interaction between $A\beta_{40}$ and CLU (Matsubara et al. 1996; Zlokovic et al. 1996; Trougakos, Gonos 2006).

Another gene, ATP-binding cassette, subfamily A (ABC1), member 7 (ABCA7), was recently identified as an AD susceptibility locus based on a significant association of rs3764650, located in intron 13 of ABCA7, and AD (Hollingworth et al. 2011; Naj et al. 2011). ABCA7 is an ATP-binding cassette transporter used to move numerous molecules across membranes and interference of ABCA7 decreases phagocytosis (Tanaka et al. 2010). ABCA7 helps maintain lipid homeostasis through its role in lipid transport across the cellular membrane (Hayashi et al. 2005; Tanaka et al. 2011). Additionally, ABCA7 expression is responsive to lipoprotein levels and type (Kaminski et al. 2000). Lipid dysfunction, changes in lipid homeostasis, and modifications of neuronal membrane homeostasis can all cause numerous diseases, including AD (Bales 2010; Di Paolo, Kim 2011; Matsuzaki et al. 2011), and therefore provide some ideas about how ABCA7 can lead to AD. rs3764560 is associated with increased risk for AD and likely disrupts, or is associated with a variant that disrupts, lipid homeostasis and/or membrane homeostasis. This makes sense given that ABCA7 is a lipid transporter and is involved in phagocytosis.

Genes Involved in Endocytosis

Endocytosis is the process a cell uses to transport molecules across the cellular membrane and into the cell. Previous studies have demonstrated a role for endocytosis in general, and clathrin-mediated endocytosis specifically, in AD (Wu, Yao 2009). APP is processed in endosomes, therefore endocytosis of APP from the cell surface is necessary for $A\beta_{42}$ production. Inhibiting clathrin-mediated endocytosis decreases levels of $A\beta_{42}$ (Wu, Yao 2009). Given the importance of endocytosis in AD it makes sense that several genes with variants associated with AD are involved in endocytosis.

First, rs744373, located upstream of bridging integrator 1 (BIN1) is associated with AD (Harold et al. 2009; Biffi et al. 2010; Seshadri et al. 2010; Hollingworth et al. 2011; Hu et al. 2011; Lee et al. 2011; Naj et al. 2011; Wijsman et al. 2011). BIN1 is located on chromosome 2 (2q14) and has at least 10 different isoforms. BIN1 has multiple functions. First, BIN1 is involved in synaptic vesicle endocytosis (Cousin, Robinson 2001; Seshadri et al. 2010). Like clathrin-mediated endocytosis, although to a lesser extent, synaptic activity-endocytosis has a role in the processing of APP (Wu, Yao 2009). Second, BIN1 decreases the formation of clathrin-coated vesicles—a necessary step in clathrin-mediated endocytosis (Simpson et al. 1999). Mutations in BIN1 could, therefore, have different effects on the risk for AD. Variants, which adversely affect BIN1's role in synaptic vesicle endocytosis, would likely be protective since they would decrease APP processing. In contrast, BIN1 variants that affect its ability to decrease the formation of clathrin-coated vesicles would increase clathrin-mediated endocytosis and subsequent APP processing resulting in increased $A\beta_{42}$ production. These variants would increase risk for AD. A single variant could, conceivably, have both effects, but since clathrin-mediated endocytosis has a larger role in APP processing the net effect would be increased risk for AD. rs744373 in BIN1 is associated with increased risk for AD.

Another gene associated with AD and endocytosis is phosphatidylinositol binding clathrin assembly protein (PICALM), which is located on chromosome 11 (11q14) and has at least four known isoforms. Harold et al. (Harold et al. 2009) identified a single variant, rs3851179, with strong association with AD controls compared to cases. This same association has been replicated several times (Reiman et al. 2007; Carrasquillo et al. 2010; Jun et al. 2010; Seshadri et

al. 2010; Hu et al. 2011; Naj et al. 2011). PICALM is involved in protein trafficking and synaptic vesicle endocytosis and may control levels of GluR2 and VAMP2 (Cousin, Robinson 2001; Harel et al. 2008; Harel, Mattson, Yao 2011). Its main function, however, is as a clathrin assembly protein, where it increases the assembly of clathrin-coated vesicles and helps regulate the amount of membrane recycling and clathrin-mediated endocytosis (Dreyling et al. 1996; Harel, Mattson, Yao 2011). The finding that rs3851179 is a protective allele against AD fits with a hypothesis that this variant decreases formation of clathrin-coated vesicles by disrupting PICALM function.

The last gene in the endocytic set is complement component (3b/4b) receptor 1 (CR1). CR1 was first identified as a risk locus for AD in 2009 (rs3818361) (Reiman et al. 2007; Harold et al. 2009; Lambert et al. 2009), with replication in several ethnic groups (Carrasquillo et al. 2010; Corneveaux et al. 2010; Jun et al. 2010; Seshadri et al. 2010; Zhang et al. 2010). CR1 is located on chromosome 1 (1q32), and has at least two known isoforms. Although an exact function for CR1 is not known, it has been suggested that CR1, working with C3b (a complement fragment in the complement cascade), plays a role in Aβ clearance (Rogers et al. 2006; Lambert et al. 2009; Zhang et al. 2010). Additionally, CR1 appears to facilitate endocytosis (Arumugam et al. 2006). rs3818361 is associated with increased risk for AD. Variants in CR1 could potentially cause AD by disrupting its Aβ clearing function or by a gain of function mutation resulting in increased endocytosis.

Lastly, an association with AD for rs9349407, in a new AD susceptibility gene, CD2-associated protein (CD2AP), was recently reported (Hollingworth et al. 2011; Naj et al. 2011). CD2AP is

located on chromosome 6 (6p12). CD2AP is responsible for regulation of the actin cytoskeleton (Ma et al. 2010; Yaddanapudi et al. 2011). Additionally, CD2AP is involved in receptor-mediated endocytosis (Kobayashi et al. 2004). Changing endocytosis can modify lipid homeostasis and APP processing, among other things, and is a plausible explanation for how rs9349407, or a variant linked with rs9349407, could cause AD.

MS4A6A and MS4A4E

Membrane-spanning 4-domains, subfamily A, member 6A (MS4A6A) and membrane-spanning 4-domains, subfamily A, member 4E (MS4A4E) were only recently identified as AD risk loci with rs610932 (MS4A6A) and rs670139 (MS4A4E) showing association with AD (Harold et al. 2009; Hollingworth et al. 2011; Naj et al. 2011). rs610932 is located in the 3'-UTR of MS4A6A and rs670139 in the intergenic region between MS4A6A and MS4A4E and each has a different association with AD. rs610932 is protective against AD, while rs670139 increases risk for AD. MS4A6A and MS4A4E are located together on chromosome 11 (11q12.1 and 11q12.2, respectively) with at least four and one known isoform(s), respectively. They are located in a cluster with other MS4A (membrane-spanning 4-domains subfamily A) subfamily genes (Ishibashi et al. 2001; Liang et al. 2001). Very little is known about the function of either of these genes.

Others

Lastly, there is one other locus associated with AD, which did not fit in any of the above categories. An association for rs3865444 in CD33 molecule (CD33) was initially identified in 2008 (Bertram et al. 2008), and was replicated (Harold et al. 2009; Bettens et al. 2010;

Hollingworth et al. 2011; Naj et al. 2011). CD33 is located on chromosome 19q13.3 with at least three known isoforms. CD33 is a myeloid antigen and is expressed in a variety of tissues and cell types and is especially prevalent in leukemia (Brinkman-Van der Linden et al. 2003).

Mitochondrial Genetics and Alzheimer's disease

As explained above, the malfunction of mitochondria in AD is well known. Whether these changes are a cause or effect of AD is unclear. Similarly, what role, if any, the mitochondrial genome has in AD risk is unknown. Numerous studies have been performed analyzing mitochondrial variation and/or haplotypes to identify sequence features in the mitochondrial genome associated with AD. While a number of these studies have identified significant associations, there is no consensus and some of these studies offer conflicting results. In Table 2 we list a summary of studies performed looking at variation in the mitochondrial genome and its role in AD.

Conclusions

Here we have reviewed known genetic risk and protective factors and CSF biomarkers of AD. The findings of research efforts thus far are substantial. However, while there has been substantial progress in the quest to understand AD, there remain numerous, important unanswered questions. How can we accurately diagnose AD early and predict disease severity in those diagnosed? Which variants are responsible for LOAD? Which genes and interactions of genes cause disease? How significant are environmental factors versus genetic factors? Answers to these and other questions are likely to be found as research builds upon current knowledge.

In this research we studied the contribution of the mitochondrial genome to risk for AD and numerous endophenotypes for AD. Our novel approach provides further evidence of a direct role of the mitochondrial genome in risk for and the pathophysiology of AD. In addition, we make observations about the general state of AD genetics and suggest novel study designs for identifying additional AD genes and markers.

References

Almeida, OP, GK Hulse, D Lawrence, L Flicker. 2002. Smoking as a risk factor for Alzheimer's disease: contrasting evidence from a systematic review of case-control and cohort studies. Addiction 97:15-28.

Anandatheerthavarada, HK, L Devi. 2007. Amyloid precursor protein and mitochondrial dysfunction in Alzheimer's disease. Neuroscientist 13:626-638.

Ankarcrona, M, F Mangialasche, B Winblad. 2010. Rethinking Alzheimer's disease therapy: are mitochondria the key? J Alzheimers Dis 20 Suppl 2:S579-590.

Armstrong, RA. 2011. The pathogenesis of Alzheimer's disease: a reevaluation of the "amyloid cascade hypothesis". Int J Alzheimers Dis 2011:630865.

Arumugam, TV, T Magnus, TM Woodruff, LM Proctor, IA Shiels, SM Taylor. 2006. Complement mediators in ischemia-reperfusion injury. Clin Chim Acta 374:33-45.

Bales, KR. 2010. Brain lipid metabolism, apolipoprotein E and the pathophysiology of Alzheimer's disease. Neuropharmacology 59:295-302.

Bales, KR, JC Dodart, RB DeMattos, DM Holtzman, SM Paul. 2002. Apolipoprotein E, amyloid, and Alzheimer disease. Mol Interv 2:363-375, 339.

Ballard, C, S Gauthier, A Corbett, C Brayne, D Aarsland, E Jones. 2011. Alzheimer's disease. Lancet.

Bell, RD, AP Sagare, AE Friedman, GS Bedi, DM Holtzman, R Deane, BV Zlokovic. 2007. Transport pathways for clearance of human Alzheimer's amyloid beta-peptide and apolipoproteins E and J in the mouse central nervous system. J Cereb Blood Flow Metab 27:909-918.

Bertram, L, C Lange, K Mullin, et al. 2008. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. Am J Hum Genet 83:623-632.

Bettens, K, N Brouwers, H Van Miegroet, A Gil, S Engelborghs, PP De Deyn, R Vandenberghe, C Van Broeckhoven, K Sleegers. 2010. Follow-up study of susceptibility loci for Alzheimer's disease and onset age identified by genome-wide association. J Alzheimers Dis 19:1169-1175.

Biffi, A, CD Anderson, RS Desikan, M Sabuncu, L Cortellini, N Schmansky, D Salat, J Rosand. 2010. Genetic variation and neuroimaging measures in Alzheimer disease. Arch Neurol 67:677-685.

Blennow, K, MJ de Leon, H Zetterberg. 2006. Alzheimer's disease. Lancet 368:387-403.

Brinkman-Van der Linden, EC, T Angata, SA Reynolds, LD Powell, SM Hedrick, A Varki. 2003. CD33/Siglec-3 binding specificity, expression pattern, and consequences of gene deletion in mice. Mol Cell Biol 23:4199-4206.

Brookmeyer, R, E Johnson, K Ziegler-Graham, HM Arrighi. 2007. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 3:186-191.

Campion, D, JM Flaman, A Brice, et al. 1995. Mutations of the presenilin I gene in families with early-onset Alzheimer's disease. Hum Mol Genet 4:2373-2377.

Carrasquillo, MM, O Belbin, TA Hunter, et al. 2010. Replication of CLU, CR1, and PICALM associations with alzheimer disease. Arch Neurol 67:961-964.

Carrieri, G, M Bonafe, M De Luca, et al. 2001. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. Hum Genet 108:194-198.

Cataldo, JK, JJ Prochaska, SA Glantz. 2010. Cigarette smoking is a risk factor for Alzheimer's Disease: an analysis controlling for tobacco industry affiliation. J Alzheimers Dis 19:465-480.

Chan, DC. 2006. Mitochondria: dynamic organelles in disease, aging, and development. Cell 125:1241-1252.

Chinnery, PF, GA Taylor, N Howell, RM Andrews, CM Morris, RW Taylor, IG McKeith, RH Perry, JA Edwardson, DM Turnbull. 2000. Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. Neurology 55:302-304.

Citron, M, D Westaway, W Xia, et al. 1997. Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid beta-protein in both transfected cells and transgenic mice. Nat Med 3:67-72.

Corder, EH, AM Saunders, NJ Risch, et al. 1994. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 7:180-184.

Corder, EH, AM Saunders, WJ Strittmatter, DE Schmechel, PC Gaskell, GW Small, AD Roses, JL Haines, MA Pericak-Vance. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261:921-923.

Corneveaux, JJ, AJ Myers, AN Allen, et al. 2010. Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. Hum Mol Genet 19:3295-3301.

Cousin, MA, PJ Robinson. 2001. The dephosphins: dephosphorylation by calcineurin triggers synaptic vesicle endocytosis. Trends Neurosci 24:659-665.

Cruts, M, C Van Broeckhoven. 1998a. Molecular genetics of Alzheimer's disease. Ann Med 30:560-565.

Cruts, M, C Van Broeckhoven. 1998b. Presenilin mutations in Alzheimer's disease. Hum Mutat 11:183-190.

Cruts, M, CM van Duijn, H Backhovens, et al. 1998. Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease. Hum Mol Genet 7:43-51.

DeMattos, RB, JR Cirrito, M Parsadanian, et al. 2004. ApoE and clusterin cooperatively suppress Abeta levels and deposition: evidence that ApoE regulates extracellular Abeta metabolism in vivo. Neuron 41:193-202.

DeMattos, RB, A O'Dell M, M Parsadanian, JW Taylor, JA Harmony, KR Bales, SM Paul, BJ Aronow, DM Holtzman. 2002. Clusterin promotes amyloid plaque formation and is critical for neuritic toxicity in a mouse model of Alzheimer's disease. Proc Natl Acad Sci U S A 99:10843-10848.

Devi, L, BM Prabhu, DF Galati, NG Avadhani, HK Anandatheerthavarada. 2006. Accumulation of amyloid precursor protein in the mitochondrial import channels of human Alzheimer's disease brain is associated with mitochondrial dysfunction. J Neurosci 26:9057-9068.

Di Paolo, G, TW Kim. 2011. Linking lipids to Alzheimer's disease: cholesterol and beyond. Nat Rev Neurosci 12:284-296.

Dreyling, MH, JA Martinez-Climent, M Zheng, J Mao, JD Rowley, SK Bohlander. 1996. The t(10;11)(p13;q14) in the U937 cell line results in the fusion of the AF10 gene and CALM, encoding a new member of the AP-3 clathrin assembly protein family. Proc Natl Acad Sci U S A 93:4804-4809.

Dubois, B, HH Feldman, C Jacova, et al. 2007. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 6:734-746.

Eisenberg, DT, CW Kuzawa, MG Hayes. 2010. Worldwide allele frequencies of the human apolipoprotein E gene: climate, local adaptations, and evolutionary history. Am J Phys Anthropol 143:100-111.

Fesahat, F, M Houshmand, MS Panahi, K Gharagozli, F Mirzajani. 2007. Do haplogroups H and U act to increase the penetrance of Alzheimer's disease? Cell Mol Neurobiol 27:329-334.

Giedraitis, V, L Kilander, M Degerman-Gunnarsson, J Sundelof, T Axelsson, AC Syvanen, L Lannfelt, A Glaser. 2009. Genetic analysis of Alzheimer's disease in the Uppsala Longitudinal Study of Adult Men. Dement Geriatr Cogn Disord 27:59-68.

Goate, A, MC Chartier-Harlin, M Mullan, et al. 1991. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349:704-706.

Guerreiro, RJ, J Beck, JR Gibbs, et al. 2010. Genetic variability in CLU and its association with Alzheimer's disease. PLoS One 5:e9510.

Hardy, JA, GA Higgins. 1992. Alzheimer's disease: the amyloid cascade hypothesis. Science 256:184-185.

Harel, A, MP Mattson, PJ Yao. 2011. CALM, a clathrin assembly protein, influences cell surface GluR2 abundance. Neuromolecular Med 13:88-90.

Harel, A, F Wu, MP Mattson, CM Morris, PJ Yao. 2008. Evidence for CALM in directing VAMP2 trafficking. Traffic 9:417-429.

Harold, D, R Abraham, P Hollingworth, et al. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41:1088-1093.

Hayashi, M, S Abe-Dohmae, M Okazaki, K Ueda, S Yokoyama. 2005. Heterogeneity of high density lipoprotein generated by ABCA1 and ABCA7. J Lipid Res 46:1703-1711.

Heckmann, JM, WC Low, C de Villiers, S Rutherfoord, A Vorster, H Rao, CM Morris, RS Ramesar, RN Kalaria. 2004. Novel presenilin 1 mutation with profound neurofibrillary pathology in an indigenous Southern African family with early-onset Alzheimer's disease. Brain 127:133-142.

Hendrie, HC, KS Hall, S Hui, et al. 1995. Apolipoprotein E genotypes and Alzheimer's disease in a community study of elderly African Americans. Ann Neurol 37:118-120.

Herrup, K. 2010. Reimagining Alzheimer's disease--an age-based hypothesis. J Neurosci 30:16755-16762.

Hollingworth, PD HaroldR Sims, et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 43:429-435.

Hu, X, E Pickering, YC Liu, S Hall, H Fournier, E Katz, B Dechairo, S John, P Van Eerdewegh, H Soares. 2011. Meta-analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. PLoS One 6:e16616.

Hudson, G, R Sims, D Harold, et al. 2012. No consistent evidence for association between mtDNA variants and Alzheimer disease. Neurology.

Hutton, M, F Busfield, M Wragg, et al. 1996. Complete analysis of the presenilin 1 gene in early onset Alzheimer's disease. Neuroreport 7:801-805.

Ishibashi, K, M Suzuki, S Sasaki, M Imai. 2001. Identification of a new multigene four-transmembrane family (MS4A) related to CD20, HTm4 and beta subunit of the high-affinity IgE receptor. Gene 264:87-93.

Janssen, JC, JA Beck, TA Campbell, A Dickinson, NC Fox, RJ Harvey, H Houlden, MN Rossor, J Collinge. 2003. Early onset familial Alzheimer's disease: Mutation frequency in 31 families. Neurology 60:235-239.

Jiang, Q, CY Lee, S Mandrekar, et al. 2008. ApoE promotes the proteolytic degradation of Abeta. Neuron 58:681-693.

Jun, G, AC Naj, GW Beecham, et al. 2010. Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes. Arch Neurol 67:1473-1484.

Kamboh, MI, RL Minster, FY Demirci, M Ganguli, ST Dekosky, OL Lopez, MM Barmada. 2010. Association of CLU and PICALM variants with Alzheimer's disease. Neurobiol Aging.

Kaminski, WE, E Orso, W Diederich, J Klucken, W Drobnik, G Schmitz. 2000. Identification of a novel human sterol-sensitive ATP-binding cassette transporter (ABCA7). Biochem Biophys Res Commun 273:532-538.

Kobayashi, S, A Sawano, Y Nojima, M Shibuya, Y Maru. 2004. The c-Cbl/CD2AP complex regulates VEGF-induced endocytosis and degradation of Flt-1 (VEGFR-1). FASEB J 18:929-931.

Koistinaho, M, S Lin, X Wu, et al. 2004. Apolipoprotein E promotes astrocyte colocalization and degradation of deposited amyloid-beta peptides. Nat Med 10:719-726.

Kovacs, DM, HJ Fausett, KJ Page, et al. 1996. Alzheimer-associated presenilins 1 and 2: neuronal expression in brain and localization to intracellular membranes in mammalian cells. Nat Med 2:224-229.

Kruger, J, R Hinttala, K Majamaa, AM Remes. 2010. Mitochondrial DNA haplogroups in early-onset Alzheimer's disease and frontotemporal lobar degeneration. Mol Neurodegener 5:8.

LaDu, MJ, JA Shah, CA Reardon, GS Getz, G Bu, J Hu, L Guo, LJ van Eldik. 2000. Apolipoprotein E receptors mediate the effects of beta-amyloid on astrocyte cultures. J Biol Chem 275:33974-33980.

Lakatos, A, O Derbeneva, D Younes, et al. 2010. Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort. Neurobiol Aging 31:1355-1363.

Lambert, JC, S Heath, G Even, et al. 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094-1099.

Lee, JH, R Cheng, S Barral, C Reitz, M Medrano, R Lantigua, IZ Jimenez-Velazquez, E Rogaeva, PH St George-Hyslop, R Mayeux. 2011. Identification of novel loci for Alzheimer disease and replication of CLU, PICALM, and BIN1 in Caribbean Hispanic individuals. Arch Neurol 68:320-328.

Levy-Lahad, E, W Wasco, P Poorkaj, et al. 1995. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 269:973-977.

Liang, Y, TR Buckley, L Tu, SD Langdon, TF Tedder. 2001. Structural organization of the human MS4A gene cluster on Chromosome 11q12. Immunogenetics 53:357-368.

Lindsay, J, D Laurin, R Verreault, R Hebert, B Helliwell, GB Hill, I McDowell. 2002. Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. Am J Epidemiol 156:445-453.

Ma, Y, H Yang, J Qi, et al. 2010. CD2AP is indispensable to multistep cytotoxic process by NK cells. Mol Immunol 47:1074-1082.

Maestre, G, R Ottman, Y Stern, B Gurland, M Chun, MX Tang, M Shelanski, B Tycko, R Mayeux. 1995. Apolipoprotein E and Alzheimer's disease: ethnic variation in genotypic risks. Ann Neurol 37:254-259.

Mahley, RW. 1988. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. Science 240:622-630.

Mancuso, M, M Nardini, D Micheli, et al. 2007. Lack of association between mtDNA haplogroups and Alzheimer's disease in Tuscany. Neurol Sci 28:142-147.

Mancuso, M, D Orsucci, G Siciliano, L Murri. 2008. Mitochondria, mitochondrial DNA and Alzheimer's disease. What comes first? Curr Alzheimer Res 5:457-468.

Maruszak, A, JA Canter, M Styczynska, C Zekanowski, M Barcikowska. 2009. Mitochondrial haplogroup H and Alzheimer's disease--is there a connection? Neurobiol Aging 30:1749-1755.

Matsubara, E, C Soto, S Governale, B Frangione, J Ghiso. 1996. Apolipoprotein J and Alzheimer's amyloid beta solubility. Biochem J 316 (Pt 2):671-679.

Matsuzaki, T, K Sasaki, J Hata, Y Hirakawa, K Fujimi, T Ninomiya, SO Suzuki, S Kanba, Y Kiyohara, T Iwaki. 2011. Association of Alzheimer disease pathology with abnormal lipid metabolism: the Hisayama Study. Neurology 77:1068-1075.

Moehlmann, T, E Winkler, X Xia, et al. 2002. Presenilin-1 mutations of leucine 166 equally affect the generation of the Notch and APP intracellular domains independent of their effect on Abeta 42 production. Proc Natl Acad Sci U S A 99:8025-8030.

Molsa, PK, RJ Marttila, UK Rinne. 1986. Survival and cause of death in Alzheimer's disease and multi-infarct dementia. Acta Neurol Scand 74:103-107.

Molsa, PK, RJ Marttila, UK Rinne. 1995. Long-term survival and predictors of mortality in Alzheimer's disease and multi-infarct dementia. Acta Neurol Scand 91:159-164.

Naj, ACG JunGW Beecham, et al. 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 43:436-441.

Newell, KL, BT Hyman, JH Growdon, ET Hedley-Whyte. 1999. Application of the National Institute on Aging (NIA)-Reagan Institute criteria for the neuropathological diagnosis of Alzheimer disease. J Neuropathol Exp Neurol 58:1147-1155.

Noguchi, S, K Murakami, N Yamada. 1993. Apolipoprotein E genotype and Alzheimer's disease. Lancet 342:737.

Onyango, I, S Khan, B Miller, R Swerdlow, P Trimmer, P Bennett, Jr. 2006. Mitochondrial genomic contribution to mitochondrial dysfunction in Alzheimer's disease. J Alzheimers Dis 9:183-193.

Patterson, C, J Feightner, A Garcia, C MacKnight. 2007. General risk factors for dementia: a systematic evidence review. Alzheimers Dement 3:341-347.

Patterson, C, JW Feightner, A Garcia, GY Hsiung, C MacKnight, AD Sadovnick. 2008. Diagnosis and treatment of dementia: 1. Risk assessment and primary prevention of Alzheimer disease. CMAJ 178:548-556.

Podewils, LJ, E Guallar, LH Kuller, LP Fried, OL Lopez, M Carlson, CG Lyketsos. 2005. Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. Am J Epidemiol 161:639-651.

Priller, C, T Bauer, G Mitteregger, B Krebs, HA Kretzschmar, J Herms. 2006. Synapse formation and function is modulated by the amyloid precursor protein. J Neurosci 26:7212-7221.

Pyle, A, T Foltynie, W Tiangyou, et al. 2005. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann Neurol 57:564-567.

Querfurth, HW, FM LaFerla. 2010. Alzheimer's disease. N Engl J Med 362:329-344.

Raux, G, L Guyant-Marechal, C Martin, J Bou, C Penet, A Brice, D Hannequin, T Frebourg, D Campion. 2005. Molecular diagnosis of autosomal dominant early onset Alzheimer's disease: an update. J Med Genet 42:793-795.

Reiman, EM, K Chen, X Liu, et al. 2009. Fibrillar amyloid-beta burden in cognitively normal people at 3 levels of genetic risk for Alzheimer's disease. Proc Natl Acad Sci U S A 106:6820-6825.

Reiman, EM, JA Webster, AJ Myers, et al. 2007. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. Neuron 54:713-720.

Ridge, P, T Maxwell, C Corcoran, M Norton, J Tschanz, E O'Brien, R Kerber, R Cawthon, R Munger, K JSK. 2012. Mitochondrial genomic analysis of late onset Alzheimer's disease reveals protective haplogroups H6A1A/H6A1B: the Cache County Study on Memory in Aging. PLoS One 7:e45134.

Rogers, J, R Li, D Mastroeni, et al. 2006. Peripheral clearance of amyloid beta peptide by complement C3-dependent adherence to erythrocytes. Neurobiol Aging 27:1733-1739.

Rudzinski, LA, RM Fletcher, DW Dickson, R Crook, ML Hutton, J Adamson, NR Graff-Radford. 2008. Early onset familial Alzheimer Disease with spastic paraparesis, dysarthria, and seizures and N135S mutation in PSEN1. Alzheimer Dis Assoc Disord 22:299-307.

Santoro, A, V Balbi, E Balducci, et al. 2010. Evidence for sub-haplogroup h5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease. PLoS One 5:e12037.

Scarmeas, N, Y Stern, R Mayeux, JA Luchsinger. 2006. Mediterranean diet, Alzheimer disease, and vascular mediation. Arch Neurol 63:1709-1717.

Schellenberg, GD, TD Bird, EM Wijsman, et al. 1992. Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science 258:668-671.

Schrijvers, EM, PJ Koudstaal, A Hofman, MM Breteler. 2011. Plasma clusterin and the risk of Alzheimer disease. JAMA 305:1322-1326.

Seshadri, S, AL Fitzpatrick, MA Ikram, et al. 2010. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA 303:1832-1840.

Sherrington, R, S Froelich, S Sorbi, et al. 1996. Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant. Hum Mol Genet 5:985-988.

Sherrington, R, EI Rogaev, Y Liang, et al. 1995. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375:754-760.

Simpson, F, NK Hussain, B Qualmann, RB Kelly, BK Kay, PS McPherson, SL Schmid. 1999. SH3-domain-containing proteins function at distinct steps in clathrin-coated vesicle formation. Nat Cell Biol 1:119-124.

Steiner, H, E Winkler, D Edbauer, S Prokop, G Basset, A Yamasaki, M Kostka, C Haass. 2002. PEN-2 is an integral component of the gamma-secretase complex required for coordinated expression of presenilin and nicastrin. J Biol Chem 277:39062-39065.

Swerdlow, RH, JM Burns, SM Khan. 2010. The Alzheimer's disease mitochondrial cascade hypothesis. J Alzheimers Dis 20 Suppl 2:S265-279.

Swerdlow, RH, SM Khan. 2004. A "mitochondrial cascade hypothesis" for sporadic Alzheimer's disease. Med Hypotheses 63:8-20.

Swerdlow, RH, SM Khan. 2009. The Alzheimer's disease mitochondrial cascade hypothesis: an update. Exp Neurol 218:308-315.

Takasaki, S. 2009. Mitochondrial haplogroups associated with Japanese Alzheimer's patients. J Bioenerg Biomembr 41:407-410.

Tanaka, N, S Abe-Dohmae, N Iwamoto, ML Fitzgerald, S Yokoyama. 2010. Helical apolipoproteins of high-density lipoprotein enhance phagocytosis by stabilizing ATP-binding cassette transporter A7. J Lipid Res 51:2591-2599.

Tanaka, N, S Abe-Dohmae, N Iwamoto, S Yokoyama. 2011. Roles of ATP-binding cassette transporter A7 in cholesterol homeostasis and host defense system. J Atheroscler Thromb 18:274-281.

Thambisetty, M, Y An, A Kinsey, et al. 2012. Plasma clusterin concentration is associated with longitudinal brain atrophy in mild cognitive impairment. Neuroimage 59:212-217.

Thambisetty, M, A Simmons, L Velayudhan, et al. 2010. Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. Arch Gen Psychiatry 67:739-748.

Thinakaran, G, EH Koo. 2008. Amyloid precursor protein trafficking, processing, and function. J Biol Chem 283:29615-29619.

Trougakos, IP, ES Gonos. 2006. Regulation of clusterin/apolipoprotein J, a functional homologue to the small heat shock proteins, by oxidative stress in ageing and age-related diseases. Free Radic Res 40:1324-1334.

Tycko, B, L Feng, L Nguyen, et al. 1996. Polymorphisms in the human apolipoprotein-J/clusterin gene: ethnic variation and distribution in Alzheimer's disease. Hum Genet 98:430-436.

Ueki, A, M Kawano, Y Namba, M Kawakami, K Ikeda. 1993. A high frequency of apolipoprotein E4 isoprotein in Japanese patients with late-onset nonfamilial Alzheimer's disease. Neurosci Lett 163:166-168.

van der Walt, JM, YA Dementieva, ER Martin, et al. 2004. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. Neurosci Lett 365:28-32.

Wang, HX, A Karp, B Winblad, L Fratiglioni. 2002. Late-life engagement in social and leisure activities is associated with a decreased risk of dementia: a longitudinal study from the Kungsholmen project. Am J Epidemiol 155:1081-1087.

Wijsman, EM, ND Pankratz, Y Choi, et al. 2011. Genome-wide association of familial late-onset Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. PLoS Genet 7:e1001308.

Wu, F, PJ Yao. 2009. Clathrin-mediated endocytosis and Alzheimer's disease: an update. Ageing Res Rev 8:147-149.

Yaddanapudi, S, MM Altintas, AD Kistler, et al. 2011. CD2AP in mouse and human podocytes controls a proteolytic program that regulates cytoskeletal structure and cellular survival. J Clin Invest 121:3965-3980.

Young-Pearse, TL, J Bai, R Chang, JB Zheng, JJ LoTurco, DJ Selkoe. 2007. A critical function for beta-amyloid precursor protein in neuronal migration revealed by in utero RNA interference. J Neurosci 27:14459-14469.

Yu, JT, L Li, QX Zhu, Q Zhang, W Zhang, ZC Wu, J Guan, L Tan. 2010. Implication of CLU gene polymorphisms in Chinese patients with Alzheimer's disease. Clin Chim Acta 411:1516-1519.

Zhang, Q, JT Yu, QX Zhu, W Zhang, ZC Wu, D Miao, L Tan. 2010. Complement receptor 1 polymorphisms and risk of late-onset Alzheimer's disease. Brain Res 1348:216-221.

Zlokovic, BV, CL Martel, E Matsubara, JG McComb, G Zheng, RT McCluskey, B Frangione, J Ghiso. 1996. Glycoprotein 330/megalin: probable role in receptor-mediated transport of apolipoprotein J alone and in a complex with Alzheimer disease amyloid beta at the blood-brain and blood-cerebrospinal fluid barriers. Proc Natl Acad Sci U S A 93:4229-4234.

Zsurka, G, J Kalman, A Csaszar, I Rasko, Z Janka, P Venetianer. 1998. No mitochondrial haplotype was found to increase risk for Alzheimer's disease. Biol Psychiatry 44:371-373.

Table 1, Chapter 1. Late onset Alzheimer's disease associated genes/variants

Variant	Gene	Abbreviation	Risk/Protective
rs7412	Apolipoprotein E	APOE	Risk
rs429358	Apolipoprotein E	APOE	Protective
rs744373	Bridging Integrator 1	BIN1	Risk
rs11136000	Clusterin	CLU	Protective
rs3764650	ATP-binding cassette, sub-family A (ABC1), member 7	ABCA7	Risk
rs3818361	Complement component (3b/4b) receptor 1 (Knops blood group)	CR1	Risk
rs3851179	Phosphatidylinositol binding clathrin assembly protein	PICALM	Protective
rs610932	Membrane-spanning 4-domains, subfamily A, member 6A	MS4A6A	Protective
rs3865444	CD33 molecule	CD33	Protective
rs670139	Membrane-spanning 4-domains, subfamily A, member 4E	MS4A4E	Risk
rs9349407	CD2-associated protein	CD2AP	Risk

Each of the top variants associated with late onset Alzheimer's disease from meta-analysis done by the Alzheimer Research Forum are listed here, together with the specific associated variant, and whether the variant increases risk or provides protection.

Table 2, Chapter 1. Mitochondrial variation/haplogroups associated with AD

Haplogroup	Dataset	Effect	Ethnicity	# cases / controls
B4C1 (Takasaki 2009)	Selected SNPs	Risk	Japanese	96 / 384
G2A (Takasaki 2009)	Selected SNPs	Risk	Japanese	96 / 384
HV (Maruszak et al. 2009)	Haplogroups, SNPs	Risk	Polish	222 / 252
H (Fesahat et al. 2007)	HVS-I sequence	Risk	Iranian	30 / 100
H5 / H5A (Santoro et al. 2010)	D-loop sequence, restriction analysis	Risk	Italian	936 / 776
H6A1A / H6A1B (Ridge et al. 2012)	Full mtDNA sequences	Protective	Caucasian	101 / 632
K (Carrieri et al. 2001)	Haplogroups	Protective	Italian	N/A*
N9B1 (Takasaki 2009)	Selected SNPs	Risk	Japanese	96 / 384
U (van der Walt et al. 2004; Fesahat et al. 2007)	HVS-I sequence, 10 SNPs	Risk	Iranian, Caucasian	30 / 100, 989 / 328**
U (Carrieri et al. 2001; van der Walt et al. 2004)	Haplogroups, 10 SNPs	Protective	Italian, Caucasian	N/A*, 989 / 328**
UK (Lakatos et al. 2010)	138 SNPs	Risk	Caucasian	170 / 188
None (Zsurka et al. 1998)	4 SNPs	None	Unknown	70 / 80
None (Chinnery et al. 2000)	European Haplogroups	None	Unknown	185 / 179
None (Pyle et al. 2005)	U, K, J, and T haplogroups	None	English	185 / 447
None (Mancuso et al. 2007)	European Haplogroups	None	Tuscan	209 / 191
None (Kruger et al. 2010)	Haplogroups	None	Finnish	128 / 99***
None (Hudson et al. 2012)	138 SNPs	None	Caucasian	3250 / 1221

* The authors showed that haplogroups U and K neutralized the risk of the APOE e4 allele

** The authors demonstrated an increased risk for AD for males with haplogroup U, and decreased risk for females with haplogroup U

*** These were early onset AD cases

Manuscript published in PLoS One. 2012; 7(9):e45134

Chapter 2

Mitochondrial genomic analysis of late onset Alzheimer's disease reveals protective haplogroups H6A1A/H6A1B: The Cache County Study on Memory in Aging

Perry G. Ridge[1,2], Taylor J. Maxwell[3], Christopher D. Corcoran[4, 5], Maria C. Norton[5,6,7], JoAnn T. Tschanz[5,7], Elizabeth O'Brien[8], Richard A. Kerber[8], Richard M. Cawthon[9], Ronald G. Munger[5,10], John S.K. Kauwe[1]

[1] Department of Biology, Brigham Young University, Provo, Utah

[2] ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah

[3] Human Genetics Center, University of Texas School of Public Health, Houston, Texas

[4] Department of Mathematics and Statistics, Utah State University, Logan, Utah

[5] Center for Epidemiologic Studies, Utah State University, Logan, Utah

[6] Department of Family Consumer and Human Development, Utah State University, Logan, Utah

[7] Department of Psychology, Utah State University, Logan, Utah

[8] Department of Epidemiology and Population Health, University of Louisville, Louisville, Kentucky

[9] Department of Human Genetics, University of Utah, Salt Lake City, Utah

[10] Department of Nutrition, Dietetics, and Food Sciences, Utah State University, Logan, Utah

Abstract

Background: Alzheimer's disease (AD) is the most common cause of dementia and AD risk clusters within families. Part of the familial aggregation of AD is accounted for by excess maternal vs. paternal inheritance, a pattern consistent with mitochondrial inheritance. The role of specific mitochondrial DNA (mtDNA) variants and haplogroups in AD risk is uncertain.

Methodology/Principal Findings: We determined the complete mitochondrial genome sequence of 1007 participants in the Cache County Study on Memory in Aging, a population-based prospective cohort study of dementia in northern Utah. AD diagnoses were made with a multi-stage protocol that included clinical examination and review by a panel of clinical experts. We used TreeScanning, a statistically robust approach based on haplotype networks, to analyze the mtDNA sequence data. Participants with major mitochondrial haplotypes H6A1A and H6A1B showed a reduced risk of AD ($p = 0.017$, corrected for multiple comparisons). The protective haplotypes were defined by three variants: m.3915G>A, m.4727A>G, and m.9380G>A. These three variants characterize two different major haplogroups. Together m.4727A>G and m.9380G>A define H6A1, and it has been suggested m.3915G>A defines H6A. Additional variants differentiate H6A1A and H6A1B; however, none of these variants had a significant relationship with AD case-control status.

Conclusions: Our findings provide evidence of a reduced risk of AD for individuals with mtDNA haplotypes H6A1A and H6A1B. These findings are the results of the largest study to date with complete mtDNA genome sequence data, yet the functional significance of the associated haplotypes remains unknown and replication in others studies is necessary.

Introduction

Alzheimer's disease (AD) is a complex disorder and is the most common form of dementia (Querfurth, LaFerla 2010). After age, family history is the single greatest risk factor for AD. AD can be classified into early and late onset forms. Mutations in three genes: PSEN1/2 and APP are known to cause early onset AD in an autosomal dominant manner (Cruts, Van Broeckhoven 1998; Janssen et al. 2003). The majority of AD cases, however, are late onset (LOAD) and the APOE ε4 allele is the strongest known genetic risk factor. Many additional genetic polymorphisms have been identified, though with substantially lower risk estimates (Lambert et al. 2009; Carrasquillo et al. 2010; Corneveaux et al. 2010; Jun et al. 2010; Kamboh et al. 2010; Querfurth, LaFerla 2010; Hollingworth et al. 2011; Naj et al. 2011). LOAD appears to be inherited and/or sporadic and there is evidence of a maternal inheritance pattern (Cupples et al. 2004). Current estimates suggest that more than 20% of inherited LOAD cases are maternally inherited (Mosconi et al. 2010).

Analyses of families with inherited LOAD have repeatedly reported a greater incidence of AD in children with affected mothers than with affected fathers. Among individuals affected with AD who have one affected parent, the mother is 1.8 to 3.8 times more likely to be affected than the father (Heyman et al. 1983; Farrer et al. 1991; Duara et al. 1993; Edland et al. 1996). When extended to families with multiple affected siblings and a single affected parent, the ratio of affected mothers to fathers increased to 9:1 (Edland et al. 1996). While no biological mechanisms for maternal inheritance were demonstrated, results of these studies, observed in multiple datasets, strongly suggest maternal inheritance.

Imaging studies provide additional evidence of maternal inheritance of AD. These studies have identified decreased glucose metabolism and atrophy in brain regions affected in AD. Similar to the studies cited above, individuals with a paternal, maternal, or no family history of AD were compared. First, progressive gray matter atrophy was only observed in people with a maternal family history of AD. These same individuals had greater atrophy in the precuneus and parahippocampal gyrus (regions known to be affected in AD) than those with a paternal or no family history of AD (Honea et al. 2011). Next, other studies compared reductions in glucose metabolism in the brain for each of the three groups listed above. Subjects with a paternal family history of LOAD had decreased glucose metabolism similar to those with no family history; however, individuals with a maternal family history of LOAD had significantly decreased glucose metabolism compared to the other groups (Mosconi et al. 2007; Mosconi et al. 2009). Additionally, similar to the atrophy studies, lowered glucose metabolism was concentrated in the same brain regions known to have impaired glucose metabolism in AD (posterior cingulate cortex/precuneus, parieto-temporal, and medial temporal cortices) (Mosconi et al. 2007; Mosconi et al. 2009). The increased incidence of AD or risk for AD-related phenotypes among individuals with a maternal family history of AD, compared to people with no family history or a paternal family history of AD, strongly support a maternal inheritance pattern for LOAD.

Maternal inheritance occurs by several mechanisms including disease susceptibility genes on the X-chromosome, maternal specific genetic imprinting, or by mitochondrial genetic effects. We investigated the role of mitochondrial sequence variants in maternal transmission of LOAD. Mitochondrial malfunction is a plausible explanation for a number of AD phenotypes, including the decreased glucose metabolism in specific brain regions discussed above. Numerous

mitochondrial modifications in patients with AD have been reported; these include morphological changes (Chan 2006; Swerdlow, Burns, Khan 2010), alterations in the enzymes of the electron transport chain, including cytochrome c oxidase (Du et al. 2010; Santos et al. 2010), changes in the mitochondrial proteome (Chou et al. 2011), and reduced numbers of mitochondria (Santos et al. 2010). Beta-amyloid plaques aggregate within mitochondria (Devi et al. 2006; Anandatheerthavarada, Devi 2007) and it has been hypothesized that changes in mitochondrial function facilitate Aβ deposition and tau phosphorylation (Silva et al. 2011). These observations have led investigators to ask whether mitochondrial dysfunction is a cause or effect of plaque aggregation.

The mitochondrial cascade hypothesis (Swerdlow, Khan 2004) posits that a decline in mitochondrial number and function is a cause of neurodegeneration. Briefly, it is known that mitochondrial function declines with age and in conjunction with certain morphological changes (Chan 2006; Swerdlow, Burns, Khan 2010). As mitochondrial function declines with age, hypothesized consequences are increased tau phosphorylation and beta-amyloid amyloidosis in brain tissue. In contrast, in familial forms of AD, Aβ aggregation and tau phosphorylation are hypothesized to occur before mitochondrial malfunction and lead to the mitochondrial dysmorphology and dysfunction characteristic of AD (Swerdlow, Khan 2004; Swerdlow, Khan 2009; Swerdlow, Burns, Khan 2010). Other evidence suggests that Aβ aggregation directly causes mitochondrial malfunction (Cho et al. 2009; Pagani, Eckert 2011) or that Aβ and tau interact to increase oxidative stress and impede mitochondrial function (Quintanilla et al. 2012).

Mitochondrial malfunction can be caused by numerous factors, one of which is inherited sequence variation in the mitochondrial genome. To date, many studies have been published analyzing the association between mitochondrial haplogroups or specific mitochondrial sequence variants, and AD. The results have been mixed, confusing, and at times contradictory. The majority of studies have not identified any associations (Zsurka et al. 1998; Chinnery et al. 2000; Pyle et al. 2005; Mancuso et al. 2007; Kruger et al. 2010; Hudson et al. 2012), but some have reported significant associations. Haplogroups H and U (or sub-haplogroups of H and U) have been associated with both increased and decreased risk of AD (Carrieri et al. 2001; van der Walt et al. 2004; Fesahat et al. 2007; Santoro et al. 2010; Coskun et al. 2011) and different effects for men and women. The UK and HV clusters, as well as haplogroups J, G2A, B4C1, and N9B1, have been associated with increased risk for AD (Chagnon et al. 1999; Maruszak et al. 2009; Takasaki 2009a; Takasaki 2009b; Lakatos et al. 2010), while haplogroups K and T are thought to be protective (Chagnon et al. 1999; Carrieri et al. 2001). No consensus has been reached and no previous studies have reported on large population-based samples with complete mtDNA genome sequence data. Here we present the largest analysis to date of mitochondrial haplotypes and associated risks for LOAD based on fully sequenced mitochondrial genomes.

Materials and Methods

Ethics Statement

All study procedures were approved by the Institutional Review Boards of Brigham Young University, Utah State University, Duke University, and Johns Hopkins University. Written consent was obtained for each individual. To verify a subject's capacity to consent, subjects attempted the Modified Mini-Mental State Exam (3MS). If there was an indication of poor

cognitive ability as determined by poor performance on the entire test (scoring below a designated total of 60 points), poor performance on temporal or spatial orientation, or clear difficulty in understanding the nature of the interview, the visit was discontinued and informed consent was obtained from a responsible caregiver—often the next-of-kin. We re-consented subjects/caregivers at each study visit and procedure.

Sample Acquisition and Sequencing

The Cache County Study on Memory in Aging was initiated in 1994 to investigate the occurrence of dementia and associations with APOE genotype, environmental exposures, and cognitive function. A cohort comprised of 5,092 Cache County, Utah, residents was established and followed continually for 12 years. The cohort represents a 90% sample of all residents of Cache County who were aged 65 and older in 1994. Over the 12-year follow-up period, data were collected in four (triennial) waves. Data collected in each wave included basic demographic information, family and medical histories, and a multistage dementia assessment screen; in addition, a more detailed clinical assessment was done for: a) those who screened positive for AD according to the multistage screening protocol for dementia; or b) were randomly selected, according to age, gender, and APOE genotype, to complete all stages of assessment (Breitner et al. 1999). Diagnoses of dementia were based on expert clinical assessments, standard MRI, and laboratory studies. Diagnoses of Probable or Possible AD were based on NINCDS-ADRDA criteria (McKhann et al. 1984). Those without dementia were diagnosed as "non cases" following a clinical assessment (if the individual was a designated subsample member), or if lacking a clinical assessment, screened negative at each stage of dementia screening and evaluation. Any participant who screened positive at any screening stage, but failed to complete

the next stage of screening was removed from the analyses. For individuals without clear non-demented or Alzheimer's disease diagnosis case-control status was set to missing or unknown.

The Utah population is the source of most of the Centre d'Etude du Polymorphisme Humain (CEPH) families. The CEPH families have been used in countless genetic studies to represent Caucasians worldwide, or for example, to represent the "European" sample assayed by the HapMap project (2003). The genetic structure of the Utah population is broadly representative of other U.S. populations of northern European ancestry characterized by very little inbreeding. Utah's founding pioneer group was relatively large, migrated from many different points of European origin, and overall, was largely unrelated (Jorde 1982; Jorde, Morgan 1987; O'Brien et al. 1994).

Cache County study participants were linked to the Utah Population Data Base and each participant was assigned to a unique matrilineage with a common maternal founder. The matrilineages were rank ordered by size (total number of known members) and a single Cache County Study participant was randomly selected from each of the Cache County matrilineages. The number of matrilineages sequenced was limited by the available funding; we started with the largest matrilineage and worked down the list. The participants selected for mtDNA sequencing were selected independent of their cognitive or dementia status. 274 matrilineages were represented by this dataset. As a result, the sequenced mitochondrial genomes also represent as many different major mitochondrial haplogroups and clusters as possible (Table 1). Selection was made blind to case-control status. 287 samples were sent to Family Tree DNA (www.familytreedna.com) for Sanger sequencing of the mitochondrial genomes. Family Tree

DNA applied QC criteria, and 285 of 287 passed QC. Family Tree DNA also reported which of the known mitochondrial haplogroups or clusters each of these individuals belonged to. We then identified each of the unique haplotypes, and their frequencies, in our dataset. In many instances several of the unique haplotypes in our dataset all belonged to the same major mitochondrial haplogroup/cluster (or sub-haplogroup). In this manuscript major mitochondrial haplogroups/clusters and sub-haplogroups refer to known mitochondrial haplotypes/clusters (i.e. H, H6A1A, HV, J, etc.) and haplotypes refer to the distinct mitochondrial haplotypes identified in our dataset. Among the 285 individuals whose mitochondrial genomes were sequenced were eight cases, 197 controls, and 80 individuals with missing AD case-control status.

Analyses

Following sequencing we were provided with a list of mtDNA sequence variants for each of the patient samples. Whole mtDNA sequences were generated for each person using Java programs to map variant sites back to a reference mtDNA sequence (NC_012920), and MitoMap annotations were used for describing gene coordinates (MITOMAP 2011). A number of sites were heteroplasmic, and for these we considered only the most common of the two possible alleles. Our dataset contained a total of 285 full mitochondrial sequences of high quality. Using extended pedigree information from the Utah Population Database (Skolnick 1980), we identified individuals sharing maternal lineage membership with those who were directly genotyped. Based on these pedigree (matrilineal) relationships, we were able to impute an additional 722 genome sequences for a total of 1007 full mitochondrial genome sequences.

ClustalW (Thompson, Gibson, Higgins 2002) was used to align the mitochondrial genomes. Using the 285 genotyped individuals, we inferred a haplotype network using TCS (Clement, Posada, Crandall 2000). Genotype-phenotype associations were evaluated using an evolution-based method known as TreeScanning (Posada, Maxwell, Templeton 2005; Templeton et al. 2005) that makes use of haplotype networks. Such networks provide an *a priori* framework from which to pool haplotypes based on common descent with the assumption that a mutation causing a phenotypic effect is embedded within the same historical structure represented by the haplotype network (Templeton, Boerwinkle, Sing 1987). Practically, TreeScanning uses each branch of the haplotype network to define a set of bi-allelic partitions. For each branch, haplotypes are pooled together into one of two allelic classes depending on which side of the branch they occur. Individual genotypes are then determined by which "alleles" they have. Being haploid, there are only two genotypes possible (i.e. no heterozygote class). This is done for each branch in the tree, resulting in a set of correlated tests. Significance levels were corrected for using a permutation analog of the step-down, stepwise Bonferroni method (Westfall, Young 1993), which takes into account the correlations among tests, and allows for more than one significant test. A second round of TreeScanning conditional on a significant branch from the first round is used to determine if other significant branches represent the same or different effects, or possibly additional associations masked by effects in the previous round of TreeScanning. The original TreeScanning method (Templeton et al. 2005) dealt with univariate continuous data.

For our case-control data we used logistic regression (Nowotny et al. 2005; Grupe et al. 2006; Li et al. 2006) with age, gender, familial risk, and APOE genotype as covariates. Familial risk

scores were computed for each study subject as the weighted sum of biological relatives who were diagnosed with AD divided by the weighted sum of biological relatives who were members of the at-risk cohort, using the coefficient of relationship (twice the kinship coefficient) as a weighting function (Kerber 1995). Each analysis was performed with 10,000 permutations. Only tests with at least two relevant genotypic classes, each containing five or more individuals, were tested. This was done to exclude tests with little power to detect associations and to increase statistical power overall by reducing the total number of tests. Significance was inferred if the multiple-test-corrected p-value was less than 0.05.

Results

Haplotype Network and mtDNA Variation

We observed 249 different haplotypes in our sample of 1007 full mitochondrial genomes (mtDNA). The majority of haplotypes (152 of 249) were observed in three or fewer individuals and the two most frequently observed haplotypes consisted of 39 and 32 individuals, respectively. The most frequently observed haplotype (39 individuals) was the root of our haplotype network. Our network (Supplementary Figure 1) contained one unresolved loop. This loop was left in the network and the ambiguity factored into all subsequent analyses.

We identified 899 single nucleotide variants (SNVs), 26 insertions, and 20 deletions in our dataset. The most frequently observed SNVs occurred in 281 genomes (m.263A>G, m.8860A>G, and m.15326A>G), and three more SNVs were observed in 280 genomes (m.750A>G, m.1438A>G, and m.4769A>G). Compared to the reference sequence (NC_012920), the average

number of variants per individuals was 25.3; one individual had the highest number of variants (52), and one individual the fewest (2).

Lastly, we analyzed the distribution of major mitochondrial haplogroups within our dataset (Table 1, Supplementary Table 1). Individuals in our dataset corresponded by haplotype to 102 major mitochondrial haplogroups/clusters (or sub-haplogroups), and as expected, the majority (987 of 1007) belonged to European based major haplogroups.

mtDNA Variation Associated with Alzheimer's disease Risk

In a haplotype network each segment of a branch (and a branch can have multiple segments) corresponds to a specific sequence feature, and each individual appearing below a segment of the network shares the particular feature that corresponds to the branch segment. In our study, two different branches were significantly associated with AD (see Tables 2 and 3) in the first round of TreeScanning. As shown in Figure 1, the clade defined by branch 269 is nested within the clade defined by branch 270; as a result, these two highly correlated branches represent a single significant association. The clade defined by branch 270 consists of 38 individuals and 11 distinct haplotypes. These individuals were 73.85 (standard deviation 7.55) years old on average, and 13 of 38 were male. Among the 38 individuals of branch 270 there were seven groups of siblings, one group of three siblings, and six sets of sibling pairs; 23 individuals had no siblings in the dataset. We classified each of the individuals into major mitochondrial haplogroups. Of the 38 branch 270 individuals, one belonged to major haplogroup H, 12 belonged to H6A1A, and 25 belonged to H6A1B. None of the 38 individuals of branch 270 tested positive for AD, an absence

of cases that makes this group of individuals, branch 270, significantly different in comparison to all other branches of the haplotype network (corrected p-value 0.016).

Since the clades defined by the two branches (269 and 270) are statistically indistinguishable, so too are the three variants (Figure 1) corresponding to these two branches. The sequence variant, m.3915G>A (p.G213G), that defines branch 270, is in the NADH dehydrogenase I gene (ND1) and has an estimated population frequency of ~1% (Saxena et al. 2006). This branch 270 variant (m.3915G>A) has been observed in three of the major mitochondrial haplogroups: B (Asian), H (European), and L3 (African) (Herrnstadt et al. 2002). Branch 269 is two segments long and thus is defined by two variants: m.4727A>G (p.M86M) is located in the NADH dehydrogenase 2 gene (ND2); and m.9380G>A (p.W58W) is located in the cytochrome c oxidase III gene (COX3). Both are synonymous polymorphisms that have been observed previously (Herrnstadt et al. 2002). m.9380G>A differentiates major haplogroup H6A1 (from H6A), and m.4727A>G is also specific to H6A1 (Brandstatter et al. 2006).

Our data contained seven sib groups (6 pairs and a trio) among the individuals contained in branches 269 and 270 as well as more distant relationships among some individuals without sibs. In order to verify that family risk for AD was not the basis of the observed associations, we added an additional covariate, family risk for AD (see Methods), to our model and reran our analyses. Even after controlling for family risk, these two contrasts remained significant (corrected p-value = 0.0174 and 0.0213 for 270 and 269, respectively).

Discussion

Here we present evidence for decreased risk of AD for complete mitochondrial haplotypes categorized within haplogroups H6A1A and H6A1B. Results from analyses, whether or not controlling for relatedness, were consistent; therefore our approach is robust when considering the familial relationships within the sample. The protective haplotypes were defined by three variants: m.3915G>A, m.4727A>G, and m.9380G>A. These three variants characterize two different major haplogroups. Together m.4727A>G and m.9380G>A define H6A1, and it has been suggested m.3915G>A defines H6A (Brandstatter et al. 2006), although that is not certain as it has been observed outside major haplogroup H6A (Herrnstadt et al. 2002). Additional variants differentiate H6A1A and H6A1B; however, none of these variants had a significant relationship with AD case-control status.

We know of no known functional explanation for these three variants or two major haplogroups that provide protection against AD. Each of the three variants is synonymous, and none of the three are located close to the 5' end of a gene. All three, however, are located in genes important for oxidative phosphorylation. ND1 and ND2 are components of NADH dehydrogenase, or Complex I, the first and largest enzyme in the electron transport chain (Brandt 2006). Lin and colleagues reported mutations in codon 331 of ND2 in the brains of AD patients; however, this hasn't been replicated and the sample size was very small (19 cases and 11 controls) (Lin et al. 1992). The observed variant in ND2 in our dataset was in codon 86 of ND2. The third variant was located in COX3, a critical transmembrane component of cytochrome c oxidase (complex IV), the final enzyme in the electron transport chain (Khalimonchuk, Rodel 2005). Given the evidence of changes in the electron transport chain in AD patients (Du et al. 2010; Santos et al.

2010), variants affecting the efficiency of respiration could plausibly contribute to or protect against AD.

There is no consensus on the effect of specific mitochondrial haplogroups for risk of AD and there have been mixed results for haplogroup H (and cluster HV). In one study haplogroup H was used as the reference group (since it was the most common in the dataset), and males with haplogroup U had increased risk for AD, while haplogroup U females had decreased risk compared to haplogroup H (van der Walt et al. 2004). These data are consistent with a possible protective effect for haplogroup H in males, although the study was based on only 10 SNPs and consequently did not provide more resolution. Another study found significant enrichment for AD cases in haplogroup H, but only a portion of the D-loop was sequenced and the sample size was small (Fesahat et al. 2007). In yet another study of D-loop sequences and specific coding region markers, H5 individuals were enriched for AD cases compared to other haplogroups (Santoro et al. 2010). Maruszak et al (Maruszak et al. 2009) found an association for the HV cluster with AD, but concluded that further analysis is necessary in order to understand the precise relationship of different haplotypes with AD. Lastly, the GERAD Consortium, in a very large study (~3000 cases and ~1000 age matched controls), tested 138 mtDNA markers, but found no significant associations with AD (Hudson et al. 2012). While the GERAD study had a very large sample size it had insufficient markers to test the groups we tested. One possible cause of the ambiguity in previous research is that in most cases only major mitochondrial haplogroups and/or clusters were analyzed; however, within each major haplogroup are many different haplotypes. By using the TreeScanning method we were able to overcome this challenge and identify which sub-haplogroups specifically provided protection against AD.

We have presented the results of association analyses using full mitochondrial sequences. Use of this complete mitochondrial genomic data allowed more precise identification of sequence variants for use in association studies. Furthermore, we were able to map these individuals back to two previously defined haplogroups: H6A1A and H6A1B. Our findings provide evidence of a reduced risk of AD for individuals of the two haplogroups defined by the three reported variants. This is the largest study to date using full mitochondrial genome sequences to look for association with LOAD. These data lack replication from an independent study and warrant validation in other datasets with a more diverse sampling of major haplogroups.

Acknowledgments

The authors thank the participants and staff of the Dementia Progression Study, the Utah Population Database, and the Cache County Study on Memory in Aging for their important contributions to this work. Additionally, the authors acknowledge the assistance of Drs. David Ward and Ned Weinshenker. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

The International HapMap Consortium. 2003. The International HapMap Project. Nature 426:789-796.

Anandatheerthavarada, HK, L Devi. 2007. Amyloid precursor protein and mitochondrial dysfunction in Alzheimer's disease. Neuroscientist 13:626-638.

Brandstatter, A, A Salas, H Niederstatter, C Gassner, A Carracedo, W Parson. 2006. Dissection of mitochondrial superhaplogroup H using coding region SNPs. Electrophoresis 27:2541-2550.

Brandt, U. 2006. Energy converting NADH:quinone oxidoreductase (complex I). Annu Rev Biochem 75:69-92.

Breitner, JC, BW Wyse, JC Anthony, et al. 1999. APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. Neurology 53:321-331.

Carrasquillo, MM, O Belbin, TA Hunter, et al. 2010. Replication of CLU, CR1, and PICALM associations with alzheimer disease. Arch Neurol 67:961-964.

Carrieri, G, M Bonafe, M De Luca, et al. 2001. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. Hum Genet 108:194-198.

Chagnon, P, M Gee, M Filion, Y Robitaille, M Belouchi, D Gauvreau. 1999. Phylogenetic analysis of the mitochondrial genome indicates significant differences between patients with Alzheimer disease and controls in a French-Canadian founder population. Am J Med Genet 85:20-30.

Chan, DC. 2006. Mitochondria: dynamic organelles in disease, aging, and development. Cell 125:1241-1252.

Chinnery, PF, GA Taylor, N Howell, RM Andrews, CM Morris, RW Taylor, IG McKeith, RH Perry, JA Edwardson, DM Turnbull. 2000. Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. Neurology 55:302-304.

Cho, DH, T Nakamura, J Fang, P Cieplak, A Godzik, Z Gu, SA Lipton. 2009. S-nitrosylation of Drp1 mediates beta-amyloid-related mitochondrial fission and neuronal injury. Science 324:102-105.

Chou, JL, DV Shenoy, N Thomas, PK Choudhary, FM Laferla, SR Goodman, GA Breen. 2011. Early dysregulation of the mitochondrial proteome in a mouse model of Alzheimer's disease. J Proteomics 74:466-479.

Clement, M, D Posada, KA Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657-1659.

Corneveaux, JJ, AJ Myers, AN Allen, et al. 2010. Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. Hum Mol Genet 19:3295-3301.

Coskun, P, J Wyrembak, S Schriner, HW Chen, C Marciniack, F Laferla, DC Wallace. 2011. A mitochondrial etiology of Alzheimer and Parkinson disease. Biochim Biophys Acta.

Cruts, M, C Van Broeckhoven. 1998. Presenilin mutations in Alzheimer's disease. Hum Mutat 11:183-190.

Cupples, LA, LA Farrer, AD Sadovnick, N Relkin, P Whitehouse, RC Green. 2004. Estimating risk curves for first-degree relatives of patients with Alzheimer's disease: the REVEAL study. Genet Med 6:192-196.

Devi, L, BM Prabhu, DF Galati, NG Avadhani, HK Anandatheerthavarada. 2006. Accumulation of amyloid precursor protein in the mitochondrial import channels of human Alzheimer's disease brain is associated with mitochondrial dysfunction. J Neurosci 26:9057-9068.

Du, H, L Guo, S Yan, AA Sosunov, GM McKhann, SS Yan. 2010. Early deficits in synaptic mitochondria in an Alzheimer's disease mouse model. Proc Natl Acad Sci U S A 107:18670-18675.

Duara, R, RF Lopez-Alberola, WW Barker, DA Loewenstein, M Zatinsky, CE Eisdorfer, GB Weinberg. 1993. A comparison of familial and sporadic Alzheimer's disease. Neurology 43:1377-1384.

Edland, SD, JM Silverman, ER Peskind, D Tsuang, E Wijsman, JC Morris. 1996. Increased risk of dementia in mothers of Alzheimer's disease cases: evidence for maternal inheritance. Neurology 47:254-256.

Farrer, LA, RH Myers, L Connor, LA Cupples, JH Growdon. 1991. Segregation analysis reveals evidence of a major gene for Alzheimer disease. Am J Hum Genet 48:1026-1033.

Fesahat, F, M Houshmand, MS Panahi, K Gharagozli, F Mirzajani. 2007. Do haplogroups H and U act to increase the penetrance of Alzheimer's disease? Cell Mol Neurobiol 27:329-334.

Grupe, A, Y Li, C Rowland, et al. 2006. A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease. Am J Hum Genet 78:78-88.

Herrnstadt, C, JL Elson, E Fahy, et al. 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152-1171.

Heyman, A, WE Wilkinson, BJ Hurwitz, D Schmechel, AH Sigmon, T Weinberg, MJ Helms, M Swift. 1983. Alzheimer's disease: genetic aspects and associated clinical disorders. Ann Neurol 14:507-515.

Hollingworth, PD HaroldR Sims, et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 43:429-435.

Honea, RA, RH Swerdlow, ED Vidoni, JM Burns. 2011. Progressive regional atrophy in normal adults with a maternal history of Alzheimer disease. Neurology 76:822-829.

Hudson, G, R Sims, D Harold, et al. 2012. No consistent evidence for association between mtDNA variants and Alzheimer disease. Neurology.

Janssen, JC, JA Beck, TA Campbell, A Dickinson, NC Fox, RJ Harvey, H Houlden, MN Rossor, J Collinge. 2003. Early onset familial Alzheimer's disease: Mutation frequency in 31 families. Neurology 60:235-239.

Jorde, LB. 1982. The genetic structure of the Utah Mormons: migration analysis. Hum Biol 54:583-597.

Jorde, LB, K Morgan. 1987. Genetic structure of the Utah Mormons: isonymy analysis. Am J Phys Anthropol 72:403-412.

Jun, G, AC Naj, GW Beecham, et al. 2010. Meta-analysis confirms CR1, CLU, and PICALM as alzheimer disease risk loci and reveals interactions with APOE genotypes. Arch Neurol 67:1473-1484.

Kamboh, MI, RL Minster, FY Demirci, M Ganguli, ST Dekosky, OL Lopez, MM Barmada. 2010. Association of CLU and PICALM variants with Alzheimer's disease. Neurobiol Aging.

Kerber, RA. 1995. Method for calculating risk associated with family history of a disease. Genet Epidemiol 12:291-301.

Khalimonchuk, O, G Rodel. 2005. Biogenesis of cytochrome c oxidase. Mitochondrion 5:363-388.

Kruger, J, R Hinttala, K Majamaa, AM Remes. 2010. Mitochondrial DNA haplogroups in early-onset Alzheimer's disease and frontotemporal lobar degeneration. Mol Neurodegener 5:8.

Lakatos, A, O Derbeneva, D Younes, et al. 2010. Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort. Neurobiol Aging 31:1355-1363.

Lambert, JC, S Heath, G Even, et al. 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094-1099.

Li, Y, A Grupe, C Rowland, et al. 2006. DAPK1 variants are associated with Alzheimer's disease and allele-specific expression. Hum Mol Genet 15:2560-2568.

Lin, FH, R Lin, HM Wisniewski, YW Hwang, I Grundke-Iqbal, G Healy-Louie, K Iqbal. 1992. Detection of point mutations in codon 331 of mitochondrial NADH dehydrogenase subunit 2 in Alzheimer's brains. Biochem Biophys Res Commun 182:238-246.

Mancuso, M, M Nardini, D Micheli, et al. 2007. Lack of association between mtDNA haplogroups and Alzheimer's disease in Tuscany. Neurol Sci 28:142-147.

Maruszak, A, JA Canter, M Styczynska, C Zekanowski, M Barcikowska. 2009. Mitochondrial haplogroup H and Alzheimer's disease--is there a connection? Neurobiol Aging 30:1749-1755.

McKhann, G, D Drachman, M Folstein, R Katzman, D Price, EM Stadlan. 1984. 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 34:939-944.

MITOMAP. 2011. MITOMAP: A Human Mitochondrial Genome Database, http://www.mitomap.org.

Mosconi, L, V Berti, RH Swerdlow, A Pupi, R Duara, M de Leon. 2010. Maternal transmission of Alzheimer's disease: prodromal metabolic phenotype and the search for genes. Hum Genomics 4:170-193.

Mosconi, L, M Brys, R Switalski, R Mistur, L Glodzik, E Pirraglia, W Tsui, S De Santi, MJ de Leon. 2007. Maternal family history of Alzheimer's disease predisposes to reduced brain glucose metabolism. Proc Natl Acad Sci U S A 104:19067-19072.

Mosconi, L, R Mistur, R Switalski, M Brys, L Glodzik, K Rich, E Pirraglia, W Tsui, S De Santi, MJ de Leon. 2009. Declining brain glucose metabolism in normal individuals with a maternal history of Alzheimer disease. Neurology 72:513-520.

Naj, ACG JunGW Beecham, et al. 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 43:436-441.

Nowotny, P, AL Hinrichs, S Smemo, et al. 2005. Association studies between risk for late-onset Alzheimer's disease and variants in insulin degrading enzyme. Am J Med Genet B Neuropsychiatr Genet 136B:62-68.

O'Brien, E, AR Rogers, J Beesley, LB Jorde. 1994. Genetic structure of the Utah Mormons: comparison of results based on RFLPs, blood groups, migration matrices, isonymy, and pedigrees. Hum Biol 66:743-759.

Pagani, L, A Eckert. 2011. Amyloid-Beta interaction with mitochondria. Int J Alzheimers Dis 2011:925050.

Posada, D, TJ Maxwell, AR Templeton. 2005. TreeScan: a bioinformatic application to search for genotype/phenotype associations using haplotype trees. Bioinformatics 21:2130-2132.

Pyle, A, T Foltynie, W Tiangyou, et al. 2005. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann Neurol 57:564-567.

Querfurth, HW, FM LaFerla. 2010. Alzheimer's disease. N Engl J Med 362:329-344.

Quintanilla, RA, PJ Dolan, YN Jin, GV Johnson. 2012. Truncated tau and Abeta cooperatively impair mitochondria in primary neurons. Neurobiol Aging 33:619 e625-635.

Santoro, A, V Balbi, E Balducci, et al. 2010. Evidence for sub-haplogroup h5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease. PLoS One 5:e12037.

Santos, RX, SC Correia, X Wang, G Perry, MA Smith, PI Moreira, X Zhu. 2010. Alzheimer's disease: diverse aspects of mitochondrial malfunctioning. Int J Clin Exp Pathol 3:570-581.

Saxena, R, PI de Bakker, K Singer, et al. 2006. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. Am J Hum Genet 79:54-61.

Silva, DF, AR Esteves, CR Oliveira, SM Cardoso. 2011. Mitochondria: the common upstream driver of amyloid-beta and tau pathology in Alzheimer's disease. Curr Alzheimer Res 8:563-572.

Skolnick, M. 1980. The Utah genealogical database: A resource for genetic epidemiology. In: JLJ Cairns, M Skolnick, editors. Banbury Report No 4; Cancer Incidence in Defined Populations. New York: Cold Spring Harbor. p. 285-297.

Swerdlow, RH, JM Burns, SM Khan. 2010. The Alzheimer's disease mitochondrial cascade hypothesis. J Alzheimers Dis 20 Suppl 2:S265-279.

Swerdlow, RH, SM Khan. 2004. A "mitochondrial cascade hypothesis" for sporadic Alzheimer's disease. Med Hypotheses 63:8-20.

Swerdlow, RH, SM Khan. 2009. The Alzheimer's disease mitochondrial cascade hypothesis: an update. Exp Neurol 218:308-315.

Takasaki, S. 2009a. Mitochondrial haplogroups associated with Japanese Alzheimer's patients. J Bioenerg Biomembr 41:407-410.

Takasaki, S. 2009b. Mitochondrial haplogroups associated with Japanese centenarians, Alzheimer's patients, Parkinson's patients, type 2 diabetic patients and healthy non-obese young males. J Genet Genomics 36:425-434.

Templeton, AR, E Boerwinkle, CF Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in Drosophila. Genetics 117:343-351.

Templeton, AR, T Maxwell, D Posada, JH Stengard, E Boerwinkle, CF Sing. 2005. Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. Genetics 169:441-453.

Thompson, JD, TJ Gibson, DG Higgins. 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics Chapter 2:Unit 2 3.

van der Walt, JM, YA Dementieva, ER Martin, et al. 2004. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. Neurosci Lett 365:28-32.

Westfall, P, SS Young. 1993. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustments. New York City, NY: Wiley-Interscience.

Zsurka, G, J Kalman, A Csaszar, I Rasko, Z Janka, P Venetianer. 1998. No mitochondrial haplotype was found to increase risk for Alzheimer's disease. Biol Psychiatry 44:371-373.

Figure Legends

Figure 1. Branches 269 and 270. Here we show branches 269 and 270 from our network (Supplementary Figure 1). Branch 270 consists of a single segment and 269 two segments, where each segment corresponds to a single variant. The two branches are labeled along with the variants corresponding to these branches. Together these two branches represent a single significant association. The other branches/variants are not labeled since they were not significantly associated with Alzheimer's disease case-control status. The nodes corresponding to major mitochondrial haplogroups H6A1A and H6A1B are enclosed by red rectangles and are labeled. Finally, observed haplotypes within our dataset are represented by the blue ellipses.

Table 1, Chapter 2. Distribution of major mtDNA haplogroups/clusters

Major Haplogroup	Number	Cases	Controls	Missing[1]	Ethnicities (Herrnstadt et al. 2002; MITOMAP 2011)
H	424	55	264	105	European
U	147	12	88	47	European
T	121	7	74	40	European
J	99	12	60	27	European
K	95	8	65	22	European
V	34	1	24	9	European
I	21	2	10	9	European
W	20	1	15	4	European
HV	18	2	14	2	European
X	8	0	8	0	European
C	5	1	2	2	Asian
L	4	0	2	2	African
Missing[2]	11	0	6	5	

Here we report the number of individuals belonging to each of the major haplogroups represented in our dataset along with case-control status.

[1]Missing case-control status

[2]Missing major haplogroup

Table 2, Chapter 2. Demographic characteristics of selected participants in the Cache County Study of Memory in Aging

	Number	Age (St dev)	Gender (male/female)	Cases/Controls/Missing
Total Individuals	1007	75.64 (7.50)	442/565	101/632/274
Branch 270	44	74.56 (8.11)	14/30	0/38/6
Branch 269	42	74.62 (8.29)	13/29	0/37/5

Here we report demographic information for total individuals and for individuals in each of two contrasts in our dataset, including the number of cases, controls, and those with no phenotype. Ages were based on collection date of blood.

Table 3, Chapter 2. Information about the significant contrasts

Contrast	p-values no family risk		p-values with family risk		Defining variant(s)
	nominal	corrected	nominal	corrected	
Branch 270	1.0E-04	0.016	3.0E-04	0.017	m.3915G>A (p.G213G)
Branch 269	1.0E-04	0.018	3.0E-04	0.021	m.4727A>G (p.M86M)
					m.9380G>A (p.W58W)

Reported here are the two contrasts. Two sets of p-values are reported. First, uncorrected and corrected p-values without the family risk for AD covariate, and p-values after adding the family risk covariate. These two contrasts are not independent and effectively represent the same single significant association.

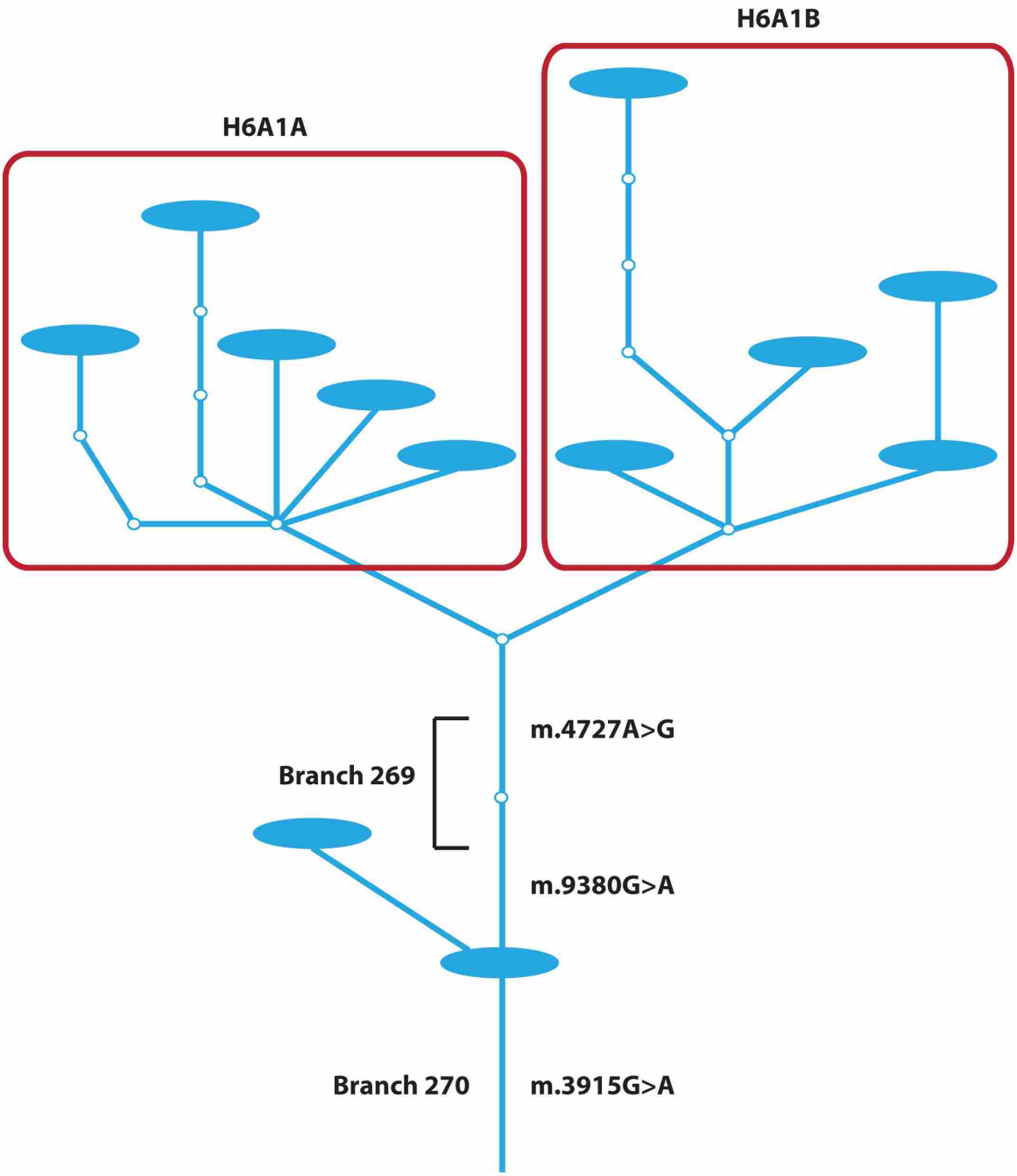


Figure 1, Chapter 2. Branches 269 and 270. Here we show branches 269 and 270 from our network (Supplementary Figure 1). Branch 270 consists of a single segment and 269 two segments, where each segment corresponds to a single variant. The two branches are labeled along with the variants corresponding to these branches. Together these two branches represent a single significant association. The other branches/variants are not labeled since they were not significantly associated with Alzheimer's disease case-control status. The nodes corresponding to major mitochondrial haplogroups H6A1A and H6A1B are enclosed by red rectangles and are labeled. Finally, observed haplotypes within our dataset are represented by the blue ellipses.

Manuscript submitted to BMC Medical Genomics

Chapter 3

Mitochondrial genomic variation associated with higher mitochondrial copy number: The Cache County Study on Memory Health and Aging

Perry G. Ridge[1,2], Taylor J. Maxwell[3], Spencer J. Foutz[1], Matthew H. Bailey[1], Christopher D. Corcoran[4,5], JoAnn T. Tschanz[5,6], Maria C. Norton[5,6,7], Ronald G. Munger[5,8], Elizabeth O'Brien[9], Richard A. Kerber[9], Richard M. Cawthon[10], John S.K. Kauwe[1]

[1]Department of Biology, Brigham Young University, Provo, UT

[2]ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT

[3]Human Genetics Center, University of Texas School of Public Health, Houston, TX

[4]Department of Mathematics and Statistics, Utah State University, Logan, UT

[5]Center for Epidemiologic Studies, Utah State University, Logan, UT

[6]Department of Psychology, Utah State University, Logan, UT

[7]Department of Family Consumer and Human Development, Utah State University, Logan, UT

[8]Department of Nutrition, Dietetics, and Food Sciences, Utah State University, Logan, UT

[9]Department of Epidemiology and Population Health, University of Louisville, Louisville, KY

[10]Department of Human Genetics, University of Utah, Salt Lake City, UT

Abstract

Background: The mitochondria are essential organelles and are the location of cellular respiration, which is responsible for the majority of ATP production. Each cell contains multiple mitochondria, and each mitochondrion contains multiple copies of its own circular genome. The ratio of mitochondrial genomes to nuclear genomes is referred to as mitochondrial copy number. Decreases in mitochondrial copy number are known to occur in many tissues as people age, and in certain diseases. The regulation of mitochondrial copy number by nuclear genes has been studied extensively. While mitochondrial variation has been associated with longevity and some of the diseases known to have reduced mitochondrial copy number, the role that the mitochondrial genome itself has in regulating mitochondrial copy number remains poorly understood.

Results: We analyzed the complete mitochondrial genomes from 1007 individuals randomly selected from the Cache County Study on Memory Health and Aging utilizing the inferred evolutionary history of the mitochondrial haplotypes present in our dataset to identify sequence variation and mitochondrial haplotypes associated with changes in mitochondrial copy number. Three variants belonging to mitochondrial haplogroups U5A1 and T2 were significantly associated with higher mitochondrial copy number in our dataset.

Conclusions: We identified three variants associated with higher mitochondrial copy number and suggest several hypotheses for how these variants influence mitochondrial copy number by interacting with known regulators of mitochondrial copy number. Our results are the first to report sequence variation in the mitochondrial genome that causes changes in mitochondrial

copy number. The identification of these variants that increase mtDNA copy number has important implications in understanding the pathological processes that underlie these phenotypes.

Key Words

Mitochondrial copy number, mitochondrial disease, mitochondrial genome, mitochondrial genetics, aging, haplotypes

Introduction

Mitochondria are the location of the citric acid or Krebs Cycle, which produces the majority of ATP for cellular work. Each cell has multiple mitochondria and each mitochondrion contains one or more copies of its own circular genome (mtDNA), which is 16569 bases in length and encodes 37 genes. Mitochondria are necessary for survival and malfunctioning mitochondria are the cause of a variety of diseases (Vilming et al. 1986; Shoffner et al. 1990; Ballinger et al. 1994; Jun, Brown, Wallace 1994; Taylor et al. 2002; Konradi et al. 2004; Swerdlow, Khan 2004; Swerdlow, Khan 2009; Swerdlow, Burns, Khan 2010; Bianchi et al. 2011; Sequeira et al. 2012). Mitochondrial diseases tend to affect the CNS or muscle tissue because of the high energy needs of these tissues (Finsterer 2007). Mitochondrial diseases have been well studied and can be the result of genetic variation in the mitochondrial and/or nuclear genomes. Pathogenic nuclear mutations are inherited in a typical Mendelian pattern and can present with a dominant, recessive, or X-linked dominant or recessive inheritance pattern. Examples of mitochondrial diseases caused by mutations in the nuclear genome include Friedrich's ataxia (Koutnikova et al. 1997), Wilson's disease (Gu et al. 2000), and Barth syndrome (Schlame et al. 2002).

In contrast, mitochondrial diseases caused by variation in the mtDNA are not as straightforward. Mitochondria are maternally inherited, so mitochondrial disease caused by these variants will display maternal inheritance. However, in most cases both normal and pathogenic mtDNA are inherited together and the mix can vary from predominantly wild type to predominantly pathogenic. Depending on the severity of the mutation, proportion of wild type versus affected mitochondria, and the specific tissue, there may or may not be a disease phenotype. Over the course of life the proportion of diseased mitochondria can change, possibly reaching a critical

threshold at which the disease phenotype is expressed. Alternatively, a constant proportion of diseased mitochondria might contribute to disease only when present in combination with one or more additional factors (e.g. stresses of various kinds, and/or aging). In addition to inherited mtDNA variation, mtDNA is prone to somatic mutations (Larsson 2010), and if affected mtDNA are propagated they can eventually reach a threshold at which mitochondrial function is insufficient to support normal cellular functions and disease appears. Some examples of disorders caused by mtDNA mutations are Kearns-Sayre syndrome (Bianchi et al. 2011), diabetes mellitus and deafness (Ballinger et al. 1994), Leber's hereditary optic neuropathy (Jun, Brown, Wallace 1994), Leigh Syndrome (Taylor et al. 2002), and Myoclonic Epilepsy with Ragged Red Fibers (a.k.a. MERRF syndrome) (Shoffner et al. 1990).

Additionally, mitochondria have a role in aging. The free-radical theory of aging, or mitochondrial free radical theory of aging, hypothesizes that aging occurs as damage from reactive oxygen species (ROS) accumulates. ROS are produced in the electron transport chain (Adam-Vizi 2005) and readily oxidize DNA and RNA, amino acids, and fatty acids (Cooke et al. 2003; Stadtman, Levine 2003; Li, Wu, Deleo 2006). Damage from ROS can accumulate with time resulting in cellular dysfunction, and death (Vendelbo, Nair 2011).

MtDNA copy number, or the cellular ratio of mitochondrial genomes to nuclear genomes, decreases with age in some, but not all, tissues (Barazzoni, Short, Nair 2000; Miller et al. 2003; Cree et al. 2008; Hartmann et al. 2011) and mtDNA copy number variation has been associated with numerous phenotypes (Hirai et al. 2001; Lamson, Plaza 2002; Lebedeva, Shadel 2007; Yu et al. 2007; Blokhin et al. 2008; Lee et al. 2009; Lee et al. 2010b; Santos et al. 2010;

Suomalainen, Isohanni 2010; Yu, Wan, Zou 2010; Lynch et al. 2011; Coskun et al. 2012; Thyagarajan et al. 2012). MtDNA copy number is tissue dependent (Clay Montier, Deng, Bai 2009) and varies with age and the energy needs of the cell (Barazzoni, Short, Nair 2000; Miller et al. 2003; Hassanin et al. 2009).

Several nuclear genes regulate mtDNA copy number. First, there is substantial evidence that mitochondrial transcription factor A (*TFAM*) regulates mtDNA copy number (Ekstrand et al. 2004; Matsushima, Goto, Kaguni 2010; Carling, Cree, Chinnery 2011; Guo et al. 2011). The *Mec1*/*Rad53* (yeast) pathway has been implicated in controlling mtDNA copy number, and mtDNA levels can be controlled by any of several genes responsible for regulating the pathway (Taylor et al. 2005). *p53* deficient cells or mutated *p53* leads to decreased levels of mtDNA (Kulawiec, Ayyasamy, Singh 2009). Two common nuclear SNPs in signal transducer and activator of transcription 3 (*STAT3*) were significantly associated with mtDNA levels in leukocytes (Gianotti et al. 2011). Both the Ras pathway and *p66Shc* likely have roles in regulating mtDNA copy number (Trinei et al. 2006). *MnSOD* prevents decreases in mtDNA levels by preventing a decrease in mtDNA replication proteins (Tewari, Santos, Kowluru 2012). And finally, overexpression of *Twinkle* increases mtDNA copy number (Ylikallio et al. 2010).

The direct role for the mitochondrial genome regulating levels of mtDNA has not been studied extensively. Here we conduct a genetic association study of full mitochondrial genome data and mtDNA copy number in individuals from the Cache County Study on Memory Health and Aging. Our results identify association between mitochondrial haplogroups U5A1 and T2 and increased mtDNA copy numbers.

Materials and Methods

Ethics Statement

All study procedures were approved by the Institutional Review Boards of Brigham Young University, Utah State University, Duke University, and Johns Hopkins University. Written consent was obtained for each individual. To verify a subject's capacity to consent, subjects attempted the Modified Mini-Mental State Exam (3MS). If there was an indication of poor cognitive ability as determined by poor performance on the entire test (scoring below a designated total of 60 points), poor performance on temporal or spatial orientation, or clear difficulty in understanding the nature of the interview, the visit was discontinued and informed consent was obtained from a responsible caregiver—often the next-of-kin. We re-consented subjects/caregivers at each study visit and procedure.

Sample Acquisition and Sequencing

Samples for this study were selected from the Cache County Study on Memory Health and Aging (Breitner et al. 1999). This study was initiated in 1994 to investigate associations of genetic and environmental factors with cognitive function. In 1994, the 5,092 individuals enrolled in the study from Cache County, Utah, represented 90% of all Cache County, Utah, residents who were 65 or older. The cohort was followed for 12 years and data (medical histories, demographics, and a multistage dementia assessment) were collected in four triennial waves. The Utah population is similar to other U.S. populations of northern European ancestry characterized by very little inbreeding. The founding group of Utah's population was unrelated and migrated from various locations in Europe (Jorde 1982; Jorde, Morgan 1987; O'Brien et al. 1994).

The Utah Population database has complete pedigree information going back 14 generations to the original Utah Founders. Using this information we identified individuals from the Cache County Study with the same maternal line of inheritance (matrilineage). We randomly selected one individual from each matrilineage, selecting individuals from the largest matrilineages first to maximize our ability to infer mitochondrial genomic information. Given our resources, we were able to sequence a representative sample from 274 of the 3151 matrilineages that exist in the Cache County Study samples. The sequenced mitochondrial genomes represent many different major mitochondrial haplogroups (Table 1). 287 samples were sent to Family Tree DNA (www.familytreedna.com) for Sanger sequencing of the mitochondrial genomes. Two samples failed quality control at Family Tree DNA. Based on maternal inheritance of the mtDNA we inferred that individuals who share matrilineal relationships have the same mtDNA. Using this we inferred the status of full mitochondrial genome sequence for 722 additional individuals for a total of 1007 individuals, not accounting for de novo mutation. The extensive pedigree data in the UPDB allows identification of shared maternal lineages for very distant relationships. As this was a population-based study it is one generation in depth, but there are extended familial relationships, even very distant cousins. Ridge et al (Ridge et al. 2012) contains additional details about the sequencing and inference of the mtDNA status in this dataset.

Measurement of mtDNA copy number

Relative quantitation of the ratio of the copy number of the mitochondrial genome to the copy number of the nuclear single copy gene beta-globin, as compared to that ratio in a reference

DNA sample, was determined by monochrome multiplex quantitative polymerase chain reaction (QPCR). Buccal sample cell lysates were diluted in water (containing yeast total RNA as carrier, at 2.5 ng per microliter) to a final total cellular DNA concentration of approximately 1 ng per 10 microliters. QPCR was carried out in 25 microliter reactions, containing 10 microliters of the diluted buccal lysate and 15 microliters of QPCR reagent mix with primers.

The QPCR reagent mix, without primers, was exactly as described by Cawthon (Cawthon 2009). The primers for mtDNA amplification were mt3257u, 5'- GCAGAGCCCGGTAATCGCA-3', and mt3272d, 5'-TAAGAAGAGGAATTGAACCTCTGACTGTAA-3'. The mt3272d primer has previously been shown to be specific to mtDNA and unable to amplify any nuclear-embedded mtDNA-like sequences (numts) from rho 0 cell line DNA (Parfait et al. 1998) (Rho 0 cell lines are mtDNA-free). The primers for the beta-globin gene were hbgugc2, 5'-CGGCGGCGGGCGGCGCGGGCTGGGCGGCTTCATCCACGTTCACCTTG-3', and hbgdgc2, 5'-GCCCGGCCCGCCGCGCCCGTCCCGCCGGAGGAGAAGTCTGCCGTT-3'. Both beta-globin primers contained 5' GC-clamp (non-templated) sequences that confer a high melting temperature on their amplicon. Each of the four primers was present at a final concentration of 900 nM.

The thermal profile for QPCR began with 95 degrees C for 15 minutes to activate the hot-start polymerase and fully denature the DNA; followed by 35 cycles of: 94 degrees for 15 sec, 62 degrees for 20 sec, 72 degrees for 15 sec with signal acquisition (to read the mtDNA amplification signal), 84 degrees for 10 sec, and 88 degrees for 15 sec (to read the beta-globin signal). In this monochrome multiplex QPCR (MMQPCR) strategy, first described by Cawthon

(Cawthon 2009), the higher copy number target (in this case mtDNA) has its amplification signal collected over a cycle range in which the lower copy number target's (in this case the beta-globin genes) amplification signal is still at baseline, and the lower copy number target's amplificaton signal is collected in later cycles, at a temperature that is sufficiently high to completely melt the amplicon of the higher copy number target, driving its signal to baseline so that the signal from the high melting amplicon can be cleanly read. All QPCR runs were done on Bio-Rad MyiQ real-time machines, using the manufacturer's accompanying software. The Standard Curve method for relative quantitation was used, with 36 ng of a reference DNA sample as the high end, and four additional standard concentrations obtained via 3-fold serial dilutions from the high end. Each subject's buccal lysate was assayed in triplicate. The average of the three measurements for each sample was used in this study (Supplementary Table 1). DNA is not available from other tissue for the majority of these samples.

Sequence and Statistical Analyses

We used ClustalW (Thompson, Gibson, Higgins 2002) to align the mitochondrial genomes and inferred a haplotype network using TCS (Clement, Posada, Crandall 2000) and the 285 sequenced mitochondrial genomes. In a haplotype network, segments of branches correspond to a single sequence feature (single nucleotide variant, indel, etc.), and nodes in the network correspond to haplotypes. Branches, comprised of one or more segments, connect observed nodes, while clades are comprised of one or more observed nodes, and are defined by a branch.

Genotype-phenotype associations were evaluated using an evolution-based method known as TreeScanning (Posada, Maxwell, Templeton 2005; Templeton et al. 2005) that makes use of

haplotype networks. Haplotype networks provide a framework from which to select evolutionarily related haplotypes to pool together for comparison. Additional details about the application of TreeScanning to this dataset can be found in Ridge et al (Ridge et al. 2012). The null hypothesis of TreeScanning is that the phenotype does not differ in distribution across the genotypes derived from allelic classes defined by the branches of the haplotype network. Each branch partitions the haplotypes into bi-allelic pools from which genotypes are constructed and treated as a separate test. Because we have multiple tests that are correlated we obtained multiple-test corrected p-values by a permutation analog of the sequential step-down Bonferroni (Westfall, Young 1993) with 10,000 permutations. If significant branches are found in the first round of TreeScanning, a second round of TreeScanning is performed that can detect phenotypic heterogeneity within the allelic classes of the significant branch. This is accomplished by creating a three-allele system and using conditional permutations that hold one of the alleles constant while subdividing the other class into two alleles (Templeton et al. 2005).

For these analyses we tested for association with mitochondrial copy number after adjusting for age, gender, and familial relationships. Familial adjustment scores, which quantify the variance in mtDNA copy number that is due to familial relationships between individuals in the dataset, were computed using the method developed by Kerber (Kerber 1995) (modified for a continuous trait). For each individual we summed the product of the mtDNA copy number and the kinship coefficient across all other individuals in the sample. This sum is then divided by the total number of samples in the dataset. Finally, we divide by the mtDNA copy number of the individual, yielding a value, which represents the relationship between mtDNA copy number and relatedness to other individuals in the dataset. Following is the equation for calculating the

familial adjustment score:

$$\text{familial adjustment score} = \frac{\frac{\sum_{j=1}^{N} copy\ number\ j * f(individual, j)}{N}}{individual\ copy\ number}$$

where *N* is the number of individuals in the cohort and *f(individual, j)* is the kinship coefficient between the individual for whom we are calculating a familial adjustment score and individual *j*. Using this equation we calculated a familial adjustment score for each individual in the dataset. Inclusion of this score as a covariate in our analyses removes variance in mtDNA copy number that is due to relatedness between individuals, making it possible to test for association independently of pedigree relationships in the data. This adjustment addresses both maternal and paternal relationships in the data, thus correcting for possible nuclear genomic confounds as well. Each analysis was performed with 10,000 permutations. Only tests with at least two relevant genotypic classes, each containing five or more individuals, were tested. Significance was inferred if the multiple-test-corrected p-value was less than 0.05.

Bioinformatic Analyses of Variants

In order to determine the functional impact of variants of interest we applied *in silico* functional prediction algorithms, analyzed pathways, examined protein sequence conservation, and identified conserved domains. We obtained protein sequences from NCBI using blast (Altschul et al. 1990), aligned and analyzed them using the CLCViewer (http://clcbio.com/), identified conserved domains using the NCBI conserved domain database (Marchler-Bauer et al. 2011), identified pathways using Ingenuity (http://Ingenuity.com/), and obtained functional predictions from polyphen-2 (Adzhubei et al. 2010) and SIFT (Kumar, Henikoff, Ng 2009a; Kumar, Henikoff, Ng 2009b) webservers.

Results

Haplotype Network and mtDNA Variation

We sequenced 285 complete mitochondrial genomes from individuals in the Cache County Study on Memory Health and Aging and imputed 722 additional full mitochondrial genomes using maternal lineages for a full dataset of 1007 full mitochondrial genomes. We built our network using the 285 genotyped individuals (Supplementary Figures 1-2). Our network contained 249 different haplotypes and the majority of haplotypes (152 of 249) were observed in three or fewer individuals with the two most frequently observed haplotypes observed in 39 and 32 individuals, respectively. Our network contained one unresolved loop and the ambiguity was factored into subsequent analyses.

We identified 899 single nucleotide variants (SNVs), 26 insertions, and 20 deletions in our dataset. The most frequently observed SNVs occurred in 281 genomes (m.263A>G, m.8860A>G, and m.15326A>G), and three more SNVs were observed in 280 genomes (m.750A>G, m.1438A>G, and m.4769A>G). Compared to the reference sequence (NC_012920), each person had an average of 25.3 variants (52 variants were the most identified in a single individual and 2 variants the fewest, each extreme observed in one person).

The distribution of major mitochondrial haplogroups within our dataset is reported in Table 1 (major mitochondrial haplogroups/clusters) and Supplementary Table 2 (major mitochondrial haplogroups and sub-haplogroups). We had individuals from 102 major mitochondrial haplogroups/clusters (or sub-haplogroups) in our dataset. As expected, the majority (987 of 1007) of individuals in our dataset belonged to European based major mitochondrial

haplogroups. We identified three different branches, corresponding to two different clades, significantly associated with mtDNA copy number.

Branches 124 and 121 are Associated with mtDNA Copy Number

First, branches 124 and 121, p-values of 8.0e-4 and 0.0043 (multi-test corrected p-values), respectively (Table 2, Figure 1), were associated with higher mtDNA copy number. The clade defined by branch 121 is wholly contained within branch 124 (Figure 2); therefore these two branches are highly correlated and represent the same effect. Branch 124 is defined by a single variant (Table 3), *m.9667A>G*. This is a missense variant, *p.Asn154Ser*, located in cytochrome C oxidase 3 (*COXIII*). Branch 121 is defined by two variants (Table 3), *m.12582A>G* and *m.12879T>C*, both synonymous variants in NADH dehydrogenase 5 (*ND5*).

Since these two branches correspond to a single effect and branch 121 is wholly contained within branch 124, we consider only the clade defined by branch 124 from this point forward. This clade contains 14 individuals for whom we have mtDNA copy number measurements. Pairwise kinship coefficients are reported for these individuals in Supplementary Table 3. Individuals in this clade have a mtDNA copy number nearly 50% higher (3.81 compared to 2.69, p-value 8.0e-4) than individuals in the rest of the dataset.

All of the individuals in the clade defined by branch 124 belong to major mitochondrial haplogroup U5A1, and have one of four different haplotypes (represented by nodes in Figure 2). Nine other individuals (five different haplotypes) in the dataset also belong to U5A1. These individuals are located in adjacent clades to the one defined by branch 124 and have significantly

lower mitochondrial copy numbers than the other U5A1 individuals (p-value 0.0082). The contrast of all U5A1 individuals against the rest of the dataset was nominally significant (p-value 0.0019). While no d-loop variants define branch 124, *m.16399A>G*, a d-loop variant, is only found in the U5A1 individuals in our dataset and in general appears to be found in all U5A1 individuals (van Oven, Kayser 2009).

Branch 50 is Associated with mtDNA Copy Number

Branch 50 is the third branch significantly associated higher mtDNA copy number (p-value 0.015, multi-test corrected p-value, Table 2, Figure 1). This represents a statistically separate effect as we controlled for the effect of branch 124 in our analyses (just as we controlled for branch 50 in our analyses of branch 124). Eight sequence features define branch 50: seven single nucleotide variants and one nine base pair deletion (Table 3). Six of the eight features are intergenic or synonymous, but the other two are both missense variants. *m.5277T>C* (*p.Phe270Leu*) is a missense variant in NADH dehydrogenase 2 (*ND2*) and *m.6489C>A* (*p.Leu196Ile*) is a missense variant in cytochrome C oxidase 1 (*COXI*).

In the clade defined by branch 50 there are 12 individuals with mtDNA copy number measurements. Pairwise kinship coefficients are reported for these individuals in Supplementary Table 4. The average mtDNA copy number for individuals in this clade is 3.64 and is significantly higher than the average for the rest of the dataset (2.69, p-value 0.015). Individuals in this clade belong to major mitochondrial haplogroup T2 and all have the exact same haplotype. There were no other T2 individuals in the rest of our dataset; however, there were T2A, T2B, T2C, and T2E individuals. The contrast between T2 and all T2 sub-haplogroups

(T2A, T2B, T2C, and T2E) and the rest of the data was nominally significant, p-value 0.019, and the contrast of T2B individuals alone against the rest of the dataset was nominally significant, p-value 0.0062.

Bioinformatic Analyses of m.9667A>G, branch 124

m.9667A>G is the defining sequence change between the U5A1 individuals in our dataset who had significantly higher mtDNA copy number levels from the other U5A1 individuals in our dataset whose concentration measurements were not statistically different from the rest of the dataset. *m.9667A>G* causes an amino acid substitution, asparagine to serine, at position 154 of *COXIII*, which is located in an 11 residue stretch between transmembrane domains. Since this is a missense mutation, we sought to determine if it changes or inhibits *COXIII* and/or the cytochrome c oxidase complex. We compared *COXIII* sequences in organisms from humans through yeast by aligning a 41-residue stretch of *COXIII*. In Figure 3, position 154 of *COXIII* (the position of the amino acid substitution corresponding to *m.9667A>G*) is in position 21 of the alignment. As seen in Figure 3, two different amino acids appear in this position: asparagine and glycine. Asparagine and glycine are both uncharged amino acids; however, asparagine is polar, whereas glycine is nonpolar. The substituted amino acid, serine, is polar and similar in size to asparagine (asparagine 132.1 g/mol, glycine 75.1 g/mol, and serine 105.1 g/mol).

We further analyzed the effect of this substitution on *COXIII* by using in silico algorithms that predict the effect of amino acid substitutions on protein function using a variety of criteria such as conservation, amino acid biochemical properties, known domains/structures of the protein,

etc. Polyphen-2 predicted the substitution to be benign and SIFT predicted a pathogenic mutation, but noted that its prediction was of very low confidence.

Lastly, we looked at possible interactions of *COXIII* with known regulators (listed in the Introduction) of mtDNA copy number to identify mechanisms *m.9667A>G* could cause the increased copy number. We found common regulators of both *COXIII* and the mtDNA copy number regulators, and we found ways that these regulators could affect *COXIII* expression; however, we identified no pathways by which *COXIII* could regulate mtDNA copy number by known mechanisms (Figure 4).

Bioinformatic Analyses of m.5277T>C and m.6489C>A, branch 50

It is more difficult to say which variants are causing the increase in mtDNA copy number for the clade-defined branch 50 since this branch consists of eight different sequence features. We chose to focus our analyses on two of the features: *m.5277T>C* and *m.6489C>A* since these two variants are missense variants and the six others features are either synonymous or intergenic changes.

First, *m.5277T>C* results in a phenylalanine to leucine change in *ND2*. Position 270 of *ND2* is column 21 in Figure 5. At this position, primates have phenylalanine and other species before have leucine. *p.Phe270Leu* changes the human sequence back to the historical residue. Polyphen-2 and SIFT predict that this substitution is benign and tolerated, respectively.

Next, *m.6489C>A* causes a leucine to isoleucine change at position 196 of *COXI*. This region of *COXI* is highly conserved. Position 196 is leucine in every species we examined from humans to yeast except nematodes that have valine at this position (Figure 6). Polyphen-2 predicts that this substitution is probably damaging, and SIFT also predicts that this substitution affects function, but it is a low confidence prediction. Lastly, we identified pathways in which *COXI* malfunction could cause an increase in mtDNA copy number. First we analyzed pathways for all nuclear genes known to modify mtDNA copy number and found no obvious pathways for genes other than *p53* and *TFAM*. We identified several pathways in which *COXI* malfunction could change mtDNA copy number, the majority of which function through intermediate genes activated by reactive oxygen species (Figure 7).

Discussion

Using 1007 full mitochondrial genome sequences we have identified sequence variation in mtDNA that affects mtDNA copy number. Two different clades were significantly associated with higher mtDNA copy number. Each of these clades represents statistically separate effects. The first was defined by branch 124 and consisted of individuals with haplogroup U5A1, and is defined by *m.9667A>G* (*p.Asn154Ser*). This variant has also been reported in D2A1, D4M1, and J1B2A haplogroups (van Oven, Kayser 2009); however, no individuals in our dataset belong to these haplogroups. We analyzed this substitution to determine if it likely causes *COXIII* malfunction, and then to determine whether or not it could cause the observed increase in mtDNA copy number. Our analyses suggest this substitution does not impact *COXIII* function. This conclusion is based on several lines of evidence; first, this is a high frequency, known substitution (Saxena et al. 2006), second the substitution occurs in an unconserved site (Figure

3), third asparagine and glycine, two very different amino acids, appear historically in this position and a change from asparagine to the more similar serine is likely to be tolerated, and finally this position is in a short stretch of sequence located between transmembrane domains and is not a known position of importance in the heme-copper oxidase subunit III super family, of which it is a part.

While it seems likely this variant does not disrupt *COXIII* function, it is still possible that it could alter protein-protein interactions or specific dynamics associated with the electron transport chain and ultimately lead to changes in mtDNA copy number. Our initial analyses of known regulators of mtDNA copy number with *COXIII* (Figure 4) revealed no obvious mechanism for *COXIII* to directly modify mtDNA copy number; however, Pello et al (Pello et al. 2008) reported that *m.9667A>G* causes respiratory chain assembly deficiencies in patients with Leber's hereditary optic neuropathy. *TFAM* (the main known regulator of mtDNA copy number) concentration and mtDNA copy number are proportional (Ekstrand et al. 2004); therefore, upregulators of *TFAM* increase mtDNA copy number. *TFAM* is regulated by *NRF-1* and *NRF-2*, and all three are sensitive to the energy needs of the cell (Piantadosi, Suliman 2006; Hock, Kralli 2009). Silencing of *NRF-1* is known to lead to higher levels of *TFAM* and *NRF-1* expression is known to increase in response to signals meant to increase energy production (Hock, Kralli 2009). We propose the following model for *m.9667A>G* to increase mtDNA copy number. First, *m.9667G>A* can decrease the efficiency of complex assembly and decrease overall energy production of the mitochondria, in response, *NRF-1* expression increases, which in turn increases *TFAM* expression, and *TFAM* expression increases mtDNA copy number.

The second clade significantly associated with higher mtDNA copy number is defined by branch 50. Branch 50 consists of eight sequence features, six of which are synonymous changes or located in intergenic regions. We focused our functional analyses on the two missense variants. The first is *m.5277T>C* (*p.Phe270Leu*) in *ND2*. Besides T2, this variant has also been reported in L1C1A1B individuals (van Oven, Kayser 2009) and there are no L1C1A1B individuals in our dataset. This variant is in an unconserved region immediately adjacent to a low complexity region, predicted to not affect protein function, and is not novel (Finnila, Lehtonen, Majamaa 2001; Herrnstadt et al. 2002b). These data suggest this variant is not functionally deleterious.

In contrast, there is evidence that the second of the two variants that define this clade, *m.6489C>A* (*p.Leu196Ile*), are functionally deleterious and can explain changes in mtDNA copy number. *m.6489C>A* is specific to T2 (van Oven, Kayser 2009). This variant in *COXI* occurs in a highly conserved region in both *COXI* and the heme-copper subunit I domain it is in, and is predicted to affect function. *COXI* and *COXII* form the catalytic center of cytochrome c oxidase (COX), or complex IV, in the respiratory complex. *m.6489C>A* is not a novel mutation and has been reported to lead to COX deficiency and a destabilization of complex IV (Varlamov et al. 2002). It does appear, however, that a high threshold of mutant mtDNA is required before a phenotype appears (Varlamov et al. 2002; Kleefstra et al. 2011). The variant has only been observed in mitochondrial haplogroups T2F1 (van Oven, Kayser 2009) and in our T2 individuals here.

COXI is involved in several pathways that include known regulators of mtDNA copy number (Figure 7). Mutations in mtDNA in general, and cytochrome c oxidase malfunction specifically,

lead to an increase in reactive oxygen species (ROS) (Murphy 2009). ROS increase expression and/or activate protein kinase B (*Akt*) (Lu et al. 2010; Zou et al. 2010), HIF-1 Alpha (*HIF1A*) (Gao et al. 2002; Sitkovsky et al. 2004), nuclear factor (erythroid-derived 2)-like 2 (*NFE2L2*) (Nathan, Ding 2010), *SP1* transcription factor (Carnesecchi et al. 2006), and *p53* (Lee et al. 2010a; Li et al. 2010). *p53* could also be potentially activated by *COXI* binding *CDC42* (Lassus et al. 2000; Bandyopadhyay et al. 2010). Of these genes, two suppress *TFAM* (*Akt* and *HIF1A*) (Liu et al. 2009; Krishnan et al. 2012), and three increase *TFAM* expression (*NFE2L2*, *SP1*, and *p53*) (Zaid et al. 1999; McGill, Beal 2006; Park et al. 2009).

Conclusions

As we outlined previously, mtDNA copy number is related to several important human health phenotypes including several age-related disorders. The identification of these variants that increase mtDNA copy number has important implications in understanding the pathological processes that underlie these phenotypes. We have used bioinformatics analyses to generate hypotheses for the mechanisms by which these variants influence mtDNA copy number, successfully generating several hypotheses. Future work to characterize these mechanisms will provide important insights into the effects of mitochondrial genomic variation on mtDNA copy number and broader human phenotypes.

List of abbreviations

mtDNA, mitochondrial genome; CNS, central nervous system, AD, Alzheimer's disease; ROS, reactive oxygen species; ETC, electron transport chain; SNV, single nucleotide variant; TFAM, mitochondrial transcription factor A; STAT3, signal transducer and activator of transcription 3;

D-loop, displacement loop; COX, cytochrome C oxidase; COXI, cytochrome C oxidase 1; COXIII, cytochrome C oxidase 3; ND2, NADH dehydrogenase 2; ND5, NADH dehydrogenase 5; Akt, activate protein kinase B; HIF1A, HIF-1 Alpha; NFE2L2, nuclear factor (erythroid-derived 2)-like 2; CEPH, Centre d'Etude du Polymorphisme.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

PGR, TYM, RMC, and JSKK designed analyses; PGR2, TJM, and RAK performed the analyses; CDC, JTT, MCN, RGM, EO, and RAK collected the data; RMC measured mitochondrial copy number; PGR2 and SJF wrote the paper; all authors contributed to revisions and approved of the final draft.

Acknowledgements

This work was supported by grants from NIH (R01AG11380, R01AG21136, R01AG31272, R01AG042611), the Alzheimer's Association (MNIRG-11-205368) and the Utah Science, Technology, and Research initiative (USTAR), and the Brigham Young University Gerontology Program (JSKK). The authors thank the participants and staff of the Dementia Progression Study, the Utah Population Database, and the Cache County Study on Memory Health and Aging for their important contributions to this work. Additionally, the authors acknowledge the assistance of Drs. David Ward and Ned Weinshenker. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Adam-Vizi, V. 2005. Production of reactive oxygen species in brain mitochondria: contribution by electron transport chain and non-electron transport chain sources. Antioxid Redox Signal 7:1140-1149.

Adzhubei, IA, S Schmidt, L Peshkin, VE Ramensky, A Gerasimova, P Bork, AS Kondrashov, SR Sunyaev. 2010. A method and server for predicting damaging missense mutations. Nature methods 7:248-249.

Altschul, SF, W Gish, W Miller, EW Myers, DJ Lipman. 1990. Basic local alignment search tool. Journal of molecular biology 215:403-410.

Ballinger, SW, JM Shoffner, S Gebhart, DA Koontz, DC Wallace. 1994. Mitochondrial diabetes revisited. Nature genetics 7:458-459.

Bandyopadhyay, S, CY Chiang, J Srivastava, et al. 2010. A human MAP kinase interactome. Nature methods 7:801-805.

Barazzoni, R, KR Short, KS Nair. 2000. Effects of aging on mitochondrial DNA copy number and cytochrome c oxidase gene expression in rat skeletal muscle, liver, and heart. The Journal of biological chemistry 275:3343-3347.

Bianchi, M, T Rizza, D Verrigni, et al. 2011. Novel large-range mitochondrial DNA deletions and fatal multisystemic disorder with prominent hepatopathy. Biochemical and biophysical research communications 415:300-304.

Blokhin, A, T Vyshkina, S Komoly, B Kalman. 2008. Variations in mitochondrial DNA copy numbers in MS brains. Journal of molecular neuroscience : MN 35:283-287.

Breitner, JC, BW Wyse, JC Anthony, et al. 1999. APOE-epsilon4 count predicts age when prevalence of AD increases, then declines: the Cache County Study. Neurology 53:321-331.

Carling, PJ, LM Cree, PF Chinnery. 2011. The implications of mitochondrial DNA copy number regulation during embryogenesis. Mitochondrion 11:686-692.

Carnesecchi, S, JL Carpentier, M Foti, I Szanto. 2006. Insulin-induced vascular endothelial growth factor expression is mediated by the NADPH oxidase NOX3. Experimental cell research 312:3413-3424.

Cawthon, RM. 2009. Telomere length measurement by a novel monochrome multiplex quantitative PCR method. Nucleic acids research 37:e21.

Clay Montier, LL, JJ Deng, Y Bai. 2009. Number matters: control of mammalian mitochondrial DNA copy number. Journal of genetics and genomics = Yi chuan xue bao 36:125-131.

Clement, M, D Posada, KA Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657-1659.

Cooke, MS, MD Evans, M Dizdaroglu, J Lunec. 2003. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J 17:1195-1214.

Coskun, P, J Wyrembak, SE Schriner, HW Chen, C Marciniack, F Laferla, DC Wallace. 2012. A mitochondrial etiology of Alzheimer and Parkinson disease. Biochimica et biophysica acta 1820:553-564.

Cree, LM, SK Patel, A Pyle, S Lynn, DM Turnbull, PF Chinnery, M Walker. 2008. Age-related decline in mitochondrial DNA copy number in isolated human pancreatic islets. Diabetologia 51:1440-1443.

Ekstrand, MI, M Falkenberg, A Rantanen, CB Park, M Gaspari, K Hultenby, P Rustin, CM Gustafsson, NG Larsson. 2004. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. Human molecular genetics 13:935-944.

Finnila, S, MS Lehtonen, K Majamaa. 2001. Phylogenetic network for European mtDNA. American journal of human genetics 68:1475-1484.

Finsterer, J. 2007. Hematological manifestations of primary mitochondrial disorders. Acta haematologica 118:88-98.

Gao, N, BH Jiang, SS Leonard, et al. 2002. p38 Signaling-mediated hypoxia-inducible factor 1alpha and vascular endothelial growth factor induction by Cr(VI) in DU145 human prostate carcinoma cells. The Journal of biological chemistry 277:45041-45048.

Gianotti, TF, G Castano, C Gemma, AL Burgueno, MS Rosselli, CJ Pirola, S Sookoian. 2011. Mitochondrial DNA copy number is modulated by genetic variation in the signal transducer and activator of transcription 3 (STAT3). Metabolism: clinical and experimental 60:1142-1149.

Gu, M, JM Cooper, P Butler, AP Walker, PK Mistry, JS Dooley, AH Schapira. 2000. Oxidative-phosphorylation defects in liver of patients with Wilson's disease. Lancet 356:469-474.

Guo, J, L Zheng, W Liu, X Wang, Z Wang, AJ French, D Kang, L Chen, SN Thibodeau. 2011. Frequent truncating mutation of TFAM induces mitochondrial DNA depletion and apoptotic resistance in microsatellite-unstable colorectal cancer. Cancer research 71:2978-2987.

Hartmann, N, K Reichwald, I Wittig, et al. 2011. Mitochondrial DNA copy number and function decrease with age in the short-lived fish Nothobranchius furzeri. Aging cell 10:824-831.

Hassanin, A, A Ropiquet, A Couloux, C Cruaud. 2009. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). Journal of molecular evolution 68:293-310.

Herrnstadt, C, JL Elson, E Fahy, et al. 2002a. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152-1171.

Herrnstadt, C, JL Elson, E Fahy, et al. 2002b. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. American journal of human genetics 70:1152-1171.

Hirai, K, G Aliev, A Nunomura, et al. 2001. Mitochondrial abnormalities in Alzheimer's disease. The Journal of neuroscience : the official journal of the Society for Neuroscience 21:3017-3023.

Hock, MB, A Kralli. 2009. Transcriptional control of mitochondrial biogenesis and function. Annual review of physiology 71:177-203.

Jorde, LB. 1982. The genetic structure of the Utah Mormons: migration analysis. Hum Biol 54:583-597.

Jorde, LB, K Morgan. 1987. Genetic structure of the Utah Mormons: isonymy analysis. Am J Phys Anthropol 72:403-412.

Jun, AS, MD Brown, DC Wallace. 1994. A mitochondrial DNA mutation at nucleotide pair 14459 of the NADH dehydrogenase subunit 6 gene associated with maternally inherited Leber hereditary optic neuropathy and dystonia. Proceedings of the National Academy of Sciences of the United States of America 91:6206-6210.

Kerber, RA. 1995. Method for calculating risk associated with family history of a disease. Genet Epidemiol 12:291-301.

Kleefstra, T, SB Wortmann, RJ Rodenburg, et al. 2011. Mitochondrial dysfunction and organic aciduria in five patients carrying mutations in the Ras-MAPK pathway. European journal of human genetics : EJHG 19:138-144.

Konradi, C, M Eaton, ML MacDonald, J Walsh, FM Benes, S Heckers. 2004. Molecular evidence for mitochondrial dysfunction in bipolar disorder. Archives of general psychiatry 61:300-308.

Koutnikova, H, V Campuzano, F Foury, P Dolle, O Cazzalini, M Koenig. 1997. Studies of human, mouse and yeast homologues indicate a mitochondrial function for frataxin. Nature genetics 16:345-351.

Krishnan, J, C Danzer, T Simka, et al. 2012. Dietary obesity-associated Hif1alpha activation in adipocytes restricts fatty acid oxidation and energy expenditure via suppression of the Sirt2-NAD+ system. Genes & development 26:259-270.

Kulawiec, M, V Ayyasamy, KK Singh. 2009. p53 regulates mtDNA copy number and mitocheckpoint pathway. Journal of carcinogenesis 8:8.

Kumar, P, S Henikoff, PC Ng. 2009a. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols 4:1073-1081.

Kumar, P, S Henikoff, PC Ng. 2009b. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073-1081.

Lamson, DW, SM Plaza. 2002. Mitochondrial factors in the pathogenesis of diabetes: a hypothesis for treatment. Alternative medicine review : a journal of clinical therapeutic 7:94-111.

Larsson, NG. 2010. Somatic mitochondrial DNA mutations in mammalian aging. Annual review of biochemistry 79:683-706.

Lassus, P, P Roux, O Zugasti, A Philips, P Fort, U Hibner. 2000. Extinction of rac1 and Cdc42Hs signalling defines a novel p53-dependent apoptotic pathway. Oncogene 19:2377-2385.

Lebedeva, MA, GS Shadel. 2007. Cell cycle- and ribonucleotide reductase-driven changes in mtDNA copy number influence mtDNA Inheritance without compromising mitochondrial gene expression. Cell cycle 6:2048-2057.

Lee, JE, H Park, YS Ju, M Kwak, JI Kim, HY Oh, JS Seo. 2009. Higher mitochondrial DNA copy number is associated with lower prevalence of microalbuminuria. Experimental & molecular medicine 41:253-258.

Lee, JJ, JH Lee, YG Ko, SI Hong, JS Lee. 2010a. Prevention of premature senescence requires JNK regulation of Bcl-2 and reactive oxygen species. Oncogene 29:561-575.

Lee, JW, KD Park, JA Im, MY Kim, DC Lee. 2010b. Mitochondrial DNA copy number in peripheral blood is associated with cognitive function in apparently healthy elderly women. Clinica chimica acta; international journal of clinical chemistry 411:592-596.

Li, M, X Fang, DJ Baker, L Guo, X Gao, Z Wei, S Han, JM van Deursen, P Zhang. 2010. The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis. Proceedings of the National Academy of Sciences of the United States of America 107:14188-14193.

Li, Z, J Wu, CJ Deleo. 2006. RNA damage and surveillance under oxidative stress. IUBMB Life 58:581-588.

Liu, HY, E Yehuda-Shnaidman, T Hong, J Han, J Pi, Z Liu, W Cao. 2009. Prolonged exposure to insulin suppresses mitochondrial production in primary hepatocytes. The Journal of biological chemistry 284:14087-14095.

Lu, Y, N Azad, L Wang, AK Iyer, V Castranova, BH Jiang, Y Rojanasakul. 2010. Phosphatidylinositol-3-kinase/akt regulates bleomycin-induced fibroblast proliferation and collagen production. American journal of respiratory cell and molecular biology 42:432-441.

Lynch, SM, SJ Weinstein, J Virtamo, Q Lan, CS Liu, WL Cheng, N Rothman, D Albanes, RZ Stolzenberg-Solomon. 2011. Mitochondrial DNA copy number and pancreatic cancer in the alpha-tocopherol beta-carotene cancer prevention study. Cancer prevention research 4:1912-1919.

Marchler-Bauer, A, S Lu, JB Anderson, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic acids research 39:D225-229.

Matsushima, Y, Y Goto, LS Kaguni. 2010. Mitochondrial Lon protease regulates mitochondrial DNA copy number and transcription by selective degradation of mitochondrial transcription factor A (TFAM). Proceedings of the National Academy of Sciences of the United States of America 107:18410-18415.

McGill, JK, MF Beal. 2006. PGC-1alpha, a new therapeutic target in Huntington's disease? Cell 127:465-468.

Miller, FJ, FL Rosenfeldt, C Zhang, AW Linnane, P Nagley. 2003. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. Nucleic acids research 31:e61.

Murphy, MP. 2009. How mitochondria produce reactive oxygen species. The Biochemical journal 417:1-13.

Nathan, C, A Ding. 2010. Nonresolving inflammation. Cell 140:871-882.

O'Brien, E, AR Rogers, J Beesley, LB Jorde. 1994. Genetic structure of the Utah Mormons: comparison of results based on RFLPs, blood groups, migration matrices, isonymy, and pedigrees. Hum Biol 66:743-759.

Parfait, B, P Rustin, A Munnich, A Rotig. 1998. Co-amplification of nuclear pseudogenes and assessment of heteroplasmy of mitochondrial DNA mutations. Biochemical and biophysical research communications 247:57-59.

Park, JY, PY Wang, T Matsumoto, et al. 2009. p53 improves aerobic exercise capacity and augments skeletal muscle mitochondrial DNA content. Circulation research 105:705-712, 711 p following 712.

Pello, R, MA Martin, V Carelli, et al. 2008. Mitochondrial DNA background modulates the assembly kinetics of OXPHOS complexes in a cellular model of mitochondrial disease. Human molecular genetics 17:4001-4011.

Piantadosi, CA, HB Suliman. 2006. Mitochondrial transcription factor A induction by redox activation of nuclear respiratory factor 1. The Journal of biological chemistry 281:324-333.

Posada, D, TJ Maxwell, AR Templeton. 2005. TreeScan: a bioinformatic application to search for genotype/phenotype associations using haplotype trees. Bioinformatics 21:2130-2132.

Ridge, P, T Maxwell, C Corcoran, M Norton, J Tschanz, E O'Brien, R Kerber, R Cawthon, R Munger, K JSK. 2012. Mitochondrial genomic analysis of late onset Alzheimer's disease

reveals protective haplogroups H6A1A/H6A1B: the Cache County Study on Memory in Aging. PLoS One 7:e45134.

Ruiz-Pesini, E, MT Lott, V Procaccio, JC Poole, MC Brandon, D Mishmar, C Yi, J Kreuziger, P Baldi, DC Wallace. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic acids research 35:D823-828.

Santos, RX, SC Correia, X Wang, G Perry, MA Smith, PI Moreira, X Zhu. 2010. Alzheimer's disease: diverse aspects of mitochondrial malfunctioning. International journal of clinical and experimental pathology 3:570-581.

Saxena, R, PI de Bakker, K Singer, et al. 2006. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. American journal of human genetics 79:54-61.

Schlame, M, JA Towbin, PM Heerdt, R Jehle, S DiMauro, TJ Blanck. 2002. Deficiency of tetralinoleoyl-cardiolipin in Barth syndrome. Annals of neurology 51:634-637.

Sequeira, A, MV Martin, B Rollins, et al. 2012. Mitochondrial mutations and polymorphisms in psychiatric disorders. Frontiers in genetics 3:103.

Shoffner, JM, MT Lott, AM Lezza, P Seibel, SW Ballinger, DC Wallace. 1990. Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA tRNA(Lys) mutation. Cell 61:931-937.

Sitkovsky, MV, D Lukashev, S Apasov, H Kojima, M Koshiba, C Caldwell, A Ohta, M Thiel. 2004. Physiological control of immune response and inflammatory tissue damage by hypoxia-inducible factors and adenosine A2A receptors. Annual review of immunology 22:657-682.

Stadtman, ER, RL Levine. 2003. Free radical-mediated oxidation of free amino acids and amino acid residues in proteins. Amino Acids 25:207-218.

Suomalainen, A, P Isohanni. 2010. Mitochondrial DNA depletion syndromes--many genes, common mechanisms. Neuromuscular disorders : NMD 20:429-437.

Swerdlow, RH, JM Burns, SM Khan. 2010. The Alzheimer's disease mitochondrial cascade hypothesis. J Alzheimers Dis 20 Suppl 2:S265-279.

Swerdlow, RH, SM Khan. 2004. A "mitochondrial cascade hypothesis" for sporadic Alzheimer's disease. Med Hypotheses 63:8-20.

Swerdlow, RH, SM Khan. 2009. The Alzheimer's disease mitochondrial cascade hypothesis: an update. Exp Neurol 218:308-315.

Taylor, RW, AA Morris, M Hutchinson, DM Turnbull. 2002. Leigh disease associated with a novel mitochondrial DNA ND5 mutation. European journal of human genetics : EJHG 10:141-144.

Taylor, SD, H Zhang, JS Eaton, MS Rodeheffer, MA Lebedeva, W O'Rourke T, W Siede, GS Shadel. 2005. The conserved Mec1/Rad53 nuclear checkpoint pathway regulates mitochondrial DNA copy number in Saccharomyces cerevisiae. Molecular biology of the cell 16:3010-3018.

Templeton, AR, T Maxwell, D Posada, JH Stengard, E Boerwinkle, CF Sing. 2005. Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. Genetics 169:441-453.

Tewari, S, JM Santos, RA Kowluru. 2012. Damaged mitochondrial DNA replication system and the development of diabetic retinopathy. Antioxidants & redox signaling 17:492-504.

Thompson, JD, TJ Gibson, DG Higgins. 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics Chapter 2:Unit 2 3.

Thyagarajan, B, R Wang, H Barcelo, WP Koh, JM Yuan. 2012. Mitochondrial copy number is associated with colorectal cancer risk. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.

Trinei, M, I Berniakovich, PG Pelicci, M Giorgio. 2006. Mitochondrial DNA copy number is regulated by cellular proliferation: a role for Ras and p66(Shc). Biochimica et biophysica acta 1757:624-630.

van Oven, M, M Kayser. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human mutation 30:E386-394.

Varlamov, DA, AP Kudin, S Vielhaber, et al. 2002. Metabolic consequences of a novel missense mutation of the mtDNA CO I gene. Human molecular genetics 11:1797-1805.

Vendelbo, MH, KS Nair. 2011. Mitochondrial longevity pathways. Biochim Biophys Acta 1813:634-644.

Vilming, ST, P Dietrichson, MM Isachsen, L Lovvik, A Heiberg. 1986. Late-onset hereditary myopathy with abnormal mitochondria and progressive dementia. Acta neurologica Scandinavica 73:502-506.

Westfall, P, SS Young. 1993. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustments. New York City, NY: Wiley-Interscience.

Ylikallio, E, H Tyynismaa, H Tsutsui, T Ide, A Suomalainen. 2010. High mitochondrial DNA copy number has detrimental effects in mice. Human molecular genetics 19:2695-2705.

Yu, M, Y Wan, Q Zou. 2010. Decreased copy number of mitochondrial DNA in Ewing's sarcoma. Clinica chimica acta; international journal of clinical chemistry 411:679-683.

Yu, M, Y Zhou, Y Shi, L Ning, Y Yang, X Wei, N Zhang, X Hao, R Niu. 2007. Reduced mitochondrial DNA copy number is correlated with tumor progression and prognosis in Chinese breast cancer patients. IUBMB life 59:450-457.

Zaid, A, R Li, K Luciakova, P Barath, S Nery, BD Nelson. 1999. On the role of the general transcription factor Sp1 in the activation and repression of diverse mammalian oxidative phosphorylation genes. Journal of bioenergetics and biomembranes 31:129-135.

Zou, T, W Yang, Z Hou, J Yang. 2010. Homocysteine enhances cell proliferation in vascular smooth muscle cells: role of p38 MAPK and p47phox. Acta biochimica et biophysica Sinica 42:908-915.

Figure Legends

Figure 1. Box plot comparing mitochondrial copy number between different clades. The grey dots represent the mitochondrial copy number for each member of the representative groups. The top and bottom of the boxes correspond to the 75^{th} and 25^{th} percentiles, respectively, and the line through the box is the median mitochondrial copy number for the group. The whiskers correspond to the maximum and minimum mitochondrial copy numbers for the group. Three different groups are represented here: the clades defined by branches 124 and 50, and a group containing all other individuals in the dataset. The y-axis is the mitochondrial copy number. The reported p-values are corrected.

Figure 2. Significant branches. This is a subset of the full haplotype network (Supplementary Figure 1), focused on the two significant clades defined by branches 124 and 121, which are labeled here. The blue ovals represent haplotypes observed in our dataset, and the smaller white circles are unobserved haplotypes. Only the variants that define branches 124 and 121 are labeled.

Figure 3. Multiple sequence alignment of COXIII. Position 21 in the alignment corresponds to position 154 in COXIII. Background colors correspond to the level of conservation of that position in the alignment. The darker the shade of red, the higher the conservation.

Figure 4. Pathways between COXIII and known regulators of mtDNA copy number. Here we show all the known pathways between COXIII and the different genes known to regulate or modify mtDNA copy number.

Figure 5. Multiple sequence alignment of ND2. Position 21 in the alignment corresponds to position 270 in ND2. Background colors are as described in Figure 2.

Figure 6. Multiple sequence alignment of COXI. Position 21 in the alignment corresponds to position 196 in COXI. Background colors are as described in Figure 2.

Figure 7. Pathways between COXI and known regulators of mtDNA copy number. Here we show all the known pathways between COXI and the different genes known to regulate or modify mtDNA copy number.

Table 1, Chapter 3. Distribution of major mtDNA haplogroups/clusters

Major Haplogroup	Number	Ethnicities (Herrnstadt et al. 2002a; Ruiz-Pesini et al. 2007)
H	424	European
U	147	European
T	121	European
J	99	European
K	95	European
V	34	European
I	21	European
W	20	European
HV	18	European
X	8	European
C	5	Asian
L	4	African
Missing[1]	11	

Here we report the number of individuals belonging to each of the major haplogroups represented in our dataset along with case-control status.

[1]Missing major haplogroup

Table 2, Chapter 3. Demographic information for significant contrasts

	Individuals/Missing	p-value[1]		p-value[2]		Age	Male/Female	Mean copy #
		Nominal	Corrected	Nominal	Corrected			
Whole network	1007/193	N/A	N/A	N/A	N/A	75.6	442/565	2.69
Branch 124	17/3	0	6.0e-4	0	8.0e-4	75.2	9/8	3.81
Branch 121	10/1	0	0.002	1.0e-4	0.0043	76.3	4/6	4.01
Branch 50	15/3	2.0e-4	0.017	2.0e-4	0.015	78.4	7/8	3.64

Here we report demographic information for each of the significant contrasts and for all the individuals in the dataset. The clade represented by Branch 121 is wholly contained within Branch 124, so these two contrasts represent a single effect. Branches 124 and 50 represent separate effects. Missing refers to the number of individuals for whom we have no mtDNA copy number measurement.

[1]p-values were calculated controlling only for the other significant branches

[2]p-values were calculated using age, gender, mtDNA copy number family risk score, and the other significant effects as covariates

Table 3, Chapter 3. Defining variants for the three significant contrasts

Branch	Nucleotide Change	Amino Acid Change	Gene
Branch 124	m.9667A>G	p.Asn154Ser	Cytochrome C Oxidase 3
Branch 121	m.12582A>G	p.Leu82Leu	NADH Dehydrogenase 5
	m.12879T>C	p.Gly181Gly	NADH Dehydrogenase 5
Branch 50	m.5277T>C	p.Phe270Leu	NADH Dehydrogenase 2
	m.5426T>C	p.His319His	NADH Dehydrogenase 2
	m.6489C>A	p.Leu196Ile	Cytochrome C Oxidase 1
	m.8270C>T	N/A	Intergenic
	m.del8281-8289	N/A	Intergenic
	m.14458C>T	p.Ala72Ala	NADH Dehydrogenase 6
	m.15028C>A	p.Leu94Leu	Cytochrome B
	m.15043G>A	p.Gly99Gly	Cytochrome B

There were three significant contrasts in our dataset, two of which, 124 and 121, which represent a single effect. One or more sequence features define each of the branches, and each is listed here with the resulting protein change, and the gene the feature is located in.

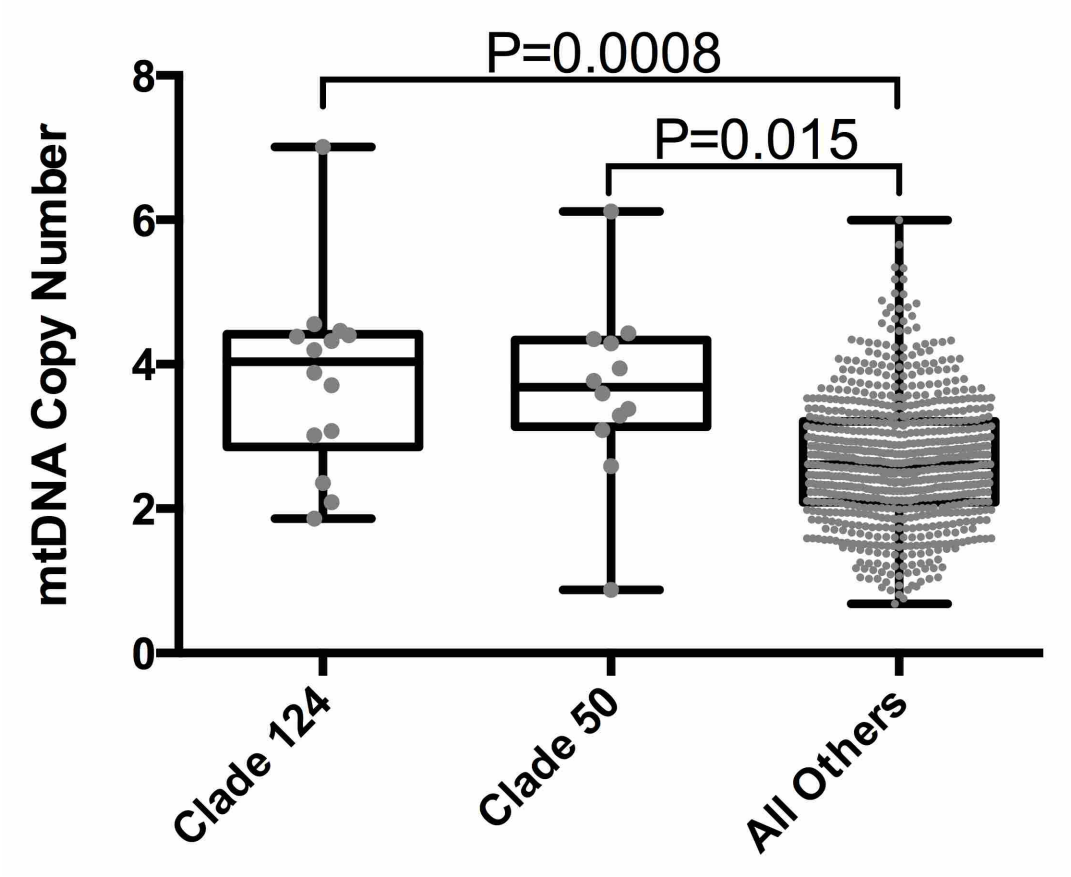


Figure 1, Chapter 3. Box plot comparing mitochondrial copy number between different clades. The grey dots represent the mitochondrial copy number for each member of the representative groups. The top and bottom of the boxes correspond to the 75th and 25th percentiles, respectively, and the line through the box is the median mitochondrial copy number for the group. The whiskers correspond to the maximum and minimum mitochondrial copy numbers for the group. Three different groups are represented here: the clades defined by branches 124 and 50, and a group containing all other individuals in the dataset. The y-axis is the mitochondrial copy number. The reported p-values are corrected.

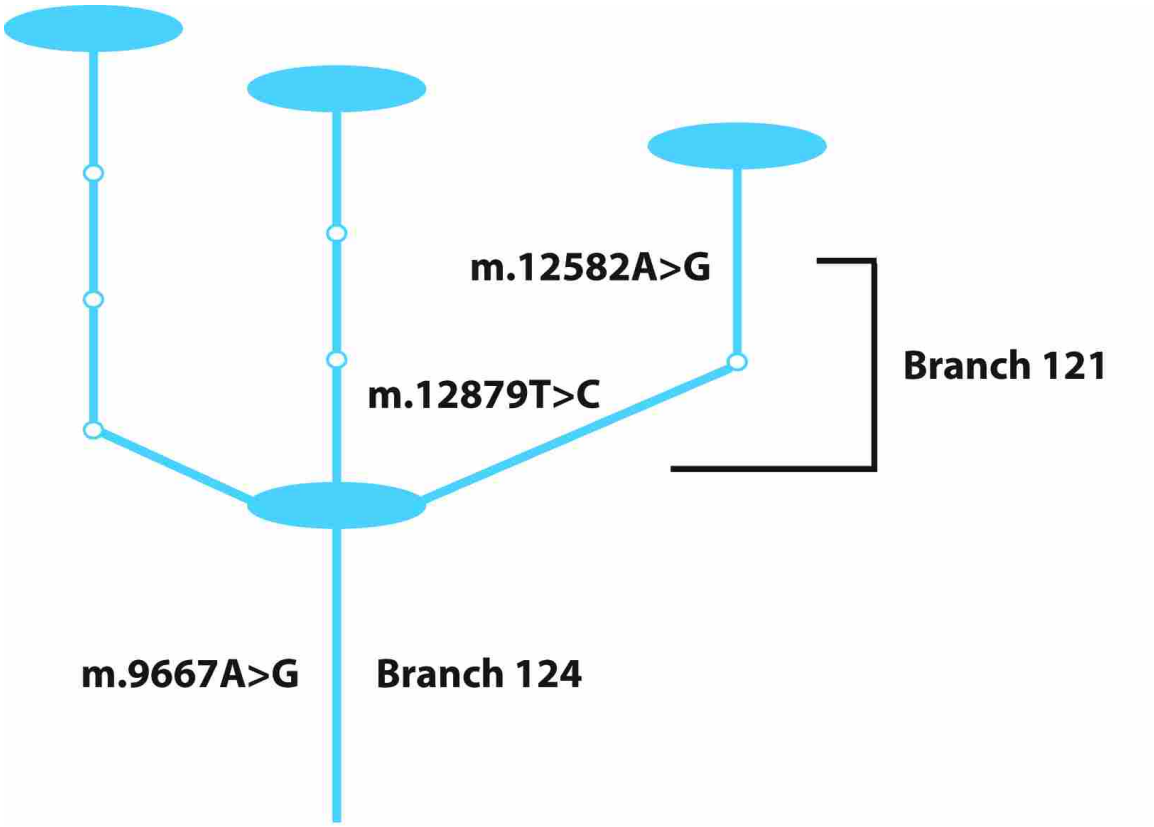


Figure 2, Chapter 3. Significant branches. This is a subset of the full haplotype network (Supplementary Figure 1), focused on the two significant clades defined by branches 124 and 121, which are labeled here. The blue ovals represent haplotypes observed in our dataset, and the smaller white circles are unobserved haplotypes. Only the variants that define branches 124 and 121 are labeled.

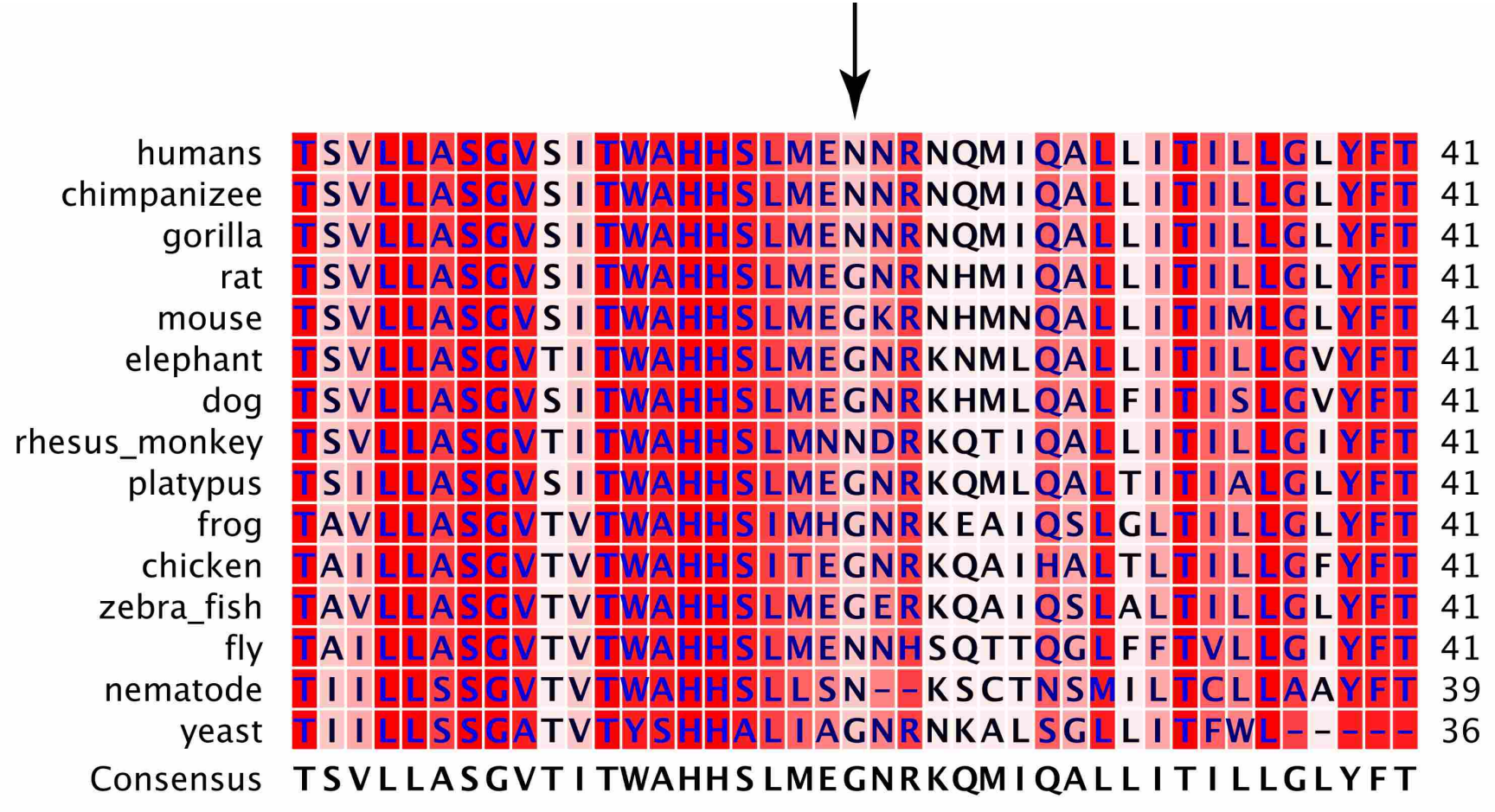


Figure 3, Chapter 3. Multiple sequence alignment of COXIII. Position 21 in the alignment corresponds to position 154 in COXIII. Background colors correspond to the level of conservation of that position in the alignment. The darker the shade of red, the higher the conservation.

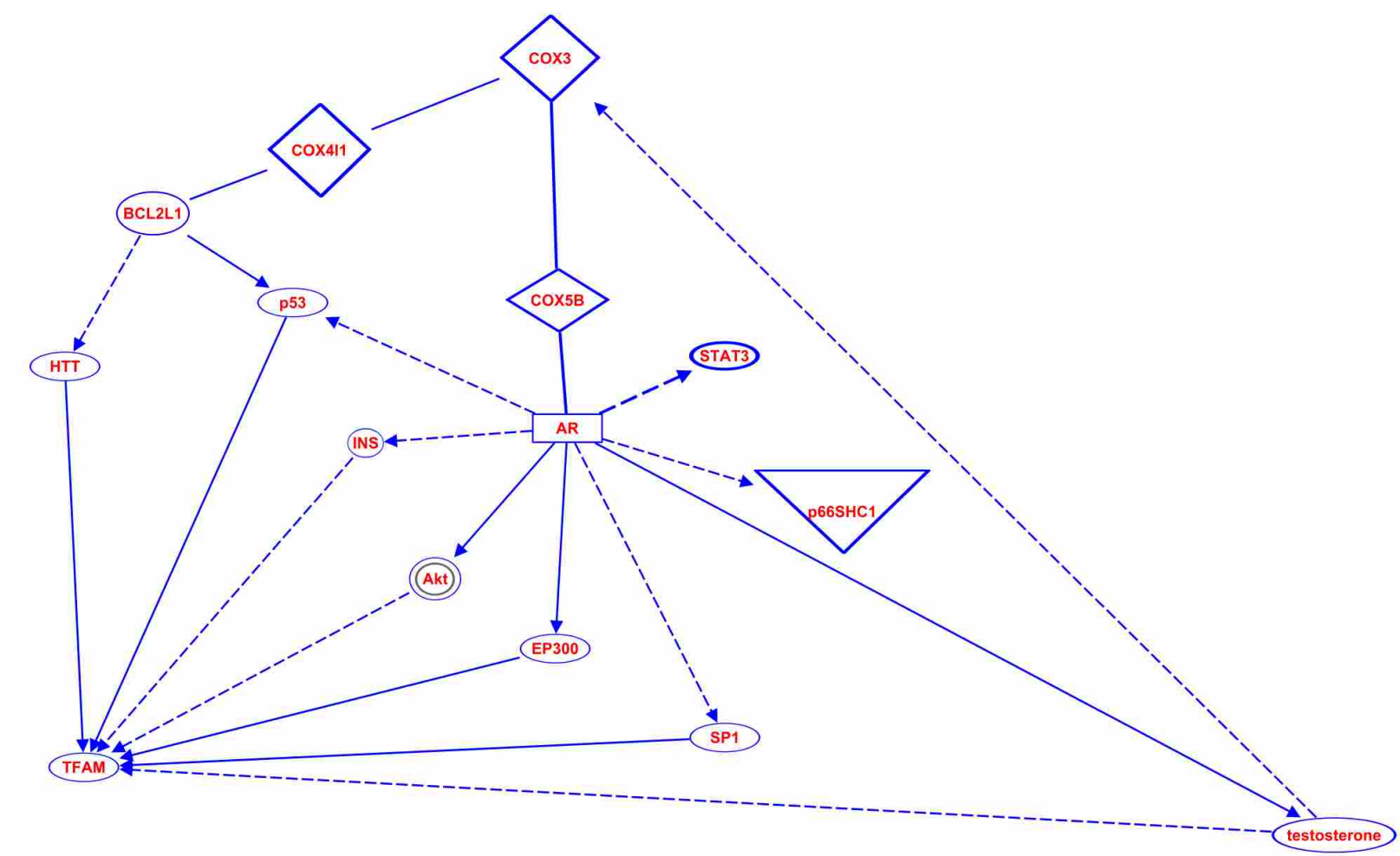


Figure 4, Chapter 3. Pathways between COXIII and known regulators of mtDNA copy number. Here we show all the known pathways between COXIII and the different genes known to regulate or modify mtDNA copy number.

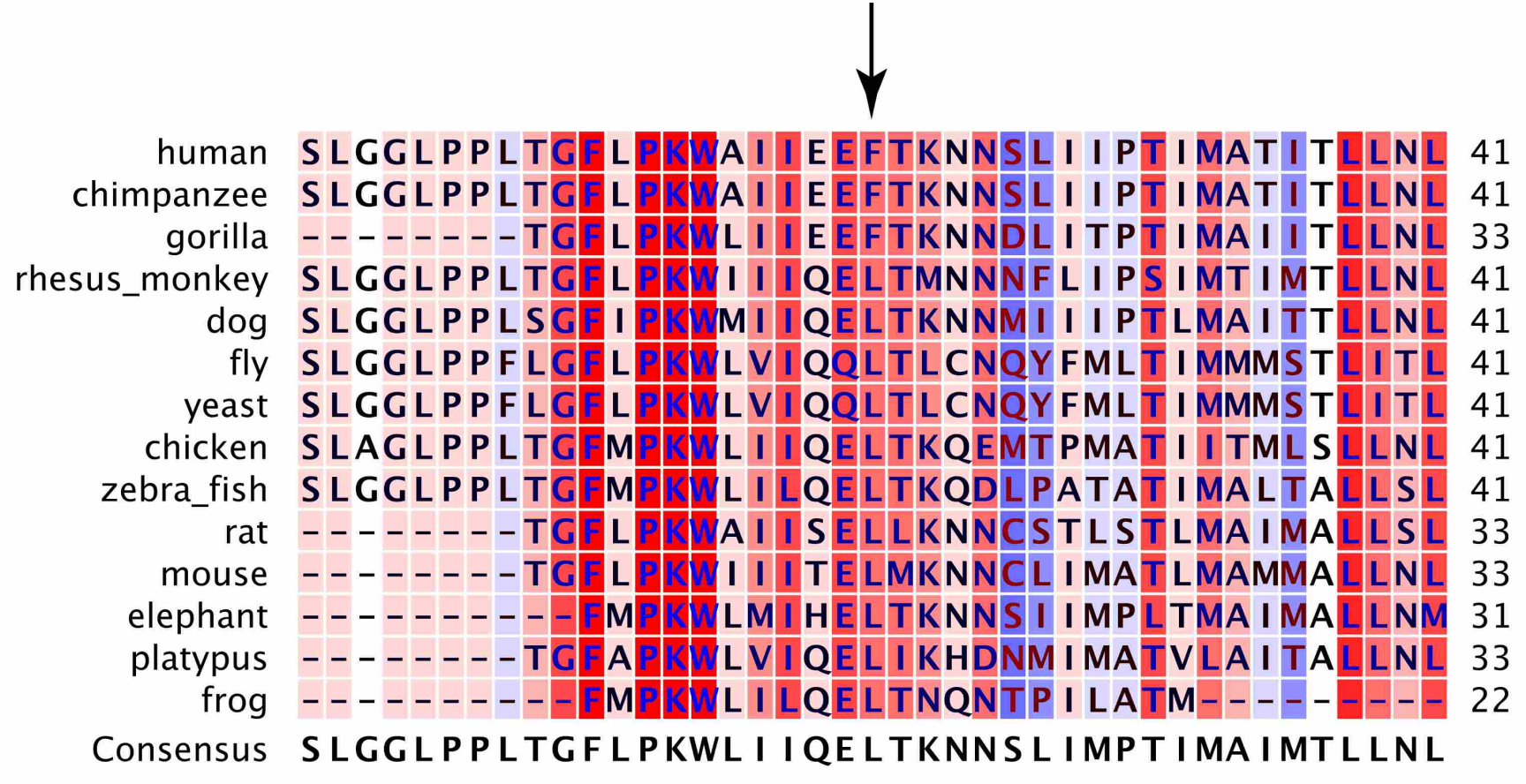


Figure 5, Chapter 3. Multiple sequence alignment of ND2. Position 21 in the alignment corresponds to position 270 in ND2. Background colors are as described in Figure 2.

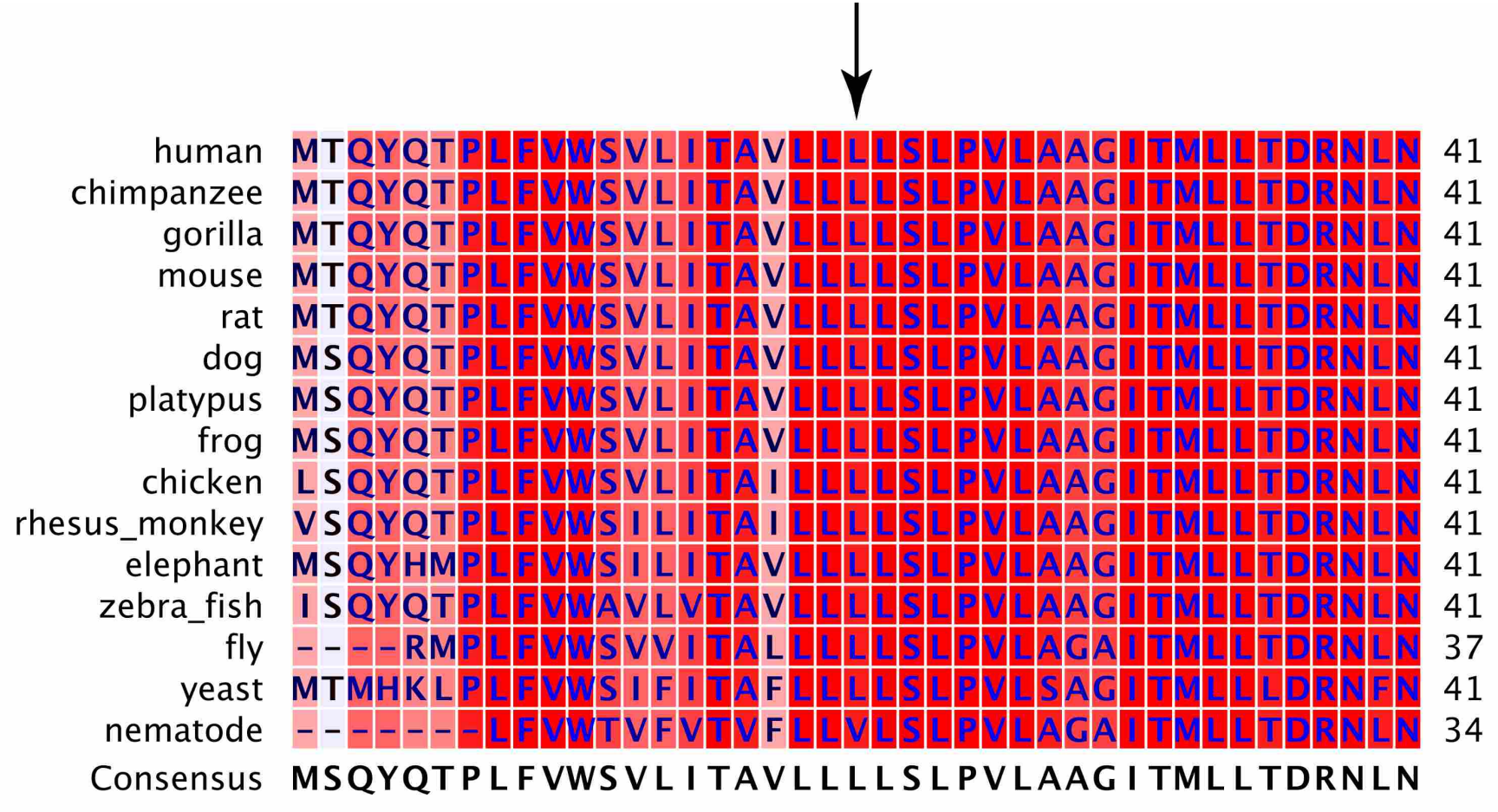


Figure 6, Chapter 3. Multiple sequence alignment of COXI. Position 21 in the alignment corresponds to position 196 in COXI. Background colors are as described in Figure 2.

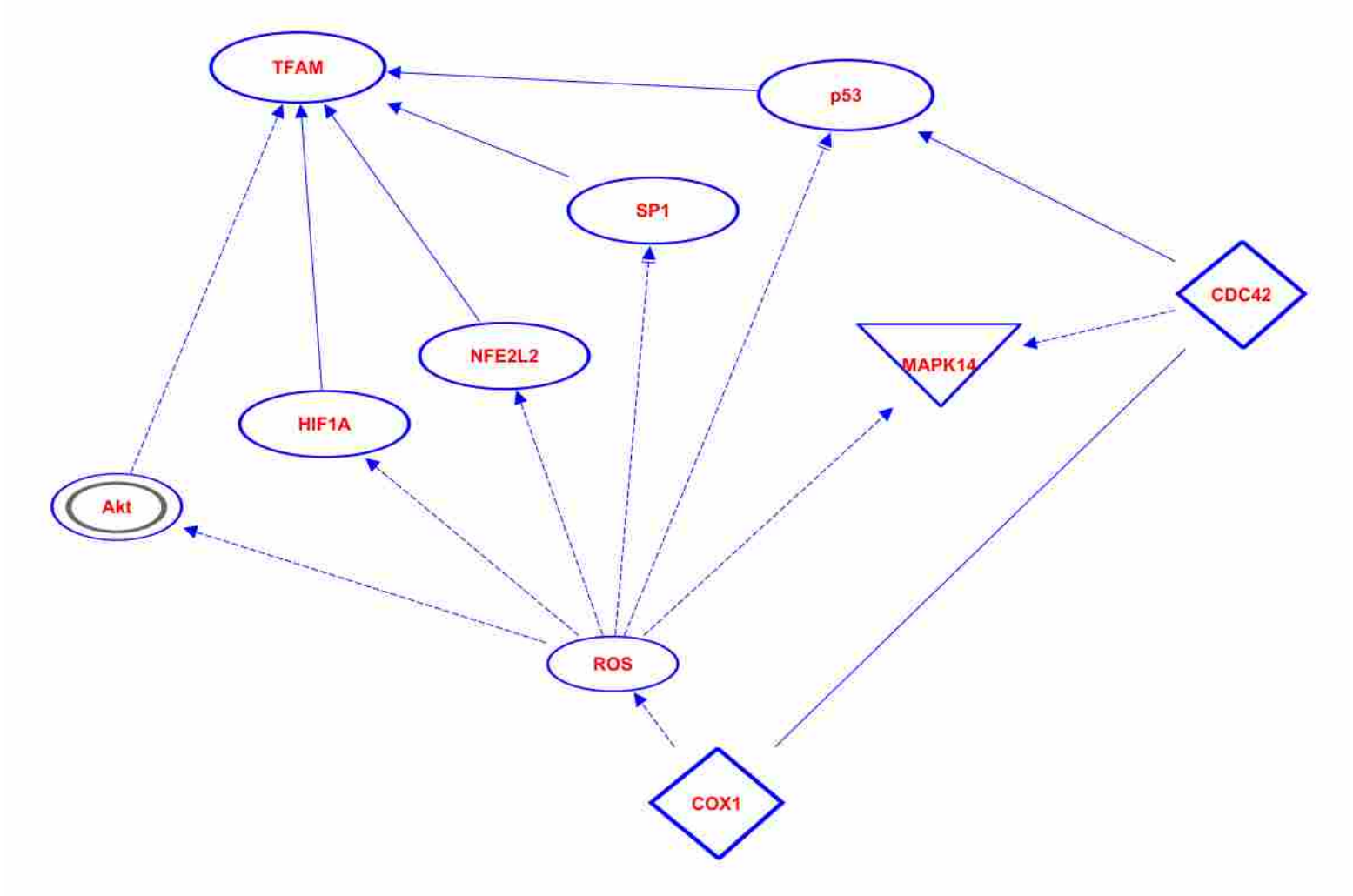


Figure 7, Chapter 3. Pathways between COXI and known regulators of mtDNA copy number. Here we show all the known pathways between COXI and the different genes known to regulate or modify mtDNA copy number.

Manuscript submitted to Neurobiology of Aging

Chapter 4

Mitochondrial haplotypes associated with biomarkers for Alzheimer's disease

Perry G. Ridge[1,2], Andre Koop[c], Taylor J. Maxwell[4], Russell H. Swerdlow[3], John S.K. Kauwe[1], Robyn A. Honea;[3] For the Alzheimer's Disease Neuroimaging Initiative[**]

[1] Department of Biology, Brigham Young University, Provo, UT

[2] ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT

[3] KU Alzheimer's Disease Center, Department of Neurology, University of Kansas School of Medicine, Kansas City, KS

[4] Human Genetics Center, University of Texas School of Public Health, Houston, TX

**Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Abstract

Various studies have suggested that the mitochondrial genome plays a role in late-onset Alzheimer's disease, although results are mixed. We used an endophenotype-based approach to further characterize mitochondrial genetic variation and its relationship to risk markers for Alzheimer's disease. We analyzed longitudinal data from non-demented, mild cognitive impairment, and late-onset Alzheimer's disease participants in the Alzheimer's Disease Neuroimaging Initiative with genetic, brain imaging, and behavioral data. We assessed the relationship of structural MRI and cognitive biomarkers with mitochondrial genome variation using TreeScanning, a haplotype-based approach that concentrates statistical power by analyzing evolutionarily meaningful groups (or clades) of haplotypes together for association with a phenotype. Four clades were associated with three different endophenotypes: whole brain volume, percent change in temporal pole thickness, and left hippocampal atrophy over two years. This is the first study of its kind to identify mitochondrial variation associated with brain imaging endophenotypes of Alzheimer's disease. Our results provide additional evidence that the mitochondrial genome plays a role in risk for Alzheimer's disease.

Search Terms

Alzheimer's disease, endophenotypes, mitochondrial genetics, mitochondrial genome, haplotype analysis

Introduction

While late-onset Alzheimer's disease (AD) is highly heritable, much of the genetic variance associated with late-onset AD is unknown. Studies of maternal inheritance, mitochondrial genome (mtDNA) mutations, and cytochrome oxidase deficits in AD provide converging evidence for the role of mitochondria in risk for Alzheimer's disease (Bonilla et al. 1999; Orth, Schapira 2001; Cardoso et al. 2004; Mosconi et al. 2009; Honea et al. 2010; Mosconi et al. 2010; Honea et al. 2012). A growing number of studies have reported associations between mitochondrial haplogroups, or specific mitochondrial DNA single nucleotide polymorphisms (SNPs) with risk for AD, however the resulting literature is mixed, either because of differences in the populations studied, sample sizes, or variations in the AD groups studied (reviewed in (Hudson et al. 2012a)).

Recently, investigators have developed neuroimaging methods to clarify the role of genes in AD pathology (Schuit et al. 2001; Podewils et al. 2005; Rovio et al. 2005; Etnier et al. 2007). The use of endophenotypes, or disease markers, has proved to be sensitive in delineating genetic effects on susceptibility to AD, as the imaging markers provide a more direct measure of the impact of the gene at the level of neuroanatomy. Volumetric magnetic resonance imaging (MRI) has been used to characterize accelerated rates of regional and whole brain atrophy and predict progression to AD before onset (Desikan et al. ; Jack et al. 2004; Desikan et al. 2009). While a host of studies have used imaging markers to test for relationships between nuclear genes and AD risk (Hibar et al. ; Hibar et al. ; Silver et al. ; Stein et al. ; Stein et al. ; Schuff et al. 2009; Biffi et al. 2010), none have used this approach to study mitochondrial genes. The current study

sought to clarify the relationship between mtDNA genetic variation and neurodegeneration, a known risk marker for AD.

Materials and Methods

Subjects

We used publically available data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu), which has undergone rigorous quality control, in the preparation of this article. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5- year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three

protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see www.adni-info.org.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). Data for the present analysis were downloaded from the ADNI web site in December 2011.

The study reported here involved 821 subjects who had genetic and cognitive data, and MRI scans at least at baseline, and some at 24 months. Of those subjects, 103 subjects were excluded from the FreeSurfer dataset (see below) for technical reasons, such as major hardware upgrades during the study (at two sites), miscalibration of image resolution, excess movement, or failure of one or more automatic processing methods. The main demographic and clinical data are summarized in Table 1. Consent was obtained according to the Declaration of Helsinki. The Ethical Committees of each Institution, in which the work was performed, approved the study.

Genotyping

Samples were genotyped using the Illumina Human610-Quad BeadChip. The Illumina Human610-Quad BeadChip consists of 550,000 polymorphic sites (SNPs), plus an additional 60,000 genetic markers including 138 mitochondrial DNA sequence polymorphic sites. The genotyping procedure for the mtDNA is described elsewhere (Potkin et al. 2009). The 138 mitochondrial SNPs are based on the AF347015.1 mtDNA reference sequence, one of 53

African sequences deposited in Genbank (Ingman, Gyllensten 2001). We first mapped the 138 mtDNA SNPs to the revised Cambridge Reference Sequence (rCRS). The 138 SNPs consist of 21 noncoding, 91 protein-coding, 4 rRNA, 11 tRNA and 1 termination sites.

ADNI Phenotypic Data

We used hippocampal and whole brain volume (WBV) data from the Anders Dale Lab (UCSD) available as part of the ADNI secondary imaging data downloads, from which we also calculated rates of hippocampal atrophy and WBV atrophy using two-year data. Details on their neuroimaging processing methods are published elsewhere (Holland et al. 2009). For normalization calculations we used total-intracranial volumes calculated automatically from FreeSurfer (Fischl et al. 1999), also available from the ADNI secondary imaging data downloads. We also chose to use several imaging biomarkers that have been elucidated in the recent literature as being sensitive to disease progression and AD genetic risk (Biffi et al. 2010). Our variables were data from the recent FreeSurfer (2012) upload on the ADNI website, and included baseline and longitudinal (2-year) percent change scores derived from cortical thickness measures in the entorhinal cortex, parahippocampal cortex, and the temporal pole. In addition, we computed a hippocampal occupancy (HOC) score as another estimate of medial temporal lobe atrophy, which may aid in differentiation of individuals with congenitally small hippocampi from those with small hippocampi due to a degenerative disorder (Heister et al. 2011). It has been shown to have higher discriminative and predictive accuracy than the standard hippocampal volume measure (Heister et al. 2011). The HOC was computed as the ratio of hippocampal volume to the sum of the hippocampal and interior lateral ventricular volumes in each hemisphere separately. Right and left HOC scores were then averaged. As a marker for disease-

related cognitive change, we used the ADAS-Cog total 11 score at baseline and 2-year change scores for longitudinal data.

Haplotype Network Estimation

We used TCS (Clement, Posada, Crandall 2000) to estimate a haplotype network using the 138 genotyped mtDNA SNPs from 821 individuals. As part of network construction, TCS collapsed all the individuals into common haplotypes. From 821 individuals there were 196 different haplotypes. The initial network contained ambiguities, and we were able to resolve all but one of these ambiguities using the following criteria: first, observed haplotypes are more likely to be descended from internal (or more ancient) haplotypes than external (or more recent) haplotypes, and second, haplotypes are more likely to descend from higher frequency haplotypes. The remaining ambiguity was factored into our analyses. The complete network can be seen in Supplementary Figure 1.

Association Analyses

Genotype-phenotype associations were evaluated using an evolution-based method known as TreeScanning (Posada, Maxwell, Templeton 2005; Templeton et al. 2005) that makes use of haplotype networks. Haplotype networks provide a framework from which to select evolutionarily related haplotypes to pool together for comparison. Additional details about the application of TreeScanning to a similar dataset can be found in Ridge et al. (Ridge et al. 2012). The null hypothesis of TreeScanning is that the phenotype does not differ in distribution across the genotypes derived from allelic classes defined by the branches of the haplotype network. Each branch partitions the haplotypes into bi-allelic pools from which genotypes are constructed

and treated as a separate test. Because we have multiple tests that are correlated we obtained multiple-test corrected p-values by a permutation analog of the sequential step-down Bonferroni (Westfall, Young 1993) with 10,000 permutations. If significant branches are found in the first round of TreeScanning, a second round of TreeScanning that can detect phenotypic heterogeneity within the allelic classes of the significant branch is performed. This is accomplished by creating a three-allele system and using conditional permutations that hold one of the alleles constant while subdividing the other class into two alleles (Templeton et al. 2005). We included age, gender, and apolipoprotein E (APOE) ε4 status as covariates for each analysis. Only analyses with at least five individuals in each group were performed and corrected p-values of 0.05 or less were considered significant. A detailed description of the methods is outlined in Ridge et al. (Ridge et al. 2012).

Mitochondrial Haplotype Assignment

In this paper we refer to haplotypes observed in our data based on the 138 genotyped mtDNA SNPs as haplotypes. Haplogroups, mitochondrial haplogroups, major mitochondrial haplogroups, and sub-haplogroups refer to known mitochondrial haplogroups (i.e. H, L, N, etc.) or their sub-haplogroups (i.e. H1A, K1, etc.).

The 138 genotyped mtDNA SNPs were insufficient to use available software for mitochondrial haplogroup assignment; therefore, we examined each of the significant contrasts in our dataset manually to identify which mitochondrial haplogroups our observed haplotypes (in only the significant clades) belong to. We used the following approach to identify haplogroups: 1) We identified all the variants common to individuals in the significant clades (in our dataset); 2)

Using phylotree.org (van Oven, Kayser 2009), we identified the different mitochondrial haplogroups a particular variant was associated with; 3) We looked for common haplogroups among all the variants in a particular haplotype (for example, all the variants are found in haplogroup U), or for smaller granularity in shared haplotypes as the network is descended (for example, an internal node is haplogroup U and one of its children is haplogroup U5); 4) We searched the primary literature for specific variants to obtain additional information about the mitochondrial haplogroups with which it is associated. In some cases we were able to narrow potential haplogroups to a probable group, while in other cases there were multiple plausible haplogroups. In the instances where we identified a likely haplogroup we report it, otherwise we report no haplogroup assignment. It should be noted that our haplogroup assignments are based on 138 SNPs and are the best estimates possible with the available data.

Results

We used TreeScanning to group the observed haplotypes in our dataset into evolutionarily meaningful groups (see methods), and tested them for association with 16 different imaging phenotypes (Supplementary Table 1) carefully selected to minimize correlations between phenotypes. Pairwise correlation coefficients for the 16 phenotypes are reported in Supplementary Table 2. We found evidence for an association between four mitochondrial clades and three of the brain imaging endophenotypes; whole brain volume, percent change in temporal pole thickness over two years, and left hippocampal atrophy over two years.

Mitochondrial Haplotypes Associated with Whole Brain Volume

The clade represented by branch 155 (Figure 1) was significantly associated (corrected p-value 0.025) with whole brain volume after correcting for age, gender, and both with and without correction for APOE ε4 status (Table 2). The clade defined by branch 155 is a large clade with 16 distinct haplotypes and 18 individuals. Individuals in this clade had lower whole brain volume (919170 versus 989295 in the rest of the network). Major haplotype groups were not distinguishable using the available SNP markers for the haplotypes in this clade (Supplementary Table 3).

Branch 184 is immediately adjacent to branch 155 (Figure 1) and the clade defined by branch 155 is contained within the clade defined by branch 184. Before multiple test correction branch 184 was associated with significantly lower levels of whole brain volume (nominal p-value 0.018, Table 2). These two clades are correlated and represent a single effect. In addition to whole brain volume, branch 184 was also associated (nominal p-value 0.0036, Table 3) with lower levels of left hippocampal atrophy over two years. The clade defined by branch 184 had, on average, 2.98% left hippocampal atrophy over two years compared to 1.18% in the rest of the dataset. There were 17 haplotypes in this clade; however, we only had left hippocampal atrophy measurements for five individuals. Additional individuals could increase the significance of the association.

Mitochondrial Haplotypes Associated with Percent Change in Temporal Pole Thickness Over Two Years

Two different clades (Table 4) were associated with higher percent change in temporal pole thickness over two years. First, the individuals in the clade defined by branch 199 (Figure 2) had an average 13.22% loss in temporal pole thickness over two years compared to 4.27% loss in the rest of the dataset (corrected p-value 0.0343). This clade consisted of three different haplotypes and five individuals. Based on the available SNPs we were able to identify the mitochondrial haplogroups (U5B1 and U5B1B2) for two of the three haplotypes in this clade. Branch 199 corresponds to m.5656A>G. This SNP has been reported to be specific to U5B (Finnila, Lehtonen, Majamaa 2001; Herrnstadt et al. 2002) or U5B1 and several other sub-haplogroups (van Oven, Kayser 2009). Individuals belonging to the second haplogroup, U5B1B2, also have m.217T>C. A d-loop variant, m.217T>C has been reported as being specific to U5B (Finnila, Lehtonen, Majamaa 2001) or specific to U5B1B2 (van Oven, Kayser 2009). Additionally, moving one branch up the network is m.3197T>C, a variant specific to U5 (Herrnstadt et al. 2002). While sources exist that suggest a wider distribution for these two variants, an analysis of each variant corresponding to each branch of the network as the network is descended to branch 199 clearly shows a trend towards sub-haplogroups of U (for complete details see Supplementary Table 4).

The second clade associated with higher percent loss of temporal pole thickness over two years (corrected p-value 0.0441) is defined by branch 9 (Figure 3). The 10 individuals in this clade had an average loss of temporal pole thickness over two years of 10% in contrast to individuals in the rest of the dataset who had an average loss in temporal pole thickness over two years of 4.079%.

This clade consists of two different haplotypes corresponding to major haplogroups K1A1B and K1A1B2A1. Assignment to these haplogroups was based on several SNPs. m.10550A>G is specific to K (Herrnstadt et al. 2002; van Oven, Kayser 2009), m.1189T>C is one of the defining variants of K1 (Herrnstadt et al. 2002), and m.15924A>G appears in several haplogroups, one of which is K1A1B. All individuals in this clade had these three SNPs. K1A1B2A1 individuals also have m.15758A>G, a variant only observed in K1A1B2A1 (van Oven, Kayser 2009). After examining the possible haplogroups for the individuals in this clade, these K sub-haplogroups are the only ones that make sense given the available SNP data (Supplementary Table 5).

In addition to percent change in temporal pole thickness, these two branches were also nominally associated with several other phenotypes. Branch 199 was nominally associated with lower levels of hippocampal occupancy (at their 2 year scan) (p-value 0.0027), baseline hippocampal occupancy score (p-value 0.0055), percent change in entorhinal cortex thickness over two years (p-value 0.0073), bilateral entorhinal cortex thickness (baseline) (p-value 0.0078), temporal pole thickness (baseline) (p-value 0.0216), and percent change in HOC over two years (p-value 0.0339). Branch 9 was nominally associated with percent change in HOC over two years (p-value 0.0164), and percent change in entorhinal cortex thickness over two years (p-value 0.02).

Discussion

Here we present evidence of association between evolutionarily related groups of haplotypes and markers of AD-related neurodegeneration. We identified groups associated with whole brain volume, temporal pole thickness, and left hippocampal atrophy over two years. Association studies of mitochondrial genes and AD risk have identified a variety of mtDNA haplogroups or

polymorphisms that may influence AD risk (Table 5). Haplogroups are defined by ancestral polymorphisms that are continent-specific, and nine primary mitochondrial haplogroups have been identified in European populations (Wallace 1994). Two of these haplogroups, specifically H and U (also UK), have been most often implicated in risk for AD (Fesahat et al. 2007; Santoro et al. 2010; Coskun et al. 2011) although a sub-clade of H has also been associated with protection from AD (Ridge et al. 2012). Our data confirm and extend studies showing that clades in haplogroups U and K may be involved in AD risk. We found that individuals from two clades belonging to either U5B1 or U5B1B2 (in the first clade, branch 199), or K1A1B or K1A1B2A1 (in the second clade, branch 9) had grater rates of temporal pole atrophy, an endophenotype of AD risk (Frisoni et al. 2009; Stein et al. 2010). Each of these clades was also associated with several other endophenotypes, and point to specific mitochondrial haplogroups controlling multiple facets of brain physiology.

We used several brain imaging endophenotypes commonly used in Alzheimer's disease to characterize disease progression, and found that temporal pole atrophy and left hippocampal atrophy were most significantly associated with mitochondrial genetic variation that has been implicated in risk for AD. The hippocampal formation is a key structure involved in learning and memory (Burgess, Maguire, O'Keefe 2002), and is highly heritable in humans (Peper et al. 2007; Stein et al. 2012). It has thus become an important phenotype for the measurement of Alzheimer's disease risk and progression, specifically, measurements of change in volume and density over time, as it is one of the first structures to degenerate (Braak, Braak 1991). Change in hippocampal volume is not only associated with risk for AD, but may also be mediated by risk genes for Alzheimer's disease such as APOE (reviewed by (Reiman 2007)), as well as affected

by risk factors such as a maternal family history of AD (Andrawis et al. 2012). In our study, the impact of mitochondrial genetic variation on hippocampal atrophy is independent of APOE ε4 status. The temporal pole, or rostralmost portion of the temporal lobe, is also known to be a location of pathology of AD (Arnold, Hyman, Van Hoesen 1994). Lower temporal lobe volume has been identified as a biological marker and risk factor for MCI and AD (Hua et al. 2008). Temporal lobe volume has also been strongly associated with a single-nucleotide polymorphism in the gene GRIN2B, which has been implicated in learning and memory as well as characteristic features of AD neurodegeneration (Stein et al. 2010). Moreover, temporal pole thickness (at baseline scan) was significantly related to genetic variation in several AD risk genes, including APOE, CNTN5 (rs10501927), and BIN1 (rs7161528) in a recent study of ADNI data (Biffi et al. 2010). Thus, our data adds to the growing literature characterizing genetic involvement in AD-related brain change by pointing towards additional mitochondrial genetic variation that may be contributing to neurodegeneration.

In addition to characterizing the impact of risk genes on brain imaging phenotypes of AD (Honea et al. 2009), we recently found that individuals with a maternal family history of AD have reductions in gray matter volume in AD-vulnerable brain regions at baseline, and that these same healthy individuals have progressive gray matter volume reductions in select AD-vulnerable brain regions over two years, independent of APOE ε4 status (Honea et al. 2010; Honea et al. 2011). In addition, we recently reported that individuals within the ADNI dataset with a maternal history of dementia had increased AD biomarkers, including increased global PiB uptake, lower cerebrospinal fluid (CSF) amyloid-β and a higher tau/amyloid-β ratio than individuals with a paternal, or no family history of AD (Honea et al. 2012). These data complement other studies of

maternal inheritance of AD, mtDNA mutations, and cytochrome oxidase deficits in AD (Parker, Filley, Parks 1990; Bonilla et al. 1999; Orth, Schapira 2001; Cardoso et al. 2004; Mosconi et al. 2009; Honea et al. 2010; Mosconi et al. 2010). The possibility mtDNA might differ between AD case and control subjects is further supported by studies of cytoplasmic hybrids (cybrids) (Swerdlow et al. 1997; Swerdlow 2007). Furthermore, AD-related increases in mtDNA mutations are found in brain regions first affected by the disease process (Hirai et al. 2001). While genetic variation in mtDNA is not fully responsible for transmission of AD, our current data adds to a growing literature pointing towards involvement of mtDNA in AD-related neurodegeneration and AD-risk, and possibly explains why individuals with a maternal family history of AD have a significantly higher predisposition towards a variety of AD biomarkers, including structural atrophy, CSF markers, and higher levels of amyloid-β burden.

Our analyses were by limited by the relatively small number of genotyped SNPs. Based on available genotype data we were able to identify likely mitochondrial haplogroups for the majority of the haplotypes in significant clades; however, without complete sequence data we cannot unambiguously assign haplotypes to major haplogroups or, in some cases, even assign probable haplogroups. Besides not being able to assign individuals to mitochondrial haplogroups, without complete mtDNA sequencing it is impossible to identify likely functional variants, whereas with full mtDNA sequencing and analysis with TreeScanning, which "cuts" one branch of the haplotype network at a time, we would know which variants are most likely to explain the observed phenotypic differences. Another limitation is that the clades associated with significant changes in imaging phenotypes were comprised of limited numbers of subjects. Although we used corrected p-values and stringent permutation testing, and our data appear to

complement a pattern already in the literature, it will be important to replicate these findings in a larger sample with imaging and mitochondrial and genetics data. Finally, it is also important to note that the imaging phenotypes we studied may not only point to risk for AD, but also denote accelerated aging effects that may be more general in nature and not specific to AD.

In conclusion, using 138 genotyped SNPs in the ADNI dataset, we estimated a haplotype network and used an evolution-based approach, TreeScanning, to look for haplotypes or clades associated with 16 different imaging phenotypes. We found clades associated with significant changes in three phenotypes: whole brain volume, temporal pole thickness, and left hippocampal atrophy over two years. In addition to these three phenotypes these same clades were nominally associated with additional phenotypes, providing suggestive evidence of clades with mtDNA variants driving multiple physiological changes in several different regions of the brain related to AD. AD is a heterogenous disease and these data provide additional evidence of a role for mtDNA in the risk of AD. Future studies should focus on denser SNP genotyping, preferably whole mtDNA sequencing, and be performed in families showing evidence for a maternal transmission of AD.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Amorfix Life Sciences Ltd.;

AstraZeneca; Bayer HealthCare; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514. JSKK is supported by the National Institutes of Health (R01AG042611), the Alzheimer's Association (MNIRG-11-205368) and the Brigham Young University Gerontology Program. RAH is supported by NIH grants K01 AG035042 and P30 AG010129. RHS is supported by the University of Kansas Alzheimer's Disease Center (P30AG035982).

Disclosure statement

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can

be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. The authors report no disclosures.

References

Andrawis, JP, KS Hwang, AE Green, J Kotlerman, D Elashoff, JH Morra, JL Cummings, AW Toga, PM Thompson, LG Apostolova. 2012. Effects of ApoE4 and maternal history of dementia on hippocampal atrophy. Neurobiology of aging 33:856-866.

Arnold, SE, BT Hyman, GW Van Hoesen. 1994. Neuropathologic changes of the temporal pole in Alzheimer's disease and Pick's disease. Archives of neurology 51:145-150.

Biffi, A, CD Anderson, RS Desikan, M Sabuncu, L Cortellini, N Schmansky, D Salat, J Rosand. 2010. Genetic variation and neuroimaging measures in Alzheimer disease. Arch Neurol 67:677-685.

Bonilla, E, K Tanji, M Hirano, TH Vu, S DiMauro, EA Schon. 1999. Mitochondrial involvement in Alzheimer's disease. Biochim Biophys Acta 1410:171-182.

Braak, H, E Braak. 1991. Neuropathological stageing of Alzheimer-related changes. Acta neuropathologica 82:239-259.

Burgess, N, EA Maguire, J O'Keefe. 2002. The human hippocampus and spatial and episodic memory. Neuron 35:625-641.

Cardoso, SM, I Santana, RH Swerdlow, CR Oliveira. 2004. Mitochondria dysfunction of Alzheimer's disease cybrids enhances Abeta toxicity. J Neurochem 89:1417-1426.

Carrieri, G, M Bonafe, M De Luca, et al. 2001. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. Hum Genet 108:194-198.

Chinnery, PF, GA Taylor, N Howell, RM Andrews, CM Morris, RW Taylor, IG McKeith, RH Perry, JA Edwardson, DM Turnbull. 2000. Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. Neurology 55:302-304.

Clement, M, D Posada, KA Crandall. 2000. TCS: a computer program to estimate gene genealogies. Molecular ecology 9:1657-1659.

Coskun, P, J Wyrembak, S Schriner, HW Chen, C Marciniack, F Laferla, DC Wallace. 2011. A mitochondrial etiology of Alzheimer and Parkinson disease. Biochim Biophys Acta.

Coto, E, J Gomez, B Alonso, AI Corao, M Diaz, M Menendez, C Martinez, MT Calatayud, G Moris, V Alvarez. 2011. Late-onset Alzheimer's disease is associated with mitochondrial DNA 7028C/haplogroup H and D310 poly-C tract heteroplasmy. Neurogenetics 12:345-346.

Desikan, RS, HJ Cabral, CP Hess, et al. 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 132:2048-2057.

Desikan, RS, HJ Cabral, F Settecase, CP Hess, WP Dillon, CM Glastonbury, MW Weiner, NJ Schmansky, DH Salat, B Fischl. Automated MRI measures predict progression to Alzheimer's disease. Neurobiol Aging 31:1364-1374.

Etnier, JL, RJ Caselli, EM Reiman, GE Alexander, BA Sibley, D Tessier, EC McLemore. 2007. Cognitive performance in older women relative to ApoE-epsilon4 genotype and aerobic fitness. Med Sci Sports Exerc 39:199-207.

Fesahat, F, M Houshmand, MS Panahi, K Gharagozli, F Mirzajani. 2007. Do haplogroups H and U act to increase the penetrance of Alzheimer's disease? Cell Mol Neurobiol 27:329-334.

Finnila, S, MS Lehtonen, K Majamaa. 2001. Phylogenetic network for European mtDNA. American journal of human genetics 68:1475-1484.

Fischl, B, MI Sereno, RB Tootell, AM Dale. 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 8:272-284.

Frisoni, GB, A Prestia, O Zanetti, et al. 2009. Markers of Alzheimer's disease in a population attending a memory clinic. Alzheimer's & dementia : the journal of the Alzheimer's Association 5:307-317.

Heister, D, JB Brewer, S Magda, K Blennow, LK McEvoy. 2011. Predicting MCI outcome with clinically available MRI and CSF biomarkers. Neurology 77:1619-1628.

Herrnstadt, C, JL Elson, E Fahy, et al. 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. American journal of human genetics 70:1152-1171.

Hibar, DP, N Jahanshad, JL Stein, et al. Alzheimer's disease risk gene, GAB2, is associated with regional brain volume differences in 755 young healthy twins. Twin Res Hum Genet 15:286-295.

Hibar, DP, JL Stein, O Kohannim, et al. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. Neuroimage 56:1875-1891.

Hirai, K, G Aliev, A Nunomura, et al. 2001. Mitochondrial abnormalities in Alzheimer's disease. J Neurosci 21:3017-3023.

Holland, D, JB Brewer, DJ Hagler, C Fennema-Notestine, AM Dale. 2009. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proc Natl Acad Sci U S A 106:20954-20959.

Honea, RA, RH Swerdlow, ED Vidoni, JM Burns. 2011. Progressive regional atrophy in normal adults with a maternal history of Alzheimer disease. Neurology 76:822-829.

Honea, RA, RH Swerdlow, ED Vidoni, J Goodwin, JM Burns. 2010. Reduced gray matter volume in normal adults with a maternal family history of Alzheimer disease. Neurology 74:113-120.

Honea, RA, E Vidoni, A Harsha, JM Burns. 2009. Impact of APOE on the healthy aging brain: a voxel-based MRI and DTI study. J Alzheimers Dis 18:553-564.

Honea, RA, ED Vidoni, RH Swerdlow, JM Burns. 2012. Maternal family history is associated with Alzheimer's disease biomarkers. Journal of Alzheimer's disease : JAD 31:659-668.

Hua, X, AD Leow, S Lee, et al. 2008. 3D characterization of brain atrophy in Alzheimer's disease and mild cognitive impairment using tensor-based morphometry. NeuroImage 41:19-34.

Hudson, G, R Sims, D Harold, et al. 2012a. No consistent evidence for association between mtDNA variants and Alzheimer disease. Neurology 78:1038-1042.

Hudson, G, R Sims, D Harold, et al. 2012b. No consistent evidence for association between mtDNA variants and Alzheimer disease. Neurology.

Ingman, M, U Gyllensten. 2001. Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. J Hered 92:454-461.

Jack, CR, MM Shiung, JL Gunter, et al. 2004. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurol 62:591-600.

Kruger, J, R Hinttala, K Majamaa, AM Remes. 2010. Mitochondrial DNA haplogroups in early-onset Alzheimer's disease and frontotemporal lobar degeneration. Mol Neurodegener 5:8.

Lakatos, A, O Derbeneva, D Younes, et al. 2010. Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort. Neurobiol Aging 31:1355-1363.

Mancuso, M, M Nardini, D Micheli, et al. 2007. Lack of association between mtDNA haplogroups and Alzheimer's disease in Tuscany. Neurol Sci 28:142-147.

Maruszak, A, JA Canter, M Styczynska, C Zekanowski, M Barcikowska. 2009. Mitochondrial haplogroup H and Alzheimer's disease--is there a connection? Neurobiol Aging 30:1749-1755.

Mosconi, L, R Mistur, R Switalski, M Brys, L Glodzik, K Rich, E Pirraglia, W Tsui, S De Santi, MJ de Leon. 2009. Declining brain glucose metabolism in normal individuals with a maternal history of Alzheimer disease. Neurology 72:513-520.

Mosconi, L, JO Rinne, WH Tsui, et al. 2010. Increased fibrillar amyloid-{beta} burden in normal individuals with a family history of late-onset Alzheimer's. Proc Natl Acad Sci U S A 107:5949-5954.

Orth, M, AH Schapira. 2001. Mitochondria and degenerative disorders. Am J Med Genet 106:27-36.

Parker, WD, Jr., CM Filley, JK Parks. 1990. Cytochrome oxidase deficiency in Alzheimer's disease. Neurology 40:1302-1303.

Peper, JS, RM Brouwer, DI Boomsma, RS Kahn, HE Hulshoff Pol. 2007. Genetic influences on human brain structure: a review of brain imaging studies in twins. Human brain mapping 28:464-473.

Podewils, LJ, E Guallar, LH Kuller, LP Fried, OL Lopez, M Carlson, CG Lyketsos. 2005. Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. Am J Epidemiol 161:639-651.

Posada, D, TJ Maxwell, AR Templeton. 2005. TreeScan: a bioinformatic application to search for genotype/phenotype associations using haplotype trees. Bioinformatics 21:2130-2132.

Potkin, SG, G Guffanti, A Lakatos, et al. 2009. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. PLoS One 4:e6501.

Pyle, A, T Foltynie, W Tiangyou, et al. 2005. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann Neurol 57:564-567.

Reiman, EM. 2007. Linking brain imaging and genomics in the study of Alzheimer's disease and aging. Ann N Y Acad Sci 1097:94-113.

Ridge, P, T Maxwell, C Corcoran, M Norton, J Tschanz, E O'Brien, R Kerber, R Cawthon, R Munger, K JSK. 2012. Mitochondrial genomic analysis of late onset Alzheimer's disease reveals protective haplogroups H6A1A/H6A1B: the Cache County Study on Memory in Aging. PLoS One 7:e45134.

Rovio, S, I Kareholt, EL Helkala, M Viitanen, B Winblad, J Tuomilehto, H Soininen, A Nissinen, M Kivipelto. 2005. Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease. Lancet Neurol 4:705-711.

Santoro, A, V Balbi, E Balducci, et al. 2010. Evidence for sub-haplogroup h5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease. PLoS One 5:e12037.

Schuff, N, N Woerner, L Boreta, T Kornfield, LM Shaw, JQ Trojanowski, PM Thompson, CR Jack, Jr., MW Weiner. 2009. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132:1067-1077.

Schuit, AJ, EJ Feskens, LJ Launer, D Kromhout. 2001. Physical activity and cognitive decline, the role of the apolipoprotein e4 allele. Med Sci Sports Exerc 33:772-777.

Silver, M, E Janousova, X Hua, PM Thompson, G Montana. Identification of gene pathways implicated in Alzheimer's disease using longitudinal imaging phenotypes with sparse regression. Neuroimage 63:1681-1694.

Stein, JL, X Hua, S Lee, et al. Voxelwise genome-wide association study (vGWAS). Neuroimage 53:1160-1174.

Stein, JL, X Hua, JH Morra, et al. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. Neuroimage 51:542-554.

Stein, JL, X Hua, JH Morra, et al. 2010. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. NeuroImage 51:542-554.

Stein, JLSE MedlandAA Vasquez, et al. 2012. Identification of common variants associated with human hippocampal and intracranial volumes. Nature genetics 44:552-561.

Swerdlow, RH. 2007. Mitochondria in cybrids containing mtDNA from persons with mitochondriopathies. J Neurosci Res 85:3416-3428.

Swerdlow, RH, JK Parks, DS Cassarino, DJ Maguire, RS Maguire, JP Bennett, Jr., RE Davis, WD Parker, Jr. 1997. Cybrids in Alzheimer's disease: a cellular model of the disease? Neurology 49:918-925.

Takasaki, S. 2009. Mitochondrial haplogroups associated with Japanese Alzheimer's patients. J Bioenerg Biomembr 41:407-410.

Templeton, AR, T Maxwell, D Posada, JH Stengard, E Boerwinkle, CF Sing. 2005. Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. Genetics 169:441-453.

van der Walt, JM, YA Dementieva, ER Martin, et al. 2004. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. Neurosci Lett 365:28-32.

van Oven, M, M Kayser. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30:E386-394.

Wallace, DC. 1994. Mitochondrial DNA sequence variation in human evolution and disease. Proc Natl Acad Sci U S A 91:8739-8746.

Westfall, P, SS Young. 1993. Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustments. New York City, NY: Wiley-Interscience.

Zsurka, G, J Kalman, A Csaszar, I Rasko, Z Janka, P Venetianer. 1998. No mitochondrial haplotype was found to increase risk for Alzheimer's disease. Biol Psychiatry 44:371-373.

Figure Legends

Figure 1. Branches 184 and 155. This is a subset of the full haplotype network (Supplementary Figure 1). The large blue ovals represent observed haplotypes within our dataset and the small white nodes ancestral (and unobserved) haplotypes in the network. Each blue segment between observed or unobserved nodes corresponds to a single SNP. Only branches 184 and 155 are labeled, as these were the only branches defining groups associated with a given phenotype. Mitochondrial haplogroups were indistinguishable for the haplotypes in these clades.

Figure 2. Branch 199. The nodes and colors are as described in Figure 1. Mitochondrial haplogroups were distinguishable for two of the three nodes in the clade and so are labeled.

Figure 3. Branch 9. The nodes and colors are as described in Figure 1. Mitochondrial haplogroups were distinguishable for both nodes in the clade and are labeled.

Table 1, Chapter 4. Baseline demographic, clinical, and neuroimaging characteristics of study participants

	Samples	Age	# Males (%)	Education Level	APOE ε4 (%)	GDS	CDR	MMSE	ADAS-COG
Controls	175	76.1 (4.9)	96 (54.8)	16.2 (2.7)	55 (31.4)	1.0 (1.2)	0.00 (0.0)	29.1 (0.95)	10.18 (6.7)
MCI	316	75.4 (7.2)	204 (64.5)	15.8 (2.9)	172 (54.4)	1.54 (1.4)	0.49 (0.03)	27.1 (1.8)	11.94 (5.9)
AD	154	75.4 (7.6)	82 (53.2)	14.9 (2.9)	90 (58.4)	1.6 (1.4)	0.72 (0.23)	23.5 (2.0)	13.2 (6.3)

Values shown in parentheses are standard deviation, except where noted in the column header. Abbreviations in column headings are as follows: MCI, mild cognitive impairment; AD, Alzheimer's disease; GDS, geriatric depression scale total score; MMSE, Mini-Mental Status Exam total score; ADAS-COG, Alzheimer's disease assessment scale-cognitive subscale, total 11. Age and education level are measured in years.

Table 2, Chapter 4. Top five hits associated with whole brain volume

Contrast	p-values without APOE		p-values with APOE		Group / Network
	nominal	corrected	nominal	corrected	
155	0.001	0.038	8.0e-4	0.025	919170 / 989295
206	0.015	0.45	0.011	0.35	931207 / 988584
184	0.024	0.59	0.018	0.49	942316 / 988703
31	0.071	0.91	0.057	0.85	968186 / 989687
40	0.071	0.91	0.057	0.85	989687 / 968186

The nominal and corrected p-values are reported for the two models either using APOE ε4 as a covariate (p-values with APOE) or not using APOE ε4 as a covariate (no APOE) for the top five contrasts with whole brain volume. We also report the number of males and females in the contrast and the levels of whole brain volume in the clade represented by the contrast and the rest of the network.

Table 3, Chapter 4. Top five hits associated with left hippocampal atrophy

Contrast	p-values without APOE		p-values with APOE		Group / Network
	nominal	corrected	nominal	corrected	
184	0.0026	0.064	0.0036	0.0909	-2.98 / -1.18
146	0.0352	0.5562	0.0549	0.7031	-1.19 / -2.27
75	0.055	0.699	0.0637	0.771	-0.281 / -1.23
67	0.0822	0.8479	0.0846	0.8295	-0.76988 / -1.252
23	0.1069	0.9134	0.1332	0.9483	-1.87 / -1.19

The nominal and corrected p-values are reported for the endophenotype as described in Table 2. Values in the Group / Network column correspond to left hippocampal atrophy.

Table 4, Chapter 4. Contrasts associated with percent change in temporal pole thickness

Contrast	p-values without APOE		p-values with APOE		Group / Network
	nominal	corrected	nominal	corrected	
199	0.0018	0.0416	0.0016	0.0343	-13.22 / -4.27
9	0.002	0.0462	0.001	0.0441	-10 / -4.10

The two contrasts associated with percent change in temporal pole thickness over two years are reported for the two models described in Table 2.

Values in the Group / Network column correspond to percent change levels.

Table 5, Chapter 4. Mitochondrial haplogroups associated with Alzheimer's disease

Haplogroup / Cluster	Risk Haplogroup	Protective Haplogroup
B4C1	(Takasaki 2009)	
G2A	(Takasaki 2009)	
HV	(Maruszak et al. 2009)	
H	(Fesahat et al. 2007; Coto et al. 2011)	
H5 / H5A	(Santoro et al. 2010; Coskun et al. 2011)	
H6A1A / H6A1B		(Ridge et al. 2012)
K		(Carrieri et al. 2001)
N9B1	(Takasaki 2009)	
U	(van der Walt et al. 2004; Fesahat et al. 2007)	(Carrieri et al. 2001; van der Walt et al. 2004)
UK	(Lakatos et al. 2010)	
No Reported Association (Zsurka et al. 1998; Chinnery et al. 2000; Pyle et al. 2005; Mancuso et al. 2007; Kruger et al. 2010; Hudson et al. 2012b)		

Here we list previously reported associations (or lack thereof) of mtDNA with Alzheimer's disease.

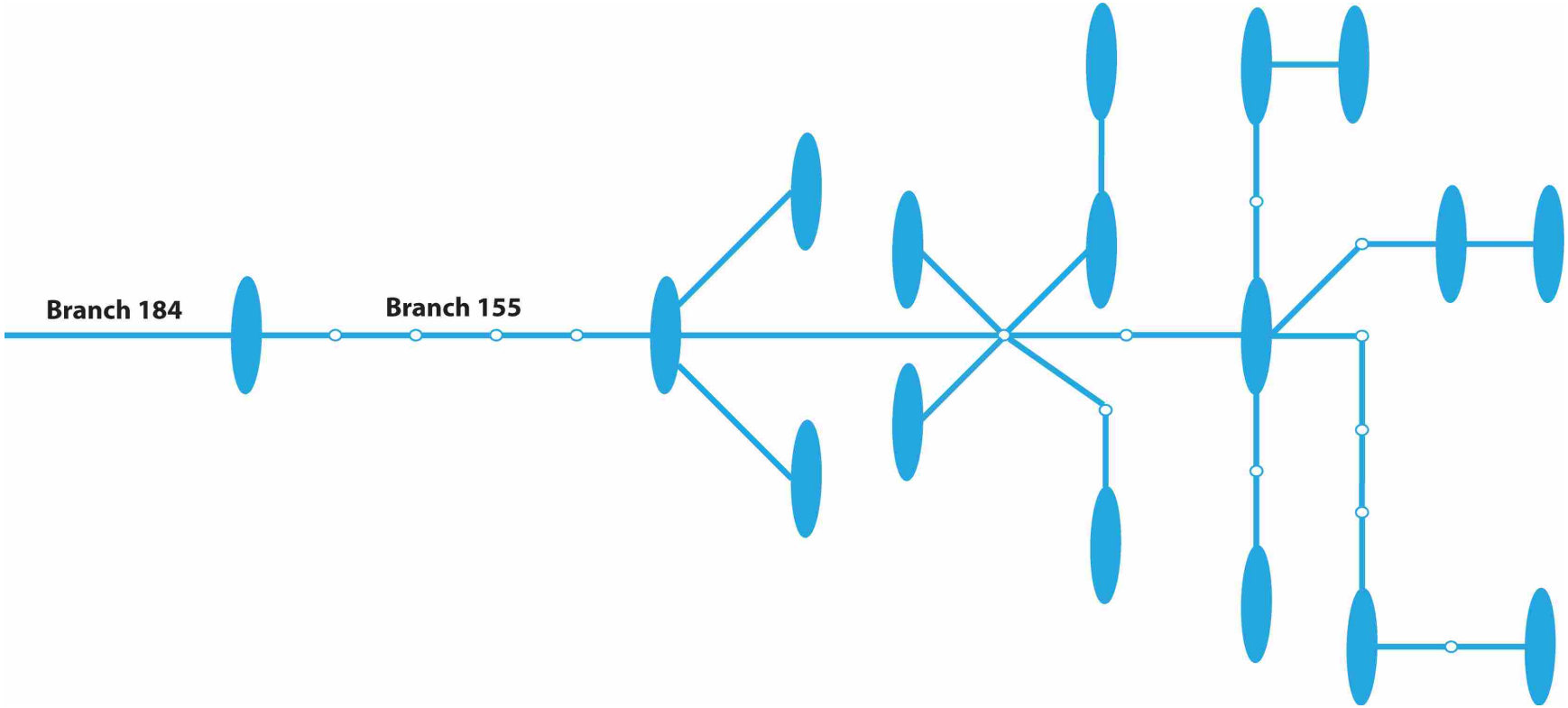


Figure 1, Chapter 4. Branches 184 and 155. This is a subset of the full haplotype network (Supplementary Figure 1). The large blue ovals represent observed haplotypes within our dataset and the small white nodes ancestral (and unobserved) haplotypes in the network. Each blue segment between observed or unobserved nodes corresponds to a single SNP. Only branches 184 and 155 are labeled, as these were the only branches defining groups associated with a given phenotype. Mitochondrial haplogroups were indistinguishable for the haplotypes in these clades.

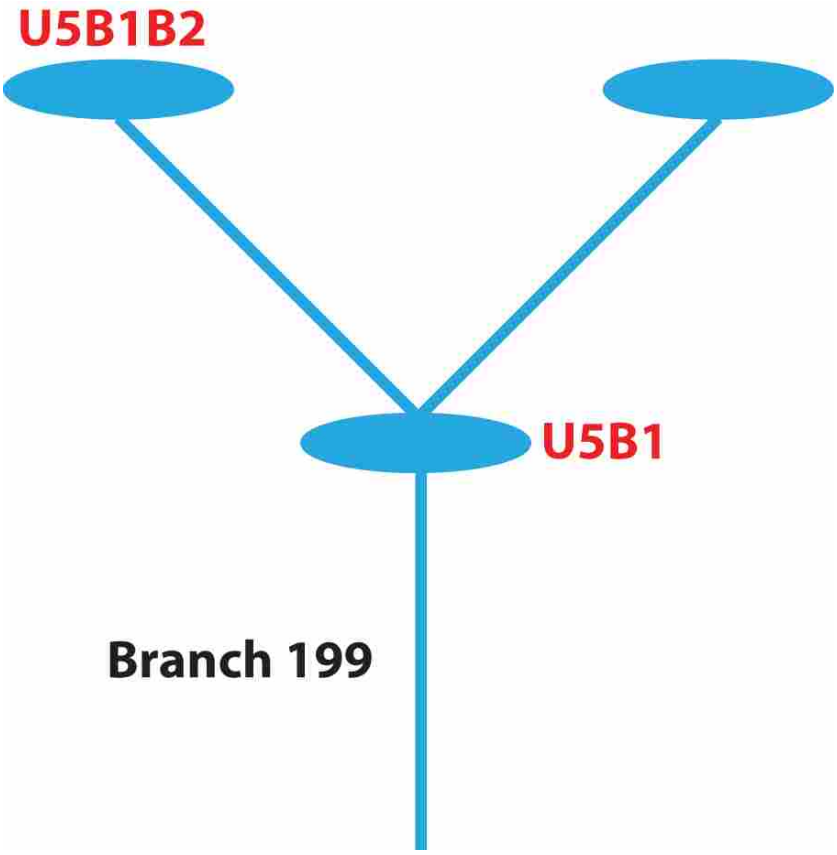


Figure 2, Chapter 4. Branch 199. The nodes and colors are as described in Figure 1. Mitochondrial haplogroups were distinguishable for two of the three nodes in the clade and so are labeled.

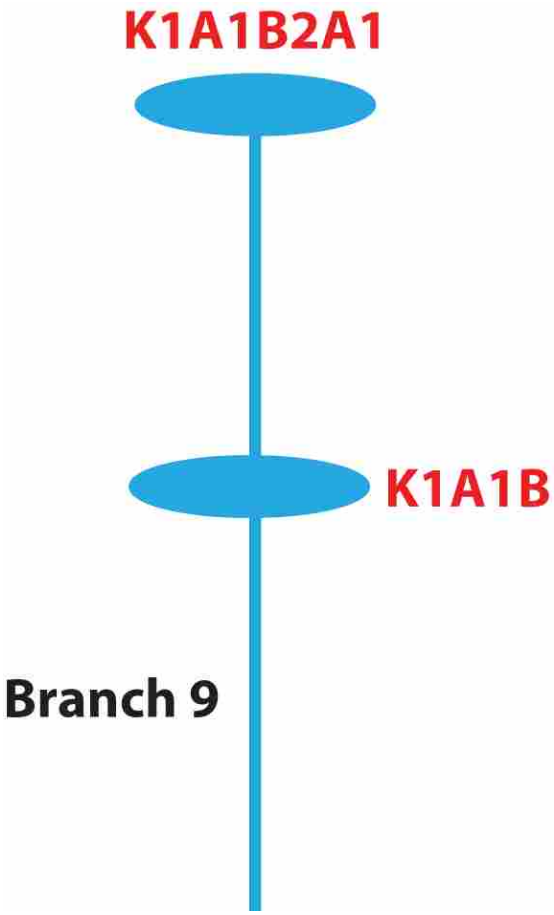


Figure 3, Chapter 4. Branch 9. The nodes and colors are as described in Figure 1. Mitochondrial haplogroups were distinguishable for both nodes in the clade and are labeled.

Manuscript submitted to Neurobiology of Aging

Chapter 5

Alzheimer's disease: finding the missing heritability

Perry G. Ridge[1,2], Shubhabrata Mukherjee[3], Paul Crane[3], John S.K. Kauwe[1], Alzheimer's Disease Genetics Consortium[**]

[1]Department of Biology, Brigham Young University, Provo, UT, USA

[2]ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT, USA

[3]Department of Medicine, University of Washington, Seattle, WA, USA

[**]Coauthors from the Alzheimer's Disease Genetics Consortium are listed in the

Abstract

Background: Alzheimer's disease (AD) is the most common form of dementia with worldwide estimates of 24-35 million people affected. As the world's population ages, it is expected that incidence will rapidly increase. Presently there are no effective treatments to prevent or slow disease progression and individuals with late stage AD require fulltime care. AD is a complex disorder influenced by both environmental and genetic factors. Recent work has identified 11 markers in 10 different loci that show replicable association with late-onset AD.

Methods: For this research we used over 2 million SNPs for each of 9,884 individuals from the Alzheimer's Disease Genetics Consortium. We used Genome-wide Complex Trait Analysis, a novel approach for measuring phenotypic variance explained by a set of SNPs.

Findings: Here we assess the phenotypic variance explained first by known late-onset AD loci, and then by over 2 million common SNPs in the Alzheimer's Disease Genetics Consortium dataset. In all, 1/3 of phenotypic variance is explained by all common SNPs. *APOE* alone explains 6% of the variance and other known markers explain another 2%. This means that more than 25% of total variance in AD status remains unexplained by known markers, but is tagged by common SNPs included on genotyping arrays or imputed with HapMap genotypes.

Interpretation: Novel AD markers likely to explain large amounts of phenotypic variance are likely to be rare and unidentifiable using genome-wide association studies. Based on our findings and the current direction of human genetics research, we suggest specific study designs for future studies to identify the remaining heritability of Alzheimer's disease.

Funding

Resources for this work were provided by grants from NIH (R01AG042611), the Alzheimer's Association (MNIRG-11-205368) and the Brigham Young University Gerontology Program.

Introduction

Alzheimer's disease (AD) is the most common form of dementia. Worldwide estimates of prevalence vary, with estimates of 24 to 35 million people affected (Brookmeyer et al. 2007; Querfurth, LaFerla 2010; Ballard et al. 2011) and with an aging population, prevalence is expected to increase to 1 in 85 people by 2050 (Brookmeyer et al. 2007).

AD is a heterogeneous disease caused by a combination of environmental and genetic factors. The most important risk factor for Alzheimer's disease is age (Herrup 2010; Querfurth, LaFerla 2010). Environmental risk factors include hypertension, use of estrogen(Patterson et al. 2008), smoking (Almeida et al. 2002; Cataldo, Prochaska, Glantz 2010), stroke, heart disease, depression, arthritis, and diabetes (Lindsay et al. 2002). In addition, certain lifestyle choices appear to decrease the risk of AD: exercise (Podewils et al. 2005), intellectual stimulation (Wang et al. 2002), and maintaining a Mediterranean diet (including fish) (Scarmeas et al. 2006; Patterson et al. 2007).

The genetics of AD are complex. Several genes are known to harbor either causative or risk variants for AD. There are two primary types of AD as defined by age. The first is early-onset, or familial AD, and the second type is late-onset AD (LOAD), or sporadic AD. Three genes, *APP* (Goate et al. 1991), *PSEN1* (Sherrington et al. 1995), and *PSEN2* (Levy-Lahad et al. 1995) are known to harbor many highly penetrant, autosomal dominantly-inherited variants, which lead to early-onset AD but account for only a small fraction of total AD cases.

LOAD accounts for 99% of AD cases and is caused by a more complex underlying genetic architecture. Genome-wide association studies (GWAS) have identified 10 different loci associated with AD (Table 1). Recent applications of next-generation sequencing (NGS) have suggested an important role of rare variants of large effect in the etiology of AD (Guerreiro et al. 2012; Jonsson et al. 2012a; Jonsson et al. 2012b). Identification of additional variants will provide information that is integral to the development, evaluation and application of effective therapeutic strategies for AD. In this study we evaluate the variance in AD status explained by known AD genes and common variants and suggest strategies for identifying additional AD genes.

Materials and Methods

Dataset

We used the Alzheimer's Disease Genetics Consortium (ADGC) dataset described in Naj et al. (Naj et al. 2011) for our analyses. Briefly, samples were genotyped using Affymetrix and Illumina SNP chips. Minimum call rates and minor allele frequencies cutoffs for SNPs genotyped on each of the two platforms were 95%/2% and 98%/1%, respectively. To have a common set of SNPs across all samples, imputation to HapMap phase 2 (release 22)(The International HapMap Consortium. 2003) was performed using MaCH (Li et al. 2010) and strand ambiguous SNPs were removed, resulting in a rectangular dataset with 2,042,114 SNPs. Only SNPs imputed with $R^2 \geq 0.50$ were included in the dataset. We added an additional two SNPs, rs7412 and rs429358, corresponding to *APOE* ε2 and ε4, respectively.

We used a compiled dataset of directly genotyped SNPs common to all 15 studies to assess cryptic relatedness and calculate principal components to account for population-specific

variations in allele distribution. We excluded strand ambiguous SNPs, resulting in a rectangular dataset with 21,880 directly observed (not imputed) SNPs in common across all the studies. We filtered SNPs with pairwise LD (r^2) < 0.20, resulting in a dataset with 17,054 SNPs. We used both PLINK (Purcell et al. 2007) and KING-ROBUST (Manichaikul et al. 2010) for relatedness analysis. KING-ROBUST provided unbiased kinship coefficient estimates for related individuals in our dataset. We excluded up to 3rd degree relatives (kinship >= 0.0442) for a final dataset containing 19,692 individuals.

Of the 19,692 individuals in the original dataset we analyzed a subset of 10,922 individuals who had complete data for the 11 markers listed in Table 1, AD case-control status, age, sex, and 10 principal components from the population stratification analysis (missingness rates for each of the covariates and case-control status are reported in Supplementary Table 1). Basic demographic information for the 10,922 individuals in the subset of the dataset used in this study is presented in Table 2. We collected chromosome length and number of genes per chromosome from the Vega database (Wilming et al. 2008).

Genetic Analyses

We used Genome-wide Complex Trait Analysis (GCTA) (Yang et al. 2011a), a tool that implements the methods described in Yang et al. (Yang et al. 2010), Lee et al. (Lee et al. 2011), and Yang et al. (Yang et al. 2011b) to estimate the phenotypic variance explained by known AD genes and tagged by SNPs on the SNP arrays. Briefly, GCTA uses a mixed linear model and treats the effects of SNPs as random effects, effectively testing all the SNPs together for effect (in contrast to GWAS, which considers each SNP individually). We used age, sex, site and 10

principal components as covariates. For the analyses in which we examined unexplained phenotypic variance, we also controlled for the 11 known AD markers (Table 1). We specified a population prevalence of LOAD at 0.13 (Alzheimer's Association 2012).

Results

We estimated the variance in AD case-control status focusing first on the 11 known AD markers (Table 1). Together these markers account for 7.8% of the phenotypic variance. Next, we separated the explained phenotypic variance up by chromosome (Figure 1). Chromosome 19 accounts for the highest proportion of phenotypic variance.

In all, the 2,042,116 SNPs in the HapMap imputed ADGC dataset explain 33.1% of phenotypic variance (genetic variance of 0.0711, standard error 0.0072). The APOE ε2 and ε4 alleles account for 5.9% of the phenotypic variance. The other 9 known high frequency SNPs identified in GWAS explain an additional 1.9%. After controlling for these 11 markers, an additional 25.3% of the total phenotypic variance (genetic variance of 0.046, standard error 0.006) is explained with as-yet unidentified variants. The remaining phenotypic variance explained by each chromosome after controlling for the 11 known markers is shown in Figure 2. SNPs on chromosomes 1, 4, 5 and 17 account for the largest percentage of remaining unexplained phenotypic variance compared to other chromosomes, each accounting for more than 2% (Figure 2). Chromosomes 9, 14, and 21 account for the least (<0.0001% each); however, there is unexplained phenotypic variance on all the autosomes. There is no relationship between explained variance and chromosome length (p-value = 0.8), or number of genes per chromosome (p-value = 0.7).

Discussion

A clear understanding of the genetic architecture of Alzheimer's disease provides the foundation of information needed to cure this terrible disease. While many large GWAS for AD have been performed and several loci that show replicable association have been identified (as referenced in Table 1), relatively little phenotypic variance is explained by these variants. Our results show that there is clearly much work to be done if we are to solve the genetic architecture of AD. GWAS with sample sizes performed to date are able to identify common variants with moderate to small effect sizes. Results of GWAS in AD and other conditions suggest there may be a large number of such variants and that increasing sample sizes will result in the identification of additional loci. However, additional loci detected with the GWAS strategy will likely have effects either similar to or smaller than SNPs already identified. The range of SNPs identifiable by current GWAS (Harold et al. 2009; Biffi et al. 2010; Carrasquillo et al. 2010; Corneveaux et al. 2010; Hollingworth et al. 2011; Hu et al. 2011; Naj et al. 2011) is marked on Figure 3 by the large box bordered by dots (the GWAS search space), with recent GWAS hits inside the labeled oval. The GWAS being conducted by the International Genomics of Alzheimer's Project represents a substantial increase in sample size and will undoubtedly identify additional common loci with small effects on AD risk. Nevertheless, it is unlikely that many common variants of even modest effect size remain to be identified.

There are still many AD variants that remain to be identified, however, and these variants exist on every autosome (Figure 2). Variants with large effects are almost certainly rare or they would have been identified in GWAS. Such variants are located in the smaller box bordered with dashes in Figure 3. To date, alleles of this type that have been identified have clear functional

effects and large effect sizes compared to associated alleles from GWAS. Detection of rare variants of large effect requires different experimental designs than GWAS such as sequencing causal loci. Exome chip array studies target known variation in coding regions, even those of very low frequency; this may prove a promising and economical approach. However, accurately genotyping variants of less than 1-2% using these arrays is quite challenging, and for variants that are present below these frequencies other approaches are required.

Two seemingly contradictory hypotheses exist about the architecture of complex disease: the common disease/common variant hypothesis and the multiple rare variant hypothesis. In the first, many common variants of small effect size collectively explain disease risk, while in the second, rare variants, some with large effect and high penetrance, explain disease risk. However, as suggested by Singleton et al. (Singleton, Hardy 2011) these two hypotheses are not mutually exclusive and the genetics of complex diseases are likely a hybrid of the two. Singleton et al. (Singleton, Hardy 2011) suggests that both common and rare variants that increase or decrease disease risk are likely to be found in the same loci and coined the phrase "pleomorphic risk loci". To date, AD genetics research has largely focused on common variants that influence disease risk, likely due to technological and financial constraints. However, the advent of next generation sequencing (NGS) and falling costs of this technology have made it possible to expand AD research to include the search for rare variants. Recently, this technology was used to identify a functional variant that protects against Alzheimer's disease in the amyloid precursor protein (*APP*) (Jonsson et al. 2012a). Additionally, two groups recently used NGS to identify additional, likely functional, variants associated with AD in the triggering receptor expressed on myeloid 2 (*TREM2*) gene (Guerreiro et al. 2012; Jonsson et al. 2012b). The *TREM2* variant is present in

about 1% of the general population and has a high odds ratio (2.9 to 5.1 depending on the dataset). Likewise, the *APP* variant is extremely rare (frequency of 0.00038), but confers a large protective effect on carriers. Larger scale applications of this technology and careful study design are likely to lead to the identification of additional variants and further explain the remaining phenotypic variance in AD.

Family-based studies are also an effective application of NGS. These studies require carefully ascertained families and accurate pedigree data and can be used to identify high effect, low frequency variants (located in the box with longer dashes in Figure 3). Family-based studies are especially powerful because large effect, low frequency disease-causing (or disease-modifying) sequence features, some of which may be unique to a single family, are likely to segregate, at least partially, with disease status. These approaches have not yet been extensively applied in AD research. Nevertheless, family-based studies utilizing large-scale genome or exome sequencing have recently been used to identify disease-causing variants in several Mendelian (Choi et al. 2009; Roach et al. 2010; Rope et al. 2011) and complex disorders (Krebs, Paisan-Ruiz 2012; Kilpinen, Barrett 2013).

It is also possible that gene-gene interactions account for much of the unexplained variance in AD status (Turton et al. 2011). These interactions are widespread and common (Mackay 2004; Shao et al. 2008) and approaches to understanding the effects of epistatic interactions exist and continue to mature (Combarros et al. 2009; Turton et al. 2011). Several interesting candidate interactions have been identified and Ebbert et al. 2013 (submitted) recently demonstrated that allowing for interactions greatly improves the diagnostic utility of the known AD markers.

Unfortunately, the complexity of this problem and the extremely large samples sizes required to perform agnostic screens for gene-gene interactions make it very difficult to conduct effective screens for these effects.

AD is a highly complex disease with substantial genetic and environmental components. Our results suggest that genetic variance accounts for ~30% of phenotypic variance, but over 75% of this phenotypic variance remains unexplained by currently identified AD genes. Future AD genetics research must leverage larger samples and novel technologies such as NGS to identify rare, high penetrant variation and gene-gene interactions that are likely to explain the remaining genetic and phenotypic variance in AD.

Genetic research in AD has followed roughly the same model as the study of other complex diseases; largely focusing on the identification of common variants of modest effect using association studies. Scientist in many disease fields have been very successful at identifying numerous associated variants (this is a small representative sample (Jenuth 2000; 2011; Ehret et al. 2011; Ferreira et al. 2011; Haiman et al. 2011; Wain et al. 2011; Bradfield et al. 2012; Jostins et al. 2012; Lill et al. 2012; Wen et al. 2012)). The transition from a focus on common variants to a focus on the identification of low frequency variants is now underway. These rare functionally relevant markers are often more easily characterized than common variants of small effect. This will lead to strong and testable hypotheses for the development of therapeutics, thus accelerating the progress toward effective prevention and treatment.

Systematic Review

www.alzgene.org maintains an extensive list of known AD genes and markers. We collected evidence for these known AD markers from www.alzgene.org and primary literature for data pertaining to common AD markers. In addition to common markers we searched for rare markers that have been identified using next-generation sequencing- or association-based approaches.

Interpretation

Our findings suggest the need to apply different study designs to the study of Alzheimer's disease. Specifically, common SNPs or markers in LD with them account for a large amount of unexplained variance in AD. Evaluation of the recently discovered markers makes it clear that the majority of detectable genetic variants are rare. Novel approaches to the identification of these variants have met with success and will likely provide additional insights into molecular mechanisms of AD pathology.

Acknowledgments

Resources for this work were provided by grants from NIH (R01AG042611), the Alzheimer's Association (MNIRG-11-205368) and the Brigham Young University Gerontology Program.

Biological samples and associated phenotypic data used in primary data analysis were stored at the Principal Investigator's institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD), at the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania, and the NIA Alzheimer's Disease Genetics Consortium Data Storage Site at the University of Pennsylvania. Contributors to the genetic analysis data included

Principal Investigators on projects that were individually funded by NIA, other NIH institutes, or private entities.

The National Institutes of Health, National Institute on Aging (NIH-NIA) supported this work through the following grants: ADGC, U01 AG032984, RC2 AG036528; NACC, U01 AG016976; NCRAD, U24 AG021886; NIA LOAD, U24 AG026395, U24 AG026390; Banner Sun Health Research Institute P30 AG019610; Boston University, P30 AG013846, U01 AG10483, R01 CA129769, R01 MH080295, R01 AG017173, R01 AG025259, R01AG33193; Columbia University, P50 AG008702, R37 AG015473; Duke University, P30 AG028377, AG05128; Emory University, AG025688; Group Health Research Institute, UO1 AG06781, UO1 HG004610; Indiana University, P30 AG10133; Johns Hopkins University, P50 AG005146, R01 AG020688; Massachusetts General Hospital, P50 AG005134; Mayo Clinic, P50 AG016574; Mount Sinai School of Medicine, P50 AG005138, P01 AG002219; New York University, P30 AG08051, MO1RR00096, UL1 RR029893, 5R01AG012101, 5R01AG022374, 5R01AG013616, 1RC2AG036502, 1R01AG035137; Northwestern University, P30 AG013854; Oregon Health & Science University, P30 AG008017, R01 AG026916; Rush University, P30 AG010161, R01 AG019085, R01 AG15819, R01 AG17917, R01 AG30146; TGen, R01 NS059873; University of Alabama at Birmingham, P50 AG016582, UL1RR02777; University of Arizona, R01 AG031581; University of California, Davis, P30 AG010129; University of California, Irvine, P50 AG016573, P50, P50 AG016575, P50 AG016576, P50 AG016577; University of California, Los Angeles, P50 AG016570; University of California, San Diego, P50 AG005131; University of California, San Francisco, P50 AG023501, P01 AG019724; University of Kentucky, P30 AG028383, AG05144; University of Michigan, P50 AG008671; University of

Pennsylvania, P30 AG010124; University of Pittsburgh, P50 AG005133, AG030653; University of Southern California, P50 AG005142; University of Texas Southwestern, P30 AG012300; University of Miami, R01 AG027944, AG010491, AG027944, AG021547, AG019757; University of Washington, P50 AG005136; Vanderbilt University, R01 AG019085; and Washington University, P50 AG005681, P01 AG03991. The Kathleen Price Bryan Brain Bank at Duke University Medical Center is funded by NINDS grant # NS39764, NIMH MH60451 and by Glaxo Smith Kline. Genotyping of the TGEN2 cohort was supported by Kronos Science. The TGen series was also funded by NIA grant AG034504 to AJM, The Banner Alzheimer's Foundation, The Johnnie B. Byrd Sr. Alzheimer's Institute, the Medical Research Council, and the state of Arizona and also includes samples from the following sites: Newcastle Brain Tissue Resource (funding via the Medical Research Council, local NHS trusts and Newcastle University), MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council),South West Dementia Brain Bank (funding via numerous sources including the Higher Education Funding Council for England (HEFCE), Alzheimer's Research Trust (ART), BRACE as well as North Bristol NHS Trust Research and Innovation Department and DeNDRoN), The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stiching Alzheimer Onderzoek), Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona. ADNI Funding for ADNI is through the Northern California Institute for Research and Education by grants from Abbott, AstraZeneca AB, Bayer Schering Pharma AG, Bristol-Myers Squibb, Eisai Global Clinical Development, Elan Corporation, Genentech, GE Healthcare, GlaxoSmithKline, Innogenetics, Johnson and Johnson, Eli Lilly and Co., Medpace, Inc., Merck and Co., Inc., Novartis AG, Pfizer Inc, F.

Hoffman-La Roche, Schering-Plough, Synarc, Inc., Alzheimer's Association, Alzheimer's Drug Discovery Foundation, the Dana Foundation, and by the National Institute of Biomedical Imaging and Bioengineering and NIA grants U01 AG024904, RC2 AG036535, K01 AG030514. We thank Drs. D. Stephen Snyder and Marilyn Miller from NIA who are *ex-officio* ADGC members. Support was also from the Alzheimer's Association (LAF, IIRG-08-89720; MP-V, IIRG-05-14147) and the US Department of Veterans Affairs Administration, Office of Research and Development, Biomedical Laboratory Research Program. P.S.G.-H. is supported by Wellcome Trust, Howard Hughes Medical Institute, and the Canadian Institute of Health Research.

ADGC Authors and Affiliations:

Marilyn S. Albert[1], Roger L. Albin[2-4], Liana G. Apostolova[5], Steven E. Arnold[6], Clinton T. Baldwin[7], Robert Barber[8], Michael M. Barmada[9], Lisa L. Barnes[10, 11], Thomas G. Beach[12], Gary W. Beecham[13, 14], Duane Beekly[15], David A. Bennett[10, 16], Eileen H. Bigio[17], Thomas D. Bird[18], Deborah Blacker[19,20], Bradley F. Boeve[21], James D. Bowen[22], Adam Boxer[23], James R. Burke[24], Joseph D. Buxbaum[25, 26, 27], Nigel J. Cairns[28], Laura B. Cantwell[29], Chuanhai Cao[30], Chris S. Carlson[31], Regina M. Carney[32], Minerva M. Carrasquillo[33], Steven L. Carroll[34], Helena C. Chui[35], David G. Clark[36], Jason Corneveaux[37], Paul K. Crane[38], David H. Cribbs[39], Elizabeth A. Crocco[40], Carlos Cruchaga[41], Philip L. De Jager[42,43], Charles DeCarli[44], Steven T. DeKosky[45], F. Yesim Demirci[9], Malcolm Dick[46], Dennis W. Dickson[33], Ranjan Duara[47], Nilufer Ertekin-Taner[33,48], Denis Evans[49], Kelley M. Faber[50], Kenneth B. Fallon[34], Martin R. Farlow[51], Lindsay A Farrer[7, 52, 76, 77, 83], Steven Ferris[53], Tatiana M. Foroud[50], Matthew P. Frosch[54], Douglas R. Galasko[55], Mary Ganguli[56], Marla Gearing[57,58], Daniel H. Geschwind[59], Bernardino Ghetti[60],

John R. Gilbert[13,14], Sid Gilman[2], Jonathan D. Glass[61], Alison M. Goate[41], Neill R. Graff-Radford[33,48], Robert C. Green[62], John H. Growdon[63], Jonathan L. Haines[64, 65], Hakon Hakonarson[66], Kara L. Hamilton-Nelson[13], Ronald L. Hamilton[67], John Hardy[68], Lindy E. Harrell[36], Elizabeth Head[69], Lawrence S. Honig[70], Matthew J. Huentelman[37], Christine M. Hulette[71], Bradley T. Hyman[63], Gail P. Jarvik[72,73], Gregory A. Jicha[74], Lee-Way Jin[75], Gyungah Jun[7,76,77], M. Ilyas Kamboh[9,78], Anna Karydas[23], John S.K. Kauwe[79], Jeffrey A. Kaye[80,81], Ronald Kim[82], Edward H. Koo[55], Neil W. Kowall[83,84], Joel H. Kramer[85], Patricia Kramer[80,86], Walter A. Kukull[87], Frank M. LaFerla[88], James J. Lah[61], Eric B. Larson[38,89], James B. Leverenz[90], Allan I. Levey[61], Ge Li[91], Andrew P. Lieberman[92], Chiao-Feng Lin[29], Oscar L. Lopez[78], Kathryn L. Lunetta[76], Constantine G. Lyketsos[93], Wendy J. Mack[94], Daniel C. Marson[36], Eden R. Martin[13,14], Frank Martiniuk[95], Deborah C. Mash[96], Eliezer Masliah[55,97], Richard Mayeux[70, 109, 110], Wayne C. McCormick[38], Susan M. McCurry[98], Andrew N. McDavid[31], Ann C. McKee[83,84], Marsel Mesulam[99], Bruce L. Miller[23], Carol A. Miller[100], Joshua W. Miller[75], Thomas J. Montine[90], John C. Morris[28, 101], Jill R. Murrell[50, 60], Amanda J. Myers[40], Adam C. Naj[13], John M. Olichney[44], Vernon S. Pankratz[102], Joseph E. Parisi[103,104], Margaret A. Pericak-Vance[13, 14], Elaine Peskind[91], Ronald C. Petersen[21], Aimee Pierce[39], Wayne W. Poon[46], Huntington Potter[30], Joseph F. Quinn[80], Ashok Raj[30], Murray Raskind[91], Eric M. Reiman[37,105-107], Barry Reisberg[53,108], Christiane Reitz[70,109,110], John M. Ringman[5], Erik D. Roberson[36], Ekaterina Rogaeva[111], Howard J. Rosen[23], Roger N. Rosenberg[112], Mary Sano[26], Andrew J. Saykin[50,113], Gerard D. Schellenberg[29], Julie A. Schneider[10,114], Lon S. Schneider[35,115], William W. Seeley[23], Amanda G. Smith[30], Joshua A. Sonnen[90], Salvatore Spina[60], Peter St George-Hyslop[111,116], Robert A. Stern[83], Rudolph E. Tanzi[63], John Q. Trojanowski[29], Juan C. Troncoso[117], Debby W. Tsuang[91], Otto Valladares[29], Vivianna M. Van Deerlin[29], Linda J. Van Eldik[118], Badri N.

Vardarajan[7], Harry V. Vinters[5,119], Jean Paul Vonsattel[120], Li-San Wang[29], Sandra Weintraub[99], Kathleen A. Welsh-Bohmer[24, 121], Jennifer Williamson[70], Randall L. Woltjer[122], Clinton B. Wright[123], Steven G. Younkin[33], Chang-En Yu[38], Lei Yu[10]

[1]Department of Neurology, Johns Hopkins University, Baltimore, Maryland, [2]Department of Neurology, University of Michigan, Ann Arbor, Michigan, [3]Geriatric Research, Education and Clinical Center (GRECC), VA Ann Arbor Healthcare System (VAAAHS), Ann Arbor, Michigan, [4]Michigan Alzheimer Disease Center, Ann Arbor, Michigan, [5]Department of Neurology, University of California Los Angeles, Los Angeles, California, [6]Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, [7]Department of Medicine (Genetics Program), Boston University, Boston, Massachusetts, [8]Department of Pharmacology and Neuroscience, University of North Texas Health Science Center, Fort Worth, Texas, [9]Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, [10]Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, [11]Department of Behavioral Sciences, Rush University Medical Center, Chicago, Illinois, [12]Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Phoenix, Arizona, [13]The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida, [14]Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida, [15]National Alzheimer's Coordinating Center, University of Washington, Seattle, Washington, [16]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, [17]Department of Pathology, Northwestern University, Chicago, Illinois, [18]Department of Neurology, University of Washington, Seattle, Washington, [19]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, [20]Department of Psychiatry,

Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, [21]Department of Neurology, Mayo Clinic, Rochester, Minnesota, [22]Swedish Medical Center, Seattle, Washington, [23]Department of Neurology, University of California San Francisco, San Francisco, California, [24]Department of Medicine, Duke University, Durham, North Carolina, [25]Department of Neuroscience, Mount Sinai School of Medicine, New York, New York, [26]Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, [27]Departments of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, [28]Department of Pathology and Immunology, Washington University, St. Louis, Missouri, [29]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, [30]USF Health Byrd Alzheimer's Institute, University of South Florida, Tampa, Florida, [31]Fred Hutchinson Cancer Research Center, Seattle, Washington, [32]Department of Psychiatry, Vanderbilt University, Nashville, Tennessee, [33]Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, [34]Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama, [35]Department of Neurology, University of Southern California, Los Angeles, California, [36]Department of Neurology, University of Alabama at Birmingham, Birmingham, Alabama, [37]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona, [38]Department of Medicine, University of Washington, Seattle, Washington, [39]Department of Neurology, University of California Irvine, Irvine, California, [40]Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida, [41]Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, Missouri, [42]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology & Psychiatry, Brigham and Women's Hospital

and Harvard Medical School, Boston, Massachusetts, [43]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, [44]Department of Neurology, University of California Davis, Sacramento, California, [45]University of Virginia School of Medicine, Charlottesville, Virginia, [46]Institute for Memory Impairments and Neurological Disorders, University of California Irvine, Irvine, California, [47]Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, Florida, [48]Department of Neurology, Mayo Clinic, Jacksonville, Florida, [49]Rush Institute for Healthy Aging, Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, [50]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, [51]Department of Neurology, Indiana University, Indianapolis, Indiana, [52]Department of Epidemiology, Boston University, Boston, Massachusetts, [53]Department of Psychiatry, New York University, New York, New York, [54]C.S. Kubik Laboratory for Neuropathology, Massachusetts General Hospital, Charlestown, Massachusetts, [55]Department of Neurosciences, University of California San Diego, La Jolla, California, [56]Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, [57]Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia, [58]Emory Alzheimer's Disease Center, Emory University, Atlanta, Georgia, [59]Neurogenetics Program, University of California Los Angeles, Los Angeles, California, [60]Department of Pathology and Laboratory Medicine, Indiana University, Indianapolis, Indiana, [61]Department of Neurology, Emory University, Atlanta, Georgia, [62]Division of Genetics, Department of Medicine and Partners Center for Personalized Genetic Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, [63]Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts, [64]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville,

Tennessee, [65]Vanderbilt Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, [66]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, [67]Department of Pathology (Neuropathology), University of Pittsburgh, Pittsburgh, Pennsylvania, [68]Institute of Neurology, University College London, Queen Square, London, [69]Sanders-Brown Center on Aging, Department of Molecular and Biomedical Pharmacology, University of Kentucky, Lexington, Kentucky, [70]Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Neurology, Columbia University, New York, New York, [71]Department of Pathology, Duke University, Durham, North Carolina, [72]Department of Genome Sciences, University of Washington, Seattle, Washington, [73]Department of Medicine (Medical Genetics), University of Washington, Seattle, Washington, [74]Sanders-Brown Center on Aging, Department Neurology, University of Kentucky, Lexington, Kentucky, [75]Department of Pathology and Laboratory Medicine, University of California Davis, Sacramento, California, [76]Department of Biostatistics, Boston University, Boston, Massachusetts, [77]Department of Ophthalmology, Boston University, Boston, Massachusetts, [78]University of Pittsburgh Alzheimer's Disease Research Center, Pittsburgh, Pennsylvania, [79]Department of Biology, Brigham Young University, Provo, Utah, [80]Department of Neurology, Oregon Health & Science University, Portland, Oregon, [81]Department of Neurology, Portland Veterans Affairs Medical Center, Portland, Oregon, [82]Department of Pathology and Laboratory Medicine, University of California Irvine, Irvine, California, [83]Department of Neurology, Boston University, Boston, Massachusetts, [84]Department of Pathology, Boston University, Boston, Massachusetts, [85]Department of Neuropsychology, University of California San Francisco, San Francisco, California, [86]Department of Molecular & Medical Genetics, Oregon Health & Science University, Portland, Oregon, [87]Department of Epidemiology, University of Washington, Seattle,

Washington, [88]Department of Neurobiology and Behavior, University of California Irvine, Irvine, California, [89]Group Health Research Institute, Group Health, Seattle, Washington, [90]Department of Pathology, University of Washington, Seattle, Washington, [91]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, [92]Department of Pathology, University of Michigan, Ann Arbor, Michigan, [93]Department of Psychiatry, Johns Hopkins University, Baltimore, Maryland, [94]Department of Preventive Medicine, University of Southern California, Los Angeles, California, [95]Department of Medicine - Pulmonary, New York University, New York, New York, [96]Department of Neurology, University of Miami, Miami, Florida, [97]Department of Pathology, University of California San Diego, La Jolla, California, [98]School of Nursing Northwest Research Group on Aging, University of Washington, Seattle, Washington, [99]Cognitive Neurology and Alzheimer's Disease Center, Northwestern University, Chicago, Illinois, [100]Department of Pathology, University of Southern California, Los Angeles, California, [101]Department of Neurology, Washington University, St. Louis, Missouri, [102]Department of Biostatistics, Mayo Clinic, Rochester, Minnesota, [103]Department of Anatomic Pathology, Mayo Clinic, Rochester, Minnesota, [104]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, [105]Arizona Alzheimer's Consortium, Phoenix, Arizona, [106]Department of Psychiatry, University of Arizona, Phoenix, Arizona, [107]Banner Alzheimer's Institute, Phoenix, Arizona, [108]Alzheimer's Disease Center, New York University, New York, New York, [109]Gertrude H. Sergievsky Center, Columbia University, New York, New York, [110]Department of Neurology, Columbia University, New York, New York, [111]Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, Ontario, [112]Department of Neurology, University of Texas Southwestern, Dallas, Texas, [113]Department of Radiology and Imaging Sciences, Indiana University, Indianapolis, Indiana, [114]Department of

Pathology (Neuropathology), Rush University Medical Center, Chicago, Illinois, [115]Department of Psychiatry, University of Southern California, Los Angeles, California, [116]Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, [117]Department of Pathology, Johns Hopkins University, Baltimore, Maryland, [118]Sanders-Brown Center on Aging, Department of Anatomy and Neurobiology, University of Kentucky, Lexington, Kentucky, [119]Department of Pathology & Laboratory Medicine, University of California Los Angeles, Los Angeles, California, [120]Taub Institute on Alzheimer's Disease and the Aging Brain, Department of Pathology, Columbia University, New York, New York, [121]Department of Psychiatry & Behavioral Sciences, Duke University, Durham, North Carolina, [122]Department of Pathology, Oregon Health & Science University, Portland, Oregon, [123]Evelyn F. McKnight Brain Institute, Department of Neurology, Miller School of Medicine, University of Miami, Miami, Florida

References

Almeida, OP, GK Hulse, D Lawrence, L Flicker. 2002. Smoking as a risk factor for Alzheimer's disease: contrasting evidence from a systematic review of case-control and cohort studies. Addiction 97:15-28.

Alzheimer's Association, As. 2012. Alzheimer's Association Annual Report: Alzheimer's disease Facts and Figures.

Ballard, C, S Gauthier, A Corbett, C Brayne, D Aarsland, E Jones. 2011. Alzheimer's disease. Lancet.

Biffi, A, CD Anderson, RS Desikan, M Sabuncu, L Cortellini, N Schmansky, D Salat, J Rosand. 2010. Genetic variation and neuroimaging measures in Alzheimer disease. Archives of neurology 67:677-685.

Bradfield, JP, HR Taal, NJ Timpson, et al. 2012. A genome-wide association meta-analysis identifies new childhood obesity loci. Nature genetics 44:526-531.

Brookmeyer, R, E Johnson, K Ziegler-Graham, HM Arrighi. 2007. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 3:186-191.

Carrasquillo, MM, O Belbin, TA Hunter, et al. 2010. Replication of CLU, CR1, and PICALM associations with alzheimer disease. Arch Neurol 67:961-964.

Cataldo, JK, JJ Prochaska, SA Glantz. 2010. Cigarette smoking is a risk factor for Alzheimer's Disease: an analysis controlling for tobacco industry affiliation. J Alzheimers Dis 19:465-480.

Choi, M, UI Scholl, W Ji, et al. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America 106:19096-19101.

Combarros, O, M Cortina-Borja, AD Smith, DJ Lehmann. 2009. Epistasis in sporadic Alzheimer's disease. Neurobiology of aging 30:1333-1349.

Corder, EH, AM Saunders, NJ Risch, et al. 1994. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet 7:180-184.

Corder, EH, AM Saunders, WJ Strittmatter, DE Schmechel, PC Gaskell, GW Small, AD Roses, JL Haines, MA Pericak-Vance. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261:921-923.

Corneveaux, JJ, AJ Myers, AN Allen, et al. 2010. Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. Hum Mol Genet 19:3295-3301.

Ehret, GBPB MunroeKM Rice, et al. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478:103-109.

Ferreira, MA, MC Matheson, DL Duffy, et al. 2011. Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet 378:1006-1014.

Goate, A, MC Chartier-Harlin, M Mullan, et al. 1991. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature 349:704-706.

Guerreiro, R, A Wojtas, J Bras, et al. 2012. TREM2 Variants in Alzheimer's Disease. The New England journal of medicine.

Haiman, CAGK ChenCM Vachon, et al. 2011. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nature genetics 43:1210-1214.

Harold, D, R Abraham, P Hollingworth, et al. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41:1088-1093.

Herrup, K. 2010. Reimagining Alzheimer's disease--an age-based hypothesis. J Neurosci 30:16755-16762.

Hollingworth, PD HaroldR Sims, et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 43:429-435.

Hu, X, E Pickering, YC Liu, S Hall, H Fournier, E Katz, B Dechairo, S John, P Van Eerdewegh, H Soares. 2011. Meta-analysis for genome-wide association study identifies multiple variants at the BIN1 locus associated with late-onset Alzheimer's disease. PLoS One 6:e16616.

The International HapMap Consortium. 2003. The International HapMap Project. Nature 426:789-796.

Jenuth, JP. 2000. The NCBI. Publicly available tools and resources on the Web. Methods in molecular biology 132:301-312.

Jonsson, T, JK Atwal, S Steinberg, et al. 2012a. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488:96-99.

Jonsson, T, H Stefansson, DS Ph, et al. 2012b. Variant of TREM2 Associated with the Risk of Alzheimer's Disease. The New England journal of medicine.

Jostins, LS RipkeRK Weersma, et al. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491:119-124.

Kilpinen, H, JC Barrett. 2013. How next-generation sequencing is transforming complex disease genetics. Trends in genetics : TIG 29:23-30.

Krebs, CE, C Paisan-Ruiz. 2012. The use of next-generation sequencing in movement disorders. Frontiers in genetics 3:75.

Lambert, JC, S Heath, G Even, et al. 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41:1094-1099.

Lee, SH, NR Wray, ME Goddard, PM Visscher. 2011. Estimating missing heritability for disease from genome-wide association studies. American journal of human genetics 88:294-305.

Levy-Lahad, E, W Wasco, P Poorkaj, et al. 1995. Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 269:973-977.

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

Lill, CM, JT Roehr, MB McQueen, et al. 2012. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS genetics 8:e1002548.

Lindsay, J, D Laurin, R Verreault, R Hebert, B Helliwell, GB Hill, I McDowell. 2002. Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. Am J Epidemiol 156:445-453.

Mackay, TF. 2004. The genetic architecture of quantitative traits: lessons from Drosophila. Current opinion in genetics & development 14:253-257.

Manichaikul, A, JC Mychaleckyj, SS Rich, K Daly, M Sale, WM Chen. 2010. Robust relationship inference in genome-wide association studies. Bioinformatics 26:2867-2873.

Naj, ACG JunGW Beecham, et al. 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat Genet 43:436-441.

Patterson, C, J Feightner, A Garcia, C MacKnight. 2007. General risk factors for dementia: a systematic evidence review. Alzheimers Dement 3:341-347.

Patterson, C, JW Feightner, A Garcia, GY Hsiung, C MacKnight, AD Sadovnick. 2008. Diagnosis and treatment of dementia: 1. Risk assessment and primary prevention of Alzheimer disease. CMAJ 178:548-556.

Podewils, LJ, E Guallar, LH Kuller, LP Fried, OL Lopez, M Carlson, CG Lyketsos. 2005. Physical activity, APOE genotype, and dementia risk: findings from the Cardiovascular Health Cognition Study. Am J Epidemiol 161:639-651.

Psychiatric GWAS Consortium Bipolar Disorder Working Group. 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nature genetics 43:977-983.

Purcell, S, B Neale, K Todd-Brown, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575.

Querfurth, HW, FM LaFerla. 2010. Alzheimer's disease. N Engl J Med 362:329-344.

Roach, JC, G Glusman, AF Smit, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636-639.

Rope, AF, K Wang, R Evjenth, et al. 2011. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. American journal of human genetics 89:28-43.

Scarmeas, N, Y Stern, R Mayeux, JA Luchsinger. 2006. Mediterranean diet, Alzheimer disease, and vascular mediation. Arch Neurol 63:1709-1717.

Shao, H, LC Burrage, DS Sinasac, et al. 2008. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. Proceedings of the National Academy of Sciences of the United States of America 105:19910-19914.

Sherrington, R, EI Rogaev, Y Liang, et al. 1995. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375:754-760.

Singleton, A, J Hardy. 2011. A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. Human molecular genetics 20:R158-162.

Turton, JC, J Bullock, C Medway, et al. 2011. Investigating statistical epistasis in complex disorders. Journal of Alzheimer's disease : JAD 25:635-644.

Wain, LVGC VerwoertPF O'Reilly, et al. 2011. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nature genetics 43:1005-1011.

Wang, HX, A Karp, B Winblad, L Fratiglioni. 2002. Late-life engagement in social and leisure activities is associated with a decreased risk of dementia: a longitudinal study from the Kungsholmen project. Am J Epidemiol 155:1081-1087.

Wen, W, YS Cho, W Zheng, et al. 2012. Meta-analysis identifies common variants associated with body mass index in east Asians. Nature genetics 44:307-311.

Wilming, LG, JG Gilbert, K Howe, S Trevanion, T Hubbard, JL Harrow. 2008. The vertebrate genome annotation (Vega) database. Nucleic acids research 36:D753-760.

Yang, J, B Benyamin, BP McEvoy, et al. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature genetics 42:565-569.

Yang, J, SH Lee, ME Goddard, PM Visscher. 2011a. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics 88:76-82.

Yang, J, TA Manolio, LR Pasquale, et al. 2011b. Genome partitioning of genetic variation for complex traits using common SNPs. Nature genetics 43:519-525.

Figure Legends

Figure 1. Unexplained Alzheimer's disease variance, by chromosome. In this figure we show phenotypic variance, by chromosome, explained by all SNPs. Error bars correspond to standard error.

Figure 2. Unexplained Alzheimer's disease variance, by chromosome, excluding known Alzheimer's disease markers. This figure is the same as Figure 1 except we have removed variance explained by known Alzheimer's disease markers. Error bars correspond to standard error.

Figure 3. Variant search space. Real and hypothetical variants are graphed by effect size (y-axis) and population frequency (x-axis). Known Alzheimer's disease SNPs are blue circles and hypothetical SNPs are red squares. The large box on the right outlined with dots, is the GWAS search space and the smaller box on the left, outlined with dashes, is the next-generation sequencing search space. SNPs identified in recent GWAS are located in the labeled oval inside the GWAS search space. Known Alzheimer's disease SNPs are those found in Table 1 as well as APP and TREM2, which are both labeled on the graph.

Table 1, Chapter 5. Late-onset Alzheimer's disease associated genes/variants

Variant	Gene	Abbreviation	Odds Ratio
rs429358	Apolipoprotein E (ε4 allele) (Corder et al. 1993)	APOE	3.685
rs7412	Apolipoprotein E (ε2 allele) (Corder et al. 1994)	APOE	0.621
rs744373	Bridging Integrator 1 (Biffi et al. 2010)	BIN1	1.166
rs11136000	Clusterin (Harold et al. 2009; Lambert et al. 2009)	CLU	0.879
rs3764650	ATP-binding cassette, sub-family A (ABC1), member 7 (Hollingworth et al. 2011)	ABCA7	1.229
rs3818361	Complement component (3b/4b) receptor 1 (Knops blood group) (Lambert et al. 2009)	CR1	1.174
rs3851179	Phosphatidylinositol binding clathrin assembly protein (Harold et al. 2009)	PICALM	0.879
rs610932	Membrane-spanning 4-domains, subfamily A, member 6A (Hollingworth et al. 2011)	MS4A6A	0.904
rs3865444	CD33 molecule (Hollingworth et al. 2011; Naj et al. 2011)	CD33	0.893
rs670139	Membrane-spanning 4-domains, subfamily A, member 4E (Hollingworth et al. 2011)	MS4A4E	1.079
rs9296559	CD2-associated protein (Hollingworth et al. 2011; Naj et al. 2011)	CD2AP	1.117

The dbSNP identification number, gene name, gene abbreviation, and odds ratio for each of the top variants from the Alzgene.org meta-analyses. The SNP in CD2AP, rs9349407, is not present in this sample, so we used rs9296559 instead as a proxy. These two SNPs are close together (1108 base pairs apart) and in very high LD (r^2=1).

Table 2, Chapter 5. Demographic information for individuals in the analysis dataset

	Count	Cases / Controls	Age (Cases / Controls)
Male	4489	2378 / 2111	74.8 (73.6 / 76.1)
Female	6433	3330 / 3103	75.3 (74.9 / 75.7)
Total	10922	5708 / 5214	75.1 (74.3 / 75.9)

We report total individuals, sex, case-control status, and average age for the 9,884 individuals analyzed in this report.

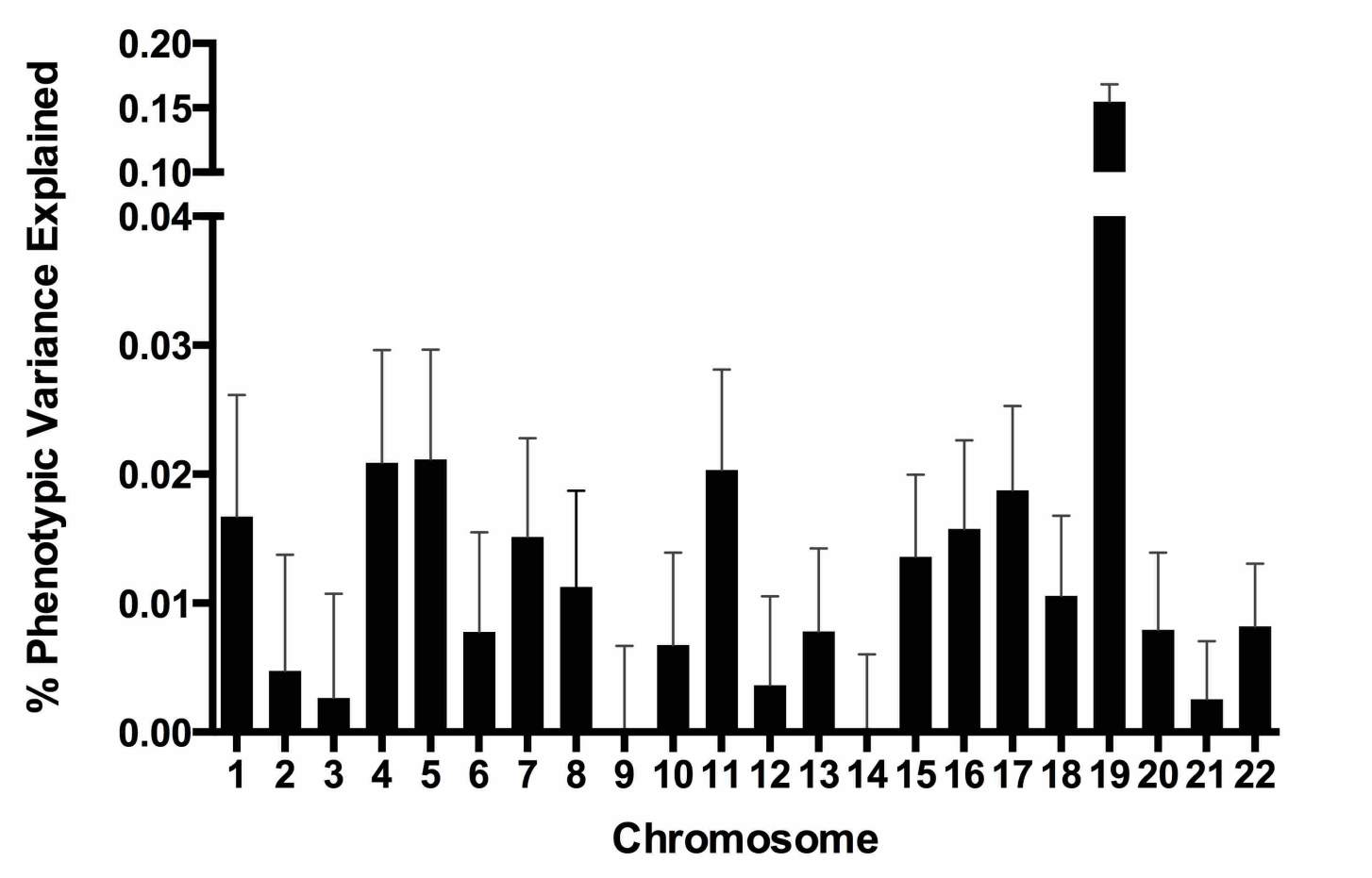


Figure 1, Chapter 5. Unexplained Alzheimer's disease variance, by chromosome. In this figure we show phenotypic variance, by chromosome, explained by all SNPs. Error bars correspond to standard error.

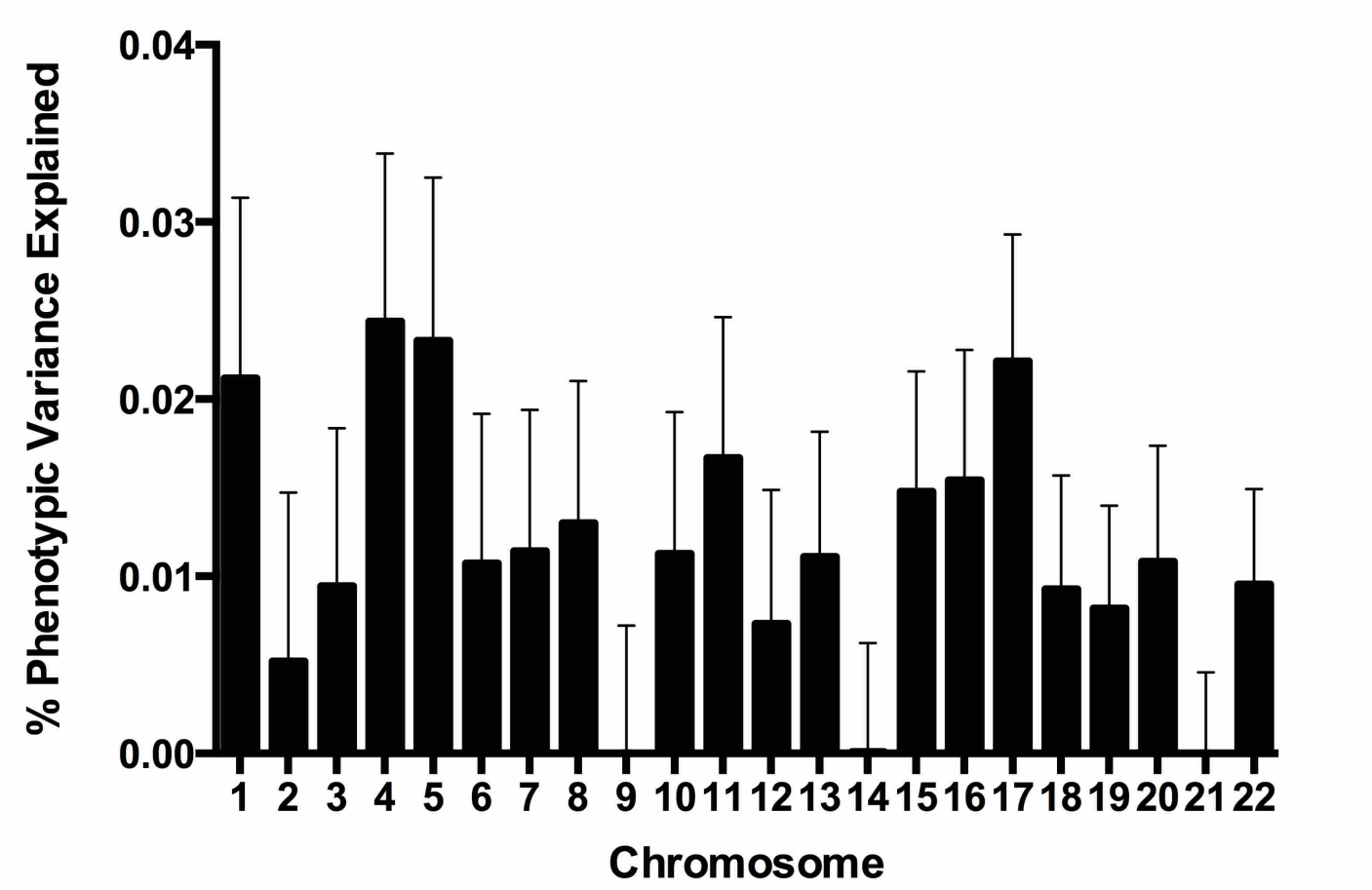


Figure 2, Chapter 5. Unexplained Alzheimer's disease variance, by chromosome, excluding known Alzheimer's disease markers. This figure is the same as Figure 1 except we have removed variance explained by known Alzheimer's disease markers. Error bars correspond to standard error.

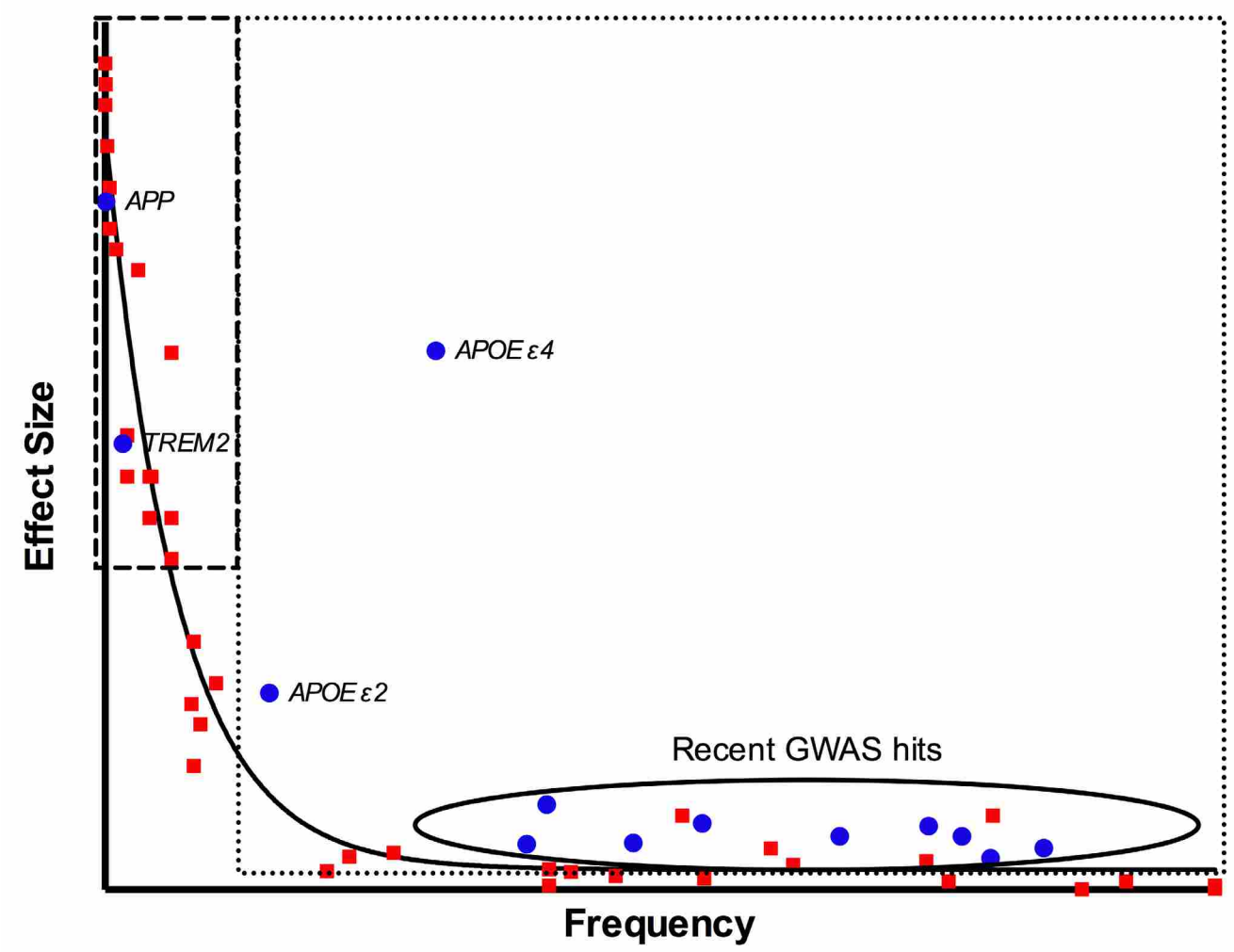


Figure 3, Chapter 5. Variant search space. Real and hypothetical variants are graphed by effect size (y-axis) and population frequency (x-axis). Known Alzheimer's disease SNPs are blue circles and hypothetical SNPs are red squares. The large box on the right outlined with dots, is the GWAS search space and the smaller box on the left, outlined with dashes, is the next-generation sequencing search space. SNPs identified in recent GWAS are located in the labeled oval inside the GWAS search space. Known Alzheimer's disease SNPs are those found in Table 1 as well as APP and TREM2, which are both labeled on the graph.

Chapter 6

Future Directions

Perry G. Ridge

The goal of the research presented in this dissertation was to further our understanding of the genetics of Alzheimer's disease (AD). In chapters 1 and 5 we pointed out that while there has been significant research into the genetics of AD, there remains significant phenotypic variance not explained by the 11 known AD disease markers. We also provided evidence that the remaining variance is explained by markers spread across many different chromosomes. Finally, we provided evidence for a role of the mitochondrial genome (mtDNA) in AD risk and physiology (Chapter 1 Table 2).

Recently there have been successes in identifying rare, large effect AD variants (Guerreiro et al. 2012; Jonsson et al. 2012a; Jonsson et al. 2012b). Like these studies, future discoveries will have to be made using innovative strategies. It is likely that the undiscovered AD variants are extremely rare. Genome wide association studies (GWAS) are limited in their power to test rare variants and are not likely to identify these disease associations. Family based studies using whole genome sequencing may be necessary to find these variants.

While applying new research tools to identify novel AD nuclear markers is of great importance, the focus of the work presented here has been the mitochondrial genome. We have applied more robust analysis methods and more complete genomic datasets than have been used in the past to study the role of mtDNA in AD. Our results add evidence for mitochondria's role in AD risk and

neurodegeneration; however, there are additional steps that should be taken to strengthen our discoveries. First, it is necessary to replicate our findings in independent datasets. In Chapters 2 and 3 we used full mtDNA from the Cache County Study on Memory Health and Aging to look for association between mitochondrial haplotypes and two phenotypes; risk for AD and mitochondrial copy number. In both cases, this is the first research that used whole mitochondrial genomes and is one of the largest samples ever to be tested. For each of these projects we anticipate the ability to use data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data for replication. Case-control status is already available for all individuals in ADNI and we are in the process of measuring mitochondrial copy number for the entire dataset. It has been announced that whole genome sequence for the full ADNI dataset will be released to qualified scientists. We will mine this data for the mitochondrial genomic data needed to validate our initial reports of association between mtDNA and both mitochondrial copy number and AD risk.

Besides replication there are a number of *in silico* bioinformatics analyses we can perform. In the studies using the Cache County data we have information about heteroplasmy; however, in these studies we ignored the heteroplasmy and selected only the most common of the two alleles. An important next step is to redo our analysis while incorporating heteroplasmy into our models. Besides the observed variants it is possible there are low-level heteroplasmic variants segregating with the phenotype. Next, we can check each of the variant sites to see if the substitution resulted in the addition or deletion of a likely (or known) protein modification site (phosphorylation, ubiquitination, etc.). Finally, while the mtDNA encodes 13 protein-coding genes, many additional nuclear genes encode proteins that are transported into the mitochondria (mitochondrial proteins). The exact number of mitochondrial proteins is not known, and is likely

tissue specific. Estimates range from ~600 to more than 1000 (Taylor et al. 2003; Elstner et al. 2008). An important next step is to analyze the mitochondrial proteins encoded in the nuclear genome for variation that explains the observed phenotypes.

In Chapter 4 we used 138 genotyped SNPs from the ADNI dataset to look for association with known neurodegenerative biomarkers of AD. Replication in this dataset is more difficult for several reasons. We found clades associated with excess neurodegeneration, but since we only used 138 SNPs it is not possible to know if the observed neurodegeneration is actually driven by the genotyped SNPs that define those clades, or if the phenotype is being driven by an ungenotyped variant. In two associated clades we were able to assign individuals to probable mitochondrial haplogroups based on the 138 SNPs; however, the data are incomplete. The next step to take for this particular study is to wait for the completion of the ADNI genomes and then redo the analyses performed here using full mtDNA. We can then work to identify other datasets for replication and apply the strategies concerning *in silico* analyses and heteroplasmy explained above.

In conclusion, AD is a heterogeneous disorder. Mitochondria are morphologically damaged and have lower oxidation rates than AD (Swerdlow, Khan 2004; Swerdlow, Burns, Khan 2010). Risk for AD is higher in individuals with a maternal family history of AD (Honea et al. 2012). Together these data suggest a role for mtDNA in the risk for AD and known AD endophenotypes (Chapter 1 Table 2). Our research has added to this body of knowledge and the continued research into these findings will provide insights to the genetics of this deadly disease.

References

Elstner, M, C Andreoli, U Ahting, I Tetko, T Klopstock, T Meitinger, H Prokisch. 2008. MitoP2: an integrative tool for the analysis of the mitochondrial proteome. Molecular biotechnology 40:306-315.

Guerreiro, R, A Wojtas, J Bras, et al. 2012. TREM2 Variants in Alzheimer's Disease. The New England journal of medicine.

Honea, RA, ED Vidoni, RH Swerdlow, JM Burns. 2012. Maternal family history is associated with Alzheimer's disease biomarkers. Journal of Alzheimer's disease : JAD 31:659-668.

Jonsson, T, JK Atwal, S Steinberg, et al. 2012a. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488:96-99.

Jonsson, T, H Stefansson, DS Ph, et al. 2012b. Variant of TREM2 Associated with the Risk of Alzheimer's Disease. The New England journal of medicine.

Swerdlow, RH, JM Burns, SM Khan. 2010. The Alzheimer's disease mitochondrial cascade hypothesis. J Alzheimers Dis 20 Suppl 2:S265-279.

Swerdlow, RH, SM Khan. 2004. A "mitochondrial cascade hypothesis" for sporadic Alzheimer's disease. Med Hypotheses 63:8-20.

Taylor, SW, E Fahy, B Zhang, et al. 2003. Characterization of the human heart mitochondrial proteome. Nature biotechnology 21:281-286.

Appendix 1

Supplementary Figures and Tables for Chapter 2

Supplementary Figure Legends

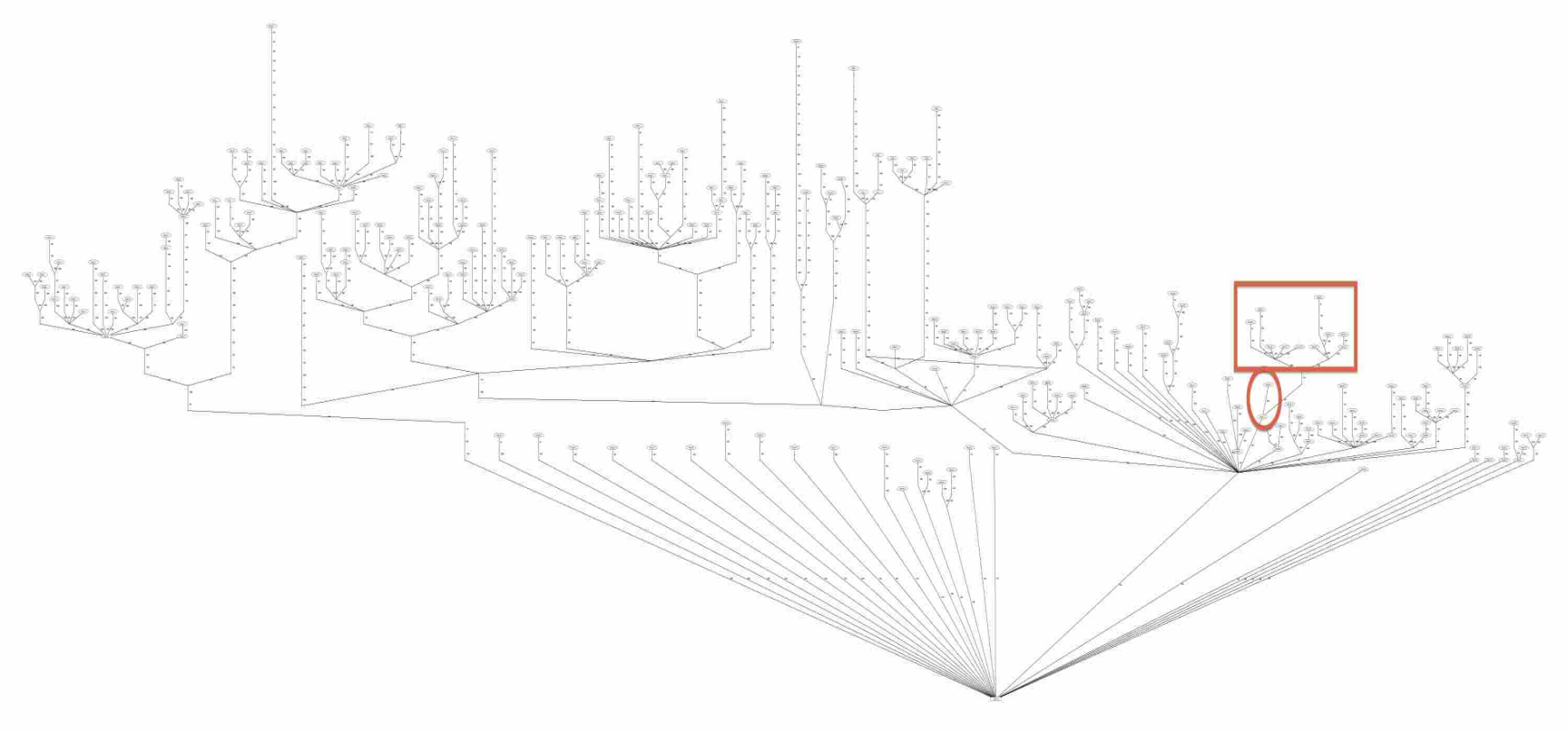
Supplementary Figure 1. Haplotype network. Here we show our entire haplotype network based on all the full mitochondrial genome sequences in our dataset, and have enclosed the clades corresponding to branches 269 (red square) and 270 (red square and circle).

Supplementary Table 1. Major mitochondrial haplogroups, clusters, and sub-haplogroups

H (424)	J (99)	U (147)	K (95)	T (121)	I (21)	W (20)	V (34)	HV (18)	X (8)	C (5)	L (4)	Missing (11)
H (49)	J1b1a (3)	U1b (2)	K1a (15)	T1 (4)	I (7)	W1 (8)	V (29)	HV (2)	X2b (5)	C4a1 (5)	L3e5 (4)	
			K1a1b					HV0b'c				
H1 (50)	J1c (2)	U2e (2)	(13)	T1a (6)	I1a1 (2)	W1c (1)	V1 (3)	(5)	X2c (3)			
		U2e1a										
H1a (17)	J1c1 (4)	(2)	K1a2 (6)	T1a1 (2)	I2 (11)	W3 (7)	V2 (2)	HV9 (11)				
H1a1	J1c1b1a		K1a3a1									
(10)	(6)	U3b (1)	(20)	T2 (15)	I4 (1)	W4 (4)						
H1b (23)	J1c2 (2)	U4a1 (3)	K1a4 (1)	T2a (9)								
		U4a1b	K1a4a1									
H1c (28)	J1c2b (3)	(3)	(17)	T2b (73)								
	J1c2b1a	U4a2	K1b1a									
H1c1 (2)	(4)	(15)	(6)	T2c (3)								
H1e (13)	J1c3 (4)	U4b1 (2)	K1b2 (1)	T2e (9)								
	J1c3a1		K1c1b									
H1g (7)	(2)	U4c (6)	(6)									
H2a2a												
(4)	J1c3b (1)	U5a (9)	K1c2 (2)									
	J1c3b1	U5a1										
H2a3 (3)	(11)	(32)	K2b (8)									
H3 (21)	J1c5 (12)	U5b1 (3)										
		U5b1b										
H3d (1)	J1c5a (4)	(7)										
		U5b1d										
H4a (29)	J1c6 (4)	(9)										
H4a1		U5b2										
(12)	J1c8 (7)	(45)										
H4a1a1	J2a1a	U5b3f										
(2)	(22)	(1)										
H4a1b	J2a1a1b											
(1)	(8)	U8a1 (5)										
H5 (13)												
H5a (10)												
H5a1												

(21)
H6a1a (13)
H6a1b (29)
H7 (10)
H7a (1)
H10 (11)
H11 (15)
H13 (24)
H13a1a (1)
H13a1a1 (2)
H15 (1)
H17 (1)

Here we have listed the haplogroup membership of all 1007 individuals in our dataset. The numbers inside the parentheses are the number of individuals with a major mitochondrial sub-haplogroup and the first column shows the membership in each of the represented major mitochondrial haplogroups in our dataset.



Supplementary Figure 1. Haplotype network. Here we show our entire haplotype network based on all the full mitochondrial genome sequences in our dataset, and have enclosed the clades corresponding to branches 269 (red square) and 270 (red square and circle).

Appendix 2

Supplementary Figures and Tables for Chapter 3

Supplementary Figure Legends

Supplementary Figure 1. Haplotype network. Our haplotype network was constructed using TCS and 285 full mitochondrial genomes. The arrows point to each of the three branches representing the significant contrasts. The blue arrow points to branch 124, the red to branch 121, and the green to branch 50.

Supplementary Figure 2. Haplotype network. We collapsed our haplotype network (Supplementary Figure 1) into nodes corresponding to the major mitochondrial haplogroups present in our network.

Supplementary Table 1. Mitochondrial copy number measurements

Cache ID	Mitochondria #	Age	Gender
4	2.8	75.28474428	female
6	1.84	76.87000329	male
9	1.99	75.69543314	female
10	2.48	68.92180484	male
11	5.18	68.77669478	female
16	2.58	69.16821816	male
17	N/A	78.0308838	female
18	N/A	74.5427664	male
19	2.61	74.14576717	female
26	1.86	70.63300843	male
30	3.36	88.6540357	female
31	2.15	68.93001862	male
32	2.72	70.19494031	female
34	3.21	86.909977	female
37	2.23	78.56477932	male
41	3.23	77.06713394	female
45	2.34	67.61033841	female
54	2.93	76.70025189	male
55	N/A	78.12944913	male
59	4.4	77.35461614	female
65	1.59	88.31727084	female
66	N/A	83.63541781	female
67	N/A	77.08082357	female
72	3.48	91.74515387	male
73	N/A	90.05585369	female
78	1.36	74.58109736	male
86	1.43	73.05333479	female
87	2.51	78.30467638	male
100	3.71	85.73266893	female
101	3.32	71.00262841	male
125	2.45	84.8154638	female
130	4.79	67.23524258	female
131	3.19	90.50761143	male
137	2.57	69.45296244	male
138	2.11	81.03165042	male
140	4.84	81.36841529	female
142	N/A	72.43182565	female
144	2.58	79.05486803	male
147	2.08	65.97579674	female
162	1.95	87.50410689	female

167	2.26	67.24619428	female
168	3.91	88.01062315	female
176	3.09	71.30653817	female
185	1.61	66.71503669	male
189	2.45	84.58274012	female
194	3.38	66.82181579	female
195	2.76	69.12167342	female
197	1.17	80.12813492	female
200	N/A	68.31398532	male
202	4.11	71.55568941	male
203	2.68	75.49008871	male
225	N/A	85.63957945	female
230	N/A	87.32614172	female
232	1.83	77.07808564	female
237	3.64	73.27510678	female
244	2.17	81.66137334	male
252	2.61	94.01215639	female
264	2.73	74.21147738	male
275	2.56	69.93483737	female
279	3.27	75.10404118	female
280	2.97	73.20939656	male
289	3.61	72.93286606	male
294	2.63	86.28572993	male
295	2.46	70.89311138	female
298	2.47	67.76092432	male
309	2.09	67.71985544	female
316	3.78	69.948527	female
333	3.03	87.86277516	female
339	N/A	81.305443	female
346	2.73	99.15124302	female
348	3.41	84.59095389	male
358	3.47	77.8501807	female
361	N/A	89.79848866	female
365	2.87	66.54528529	female
366	2.34	68.91359106	male
367	2.23	72.08410908	female
374	3.06	74.56466981	male
375	3.89	70.44682948	female
382	2.99	69.51867265	male
384	2.57	74.13207754	female
389	1.5	69.12714927	male
390	3.56	67.22976673	male
393	3.22	71.13131092	female

400	3.23	73.11083123	female
405	1.95	73.27236885	male
408	3.5	65.61986639	female
412	2.22	70.73431169	female
417	2.09	87.24126602	female
426	2.08	71.79388895	male
431	2.94	67.77735188	male
443	3.31	73.92399518	female
446	0.94	88.77176651	female
454	N/A	71.19154529	male
455	3.69	70.5864637	female
460	3.93	90.06406746	female
465	3.19	67.20786332	male
466	2.93	84.45131968	male
469	1.79	74.75632461	male
470	1.93	70.55908444	male
471	0.99	66.75610557	male
472	2.82	86.90450115	female
477	2.51	71.11488336	male
490	N/A	80.51692038	male
510	2.88	75.59960574	female
515	N/A	68.77669478	female
516	4.1	70.17851276	male
525	2.4	91.93133282	female
529	2.77	72.28671558	male
531	2.17	69.0340598	female
537	2.35	78.73453072	female
541	1.48	72.57967364	female
543	2.88	67.56105574	male
548	3.1	71.10666959	female
550	3.17	74.41134596	male
552	2.72	89.94633666	male
557	1.94	71.60223415	female
559	2.36	85.76826196	male
560	3.33	83.00295696	female
561	2.32	75.16153762	male
566	3.25	67.23250465	male
567	2.45	77.42306429	male
568	N/A	76.44014894	female
582	4.3	81.76541452	female
587	2.96	80.62917534	male
590	4.05	76.69203811	female
592	2.15	75.26831672	female

596	2.33	72.46741868	male
598	2.2	68.07030993	female
607	N/A	71.0656007	female
611	N/A	75.56948856	female
624	3.8	82.2199102	female
625	3.07	89.31113788	female
630	2.3	78.06100099	male
633	1.82	70.39480889	female
640	2.44	82.88248823	female
645	N/A	75.06023437	male
646	2.05	71.70627533	male
649	2.52	68.40433687	male
652	3.12	67.9607929	male
653	N/A	96.6925857	female
655	N/A	88.22965721	male
657	2.83	86.33227467	male
658	1.96	84.69225715	female
660	3.33	77.98433906	male
665	3.98	72.54134268	female
668	2.04	90.4610667	male
675	3.08	86.92640456	male
684	N/A	72.44551528	female
685	2.35	76.77417588	female
688	1.46	74.39218048	female
689	3.24	75.59412989	male
705	3.66	75.54484722	male
709	2.24	67.74723469	female
711	3.99	96.09571788	male
732	2.39	65.89913482	male
744	N/A	68.95739788	male
752	3.12	65.81973497	male
753	3	89.66706823	male
755	2.45	72.49753587	male
759	1.57	76.46479027	male
771	2.92	84.11181689	female
774	1.05	78.32931771	female
777	2.27	75.63519877	female
792	N/A	74.49895959	female
807	2.25	77.86660826	male
813	3.42	74.45241485	female
822	3.46	82.23633775	male
832	6.12	66.46040959	male
837	3.86	74.73168328	male

841	3.38	90.94294163	female
847	1.54	75.4462819	male
853	3.31	68.88894973	female
858	N/A	90.26667397	male
868	1.24	66.91216734	female
869	2.43	67.37213887	female
872	1.98	67.21881503	female
873	N/A	82.56488884	female
874	2.48	81.17676049	female
875	3.29	90.50213558	male
879	N/A	68.89990143	female
892	N/A	65.4829701	female
894	2.45	86.14609572	female
904	2.55	85.68064834	female
915	3.44	66.87931223	male
917	2.16	70.01971307	male
919	N/A	67.46249042	male
925	2.86	77.26974044	female
938	2.45	82.38144782	female
950	3.33	70.39754682	male
960	4.06	70.42218815	female
971	N/A	74.60026284	male
973	3.07	72.02661264	female
988	0.68	67.16679444	male
997	2.34	71.92530939	male
998	N/A	69.44474866	female
1000	2.46	68.42624028	female
1002	2.66	69.91840981	female
1005	2.92	70.72883583	female
1016	2.27	68.41255065	female
1017	2.02	66.09626547	female
1032	2.1	75.18891688	female
1053	1.73	69.81710656	male
1054	1.77	68.60694338	female
1064	1.81	69.82258241	female
1069	N/A	65.5678458	female
1074	4	81.89135911	male
1079	2	90.5568941	female
1089	N/A	68.28386814	female
1095	N/A	65.42821159	female
1119	1.98	88.03526448	female
1120	2.12	71.08476618	female
1121	2.54	67.71164166	female

1123	2.75	71.77198554	female
1162	1.46	77.61471909	male
1164	N/A	70.99989048	male
1165	1.39	72.39349469	female
1167	2.19	72.63990801	female
1188	1.74	67.1421531	female
1194	N/A	79.86803198	male
1205	2.1	72.47015661	female
1213	2.54	67.12298762	male
1223	3.21	73.80078852	male
1227	N/A	67.36392509	male
1232	2.45	66.35910634	female
1234	0.81	71.2353521	male
1235	3.53	70.24969883	male
1245	3.15	70.13744387	female
1248	2.68	69.66104479	female
1257	N/A	71.46259993	male
1259	0.76	87.48220348	female
1264	2.41	79.63530829	female
1265	3.4	68.34684043	male
1288	1.41	72.19362611	male
1294	2.93	68.10042712	male
1301	N/A	77.38473333	male
1303	1.56	72.54134268	female
1308	2.78	70.77264265	male
1309	3.5	82.08301391	male
1322	3.71	77.19581645	female
1333	2.01	82.16515168	male
1335	2.12	71.27094513	male
1336	2.8	71.07929033	female
1339	2.85	72.28397766	female
1342	1.52	66.64932647	female
1351	1.63	70.49885007	male
1358	1.94	78.08564232	male
1368	3.53	77.55722265	male
1373	2.57	95.87120797	female
1381	2.4	81.83933852	female
1383	3.5	71.74460629	female
1388	N/A	73.69127149	female
1392	1.4	89.39327565	male
1395	2.4	82.9755777	female
1398	2.45	71.49819297	female
1415	N/A	92.79925528	female

1429	N/A	75.22998576	female
1430	1.99	68.58230205	female
1436	N/A	82.30204797	female
1439	2.85	75.71186069	female
1442	2.01	71.58306867	female
1459	1.61	68.21268207	female
1470	2.87	72.71656993	female
1478	3.54	83.21377724	female
1479	3.46	82.94819844	female
1489	N/A	72.49479794	female
1497	N/A	70.63027051	female
1498	1.03	73.09987953	male
1502	1.03	71.65425474	male
1510	3.15	68.57956412	female
1539	N/A	86.01467528	female
1541	2.15	72.08137115	female
1555	N/A	80.55798927	female
1559	3.08	79.91183879	male
1575	3.37	87.14817654	female
1578	2.95	76.97678239	male
1589	2.31	65.93472785	female
1590	N/A	69.61723798	male
1601	2.67	70.40849852	male
1602	2.25	68.92180484	female
1610	2.28	71.80757858	female
1627	2.57	83.591611	male
1630	3.67	75.68448144	female
1634	2.05	67.28452524	female
1653	3.85	81.23699485	female
1655	3.14	70.17577483	male
1656	2.67	90.66914905	female
1665	2.42	89.51374439	female
1678	3.08	74.67966269	male
1695	1.97	66.67670573	female
1701	1.73	66.03329318	female
1708	2.32	77.64483627	female
1715	3.5	73.69400942	male
1716	3.4	73.94042274	female
1725	2.13	70.55634651	male
1727	3.33	84.35549228	female
1736	N/A	69.80889278	male
1745	2.19	81.88588325	male
1749	2.13	69.46665206	female

1750	2.21	71.15869018	male
1751	1.73	77.55996057	male
1764	2.67	85.40411784	female
1772	3.8	82.55941299	male
1778	N/A	72.13339174	female
1781	N/A	74.95071734	female
1785	1.57	77.88851166	male
1787	2.03	70.6083671	male
1792	1.52	67.79651736	male
1796	3.39	67.06275326	male
1797	2.31	71.94721279	female
1825	N/A	81.53269083	female
1835	N/A	81.58471142	male
1840	1.95	78.66334465	female
1855	3.53	79.80232176	female
1871	2.33	70.7041945	male
1881	2.49	71.94721279	female
1887	3.1	71.15869018	female
1893	2.49	81.40400832	female
1907	N/A	71.8732888	male
1910	3.73	89.96824006	male
1930	2.44	72.98214872	male
1948	2.9	80.09527982	male
1955	3.28	88.27346402	female
1968	N/A	66.04972073	female
1976	3.6	86.72106012	female
1977	N/A	93.83692914	female
1992	3.72	85.40685577	male
1993	1.77	78.6797722	female
2029	2.49	67.18322199	male
2034	2.71	70.68502902	female
2036	2.46	79.0411784	female
2058	2.24	74.03351221	male
2060	N/A	66.44398204	female
2065	2.46	91.80812616	male
2068	3.02	66.93407075	female
2077	N/A	83.12342569	female
2083	2.3	78.78107546	female
2086	0.98	86.78403242	female
2112	3.21	66.6849195	male
2137	1.12	74.86036579	female
2146	2.26	75.29295805	female
2150	N/A	69.93483737	female

2152	3.03	70.37016756	male
2159	2.97	81.50257365	female
2171	4.19	69.45296244	female
2175	2.71	79.34235024	male
2178	3.42	73.56532691	male
2179	N/A	71.63782718	female
2182	3.11	97.42087395	female
2189	N/A	81.11105027	male
2195	N/A	67.6842624	female
2200	2.75	82.79213668	female
2201	3.57	74.22790494	female
2202	N/A	69.75413427	female
2205	2.01	79.5832877	male
2206	N/A	93.11685467	female
2224	0.94	73.05333479	male
2229	4.23	65.52403899	male
2230	3.93	72.21552951	male
2231	2.46	66.64385062	male
2246	4.77	75.95827401	female
2257	2.16	65.58153543	female
2284	1.82	68.33041288	female
2285	1.51	70.69871865	female
2295	N/A	88.47880845	male
2298	1.3	84.17478918	female
2301	2.86	71.40236557	female
2312	3.53	68.70824663	male
2318	3.5	69.90198226	female
2324	N/A	69.28594897	male
2345	4.08	66.15376191	male
2346	2.95	78.2225386	female
2347	2.17	86.54035703	male
2370	1.59	71.84043369	male
2373	N/A	66.05793451	female
2377	3.52	87.73135473	male
2386	2.64	81.74077319	female
2391	1.22	65.5568941	female
2392	2.65	68.1195926	male
2393	3.6	94.62545176	female
2417	2.56	65.93198992	female
2420	3.19	86.09955098	female
2428	N/A	70.39754682	male
2439	3.07	70.21410579	female
2442	1.44	73.41747892	female

2449	1.91	67.16131859	female
2466	2.7	78.87690286	female
2470	2.31	70.44409156	male
2478	N/A	82.0501588	male
2484	2.9	77.7434016	male
2493	1.58	65.81973497	male
2494	3.57	69.31606615	male
2507	2.23	65.72664549	female
2521	N/A	65.61439054	female
2522	1.61	71.20523491	female
2528	3.53	73.9897054	male
2530	3.36	77.34092651	male
2532	N/A	77.1054649	female
2540	N/A	89.03186946	female
2558	5.34	74.35658745	female
2560	2.53	80.14730041	female
2576	2.15	80.93856095	male
2579	2.32	90.90187274	male
2580	2.53	66.18387909	male
2584	2.87	88.87580769	male
2588	3.52	71.24904173	female
2597	3.38	65.84985215	female
2600	2.95	70.02792684	female
2611	0.98	66.95871208	female
2612	1.86	70.48242252	male
2618	2.12	67.42142153	male
2622	2.4	66.77253313	male
2623	2.16	70.36742964	male
2624	N/A	79.03570255	male
2626	N/A	77.30807141	male
2627	2.82	69.98685796	female
2628	1.79	77.06987187	male
2629	3.15	67.41046983	male
2634	2.87	93.00186179	female
2636	N/A	70.62205673	male
2640	N/A	74.35658745	female
2645	N/A	68.42076443	female
2654	N/A	71.60223415	female
2660	2.71	73.89935385	female
2661	N/A	72.65359763	male
2663	2.42	67.28178732	female
2684	2.12	68.10590297	female
2686	3.19	87.04413536	female

2687	3.23	78.03362173	male
2694	3.39	90.35976344	female
2700	1.72	69.30511445	male
2701	N/A	82.76201949	male
2702	2.06	66.15376191	female
2710	N/A	68.71646041	female
2713	3.08	80.45121016	male
2720	1.85	82.48822692	male
2722	2.06	71.1395247	male
2723	4.17	71.49545504	female
2727	1.71	76.07052897	male
2731	4.98	74.75358668	female
2735	3.49	76.37991458	male
2752	4.24	82.38144782	male
2754	1.38	66.22768591	male
2758	2.59	89.51100646	male
2759	1.19	69.93483737	female
2765	4.35	74.87405542	male
2775	2.93	76.34705947	female
2791	3.65	72.79323185	female
2800	3.54	82.62238528	female
2806	N/A	73.38736173	female
2817	2.73	75.30938561	male
2827	N/A	69.19833534	male
2830	2.09	71.01358011	female
2837	1.72	73.80626437	female
2842	4.34	80.0186179	female
2853	N/A	69.87186508	male
2864	0.91	68.78490855	female
2868	1.38	79.52579126	male
2885	1.37	68.09221334	female
2898	2.58	92.29547695	male
2910	3.08	77.05618224	female
2916	2.33	65.60891469	male
2930	1.24	68.74110174	female
2939	3.19	67.24345636	male
2952	N/A	76.52502464	female
2953	3.32	72.17993648	female
2957	N/A	78.85773738	male
2961	N/A	74.46336655	female
2966	2.83	90.81152119	female
2973	N/A	81.37389114	female
2976	2.22	66.89573979	male

2987	3.97	67.82115869	female
2989	2.2	91.22494798	female
2995	N/A	65.98948637	female
3001	1.87	77.62293287	female
3006	4.32	84.41298872	male
3010	2	75.96101194	male
3015	N/A	71.54199978	female
3021	1.84	75.2655788	female
3036	3.07	69.82258241	male
3037	1.96	71.09297996	female
3038	2.33	72.4263498	male
3056	3.94	76.41276969	male
3060	0.87	72.74668711	female
3074	2.21	72.13886759	female
3077	N/A	78.07742854	male
3084	1.48	77.37651955	female
3096	1.78	70.84930457	female
3097	3.36	67.31738035	female
3099	1.88	69.26678348	male
3104	2.34	70.73157376	female
3108	2.81	86.27751615	female
3113	3.57	79.019275	male
3115	3.72	84.54440916	female
3123	N/A	72.85346621	male
3124	2.51	67.07918081	male
3144	3.35	78.24717994	female
3145	1.88	70.48516044	male
3155	4.71	68.3112474	male
3161	1.17	73.77614719	female
3162	2.93	92.48713175	male
3187	3.65	73.83364363	female
3188	1.6	82.06658635	male
3191	3.13	80.38823787	male
3192	1.18	71.57485489	male
3199	2.91	72.82061111	male
3202	2.7	83.03854999	male
3211	N/A	82.38966159	female
3226	5.18	75.06844814	female
3238	3.58	89.30018618	female
3243	N/A	65.82521082	female
3263	1.53	80.70857518	male
3272	2.64	69.34618333	female
3276	3.07	70.90680101	female

3292	2.66	75.88708794	female
3294	0.92	75.49282663	female
3295	3.91	81.18771219	female
3299	2.67	80.71952689	male
3306	N/A	91.3864856	female
3311	1.95	84.13645822	female
3312	4.97	87.25769357	female
3326	2.99	67.51451101	female
3328	2.09	65.37892892	female
3329	1.5	68.0566203	male
3332	1.49	78.39502793	female
3333	3.29	79.46281897	male
3339	3.7	75.83506735	male
3343	2.55	67.01894645	male
3372	N/A	70.24969883	male
3374	3.05	88.76629066	female
3378	0.94	82.15146205	female
3380	1.93	81.5080495	female
3385	2.71	82.4225167	male
3393	4.2	74.10469828	male
3398	2.37	79.86529405	female
3402	N/A	72.97667287	male
3433	N/A	87.68480999	male
3436	N/A	75.27105465	male
3438	3.12	73.7980506	male
3440	3.63	85.19603548	female
3453	2.76	72.24838462	female
3455	1.74	68.65622604	male
3465	1.87	78.28824882	female
3469	4.43	95.09911291	female
3490	3.2	79.85708028	female
3491	N/A	77.97338736	female
3514	N/A	81.95706932	female
3518	1.51	65.94294163	female
3524	2.72	87.81896835	male
3537	1.49	88.76081481	male
3539	2.49	71.38867594	male
3557	N/A	73.63651298	female
3595	2.75	71.35582083	female
3602	2.41	70.12649217	male
3604	2.37	75.60781952	female
3607	4.88	74.33468404	female
3613	2.59	68.09495126	female

3644	4.56	66.25506516	female
3645	N/A	69.50772095	female
3646	2.01	70.18946446	female
3656	2.02	67.99091009	female
3661	N/A	75.97470157	male
3685	2.94	76.59347279	female
3692	3.21	75.02464133	male
3694	N/A	72.99583835	male
3695	3.36	76.62085204	female
3701	1.76	77.69411894	female
3706	2.19	93.39064725	male
3710	1.85	83.5286387	male
3742	3.76	67.47618005	female
3751	N/A	74.88500712	male
3754	4.21	73.05881065	female
3755	3.06	81.1356916	female
3758	2.63	88.90044902	female
3760	1.12	66.96692586	female
3789	3.58	85.82849633	male
3792	2.84	77.13010623	female
3797	3.47	67.70616581	female
3802	2.55	70.21410579	male
3819	1.56	65.67736283	female
3822	3.44	69.99780966	male
3823	3.13	66.75062972	male
3825	N/A	80.68393385	male
3859	3.77	73.76519549	female
3860	2.63	90.78140401	male
3871	3.21	93.74931552	female
3878	2.41	72.63169423	male
3880	3.34	72.57693571	male
3883	3.84	85.86956522	female
3903	3.02	71.64877889	female
3912	2.37	66.16471361	female
3916	N/A	66.5671887	female
3917	N/A	67.99638594	male
3918	4.07	86.91271493	female
3922	N/A	81.46698062	female
3931	3.28	73.1710656	female
3932	1.9	74.01708466	male
3938	N/A	74.71525572	male
3940	2.43	68.41528858	male
3942	2.89	72.23743292	female

3946	2.1	85.8996824	female
3953	3.91	81.0398642	female
3982	1.4	88.21870551	male
3987	2.23	75.59412989	female
3995	N/A	72.52765305	male
3999	N/A	66.09352754	male
4004	2.58	69.49403132	female
4007	N/A	83.32055635	female
4018	2.99	76.97404446	female
4030	4.29	82.41704085	male
4037	1.24	76.26218377	male
4046	2.71	87.15912824	female
4047	3.3	69.61450005	male
4049	2.84	70.67955317	male
4059	2.15	70.48789837	female
4061	1.15	78.55930347	male
4064	2.31	80.61822363	male
4066	1.9	81.62304238	female
4067	2.46	80.74690614	male
4074	3.21	75.67900559	female
4076	1.94	70.25517468	female
4094	N/A	78.26360749	male
4102	N/A	66.74515387	female
4104	4.06	67.13667725	female
4114	1.6	69.10798379	male
4116	3.28	69.39546599	male
4123	1.84	70.8739459	male
4130	3.26	75.63519877	male
4132	2.96	68.78764648	female
4136	2.25	70.23053335	female
4137	2.36	73.82816778	male
4150	3.19	66.94228453	male
4157	2.79	79.81053554	male
4161	4.38	80.57715475	female
4162	2.36	76.08695652	female
4163	4.09	70.30719527	male
4165	3.57	70.73157376	female
4169	4.47	87.12901106	male
4174	1.74	77.85565655	male
4193	2.52	86.50750192	female
4209	2.92	87.52601029	male
4210	2.35	68.07030993	male
4213	2.99	65.89913482	female

4269	2.56	77.97886321	female
4283	N/A	88.85390428	female
4284	2.7	68.37695762	female
4289	N/A	76.25123207	female
4293	3.29	66.60004381	male
4299	N/A	76.42372139	female
4304	2.31	71.17237981	female
4313	1.8	80.07611434	male
4322	3.28	70.1292301	female
4325	3.99	72.16350893	male
4330	3.23	73.87471252	female
4332	1.7	70.54813273	male
4333	N/A	67.58295915	male
4334	N/A	68.65896397	female
4338	1.94	72.15803307	female
4347	1.97	78.46895192	male
4355	4.79	67.42689738	female
4356	3.8	79.5832877	female
4358	2.77	66.38922352	male
4367	2.23	68.47552294	female
4385	1.95	77.1903406	female
4386	N/A	76.78238966	male
4397	2.37	65.69105246	male
4398	2.34	79.9446939	female
4399	2.92	83.18639798	female
4402	3.53	67.63224181	male
4413	3.21	68.79586026	female
4414	1.73	71.60223415	male
4417	2.26	73.58996824	male
4419	1.45	74.69335232	male
4422	0.88	66.14828606	male
4426	2.67	82.22264812	female
4428	3.53	90.20370168	male
4439	3.29	65.85806593	male
4441	2.26	70.33731245	female
4444	N/A	70.14839558	female
4446	N/A	70.68502902	male
4448	2.85	84.9852152	male
4451	1.47	89.4507721	female
4454	N/A	81.0508159	female
4457	3.36	82.53750958	female
4471	N/A	77.46960902	male
4476	1.26	82.56488884	male

4481	N/A	85.02080824	female
4482	N/A	86.8251013	male
4483	2.66	88.88402147	male
4485	2.37	81.17949841	male
4486	N/A	87.74504435	female
4492	1.6	77.36282992	female
4502	N/A	70.38933304	female
4508	N/A	72.38528091	male
4509	2.43	80.41014128	male
4512	N/A	77.2012923	female
4516	2.76	75.97743949	female
4525	3.43	73.42295477	male
4534	3.67	76.11981163	female
4541	1.34	78.74000657	female
4543	2.82	80.88380243	male
4558	2.66	75.8186398	female
4560	2.82	94.73223086	female
4567	2.42	82.71273683	female
4570	3.09	66.09352754	male
4577	1.53	77.01511335	female
4582	3.63	68.79312233	female
4584	2.39	82.98379148	male
4588	2.91	87.26316942	male
4598	3.71	75.07118607	male
4600	N/A	70.47147081	male
4602	N/A	69.39546599	female
4606	2.88	69.36534881	male
4607	4.6	79.82148724	male
4610	N/A	73.18475523	female
4626	2	69.39272807	male
4627	2.61	82.5183441	male
4630	N/A	83.44102508	female
4638	4.57	65.95936918	male
4640	3.39	77.29438178	female
4650	2.37	70.8958493	female
4652	1.48	77.79542219	male
4655	3.21	68.83966707	female
4669	3.43	72.31957069	female
4670	2.38	72.10601248	male
4671	4.27	68.17708904	male
4674	N/A	83.63267988	female
4676	2.86	66.81086409	female
4680	2.15	92.45975249	female

4682	2.32	75.1067791	male
4684	2.56	74.0964845	female
4685	2.9	76.49490746	female
4688	1.51	90.23655678	male
4697	N/A	67.1421531	female
4700	1.67	69.26952141	male
4702	2.83	74.98357245	male
4706	1.21	78.44978644	male
4710	1.65	88.30084328	female
4712	4.28	82.73464024	male
4718	1.59	77.29985763	female
4720	N/A	76.76322418	female
4721	2.46	66.69039536	female
4731	N/A	71.08750411	male
4733	4.25	72.90274888	male
4735	3.43	69.80615486	female
4744	1.2	79.52852919	female
4746	4.51	65.81973497	male
4749	2.72	75.78030884	female
4752	N/A	94.50772095	female
4757	N/A	80.09801774	female
4762	2.24	72.09506078	male
4766	1.85	69.83627204	male
4776	2.4	66.33720294	male
4777	1.38	82.64155076	female
4783	1.48	67.28726317	female
4784	1.05	70.55087066	male
4785	2.95	74.86036579	male
4788	N/A	86.4527434	male
4792	1.5	67.5035593	male
4805	2.63	86.97842515	female
4808	2.73	77.78994634	female
4809	N/A	70.61658088	female
4815	2.87	79.42722593	male
4816	2.82	70.58920162	female
4817	2.86	71.38320009	female
4819	2.32	77.16296134	female
4822	3.92	75.38604753	male
4826	1.71	76.37717665	male
4839	2.33	67.91698609	male
4845	2.53	69.11345964	male
4847	N/A	82.05837258	female
4849	2.16	76.79060344	male

4862	3.32	76.27861132	female
4869	7.01	84.89760158	female
4873	N/A	86.57047421	male
4874	1.51	82.22264812	female
4875	4.46	83.52590078	female
4878	2.75	79.12331618	female
4879	4.13	90.42273574	male
4881	4.32	68.17161319	male
4883	3.32	76.36348702	female
4884	N/A	86.40619866	female
4896	4.02	66.80538824	male
4898	N/A	71.50093089	male
4900	3.76	84.6019056	female
4901	2.66	85.23710437	female
4906	2.97	70.29350564	male
4909	3.72	68.82323951	female
4911	1.78	70.38385719	male
4913	2.56	82.4964407	male
4914	3.1	79.62161866	female
4921	1.93	77.10820283	female
4934	3.26	72.99857628	female
4939	2.75	76.80703099	female
4940	3.7	75.20260651	male
4947	2.35	77.63114664	female
4962	2.28	76.25944584	male
4963	4.46	78.44157266	female
4965	N/A	89.31935166	female
4971	2.12	76.52228672	male
4975	2.75	81.34103603	male
4984	2.41	66.38374767	male
4988	6	69.44474866	female
4989	2.51	76.46479027	male
4990	3.54	71.5228343	male
4995	2.02	86.95652174	male
4996	N/A	93.69181908	female
4997	1.84	67.04906363	male
4998	3.52	87.47398971	female
5000	1.95	79.67363925	male
5002	3.51	89.69992334	female
5005	4.49	68.06209616	female
5007	2.47	69.16274231	female
5009	N/A	68.58503997	male
5011	4.15	74.12933961	male

5013	3.2	71.10119374	female
5020	2.77	78.91249589	male
5021	N/A	82.81404008	male
5025	2.78	70.92596649	female
5027	2.59	71.54473771	male
5028	N/A	68.10042712	female
5031	N/A	68.10864089	male
5032	2.02	79.498412	female
5043	1.07	66.82729164	male
5052	2.11	67.51177308	female
5053	2.79	70.57824992	male
5055	2	66.41386486	female
5057	2.62	66.88752601	male
5063	3.22	72.713832	female
5069	2.3	73.37914796	female
5070	N/A	73.82816778	female
5080	3.96	74.72073157	male
5082	1.58	80.01040412	male
5085	2.11	71.91435768	female
5088	4.31	88.69510459	male
5091	1.66	73.19570693	female
5106	1.67	73.63377505	female
5119	2.74	82.60869565	female
5123	1.03	71.12583507	female
5125	4.18	80.91118169	female
5126	1.18	72.45646698	male
5128	2.26	80.04873508	female
5132	2.26	77.32449896	female
5134	4.33	78.51823459	female
5142	2.47	67.13393933	male
5143	N/A	71.24356587	female
5159	3.63	67.1312014	male
5163	3.1	78.23622823	female
5170	2.9	66.53433359	female
5171	N/A	67.18595992	female
5173	2.55	86.85521849	female
5181	2.11	89.05651079	female
5186	N/A	69.15726645	female
5200	2.3	85.86682729	male
5207	1.61	76.46479027	female
5208	N/A	83.02759829	male
5213	4.63	94.97864418	female
5214	2.12	81.24520863	male

5216	2.83	70.37564341	male
5217	3.55	84.59642974	female
5222	2.02	75.94184646	male
5225	3.3	92.4926076	male
5228	N/A	73.95411236	female
5229	1.2	79.04665425	female
5232	1.86	88.61022889	male
5240	4.67	66.79443653	female
5253	4.02	74.04993977	female
5255	3.92	75.10130325	male
5262	2.36	84.09812726	female
5268	N/A	67.44880079	female
5269	N/A	72.47563246	male
5274	3	74.84941408	female
5279	1.82	67.43784909	male
5289	1.67	84.27335451	female
5291	2.52	67.73354507	male
5295	2.81	67.29821487	male
5302	2.21	73.2449896	female
5308	3.88	72.31683277	male
5309	1.76	70.93418027	male
5310	2.09	69.11345964	male
5311	1.53	67.9388895	female
5316	2.78	67.34749754	female
5318	1.26	71.80757858	male
5324	3.53	65.84711423	male
5338	2.66	76.29230095	male
5347	5.33	88.8347388	female
5357	2.19	69.67473442	male
5367	2.66	68.18530281	male
5369	2.65	66.16745154	female
5372	N/A	87.64921695	female
5381	1.49	93.82871537	female
5384	2.8	74.72620743	male
5389	2.84	71.92530939	male
5391	2.93	71.57485489	female
5392	2.42	66.43303034	male
5394	N/A	70.501588	male
5399	1.94	67.91151024	female
5403	2.6	66.31529953	male
5405	3.5	73.31891359	female
5422	1.57	80.11718322	female
5423	1.74	74.23338079	female

5427	N/A	81.48067024	female
5432	3.11	84.35001643	female
5439	1.98	70.0443544	male
5441	1.43	68.11411675	female
5442	2.82	74.41955974	female
5443	2.39	73.77888512	female
5454	1.81	82.65524039	female
5458	2.8	77.0096375	male
5462	2.61	82.62238528	female
5466	0.88	67.85948965	male
5469	3.94	70.65764976	male
5472	N/A	70.05804403	male
5476	2.91	81.06998138	female
5481	3.65	88.59106341	female
5493	2.32	66.84919505	female
5496	N/A	68.70550871	female
5504	2.18	89.85050925	male
5510	2.26	68.6836053	female
5513	2.46	78.15682839	male
5519	3.19	69.19559742	male
5529	3.74	80.28693462	female
5538	2.8	66.48778885	male
5542	2.5	86.93188041	female
5544	5.66	77.0945132	female
5546	1.45	83.77505202	female
5551	3.2	74.76727631	male
5553	2.39	76.6482313	female
5556	1.92	68.15792356	male
5558	N/A	69.5131968	female
5560	N/A	72.2675501	male
5561	2.74	69.60354835	female
5563	N/A	72.96298324	female
5572	N/A	78.12944913	female
5578	3.9	75.97743949	female
5579	3.21	75.18344103	female
5585	N/A	74.61395247	female
5588	1.89	73.70769905	female
5591	2.48	75.553061	male
5592	N/A	80.14730041	female
5593	N/A	75.99660497	male
5614	1.94	69.90198226	female
5615	3.32	68.45361954	male
5616	2.3	75.50104041	male

5620	3.69	79.01379915	female
5626	3.21	76.47300405	female
5637	1.45	87.20841091	female
5640	2.7	71.36677253	male
5642	N/A	93.96287373	female
5643	2.91	79.44091556	male
5644	4.08	77.09177527	female
5646	N/A	83.45197678	male
5652	N/A	83.98313438	male
5655	N/A	72.76037674	male
5693	2.35	69.41189355	male
5694	N/A	83.12068777	female
5698	N/A	71.39962764	female
5700	3.1	71.91435768	male
5704	3.49	88.73343555	female
5730	1.92	68.03745482	male
5755	4.3	103.0911182	female
5789	N/A	70.53991896	male
5792	1.63	67.66783485	female
5793	3.05	69.77603767	female
5846	N/A	90.74307305	female
5849	3.38	67.59117293	male
5850	1.58	80.33074143	male
5858	2.77	105.6729821	female
5876	3.83	90.26941189	female
5888	2.75	84.43489213	male
5900	3.08	67.00799474	male
5910	1.11	65.4939218	female
5912	N/A	66.24137553	male
5921	3.1	66.77800898	female
5935	2.51	69.11072172	female
5936	4.01	72.68645274	male
5937	N/A	65.46928047	female
5944	2.82	65.86080385	male
5968	2.97	69.03953565	female
5980	2.77	72.23743292	male
5992	N/A	79.91457672	male
5993	2.76	77.24236119	female
6004	N/A	66.86014675	female
6010	2.09	70.60015332	female
6011	2.18	74.34015989	female
6014	N/A	71.25999343	female
6029	3.3	72.47289454	female

6031	1.74	71.16964188	male
6037	2.21	81.00974702	male
6039	3.06	72.46468076	female
6046	N/A	72.78228014	male
6063	2.48	68.12506845	male
6064	3.09	71.48450334	female
6081	2.8	68.61789508	male
6083	3.05	76.60990034	male
6084	N/A	69.83900997	female
6085	2.1	66.51516811	female
6094	N/A	66.43576826	female
6096	1.51	73.30796189	female
6189	2.24	66.72051254	female
6197	2.62	76.57156938	male
6203	3.16	68.89168766	male
6224	N/A	84.10634104	female
6226	3.14	69.4064177	female
6232	4.29	86.78403242	male
6238	1.09	71.54473771	female
6253	N/A	67.65140729	female
6264	N/A	66.82181579	female

List of mitochondrial copy number measurements, age, and gender for all individuals used in our analyses.

Supplementary Table 2. Mitochondrial haplogroups

H (424)	J (99)	U (147)	K (95)	T (121)	I (21)	W (20)	V (34)	HV (18)	X (8)	C (5)	L (4)	Missing (11)
H (49)	J1b1a (3)	U1b (2)	K1a (15)	T1 (4)	I (7)	W1 (8)	V (29)	HV (2)	X2b (5)	C4a1 (5)	L3e5 (4)	
			K1a1b					HV0b'c				
H1 (50)	J1c (2)	U2e (2)	(13)	T1a (6)	I1a1 (2)	W1c (1)	V1 (3)	(5)	X2c (3)			
		U2e1a										
H1a (17)	J1c1 (4)	(2)	K1a2 (6)	T1a1 (2)	I2 (11)	W3 (7)	V2 (2)	HV9 (11)				
H1a1	J1c1b1a		K1a3a1									
(10)	(6)	U3b (1)	(20)	T2 (15)	I4 (1)	W4 (4)						
H1b (23)	J1c2 (2)	U4a1 (3)	K1a4 (1)	T2a (9)								
		U4a1b	K1a4a1									
H1c (28)	J1c2b (3)	(3)	(17)	T2b (73)								
	J1c2b1a	U4a2	K1b1a									
H1c1 (2)	(4)	(15)	(6)	T2c (3)								
H1e (13)	J1c3 (4)	U4b1 (2)	K1b2 (1)	T2e (9)								
	J1c3a1		K1c1b									
H1g (7)	(2)	U4c (6)	(6)									
H2a2a												
(4)	J1c3b (1)	U5a (9)	K1c2 (2)									
	J1c3b1	U5a1										
H2a3 (3)	(11)	(32)	K2b (8)									
H3 (21)	J1c5 (12)	U5b1 (3)										
		U5b1b										
H3d (1)	J1c5a (4)	(7)										
		U5b1d										
H4a (29)	J1c6 (4)	(9)										
H4a1		U5b2										
(12)	J1c8 (7)	(45)										
H4a1a1	J2a1a	U5b3f										
(2)	(22)	(1)										
H4a1b	J2a1a1b											
(1)	(8)	U8a1 (5)										
H5 (13)												
H5a (10)												
H5a1												

(21)
H6a1a
(13)
H6a1b
(29)
H7 (10)
H7a (1)
H10 (11)
H11 (15)
H13 (24)
H13a1a
(1)
H13a1a1
(2)
H15 (1)
H17 (1)

We have listed all the major mitochondrial haplogroups as well as sub-haplogroups in out dataset. The numbers in parenthesis represent the number of individuals in our dataset that belong to the haplogroup. There were 1007 total individuals in our dataset.

Supplementary Table 3. Kinship coefficients for the clade defined by branch 124

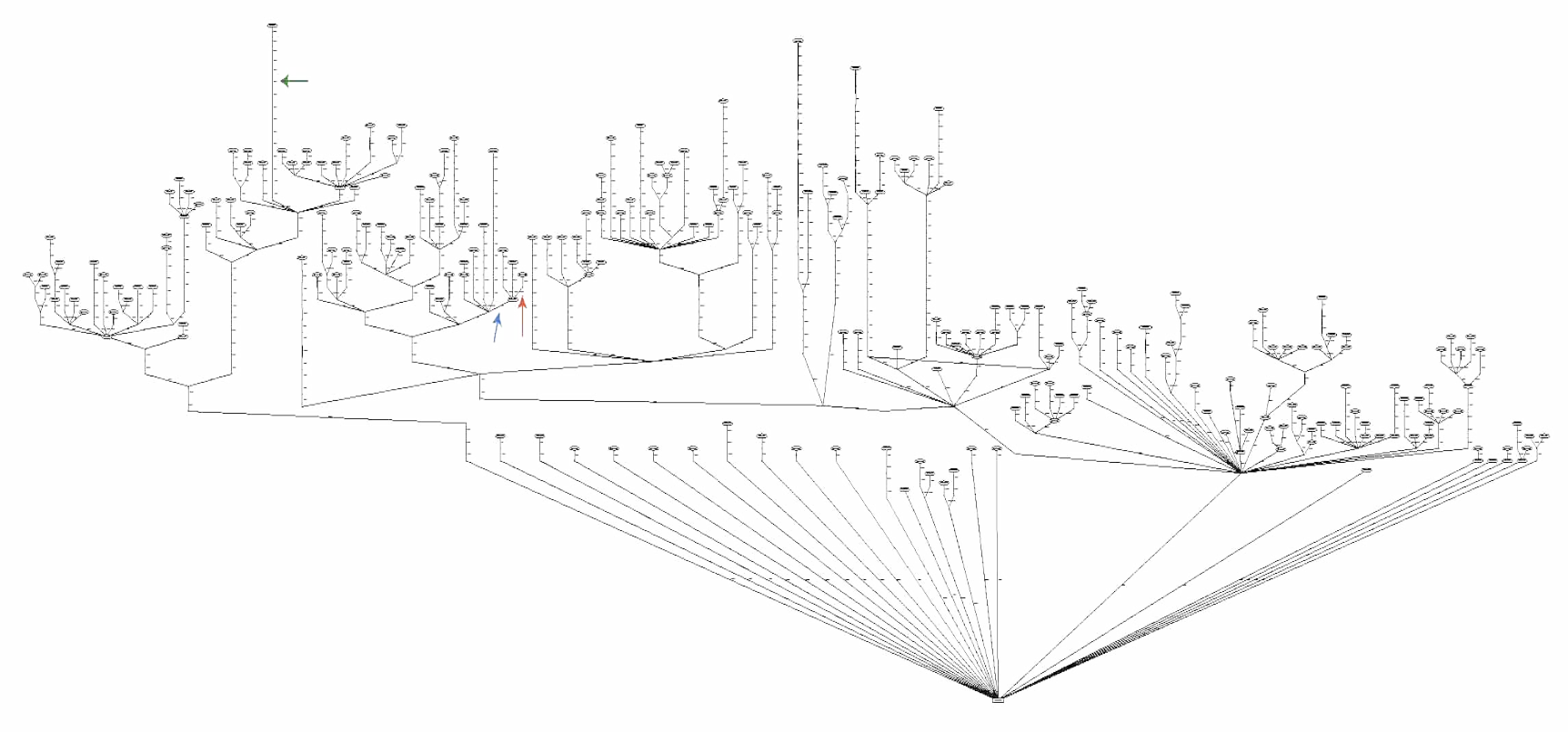
	3644	**5310**	**4137**	**5308**	**3006**	**26**	**3903**	**5900**	**1322**	**3393**	**4161**	**59**	**4875**	**4869**
3644	N/A	0	0	0	0	0	0	0	0	0	0	0	0	0
5310		N/A	0	0.25	0	0	0	0	0	0	0	0	0	0
4137			N/A	0	0	0	0	0	0	0	0	0	0	0
5308				N/A	0	0	0	0	0	0	0	0	0	0
3006					N/A	0	0	0	0	0	0	0	0	0
26						N/A	0	0	0	0	0	0.25	0	0
3903							N/A	0.25	0.25	0	0.25	0	0.25	0.25
5900								N/A	0.25	0	0.25	0	0.25	0.25
1322									N/A	0	0.25	0	0.25	0.25
3393										N/A	0	0	0	0
4161											N/A	0	0.25	0.25
59												N/A	0	0
4875													N/A	0.25
4869														N/A

We have listed the pairwise kinship coefficients for all the individuals in this clade. The IDs (row and column titles) correspond to the same IDs used in Supplementary Table 1.

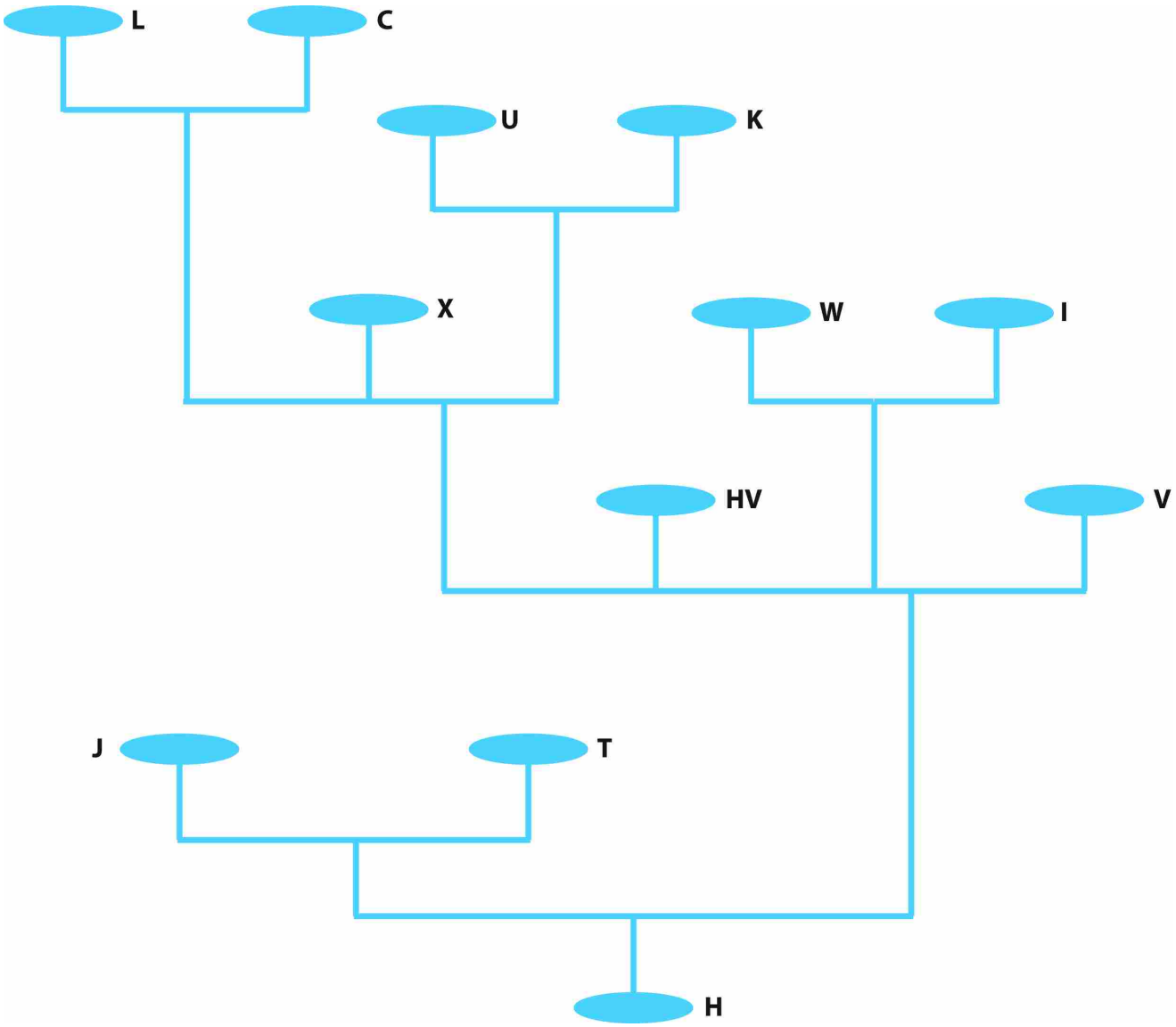
Supplementary Table 4. Kinship coefficients for the clade defined by branch 50. Supplementary Table 4

	5466	3613	176	4293	841	1976	3859	5469	6232	2765	3469	832
5466	N/A	0	0	0	0	0	0	0.25	0	0	0	0
3613		N/A	0.25	0	0	0	0	0	0	0	0	0
176			N/A	0	0	0	0	0	0	0	0	0
4293				N/A	0	0	0	0	0	0	0	0
841					N/A	0.25	0	0	0	0.25	0	0
1976						N/A	0	0	0	0.25	0	0
3859							N/A	0	0.25	0	0	0
5469								N/A	0	0	0	0
6232									N/A	0	0	0
2765										N/A	0	0
3469											N/A	0
832												N/A

We have listed the pairwise kinship coefficients for all the individuals in this clade. The IDs (row and column titles) correspond to the same IDs used in Supplementary Table 1.



Supplementary Figure 1. Haplotype network. Our haplotype network was constructed using TCS and 285 full mitochondrial genomes. The arrows point to each of the three branches representing the significant contrasts. The blue arrow points to branch 124, the red to branch 121, and the green to branch 50.



Supplementary Figure 2. Haplotype network. We collapsed our haplotype network (Supplementary Figure 1) into nodes corresponding to the major mitochondrial haplogroups present in our network.

Appendix 3

Supplementary Figures and Tables for Chapter 4

Supplementary Figure Legends

Supplementary Figure 1. Haplotype network. This is the haplotype network constructed using the 138 genotyped SNPs. The four branches, which define the clades associated with phenotypes in this study, are labeled.

Supplementary Table 1. Tested phenotypes

Phenotypes
ADAS Cog Total 11 Score Baseline
% Change in ADAS Cog Total 11 Score
Whole Brain Volume (Baseline)
Left Hippocampal Volume (Baseline)
Right Hippocampal Volume (Baseline)
Annualized Whole Brain Volume Atrophy
Annualized Left Hippocampal Volume Atrophy
Annualized Right Hippocampal Volume Atrophy
Average HOC (Baseline)
% Change HOC
Parahippocampal Cortex Thickness (Baseline)
% Change Parahippocampal Cortex Thickness
Entorhinal Cortex Thickness (Baseline)
% Change Entorhinal cortex
Temporal Pole Thickness (Baseline)
% Change Temporal Pole Thickness

This is a list of the 16 phenotypes we tested for association with mtDNA. Change measures are for 2-year longitudinal data. Volumes were normalized. HOC; Hippocampal Occupancy Score

Supplementary Table 2. Pairwise correlation coefficients

% Change in ADAS Cog Total 11 Score	
ADAS Cog Total 11 Score Baseline	
% Change Entorhinal cortex	
Entorhinal Cortex Thickness (Baseline)	
% Change Temporal Pole Thickness	
Temporal Pole Thickness (Baseline)	
% Change Parahippocampal Cortex Thickness	
Parahippocampal Cortex Thickness (Baseline)	
% Change HOC	
Average HOC (Baseline)	
Annualized Right Hippocampal Volume Atrophy	
Annualized Left Hippocampal Volume Atrophy	
Annualized Whole Brain Volume Atrophy	
Right Hippocampal Volume (Baseline)	
Left Hippocampal Volume (Baseline)	
Whole Brain Volume (Baseline)	N/A
	Correlation p-value **Whole Brain Volume (Baseline)**

Pairwise correlation coefficients are listed for each of the 16 endophenotypes we tested. We report the correlation and significance (p-value) of the observed correlations. Significant correlations are highlighted in yellow.

	N/A
N/A	0.785 0
-0.232 0	-0.177 0.002
Correlation p-value **Left Hippocampal Volume (Baseline)**	Correlation p-value **Right Hippocampal Volume (Baseline)**

		N/A	
N/A		-0.057	0.327
0.045	0.439	0.305	0
0.011	0.849	0.328	0
0.058	0.325	0.01	0.865
Correlation **Annualized Whole Brain Volume Atrophy**	p-value	Correlation **Annualized Left Hippocampal Volume Atrophy**	p-value

		N/A	
N/A		0.208	0
0.551	0	0.264	0
-0.041	0.482	0.207	0
0.276	0	0.484	0
0.246	0	0.413	0
-0.041	0.486	0.1	0.089
Correlation **Annualized Right Hippocampal Volume Atrophy**	p-value	Correlation **Average HOC (Baseline)**	p-value

		N/A	
N/A		0.172	0.003
0.569	0	0.262	0
0.268	0	0.026	0.656
0.278	0	0.089	0.131
0.548	0	0.074	0.209
0.284	0	0.189	0.001
0.251	0	0.190	0.001
0.074	0.205	0.066	0.259
Correlation	p-value	Correlation	p-value
% Change HOC		**Parahippocampal Cortex Thickness (Baseline)**	

		N/A	
N/A		0.119	0.042
0.043	0.467	0.358	0
0.526	0	0.275	0
0.212	0	0.45	0
0.087	0.138	0.029	0.623
0.115	0.05	0.119	0.042
0.44	0	0.028	0.63
0.12	0.04	0.243	0
0.064	0.275	0.195	0.001
0.073	0.214	0.195	0.001
Correlation	p-value	Correlation	p-value
% Change Parahippocampal Cortex Thickness		**Temporal Pole Thickness (Baseline)**	

		N/A	
N/A		0.214	0
0.144	0.014	0.722	0
0.378	0	0.179	0.002
0.033	0.568	0.299	0
0.477	0	0.367	0
0.273	0	0.558	0
0.052	0.372	0.132	0.024
0.057	0.329	0.212	0
0.409	0	0.099	0.092
0.071	0.225	0.403	0
0.068	0.246	0.328	0
0.123	0.035	0.18	0.002
Correlation **% Change Temporal Pole Thickness**	p-value	Correlation **Entorhinal Cortex Thickness (Baseline)**	p-value

		N/A	
N/A		0.038	0.513
0.124	0.034	-0.097	0.097
0.545	0	0.061	0.3
0.131	0.025	-0.084	0.154
0.448	0	0.025	0.673
0.01	0.86	-0.089	0.128
0.493	0	-0.071	0.223
0.198	0.001	-0.132	0.024
0.143	0.014	-0.07	0.235
0.102	0.083	-0.039	0.502
0.429	0	0.021	0.715
0.01	0.862	-0.109	0.063
0.032	0.586	-0.117	0.046
0.164	0.005	0.029	0.616
Correlation **% Change Entorhinal cortex**	p-value	Correlation **ADAS Cog Total 11 Score Baseline**	p-value

N/A	
0	0.997
-0.142	0.015
-0.097	0.099
-0.155	0.008
-0.103	0.079
-0.124	0.034
0.002	0.977
-0.202	0.001
-0.093	0.114
-0.124	0.034
-0.044	0.455
-0.155	0.008
0.06	0.306
0.069	0.236
-0.021	0.716
Correlation	p-value
% Change in ADAS Cog Total 11 Score	

Supplementary Table 3. Mitochondrial haplogroups for the clade defined by branch 155

Variant	Possible Haplogroups
m.2706A>G	L0D1, M23, M45A, D4F1, A4F, J1C3C, H, U2B
m.9950T>C	L0D1A , L3F1, M11, W1F, B5, B2
m.15301G>A*	L1C1D, L2, L3, L4, L6
m.12705C>T*	R
m.10873T>C*	N
m.10398A>G*	L1C1A, L3E1A3, N, J1C8, N8, Y, N1A, N1E, N1EI, I, R11, B4C1C, B5, R12, R21, P4, J, R0A2K1, K1
m.9540T>C*	N

Mitochondrial haplogroups associated with each of the listed variants. Variants are listed (top to bottom) as the haplotype network is descended from the root, so each variant, top to bottom, is specific to a smaller set of observed haplotypes in our dataset. Haplogroups listed in red experience a back mutation of the listed variant, so they have the ancestral allele as opposed to the variant allele.

*These variants define branch 155

Supplementary Table 4. Mitochondrial haplogroups for the clade defined by branch 199

Variant	Possible Haplogroups
m.2706A>G	L0D1, M23, M45A, D4F1, A4F, J1C3C, H, U2B
m.12372G>A	M7C1D, M12, D4H1A1, N9A, U, H4AU
m.12308A>G	U, U5A2B2
m.11467A>G	U
m.3197T>C	L3F1A1, H14B, U2E1A1A, U5
m.16270C>T	L1B, L3H1A2B, M58, M2A1, M13B2, M61, N11B, P1D1, H1BA, U5B2A1, K2B1A, U5
m.7768A>G	U5B
m.5656A>G*	L5A, M26, T2B3A, HV1B1B, U5B1
m.217T>C**	L3B1B1, H3H4, U5B1B2, U2E1, U2E2, U2E3, K1A4A1F1

Mitochondrial haplogroups associated with each of the listed variants. Variants are listed and haplogroups colored as described in Supplementary Table 2 except that highlighted (in yellow) haplogroups show the path from general (haplogroup U) to specific (haplogroups U5B1 and U5B1B2) used to assign haplogroups to the clade defined by branch 199.

*This variant defines branch 199

**This variant defines one of the branches within the clade defined by branch 199

Supplementary Table 5. Mitochondrial haplogroups for the clade defined by branch 9

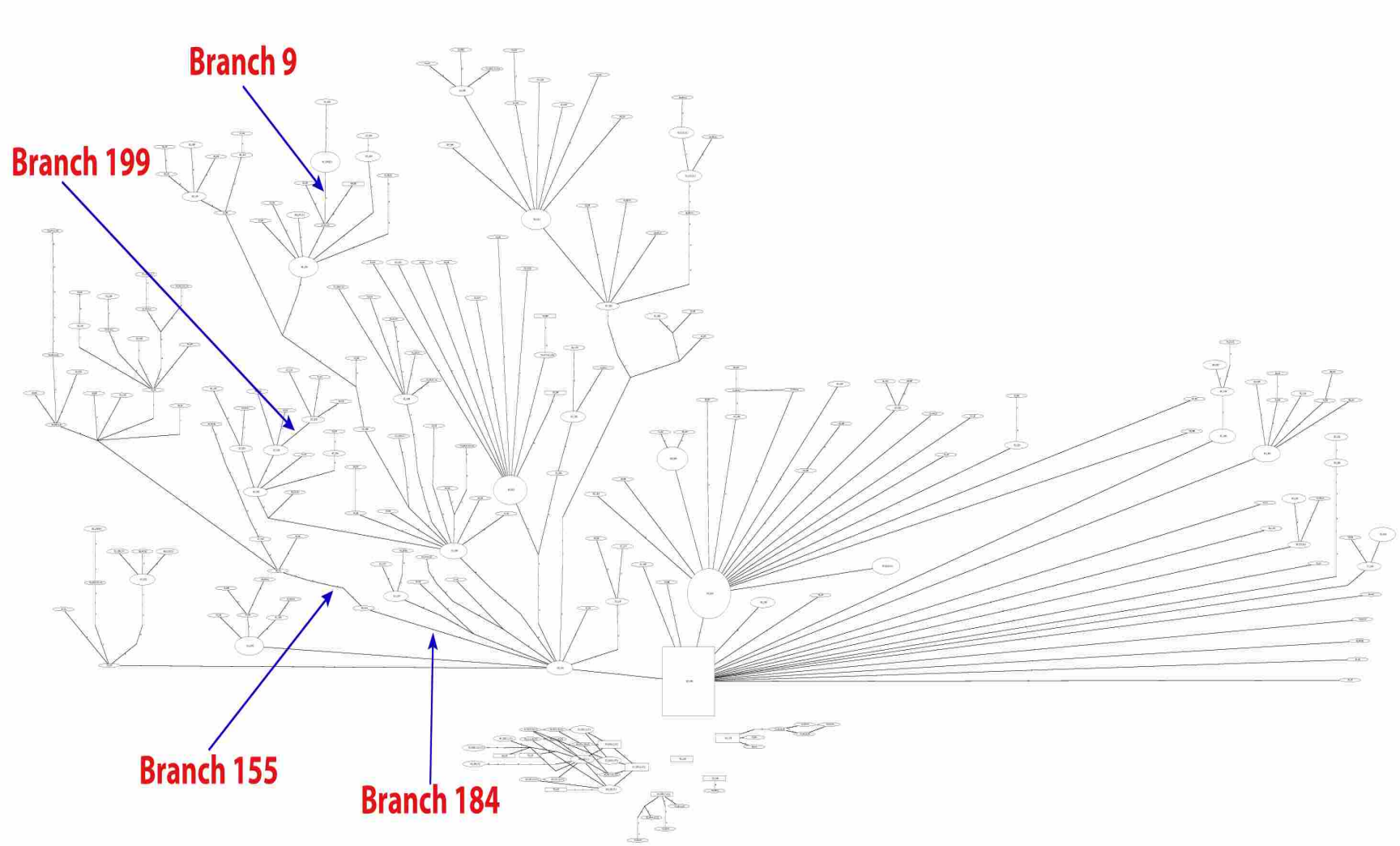
Variant	Possible Haplogroups
m.2706A>G	L0D1, M23, M45A, D4F1, A4F, J1C3C, H, U2B
m.12372G>A	M7C1D, M12, D4H1A1, N9A, U, H4AU
m.12308A>G	U, U5A2B2
m.11467A>G	U
m.9698T>C	H1B1E, U8
m.3480A>G	B5B1C, U8B, UK
m.10550A>G	K
m.14798T>C	J1C, T2G, K
m.1189T>C	U4A2F, K1
m.10398A>G*	L1C1A, L3E1A3, N, J1C8, N8, Y, N1A, N1E, N1EI, I, R11, B4C1C, B5, R12, R21, P4, J, J1C8, R0A2K1, K1
m.11914G>A	L1, L2, L3, L4, L5, L6, M20, M5A5, C, G1A1A, G3A1, M77, L1C1D, L2A1-L2A4, D1A2, D1G4, D4H3A1A2, N1C, A2F3, X2B6, T2C1A, R8A1A1C, B6A, B4B1C1, B4D, B5B1B, R23, P2, H1H1, H1N4, H1AO1, H3X1, H15A1B, U8B1A1, K1A1, K1B1C, K2A2A1
m.15924A>G**	U5A1A1, U5B2A1A2, K1A1B, L1C3B2, L2A1C2, L3E2A1A, M2A1B, C4B7, M13, M21A, M35A, D1G4, D4E1A, D4N1, N1E, I, N8, A2K1, A2X, O1A, X2G, B4A1A1B, B2J, P10, R0A2B, H1N1B, M13B2
m.15758A>G***	K1A1B2A1, L2A5, L3I1A, L3X2B, G2A1B, I2, B5B2B, H1C14, H1AT, H13A1A2B, H20B, H28A1

Mitochondrial haplogroups associated with each of the listed variants. Variants are listed and haplogroups colored as described in Supplementary Table 3 except that we are highlighting the path to K1A1B and K1A1B2A1.

*This variant shows a reverse mutation in K1; however, a review of the literature reveals there is confusion about what the ancestral allele is at this position, so the SNP chip actually found these individuals were WT (by calling the ancestral allele the variant allele)

**This variant defines branch 9

***This variant defines one of the branches within the clade defined by branch 9



Supplementary Figure 1. Haplotype network. This is the haplotype network constructed using the 138 genotyped SNPs. The four branches, which define the clades associated with phenotypes in this study, are labeled.

Appendix 4

Supplementary Table for Chapter 5

Supplementary Table 1. Missingness rates for covariants and case-control status

Covariates/Case-control Status	# of individuals missing the given covariate (% missing)
Age	1491 (7.57%)
Sex	0 (0%)
Case-control Status	0 (0%)
APOE (rs429358/rs7412)	579 (2.94%)
10 Principal Components	0 (0%)
BIN1 (rs744373)	2872 (14.58%)
CLU (rs11136000)	3009 (15.28%)
PICALM (rs3851179)	211 (1.07%)
MS4A6A (rs610932)	30 (0.15%)
CD33 (rs3865444)	392 (1.99%)
MS4A4E (rs670139)	192 (0.98%)
CD2AP (rs9296559)	2374 (12.06%)
CR1 (rs3818361)	54 (0.27%)
ABCA7 (rs3764650)	2964 (15.05%)

The Alzheimer's Disease Genetics Consortium dataset consists of 19,692 total individuals. We removed any individuals missing any of the covariates (listed here) or case-control status (included in this table). The final dataset consisted of 9,884 individuals.