NEXT GENERATION SEQUENCING ASSESSMENT OF MITOCHONDRIAL OXIDATIVE DNA DAMAGE IN COGNITIVE IMPAIRMENT: SHEDDING LIGHT ON HEALTH DISPARITIES IN MEXICAN AMERICANS

Dissertation submitted by

Danielle Marie Reid, BS

In partial fulfillment towards Doctor of Philosophy

School of Biomedical Sciences

University of North Texas Health Science Center

July 19, 2022

TABLE OF CONTENTS

APPENDIX L. Cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering population interaction effect with both sex and education.

APPENDIX M. Cellular 8oxoG “hotspot” variant count and cognitive impairment (NC vs MCI + AD) multiple linear regression model prediction considering population interaction effect with both sex and education.

APPENDIX N. Multiple linear regression results for cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) within Mexican Americans.

APPENDIX O. Multiple linear regression results for cellular 8oxoG “hotspot” variant count and cognitive impairment (NC vs MCI + AD) within Mexican Americans.

APPENDIX P. Multiple linear regression results for cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) within non-Hispanic Whites.

APPENDIX Q. Multiple linear regression results for cellular 8oxoG “hotspot” variant count and cognitive impairment (NC vs MCI + AD) within non-Hispanic Whites.

APPENDIX R. Cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype and population interaction effect with both sex and education.

APPENDIX S. Cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype as well as diabetes and cognitive status interaction.

APPENDIX T. Cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype within the Mexican American population.

APPENDIX U. Cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype within the non-Hispanic White population.

APPENDIX V. Total cellular 8oxoG variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD), *OGG1* genotype, and population interaction effect with both sex and years of education.

APPENDIX W. Total cellular 8oxoG variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD), *OGG1* genotype, and diabetes interaction effect.

APPENDIX X. Cellular 8oxoG “hotspot” variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD), *OGG1* genotype, and population interaction effect with both sex and years of education within the whole cohort.

APPENDIX Y. Cellular 8oxoG “hotspot” variant count multiple linear regression model considering cognitive impairment (NC vs. MCI + AD), *OGG1* genotype, and diabetes interaction with cognition.

APPENDIX Z. Cellular 8oxoG variant count multiple linear regression prediction model considering cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in the Mexican American population.

APPENDIX AA. Cellular 8oxoG “hotspot” variant count multiple linear regression considering cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in the Mexican American population.

APPENDIX AB. Cellular 8oxoG variant count multiple linear regression predictive model considering cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in non-Hispanic Whites.

APPENDIX AC. Multiple linear regression predictive model assessing cellular 8oxoG “hotspot” variant count with cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in non-Hispanic Whites.

APPENDIX AD. Example TSV of deletion profile for one sample.

APPENDIX AE. Example mtDNA plot of deletions (blue) and duplications (red) with varying heteroplasmy for one sample.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Nicole Phillips as this endeavor would have not been possible without her support. I have a great appreciation for her patience, encouragement, and unwavering ambition. Dr. Phillips you are truly an aspiring academic professional. I am also grateful for my dissertation committee as their help is always constructive, profound, and sincere. In no particular order, thank you kindly Dr. Robert Barber, Dr. Nathalie Sumien, Dr. Roland Thorpe, Dr. John Planz, and Dr. Scott Maddux. You are a one-of-a-kind group. I would also like to thank numerous UNTHSC colleagues and staff members that have helped me throughout my time here whether it be academic, social, and/or professional. You have all personally effected my life with your thoughtfulness, authenticity, humor, hard work, and last but not least, friendship. That being said, thank you, Dr. Harlan Jones, Dr. Caroline Rickards, Dr. Keisa Mathis, Jie, Gita, Talisa, Viviana, Courtney, Nicole H., Gretchen, Dianna, Austin, Megan W., Kumudu, Isabelle, and Ann. I would also like to extend my sincere thanks to URMC and SJFC faculty that inspired me to pursue a career in research. Your positivity and support were so empowering. Thank you, Dr. Sanjay Maggirwar, Dr. Vir Singh, Dr. Meera Singh, Dr. Edward Freeman, Dr. Noveera Ahmed, and Dr. Kirsten Picard. Many thanks to my dear friends Megan C., Tessa, Carli, Shawn, LJ, Haley, Josh, Sandra, Kristen, Maggie, and Alyssa for being awesome cheerleaders. Lastly, I would be remiss in not mentioning my family. Thank you, mom, dad, Rick, Dale, Deidre, and Dion, for your constant love, care, and support.

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AD — Alzheimer's Disease

Aβ — Amyloid Beta

BER — Base excision repair

(c)cf-mtDNA — Circulating cell-free mitochondrial DNA

CI — Cognitive Impairment

CSF — Cerebrospinal fluid

ER — Endoplasmic reticulum

ETC — Electron Transport Chain

ISC — Iron-sulfur cluster

MCI — Mild Cognitive Impairment

mtDNA — Mitochondrial DNA

mtPTP — Mitochondrial permeability transition pore

NC — Normal Controls

nDNA — Nuclear DNA

NOS — Nitric oxide synthase

NGS — Next Generation Sequencing

NHW — Non-Hispanic Whites

OS — Oxidative Stress

SOD — Superoxide dismutase

T2D — Type-2 diabetes

8oxoG — 8-oxo-7,8-dihydroguanine

CHAPTER I: INTRODUCTION TO AGING AND AGE-RELATED NEURODEGENERATION

AGING AND DISEASE

In the 1950s, Denham Harman reported a theory suggesting reactive oxygen species (ROS) generation was associated with the rate of aging[1–3]. Aging is broadly defined as a time-dependent accumulative decline in physiological processes. This decline is tightly linked to dysfunction of redox signaling and control mechanisms because of continuous loss in genomic plasticity which ultimately increases one's chances for disease and/or mortality due to free radical reactions[1–4]. More specifically, the mitochondrial theory of aging postulates that as cells age mitochondrial function declines due to accumulating mitochondrial ROS enables free radical reactions to occur and damage essential biomolecules, and as this damage aggregates the chance of disease and/or death increase[5].

Oxidation to major biomolecules has been shown to play a major role in aging and has been implicated in pathologies of age-related diseases such as cancers, pulmonary fibrosis, cardiovascular, autoimmune, and neurodegenerative diseases[6–10]. The pathophysiology of aging and disease are to no surprise similar since there's an abundance of evidence from observational studies reporting oxidative stress (OS) as a major contributing factor through altered mitochondrial function for human morbidity and mortality[4,11]. However, the main difference between aging and disease seems to be related to the key role of oxidants in stem cell biology where they appear to regulate aging and age-related reformative aptitude[1]. Physiological levels of OS that surpass normal aging conditions and processes lead to the disruption of redox signaling and control mechanisms which contribute to the development of age-related disease[4]. The aging population includes individuals aged 65 years and older. Nationally, the aging population is expected to continue to grow at a rapid rate over the next several decades from 58 million in 2021 to 88 million in 2050 (**Figure 1A**)—the Hispanic/Latino aging population is expected to increase at a more accelerated rate than other racial/ethnic groups[12–14]. Mexican Americans are the largest segment of the Hispanic/Latino population. As the aging population expands, corresponding increases in prevalence of age-related diseases, such as, cardiovascular disease, metabolic disorders, cancer, and neurodegenerative diseases will continue to burden the

healthcare system[12–14]. **Figure 1B** depicts the number of AD cases in the US by race/ethnicity through the year 2060.

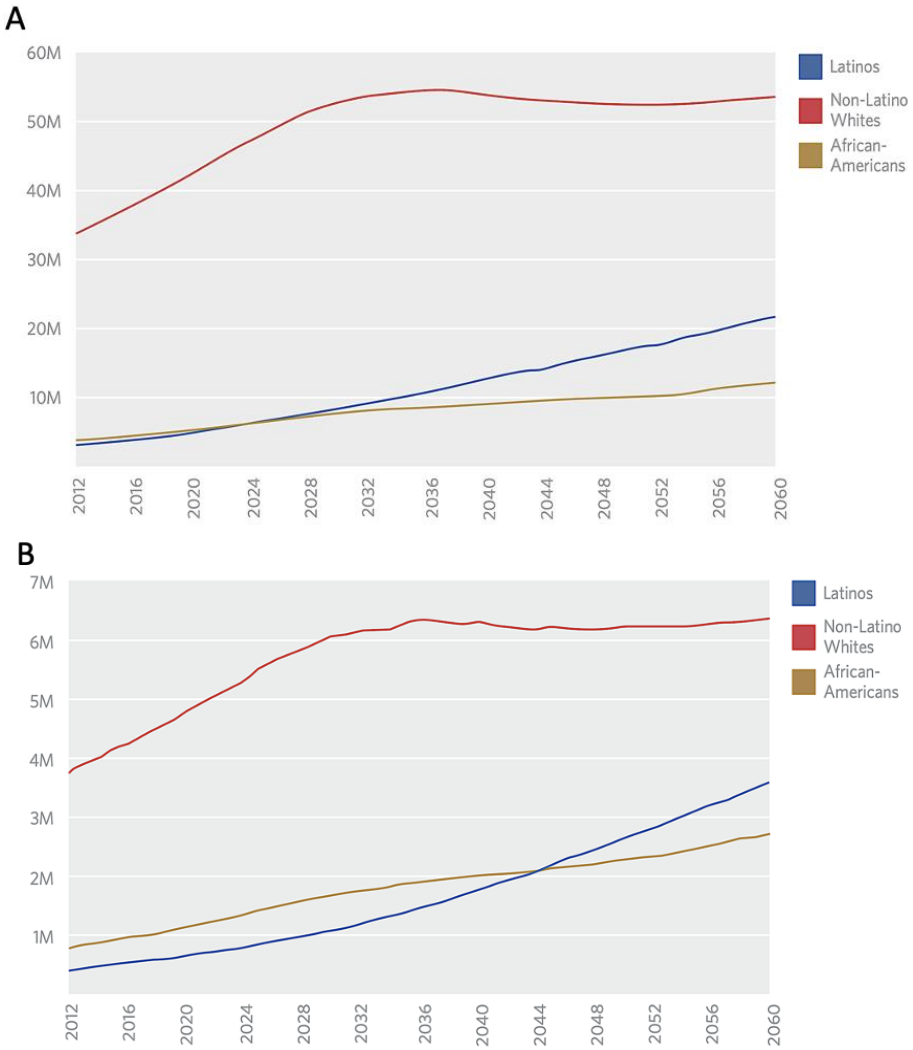


Figure 1. Projected number of US aged individuals and corresponding anticipated number of AD cases in the US through 2060. A. The number of US aged individuals from 2012 to 2060 by ethnic/racial group. Latinos show the greatest increase in US aged individuals. B. Expected number of US AD cases through 2060 by ethnic/racial group. Latinos display a larger increase in AD cases. (Figure credit: Wu, S., Vega, W. A., Resendez, J., & Jin, H. (2016). *Latinos and Alzheimer's Disease: New Numbers Behind the Crisis*.)[13].

ALZHEIMER'S DISEASE

Alzheimer's Disease (AD) is a fatal neurodegenerative disease impacting thinking, learning, and cognitive function due to neuronal damage and accumulation of amyloid beta (Aβ) plaques[12–18]. AD is the fifth leading cause of death in the aging population and the seventh leading

cause of death in the US[12,13]. Additionally, among the major causes of death—breast and prostate cancer, heart disease, stroke, and HIV have decreased or remained nearly the same, yet AD prevalence has continued to increase[12,14]. Among older adults, AD is the most common form of dementia characterized by the loss in cognitive function and behavioral competence that disrupts a person's daily activities[12,14,19].

Two classes of AD exist based on the time of onset of the disease including Early Onset Alzheimer's Disease (EOAD) and Late Onset Alzheimer's Disease (LOAD)[12,14,19]. The age of onset for EOAD is typically earlier than 65 years of age, compared to LOAD which classically develops at 65 years of age or older[12,14,19]. EOAD is a familial disease having a rare autosomal dominant inheritance pattern mainly caused by mutations in three genes—APP, PSEN1, and PSEN2, all playing a role in the production/mal-processing of Aβ peptide[12,14,15,19]. Contrary to EOAD, LOAD is sporadic, multifactorial, and genetically complex (i.e., 60-80% heritable with no single gene accounting for its heritability)[19,20]. A diverse array of environmental and genetic factors (mutations and/or polymorphisms in multiple genes) contribute to the development, progression, and severity of the disease[16,19]. Most AD cases (>90%) are late onset leaving early onset cases to represent less than 10% of total AD cases[14,21]. From here on, Alzheimer's Disease and AD will be used to refer to LOAD.

Currently, there are limitations in diagnosing individuals with AD, as an accurate diagnosis can only be made by relating clinical assessments with a brain tissue autopsy performed postmortem[22]. Although AD is the most common form of dementia, there are several other causes of dementia that are linked with different symptoms and brain abnormalities[12,14,19]. Identifying the cause of dementia can pose great difficulty because many people with dementia have mixed dementia due to brain changes associated with more than one cause of dementia[12,14]. Recent large autopsy studies revealed approximately more than 50% of individuals with Alzheimer's dementia have AD brain pathology in addition to one or more causes of dementia, for example, cerebrovascular or Lewy body disease[12,14]. Diagnostic inconsistencies such as these are recognized in the AD field and efforts have been made to improve resolution in clinical settings.

The National Institute on Aging and the Alzheimer's Association assigned three workgroups composed of individuals with balanced expertise and international representation in academia and industry to revise the 1984 criteria for AD dementia in 2009[23,24]. The purpose of the revisions was to encompass diagnostic and research criteria for the pathophysiological progression of AD because it had become recognized that cognitive deficits and AD pathology develop gradually with no distinct event denoting its onset[24,25]. This is discussed in the revised criteria establishing that there are different qualitative and quantitative clinical phases of the disease, and pathophysiological processes that manifests in each phase by incorporating biomarkers causing disease state[24]. Thus, the three workgroups were assigned the following tasks to prepare (1) diagnostic criteria for AD dementia, (2) diagnostic criteria for mild cognitive impairment (MCI) the symptomatic AD pre-dementia phase, and (3) research recommendations for studying individuals with asymptomatic AD in the preclinical phase, meaning there is evidence of early AD brain changes without symptoms of MCI or dementia[24].

The workgroups designed core clinical criteria for diagnosing MCI and AD dementia for use in all clinical settings that could be applied with and without access to biomarkers, both fluid and imaging measures, requiring specialized tests and/or procedures[24,25]. In the clinical setting dementia secondary to AD is classified into two groups—probable AD dementia and possible AD dementia[23]. Dementia is diagnosed when cognitive or behavioral symptoms interfere with a person's ability to complete routine activities assessed by patient history and objective cognitive testing[23]. Following diagnosis of dementia to provide a prognosis of probable AD dementia, the patient is assessed further by applying the core clinical criteria and using an assortment of approaches and tools for evaluation[12,14,22,23]. Formal neuropsychological evaluations, standardized mental status examinations, or informal evidence of worsening cognition provided by informants close to the patient are viable[23]. Patients showing gradual onset, and a well-defined history of worsening cognitive decline with cognitive deficits in either amnestic presentation with evidence of declining cognitive function in at least one other cognitive domain, or a non-amnestic presentation[23]. Furthermore, patients with evidence of other neurological disease/conditions, non-neurological conditions, or use of medication significantly affecting cognition need to be ruled out[23].

Similarly, core clinical criteria for diagnosis of MCI was initiated due to concerns of abnormal changes in cognitive impairment for the patients age/educational background reported by the patient, a close informant, or the observing clinician[25]. Subsequently, the patient is assessed by clinical and cognitive evaluations for evidence of mildly reduced performance in one or more cognitive domains that is greater than expected but does not significantly impact their social or occupational function, based on serial assessments or single evaluation along with medical/psychiatric history[12,14,22,25]. Upon further investigation ruling out additional systemic or brain diseases that also cause cognitive impairment the clinician can diagnose the patient with MCI that may or may not progress to AD dementia[25]. MCI due to AD can be diagnosed applying longitudinal cognitive assessments by demonstrating progressive cognitive impairment[25].

ALZHEIMER'S DISEASE PATHOPHYSIOLOGY

Prominent pathological changes observed in AD are (1) the accumulation of extracellular A peptides producing senile plaques that block cell-cell signaling at synapses and/or (2) accumulation of intracellular hyperphosphorylated tau protein resulting in neurofibrillary tangles (NFTs) that inhibit transportation of essential molecules[12,14–16,19,21]. Accumulation of Aβ plaques and NFTs generates a neurotoxic environment inducing inflammation by activating microglia and astrocytes for their clearance, as well as cellular debris[12,14,18]. These phagocytic brain cells upon activation by Aβ release pro-inflammatory intermediates comprising of cytokines, chemokines, complement proteins, and ROS[18]. Induction of the pro-inflammatory state generates a vicious cycle of oxidative stress and inflammation fueling each other[17]. The association between inflammation and oxidative stress is well documented; DNA damage has been shown to occur due to increased oxygen uptake as a combative inflammatory response leading to elevated levels of ROS[6].

For diagnostic purposes in both MCI and AD, biomarkers are complimentary to enhance the certainty of AD pathophysiological process and are not to be used for routine purposes due to several limitations[23,24]. The two main categories of biomarkers are (1) Aβ accumulation assessed by positive positron emission tomography (PET) imaging and low cerebrospinal fluid (CSF) Aβ42 initiating or upstream clinical symptoms, and (2) downstream pathophysiological

neuronal degeneration or injury assessed by elevated tau levels in the CSF (both total and phosphorylated tau), and both decreased fluorodeoxyglucose uptake and atrophy, on PET and magnetic resonance imaging (MRI) in specific topographic patterns, respectively[24,25]. Biomarkers indicative of MCI due to AD require evidence of low CSF Aβ42 and elevated CSF tau, while those indicative of AD dementia requires evidence of Aβ and tau deposition in plaques and neurofibrillary tangles, respectively[25]. The biomarkers of AD have limited applicability because (1) the core clinical criteria provides adequate diagnostic accuracy/utility in most patients, (2) more research is needed to ensure that criteria incorporating biomarkers have been designed appropriately, (3) there is deficient standardization of biomarkers in different test settings and associated with disease phases, and (4) there is varied accessibility in different settings (e.g., experienced personnel and equipment) to conduct the testing[23,25].

ALZHEIMER'S DISEASE HEALTH DISPARITY IN MEXICAN AMERICANS

The molecular and etiological events initiating AD pathologies remain to be determined[16]. The global population is diverse, however there are gaps in scientific literature in characterizing race/ethnicity-specific risk for disease development and progression[26]. First, it is important to note that health disparities in AD are also prominent in other racial/ethnic underserved groups[12,14,27]. Recently, it has been recognized that ethnic/racial factors significantly impact biological and medical risk factors for AD[26,28]. There are more non-Hispanic Whites (NHWs) living with AD than other racial/ethnic groups in the United States though per-capita Hispanics are more likely to have AD depending on their ancestry[12,14,29]. Hispanic is a broad term that encompasses individuals from cultures or countries with Spanish ancestry. Within the Hispanic population there are a variety of ethnic subgroups (i.e., Hispanic Americans or Caribbean Hispanic) attributable to geographical and cultural differences because each country of Latin America has distinct demographics, genetic structure, and migration history[29]. Hispanics are thus culturally and genetically heterogenous (admixed) due to long periods of isolation following gene exchanges between populations[29]. European, African, and Native American ancestry of varying proportions represent the Hispanic population, and previous studies indicated the overall increased risk in Hispanics may be driven by a specific ethnic group[29]. The Mexican American

(MA) population in general has little-to-no African ancestry[29]. Presence of comorbid conditions (i.e., cardiovascular disease and diabetes) may explain in part, some of the disparity in AD prevalence[12,14]. The MA population represents majority of the Hispanic population and has one of the fastest growing aging groups in the United States[28,30,31]. It is projected that the proportion of aging MAs will increase by three times and rates of AD will grow six-fold among Hispanics[28,30,31]. Regardless of the growing evidence of racial and ethnic disparities in AD risk and the growing aging population, scientific literature examining the potential differing factors for development and progression of the disease is limited[30,31].

Other than age, the APOE e4 allele has traditionally been characterized as a major risk factor for LOAD pathogenesis[12,14,15,19]. APOE encodes for the apolipoprotein E, which transports cholesterol in the bloodstream and is a major cholesterol carrier in the brain; the gene has three common alleles known as e2, e3, and e4[12,14,19]. Among the three forms of APOE, e2 is associated with decreased risk for AD compared to e3 and e4[12,14,19]. One copy of the e4 APOE allele increases one's risk for AD by three-fold and having two copies can increase the risk up to 12-fold; though these estimates are primarily based on Caucasian populations[12,14,19]. Frequency of APOE e4 differs across racial/ethnic groups; for example, the Mexican American population has a lower frequency of the APOE e4 allele and this allele also appears to have less of an effect[12,14,30,31]. In the limited literature available, research conducted by various groups indicates Mexican Americans may suffer from significant health disparities including (1) increased risk for MCI and AD, (2) missed diagnosis (3) diagnosed at more progressed stages, (4) develop MCI and AD at younger ages, (5) lack the genetic predisposition (APOE e4 allele), and (6) endure a disproportionate load of modifiable risks for AD[12,14,28,30,31]. In the Mexican American population, type-2-diabetes (T2D), depression, stroke, and obesity are common risk factors for developing cognitive impairment, although the reason for association between these comorbid conditions and cognitive impairment remains unclear[28]. This emphasizes the importance of longitudinal studies to improve diagnosis, treatment, and prevention of Alzheimer's disease in the Mexican American population.

Type-2 diabetes, a heterogenous metabolic disease, has been exemplified as having similar features in pathology to AD such as impaired glucose utilization, reduced mitochondrial

activity, and both metabolic and mitochondrial dysfunction[2,16]. T2D is characterized by hyperglycemia that is fundamentally caused by insulin resistance leading to insulin deficiency[16]. Pathophysiological features of T2D are islets of Langerhans presenting cell loss and/or dysfunction, and spontaneous islet amyloid polypeptide aggregation[16]. Islet amyloid polypeptide is a protein that is co-expressed with insulin secreted by cells and is a major component of islet amyloid[16]. Interestingly, formation of islet amyloid deposits is toxic to cells and there are reports that pancreatic amyloid islet degeneration is associated with NFT formation[16]. Recent evidence suggests insulin regulates Aβ and tau proteins, and numerous studies have established links between insulin resistance, diabetes, and AD. T2D increases one's risk for AD by two-fold and has been associated with progression of more severe forms of cognitive impairment[32]. Mexican Americans are more likely to develop T2D, which would suggest their increased risk for developing AD may be attributed to their metabolic health[33]. O'Bryant et al., conducted a study to create a serum-based biomarker profile of AD among MAs to compare with prior studies assessing markers for NHWs[31]. Top markers from their study suggested MAs exhibit a more metabolic phenotype including proteins related to obesity, insulin resistance, T2D, and metabolic syndrome, while NHWs exhibit a more inflammatory phenotype[31]. This might point to mitochondrial function as a contributing factor for the observed metabolic phenotype and differences in progression for MAs with LOAD.

MITOCHONDRIAL BIOLOGY & GENETICS

Mitochondria are essential to bioenergetic processes as they produce 90% of cellular energy to regulate cellular redox conditions, produce ROS, maintain Ca^{2+} homeostasis, synthesize and degrade biochemical intermediates, and regulate cell death via activating the mitochondrial permeability transition pore (mtPTP)[15,34]. The mitochondrial genome (i.e., mtDNA) is approximately 16.5 kilobases in size and encodes for 13 subunits of the respiratory complex, 22 tRNAs, and 2 rRNAs (12S and 16S)[15,34,35]. Remaining mitochondrial components including residual subunits of the respiratory complex, intermediary metabolizing enzymes, and mitochondrial biogenesis proteins necessary for proper function of the organelle are encoded by the nuclear genome (i.e., nuclear DNA; nDNA) and integrated into the system[15,34]. Complexes I, III, and IV are responsible for oxidizing reducing equivalents using oxygen[15,34]. Electrons flow through the

complexes to pump protons across the mitochondrial inner membrane to generate an electrical gradient[15,34]. This electrical gradient allows ATP production for energy by coupling oxidation with phosphorylation of ADP[15,34]. Within the mitochondria, ROS are predominantly generated at complexes I and III of the ETC[15].

There are 3 factors considered that can disrupt mitochondrial bioenergetic processes and cause disease variation in mtDNA sequence or nDNA-coded mitochondrial gene sequences and gene expression, or variation in calorie intake and calorie demands of the organism influenced by the environment[15,16]. Naturally, mutations to the mitochondrial genome are mixed with non-mutant mtDNA within cells, known as heteroplasmy, at varying degrees and are randomly distributed into daughter cells after replication (**Figure 2**)[5,15,34,35]. mtDNA mutations may impact protein synthesis or mitochondrial-encoded polypeptides and have been found to be clinically relevant based on three classes[34].

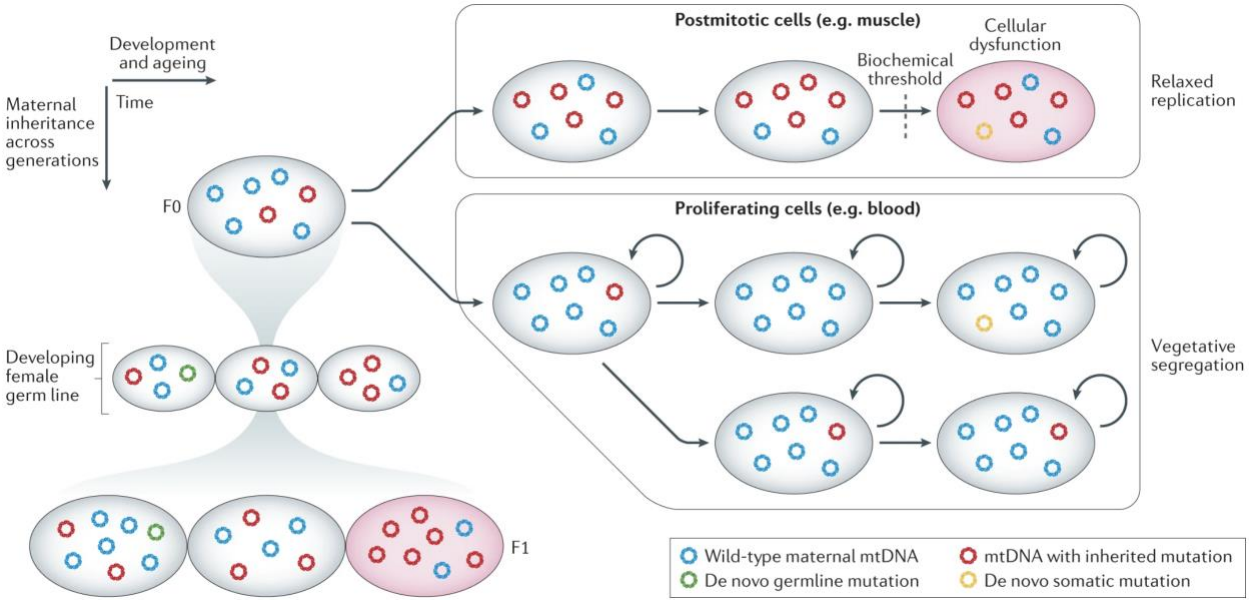


Figure 2. Elapsed development of mtDNA heteroplasmy level changes. During the human lifespan, including development and aging (horizontal axis), mutated mtDNA molecules inherited from the maternal lineage (red) can buildup in non-dividing/post-mitotic cells (top) or dividing/proliferating cells (not shown) through relaxed replication, while mutant mtDNA molecules in proliferating/dividing cells are removed (bottom) via vegetative segregation. Throughout the aging process new mutations known as de novo somatic develop (yellow). Cells with mutated mtDNA molecules above the biochemical threshold causes mitochondrial dysfunction through disruptions in oxidative phosphorylation, which can lead to cellular dysfunction and disease. Maternal genetic germline bottleneck throughout generations causes rapid changes in heteroplasmy levels due to disproportionate segregation of wild-type and

mutant mitochondrial genomes. (Image credit: Stewart, J. B., & Chinnery, P. F. (2021). *Nature Reviews Genetics*, and van den Ameele, J., Li, A. Y. Z., Ma, H., & Chinnery, P. F. (2020). *Seminars in Cell & Developmental Biology*)[35,36].

One class is characterized by recent deleterious mutations that can result into matrilineal disease due to the nature of inheriting mtDNA, and the distribution of heteroplasmic cells caused by meiotic and mitotic cell division[34]. Replicative segregation including meiotic and mitotic cell division can change the percentage of mutant and normal mtDNA approaching homoplasmy in either direction, however, mutant mtDNAs in post-mitotic cells are preferentially clonally amplified through nonselective genetic drift[15,34,35,37]. Preferential clonal expansion of mutant mtDNA in non-dividing cells results from non-selective or random replication due to increased replication frequency or the ability to replicate faster potentially caused by large deletions making the mitochondrial genome smaller in size[35]. Clonal expansion of mtDNA influences specific mutations that reach detectable levels and if the number of mutations exceed the biochemical threshold level, there will be biological effects and consequences to oxidative phosphorylation (OXPHOS)[35].

The second class is ancient adaptive variants that have accumulated down diverging maternal lineages when humans voyaged out of Africa predisposing to common diseases[34]. Mitochondrial variants with minimal to mild impact on health/fitness are constantly integrated into populations and can be selectively augmented in when advantageous in the energetic environment inhabited[34]. Succeeding accumulation of random mtDNA mutations in these subpopulations with different mitochondrial variants enriched based on their specific geographical location created haplogroups, or localized subpopulations with similar mtDNA haplotypes[34]. The haplogroup originated in Africa was then diverged into two mtDNA macrohaplogroups after initially departing from Africa[15,34]. The third class of mtDNA mutations are somatic mutations that appear throughout life in tissues[34]. Somatic mtDNA mutation rates can be modified by nuclear or mitochondrial variants and environmental factors[34,35,38]. Additionally, bioenergetic dysfunction and disease can develop from mutations in nDNA-coded mitochondrial components—those related to mitochondrial biogenesis can be destabilizing causing mtDNA deletions and/or depletion and ultimately lead to degenerative diseases[34,35].

MITOCHONDRIAL DYSFUNCTION IN AGING & ALZHEIMER'S DISEASE

In the literature, most studies that report strong correlations between age and mitochondrial function have been focused on studying skeletal muscle[5], signifying common features of aging such as reduced mitochondrial enzyme activity, respiratory capacity, and phosphocreatine recovery time, in addition to enhanced ROS production[5]. Since mitochondria harbor different ratios of non-mutant and mutant mtDNA following clonal expansion is a universal process in humans, mutations with low heteroplasmy levels will have little to no impact because naturally mitochondria are distributed in cells with varying numbers of mtDNA depending on the cell's bioenergetic needs[5,35]. It has been reported that germline cells with 60-80% mutant mtDNA surpass the biochemical threshold, at which point an individual is at high risk for developing a mitochondrial disease[5,35].

Impaired mitochondrial function has been implicated in numerous metabolic and degenerative diseases including AD. Mitochondria are essential to the brain because neuronal cells have limited glycolytic capacity to provide a constant supply of oxygen and energy for cellular/molecular needs[15,16]. Additional pathological changes exhibited in AD are reduced energy metabolism and mitochondrial dysfunction[16,21]. It is theorized that mitochondrial stress, OS, and mitochondrial dysfunction enhance pathology and play important roles in the pathogenesis of AD[21,34]. Mitochondrial stress can develop due to immense amounts of ROS, Ca^{2+} dysregulation, and/or other factors[34]. Calcium predestined for the mitochondria are released from the ER in mitochondrial-associated membranes (MAMs) containing presenilin complexes, and improper MAM Ca^{2+} regulation can cause mitochondrial ROS production leading to activation of apoptosis[34]. Aβ is an important factor capable of inducing mitochondrial dysfunction and increasing ROS production in AD[21]. Mitochondrial dysfunction and excessive levels of Aβ can activate mtPTP destroying neurons with defective mitochondria[34]. Further, Aβ decreases the activity of essential ETC enzymes and alters mitochondrial dynamics[18,21]. This has been demonstrated in the hippocampal neurons of AD patients[18]. Diminished activity of key enzymes of intermediate metabolism is a characteristic of abnormal cerebral glucose utilization and these enzymes are highly susceptible to oxidative damage[16]. Decreased pyruvate dehydrogenase activity specifically causes reduced levels of acetyl-CoA and subsequently plays a role in the

decreased synthesis of acetylcholine in presynaptic neurons[16]. Reduced production of the neurotransmitter in presynaptic neurons has been correlated with progressive mental disturbance in AD patients[16].

THE DAMAGING EFFECTS OF REACTIVE OXYGEN SPECIES AND OXIDATIVE STRESS IN ASSOCIATION WITH AGING & AD

Reactive oxygen species are partially reduced and highly reactive oxygen-containing molecules in comparison to molecular oxygen (O_2); ROS are often highly reactive molecules, otherwise termed free radicals, due to the presence of one or more unpaired electron(s), non-radicals, and singlet oxygen[39]. All aerobic cells produce ROS through enzymatic and non-enzymatic mechanisms[7,17]. ROS may also originate from exogenous sources such as pollution, tobacco, radiation, physical/chemical mutagens, chemical carcinogens, and heavy or transition metals where they metabolize into free radicals inside the body[40]. Several types of reactive oxygen species and how they are formed are represented in **Figure 3**.

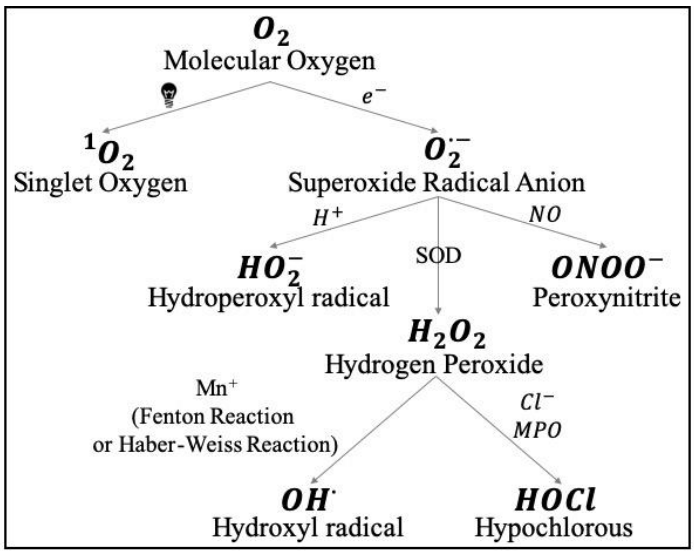


Figure 3. Several biologically relevant reactive oxygen species & their generation. Molecular oxygen is converted to singlet oxygen via photosensitization causing an energy transfer. Singlet oxygen is a ROS that oxidizes guanine to 8oxoG. Formation of superoxide radical anion is from electron transfers and enzymatic catalysis to molecular oxygen. Superoxide radical anion can be converted to hydroperoxyl radical from a proton. Conversion of superoxide radial anion to peroxynitrite is through a process of L-arginine producing nitric oxide (NO) via nitric oxide synthase (NOS). Most superoxide radical anion is converted to hydrogen peroxide via superoxide dismutase (SOD). Hydroxyl radical is formed from hydrogen peroxide via Fenton or Haber-Weiss Reactions resulting from an interaction of hydrogen peroxide with a redox-active metal ion.

Hydroxyl radical is highly reactive and are known to react mainly with phospholipids. Hydrogen peroxide converts to hypochlorous in the presence of chloride and myeloperoxidase (MPO). Hypochlorous is particularly damaging to proteins by oxidizing its amino acids.

Reactive oxygen species play a role in extraction of energy from organic molecules and are essential for cell signaling and regulation of immunological defenses and metabolism—processes vital for proper cellular function[9,21,28]. There are several intracellular sources of ROS including (1) mitochondria, (2) peroxisomes, and (3) endoplasmic reticulum (ER). Under normal conditions, ROS are generated from electron transfer reactions and in the presence of transition metal ions during processes of inflammation, respiration, and cellular metabolism, with the mitochondria being the primary source of production[7,9,10,15,41]. Peroxisomes are another endogenous source of ROS as they contain several oxidases capable of generating H_2O_2, which is then utilized by peroxisomal catalase to oxidize substrates involved in peroxidative reactions; these reactions are especially important in the liver and kidney to help detoxify molecules entering circulation[7]. Additionally, the smooth ER contains enzymes such as, cytochrome P-450 that catalyzes chain reactions to oxidize unsaturated fatty acids, detoxify lipid-soluble drugs, and other harmful metabolic products for degradation[7,9].

From various electron transfers and enzymatic catalysis reactions, molecular oxygen (O_2) can be reduced to unstable superoxide radical anion ($O_2^{\bullet-}$), which is then converted to hydrogen peroxide (H_2O_2) by acidic conditions or superoxide dismutase (SOD)[7,15]. Hydrogen peroxide is more stable, than the superoxide anion radical and is capable of diffusing through biological membranes[7,15]. Even though hydrogen peroxide is more stable and therefore less reactive, it can be decomposed to hydroxyl radical (OH) in the presence of redox-active metal ions such as, iron or copper, via a Fenton or Haber-Weiss reaction[10,15,16,39]. Iron is a transition metal that acts as a catalyst for redox reactions and helps form crucial complexes for oxygen transport and participates in cellular respiration[42]. Mitochondrial iron is necessary for heme synthesis and iron-sulfur cluster (ISC) biosynthesis, which are important for mitochondrial metabolism under normal physiological conditions[42]. Enhanced formation of ROS via Haber-Weiss and Fenton reactions can prevent mitochondrial uptake of cytosolic iron causing saturation of the iron transferrin protein and ultimately lead to elevated levels of plasma iron[9,42]. Elevated plasma iron and generation of ROS eventually result in pathological effects observed in arthritis, cirrhosis of the liver, and

depletion of beta cells in the pancreas contributing to diabetes and cardiomyopathy[42]. Friedreich's Ataxia (spinocerebellar ataxia), a neurological disease with potential to affect the heart, is caused by mitochondrial iron overload due to mutations in the gene frataxin encoding for the mitochondrial protein frataxin[42]. Frataxin binds iron and is involved in synthesis of ISCs and heme[42]. Thus, development of secondary disease could be because of iron-mediated enhancement to mitochondrial ROS production in tissues leading to mutations to the mitochondrial genome and cause progressive mitochondrial dysfunction impacting cardiac and beta cell function.

The roles of ROS can be considered paradoxical because they serve as essential biomolecules for regulating cellular functions but are toxic by-products of cellular metabolism dependent upon the concentrations of the ROS produced[7]. This contradictory phenomenon is observed with reactive nitrogen species such as nitric oxide because despite its cytotoxic effects, it functions as a signaling molecule in mediating vasodilation and is used for microbicidal killing in macrophages[7]. Aerobic cells contain antioxidant defense systems, composed of nonenzymatic and enzymatic antioxidants (e.g., glutathione, flavonoids, SOD, catalases, and glutathione peroxidase) which protect cells from ROS-related injuries by neutralizing these species to normal physiological levels[9,39]. However, an imbalance between oxidant production (i.e., prooxidants), and the antioxidant capacity of the cell to neutralize ROS and/or repair resulting oxidative damage is known as oxidative stress[6,7,17,39]. Increased levels of ROS or ROS evasion from antioxidant pathways causes damage of major biomolecules, oxidatively modifying lipids, carbohydrates, proteins, amino acids, and nucleic acids[5,8,9,39].

The link between OS and age-related diseases lies in modified biomolecule products, such as improper folding/unfolding of proteins, aliphatic side-chain accumulation, advanced oxidation protein products, and modified nucleic acid bases. This can be attributed to the fact that OS leads to subsequent oxidation of cellular components, activation of cytoplasmic and nuclear signal transduction pathways, modulation of gene expression and protein levels, and alteration of DNA polymerase activity causing cellular senescence and unsuccessful replication/transcription[6,17]. Damage to DNA can cause mutagenic lesions such as single and double strand breaks, inter/intra-strand crosslinks, DNA-protein crosslinks, sugar fragmentation products, and base oxidation.

Subsequent consequences of DNA damage are mutations, microsatellite instability, loss of heterozygosity, chromosomal aberrations, and altered gene expression[6]. Oxidative damage to DNA can be repaired through a variety of pathways including base excision repair (BER), nucleotide excision repair and mismatch repair; BER is primarily responsible for repairing bases damage by oxidation[9,39,41]. Failure of and of these repair mechanisms or persistence of oxidative damage can lead to mutagenic lesions/mutations and successive toxicity, apoptosis of cells, and/or malignancy.

Oxidation of guanine (G) via addition of hydroxide to the eighth position by singlet oxygen to 8-oxo-7,8-dihydroguanine (8oxoG) is one of the most prevalent mutagenic base modifications[9,10,17,43]. Guanine is more readily susceptible to oxidative stress due to its low oxidation potential[21,43]. There are various oxidized products of 8oxoG depending on the context; 8oxoG is the oxidized base, 8oxodG is the oxidized nucleoside (8-oxo-7,8-dihydro-2'-deoxyguanosine), and 8oxodGTP is the oxidized nucleotide (8-oxo-7,8-dihydro-2'-deoxyguanosine triphosphate)[6,9,17,41]. Oxidation of guanine causes a lack of specificity in base pairing, misreading of the modified base and adjacent nucleic acids[9,43]. During base pairing, 8oxoG takes an anti or syn conformation—anti follows Watson-Crick pairing rule by pairing with cytosine (C), and syn pairs with adenine (A).[43] Furthermore, the mutagenic property of 8oxoG or 8oxodGTP can cause transversion substitutions which can ultimately alter protein activity[6,9,17,41,43]. In events of DNA replication or synthesis, 8oxodGTP can be misincorporated opposite deoxyadenosine (dA) resulting in a A>C transversion, and 8oxoG/8oxodG results in a G>T transversion (**Figure 4**)[6,9,17,41,43].

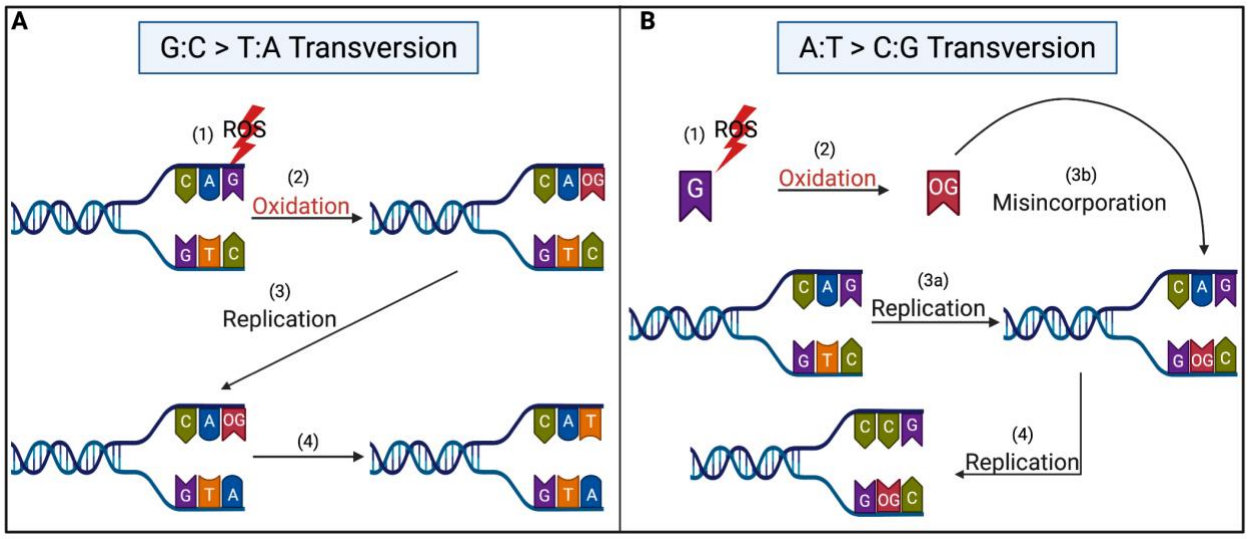


Figure 4. Overview of Mutagenesis of Oxidized Guanine in DNA. Oxidized guanine has mutagenic capabilities whether it is in native DNA (8oxoG or 8oxodG) or a free-floating nucleotide in the deoxynucleotide pool to be incorporated during replication. A. Oxidation of G in native double strand DNA (dsDNA) causes the modified base to be paired opposite deoxycytidine (dC). Subsequent replication without the removal of 8oxoG in the template strand by BER or other DNA repair pathways can result in the mispairing of dA opposite the modified base. Further replication of the dsDNA may result in deoxythymidine (dT) to be correctly paired with dA causing a G:C to T:A transversion. B. Oxidation of guanine in the deoxynucleotide pool can be misincorporated opposite dA during replication. An additional round of replication can incorporate dC opposite the oxidized base resulting in a A:T to C:G transversion. This figure was created with BioRender.com. (Figure credit: Reid, D. M., Barber, R. C., Thorpe, R. J., Sun, J., Zhou, Z., & Phillips, N. R. (2022). *Npj Aging*)[44].

The G>T transversion is possibly more common than A>C transversion because 8-oxodGTP in the nucleotide pool can be degraded into 8-oxodGMP and pyrophosphate by NDP-linked moiety X-type motif 1, and further degraded to 8-oxodG for removal[6,9,41]. Several studies have described 8-oxodG accumulation in nDNA and mitochondrial DNA (mtDNA) with age in vivo and in vitro, as well as decline in DNA repair activity[6,9].

Both modifications (8oxoG and 8oxodG) are considered biomarkers for oxidative stress and can be quantified to indicate DNA damage and repair rate[9,10]. 8oxoG is one of the most studied oxidative DNA modifications due to its highly mutagenic properties and has great clinical significance[9]. Biomarkers of oxidative stress/damage may serve as a diagnostic tool for assessing age-related disease risk and aid in the identification of therapeutic targets or evaluating therapeutic efficacy[6,17]. For example, 8oxoG and/or 8oxodG are biomarkers for COPD, cancer,

and chronic kidney disease, also, accumulation of 8oxodG in nDNA showed predictive significance for breast cancer risk[9,17]. Growing evidence has suggested 8oxoG and 8oxodG could serve as biomarkers for AD risk as well.

The brain is vastly susceptible to oxidative stress because of its high energy demand and oxygen consumption, abundance of lipids and iron, and a relatively insufficient antioxidant defense[18,21]. Accumulation of ROS modifies the function and expression of antioxidant enzymes, which has been observed in the central nervous system and peripheral tissues of AD patients[18,21]. Also, in AD brains high levels of DNA strand breaks were found in the hippocampus and cerebral cortex[21]. It is important to note that oxidative stress is a prominent contributor to Aβ aggregation and hyperphosphorylated tau, and numerous studies have provided evidence suggesting OS contributes to tau pathology because fatty acid oxidation accelerates tau polymerization[18,21,39]. Aβ peptides can bind with copper or iron to induce OS; in the hippocampus, amygdala, and other brain regions exhibiting severe AD histopathological changes showed abnormally high levels of copper and iron[21].

There is growing evidence suggesting a correlation between common pathological changes in AD and oxidative DNA damage[18,21,39]. Numerous studies have shown increased oxidative DNA and RNA damage in AD[21]. It is well established that mitochondria are predominant generators of ROS, moreover, the mitochondrial genome lacks histones and has reduced capacity to repair DNA increasing their susceptibility to OS and subsequently mitochondria are more prone to oxidative DNA damage[16]. The most common forms of oxidative damage observed in AD brains are 8oxodG and 8oxoG[21]. In the cortex and cerebellum of AD patients compared to controls, higher levels of 8oxodG were observed together with significant levels in the ventricular CSF[6]. Elevated levels of both forms of oxidatively modified guanine in AD brains have been demonstrated in nDNA and mtDNA when compared to age-match controls[6]. Mitochondrial dysfunction causes an increase in mtDNA somatic mutation rate, reduced energy metabolism, increased ROS, and intensifies the mitochondrial oxidative environment[15]. The cycle of abnormal mitochondrial function and activity increases apoptotic signals and will lead to diminished tissue function and will eventually lead to organ failure and age-related disease[15]. Altogether this evidence indicates mitochondrial-induced oxidative stress may play an important role in the

progression and pathophysiological changes in the brain of AD that relate to the endophenotype (i.e., a quantitative biological trait with relative heritability that consistently depicts the function of a distinct biological system and thus better explains the cause of death compared to the broader defined clinical phenotype[45–47]) observed in AD because neurons and mitochondria are sensitive to oxidative stress inducing detrimental cellular features such as, mitochondrial dysfunction, metal toxicity, and inflammation (**Figure 5**)[18].

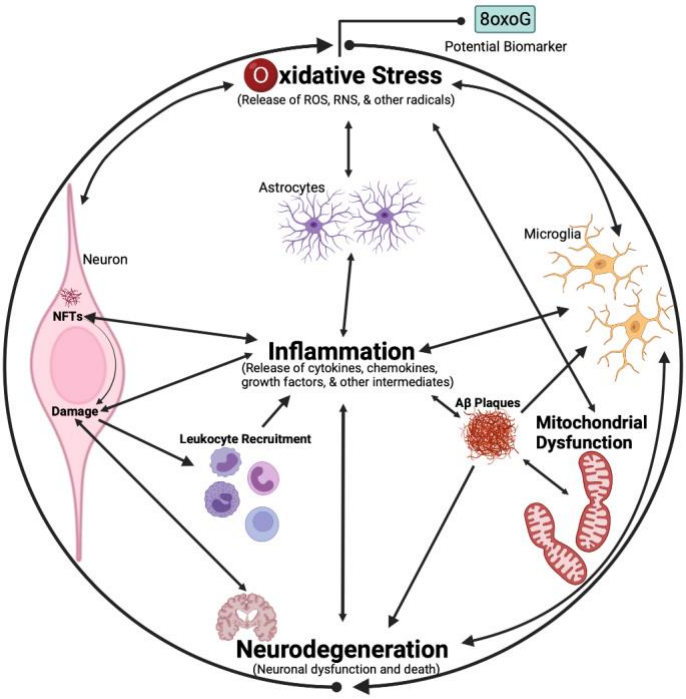


Figure 5. Cellular and molecular features of AD and their association to oxidative stress, inflammation, and neurodegeneration. In the aging brain, increased ROS, mitochondrial damage, and abnormal mitochondrial bioenergetics contribute to the developing toxic environment. Oxidative stress can be induced by several factors, such as, Aβ plaques, neurofibrillary tangles, microglia, astrocytes, etc. Brain inflammation is at the center as it is affected by plaques and tangles activating neuronal phagocytes, in addition to several other molecular factors. The induction of the pro-inflammatory state generates a vicious cycle of oxidative stress and inflammation fueling each other that contributes to neurodegeneration. 8oxoG is frequently observed in neurodegeneration and may serve as a biomarker for assessing disease risk, presence and/or progression of disease, and/or a therapeutic target. This figure was created with BioRender.com.

LIMITATIONS OF CURRENT DETECTION METHODS FOR MODIFIED BASES/OXIDATIVE DNA DAMAGE

The Alzheimer's Association highlights the need for simple and inexpensive tests that could aid in the diagnosis and/or assess the risk for development of AD[12]. In 2019, the Alzheimer's Association reported that there were three PET radiotracers approved by the U.S. Food and Drug Administration to aid clinicians in diagnosing AD; however, at the time they couldn't be used to provide a conclusive diagnosis in clinical settings in addition to other diagnostic criteria[14]. Substantial advancement has undergone in the past few years regarding the evaluation of AD hallmark pathology, which now allows for the identification of elevated levels of Aβ and phosphorylated tau in the CSF, and brain imaging via PET to locate accumulated Aβ and phosphorylated tau[12]. Biomarkers are measurable biological factors that are specific features related to the disease pathophysiological processes and can denote the manifestation or absence of a disease, risk for disease development, or disease progression.[12,14,24] Oxidative stress and subsequent oxidative damage has been observed in most diseases and their contribution to disease pathology can vary between diseases[6]. Since mitochondrial oxidative damage to guanine appears to play a role in AD, it may be useful for distinguishing phenotypic differences observed in certain populations, and a source of relevant modifiable risk. Biomarkers of oxidative damage, specifically 8oxoG and 8oxodG, may serve as a new source for clinical and experimental studies to transcend current problems with risk assessment, diagnosis, treatment, and elucidating the molecular mechanism contributing to AD pathology and progression.

There are several methods to detect and quantify global oxidatively damaged DNA such as, quantitative polymerase chain reaction (qPCR), in situ imaging, immunological techniques, high performance liquid chromatography (HPLC), liquid chromatography-mass spectroscopy (LC-MS/MS), and gas chromatography-mass spectroscopy (GC-MS)[6,8–10,41]. However, these methods have considerable shortcomings and lack of reproducibility among the techniques due to dissimilar background level detection of oxidation[8,10,41]. Immunological techniques, such as ELISA, seemingly have poor sensitivity and quantification ability of oxidized guanine compared to other methods currently used in the field. These techniques require raising antibodies against the modified base of interest and antibodies against 8oxoG have been found to be particularly cross-reactive with similar canonical bases, resulting overestimating levels of oxidized bases[10].

Furthermore, the antibodies lack specificity since most oxidized bases are poor antigens since their oxidation results in a subtle chemical change[10].

Chromatographic techniques serve as a better option to immunoassays because they have a higher sensitivity and can structurally identify modifications. Unfortunately, sample preparation for chromatographic techniques typically involves intricate DNA extraction, enzymatic digestions, and separation steps for isolation of oxidized guanine, further complicating the process, reducing throughput, and requiring experienced professionals to perform analysis[9,41]. HPLC can detect and measure 8oxoG; however, columns utilized for the separation process are not suitable for modified nucleosides, but with the addition of electrochemical (EC) detection 8oxodG quantification is achieved[6,10]. HPLC-EC has been thought to generate DNA mutations during sample preparation and, despite improvements made to the technique, artificial oxidative damage is still observed[6]. MS has been applied to previous HPLC and HPLC-EC techniques however it increased the cost of analysis and complicated protocols by introducing various labeled internal standards and artificial oxidation reactions to nucleosides during preparation[9,10]. Alternatively, GC-MS can be employed with comparable results to HPLC techniques by introducing known amounts of 8oxoG in samples or with pre-purification through HPLC or immunoaffinity to help reduce artifactual oxidation of traditional bases during sample preparation[6,8,10]. For the detection of 8oxoG, HPLC-MS/MS has shown great applicability and is considered the gold standard; however, this method is lengthy in time, has a considerable amount of preparation steps, which of some require prominent expertise and costly equipment[48]. These methods for detecting and quantifying DNA damage are deficient in identifying specific genome locations of the damage, therefore researchers investigating links between DNA damage and clinical phenotypes lose appreciation for the operable deleterious effects of the damage because the location can elucidate between protected and unprotected genomic regions and undiscovered mechanisms supporting distribution of damage and/or repair in the genome[48,49].

In the past several years, rapid and continuous advances in deep-sequencing technology have revolutionized the ability to interrogate the genome at the nucleotide level. Recent development and release of third-generation sequencing platforms have enabled researchers to

generate larger amounts of data at faster rates than even before[50,51]. With this new technology, new approaches have been developed in detecting modified bases at a greater resolution. Newley developed methods for localized detection of 8oxoG and its repair intermediates within 1kb follow DNA damage enrichment via arrays or sequencing by pull-down or chemical biology[48,49]. Reviews of these techniques discuss the different approaches and their strengths and weaknesses[48,49]. Currently there are approximately seven techniques that employ different affinity enrichment assays that are NGS-based including: OG-Seq (hyperoxidation and biotin-tag pull-down)[52], OGG1-AP-Seq (in vitro enzymatic excision of 8oxoG and biotin-tag pull-down)[53], AP-Seq (biotin-tag pull-down)[53], OxiDIP-Seq (ssDNA immunoprecipitation-based enrichment assay)[54], Click-code-Seq (enzymatic excision of 8oxoG and Click-tag pull-down)[55], snAP-Seq (hydrazine-*iso*-Pictect-Spengler- and biotin-tagged pull-down via Click chemistry)[56], and enTRAP-Seq (8oxoG affinity enrichment by His-tagged OGG1 K249Q mutant and immobilized metal affinity chromatography[57])[48,49]. Unfortunately, these techniques have additional limitations such as their several hundred base pair resolution is dependent on DNA fragment size[49] and approaches using sequencing by enrichment require substantial specificity to bind and pull-down the modification[58]. Furthermore, among these techniques, Click-code-Seq and snAP-Seq are the only two approaches that reach the single nucleotide resolution, but reagents used for these approaches are not commercially available[49].

Although single molecule sequencing methods have the advantage of being able to directly read nucleotides and their modified derivatives within DNA, in this project we did not utilize these particular methods due to their complexity, cost, and necessity for enhanced specificity to the base lesion 8oxoG. There are two approaches available for NGS including short-read sequencing and long-read sequencing[59]. In the context of investigating transversion mutations indicative of 8oxoG, short-read sequencing is advantageous because the utility is lower in cost and greater accuracy which are important for variant detection is and population-based studies with large amounts of samples[59]. A previous study conducted in 2018 by Kauppila, et.al., demonstrated the feasibility in taking advantage of the mutagenicity of 8oxoG to detect and quantify oxidative DNA damage by investigating the 8oxoG transversion mutation G>T after sequencing[60]. Regardless of all the variance between methods utilized to locate 8oxoG, evidence

throughout the literature generally indicates that the distribution of genomic 8oxoG is not random and that it may be related to transcriptional activation, and chromosomal and chromatin structures, wherein open regulatory regions are especially vulnerable to oxidation[48].

PROJECT OVERVIEW

Genetic studies of AD have revealed multi-genetic risk factors for developing the disease as well as endophenotypic differences in AD manifestation. However, genetic risk factors cannot completely explain the development of AD which further indicates the heterogeneous nature of AD pathophysiology. Studies characterizing the molecular and cellular conditions effected by AD point to oxidative stress, inflammation, altered metabolism, and mitochondrial dysfunction as underlying features of the disease. Within the Mexican American population metabolic health is of great concern and may be a contributing factor for advanced risk to AD. Mexican Americans have a greater prevalence of age-related diseases as T2D and AD[2]. Both conditions have similar pathology and display abnormal mitochondrial function and activity. The association between comorbidities burdening the MA population and cognitive decline is unclear, as well as the cause of their observed metabolic profile in individuals diagnosed with AD. Assessment of blood-based mitochondrial dysfunction via identifying oxidative mutations could help our understanding in disease risk, severity, and manifestation for Mexican Americans.

Hypothesis: Mexican Americans will exhibit elevated levels of oxidative damage due to the number of metabolic comorbidities that affect the population.

This hypothesis will be tested using two blood fractions (buffy coat PBMCs and plasma) from participants of TARCC to characterize mitochondria function in cognitively impaired aged Mexican Americans. The following chapters investigates mitochondrial health from buffy coat and plasma separately.

INNOVATION

The work described here contributes to the limited literature characterizing Alzheimer's disease in the Mexican American population. We investigate underlying factors possibly contributing to neurodegeneration caused by AD which may help explain the missing heritability

and additional features of the health disparity in MAs. This work is unique because we focus on utilizing blood-based indices of mitochondrial dysfunction via exploiting the mutagenicity of 8oxoG. Transversion substitution mutations result when 8oxoG is left unrepaired. Mapping these transversions is well-documented in various cancer types to determine mutational signatures that may relate to disease; however, employment of this approach is sparse among AD studies. Here we investigate all possible transversion substitutions specifically indicative of 8oxoG within the mitochondrial genome to determine its potential role in aging MAs brain health compared to their non-Hispanic White counterparts.

CHAPTER II: MITOCHONDRIAL DNA MUTATIONAL LOAD INDICATIVE OF 8OXOG OXIDATIVE DAMAGE FROM BUFFY COAT IN MEXICAN AMERICANS WITH COGNITIVE IMPAIRMENT

PUBLISHED IN NPJ AGING, 2022:
REID, D. M., BARBER, R. C., THORPE, R. J., SUN, J., ZHOU, Z., & PHILLIPS, N. R. (2022). MITOCHONDRIAL DNA OXIDATIVE MUTATIONS ARE ELEVATED IN MEXICAN AMERICAN WOMEN POTENTIALLY IMPLICATING ALZHEIMER'S DISEASE. NPJ AGING, 8(2). HTTPS://DOI.ORG/10.1038/S41514-022-00082-1

INTRODUCTION

The US aging population, characterized as individuals 65 years of age and older, is expanding at a rapid rate and is expected to grow for the next several decades reaching 88 million by 2050[27]. Correspondingly, the prevalence and number of age-related diseases, including Alzheimer's disease (AD), are anticipated to increase and further burden our healthcare system[27]. AD is the sixth leading cause of death in the US, and while the prevalence of other leading causes of death in the US have decreased or remained about the same, the number of deaths due to AD has significantly increased from the year 2000 to 2019[27].

Alzheimer's disease is a fatal neurodegenerative disease attributed to neuronal damage and accumulating amyloid-β (Aβ) plaques in the brain, first implicating thinking, learning, and cognitive function[15,27]. The late-onset class of AD is sporadic, multifactorial, genetically complex (i.e., AD heritability has been estimated between 60% and 80% and is highly polygenic)[19], and represents the majority (~90%) of total AD cases[21]. Moreover, it has been established that the progression of AD operates on a continuum from asymptomatic to AD-related dementia, with no distinct event denoting its onset; this progression is reflective of underlying accumulations of systemic and brain-specific pathology[23,25,27]. Early in progression there are two stages known as preclinical AD and mild cognitive impairment (MCI) due to Alzheimer's disease that identifies individuals with AD brain changes without and with associated symptoms, respectively[27].

Two of the most prominent pathological changes observed in AD are the accumulation of extracellular Aβ peptides producing senile plaques that block cell-cell signaling at synapses and

the accumulation of intracellular hyperphosphorylated tau protein resulting in neurofibrillary tangles that inhibit transportation of essential molecules[15,16,19,21,27]. However, the molecular and etiological events initiating AD pathologies remain to be determined[16]. Additional pathological changes exhibited in AD are mitochondrial dysfunction, chronic inflammation, and excess oxidative stress (OS)[16–18,21]. Mitochondrial stress, OS, and mitochondrial dysfunction are theorized to enhance AD pathology and play important roles in its pathogenesis[21,34] and impaired mitochondrial function has been implicated in both AD and metabolic disease[2,15,16,18,21].

Type-2 diabetes (T2D) has been shown to share similar pathological features with AD such as impaired glucose utilization, reduced mitochondrial activity, and both metabolic and mitochondrial dysfunction[2,16]. T2D is characterized by hyperglycemia caused by insulin resistance leading to insulin deficiency[16]. Pathophysiological features of T2D include islets of Langerhans cells presenting β cell loss and/or dysfunction, and spontaneous islet amyloid polypeptide aggregation[16]. Further, there are reports that insulin regulates Aβ and tau protein metabolism, and there are numerous reviews discussing the established connections between insulin resistance, diabetes, and AD[16,61–63].

Besides the existing general AD healthcare problems[27], there are gaps in the scientific literature characterizing race/ethnicity-specific risk for disease development and progression of Alzheimer's disease[26]. Recently, it has been recognized that ethnic/racial factors significantly impact biological and medical risk factors for AD[26–28]. In the US, there are more non-Hispanic Whites (NHWs) living with AD than other racial/ethnic groups, although per-capita Hispanics are more likely to have AD[27,29]. Hispanic is a broad term, as this population encompasses a variety of ethnic subgroups that exhibit geographical and cultural differences[29]. The Hispanic population is represented by varying proportions of European, African, and Native American ancestry, and previous studies indicated the overall increased risk in Hispanics may be driven by a specific ethnic subgroup[29]. The presence of comorbid conditions (e.g., cardiovascular disease and diabetes) may explain in part, some of the disparity in AD prevalence[27]. In the US, Mexican Americans (MAs) represent majority of the Hispanic population, has one of the fastest growing aging groups, and it is projected that by 2050 the number of aging MAs will triple, while rates of AD will grow six-fold among Hispanics[28,30,31].

AD pathophysiology in the MA population seems to be distinct from NHWs. For example, the APOE (apolipoprotein E) allele e4, which confers the largest risk for AD in NHWs, is far less significant in MAs. This may be in part due to the decreased frequency of the e4 allele, combined with a smaller effect size[27,30,31]. Correspondingly, a recent study determined that APOE e4 allele carrier status did not confer risk for MCI in MAs[64]. MAs clearly suffer from significant AD health disparities when compared to NHWs, including (1) earlier onset (~10 yrs) of cognitive impairment, (2) higher rates of missed diagnosis, (3) later diagnosis, and (4) increased prevalence of modifiable risk factors[27,28,30,31]. Depression, stroke, T2D, and obesity in the Mexican American population are common risk factors for developing cognitive impairment that are more common in MAs, although the etiology remains unclear[28,65]. Lifestyle and/or metabolic health may contribute directly to age-related neurodegeneration[27]. Combined, these data emphasize the importance of conducting further studies to improve the diagnosis, treatment, and prevention of Alzheimer's disease in the Mexican American population.

Recently, there is growing evidence suggesting a correlation between common pathological changes in AD and oxidative damage to nucleic acids[18,21,39]. Mitochondria are highly vulnerable to oxidative DNA damage because they are predominant generators of reactive oxygen species (ROS), and their mitochondrial genome lacks histones and has reduced capacity to repair DNA[16]. Oxidative stress is a prominent contributor to Aβ aggregation and hyperphosphorylated tau, and numerous studies have provided evidence suggesting OS contributes to tau pathology because fatty acid oxidation accelerates tau polymerization[18,21,39]. In the central nervous system and peripheral tissues of AD patients, accumulation of ROS modifies the function and expression of antioxidant enzymes[18,21]. Also, high levels of DNA strand breaks were found in the hippocampus and cerebral cortex of AD brains[21].

Mitochondrial dysfunction causes an increased mtDNA somatic mutation rate, reduced energy metabolism, increased ROS, and intensifies the mitochondrial oxidative environment[15]. The most common forms of oxidative damage observed in AD brains are 8-oxo-2'-deoxyguanine (8oxodG) and 8-oxo-guanine (8oxoG)[21]. In the cortex and cerebellum of AD patients compared to controls, significantly higher levels of 8oxodG were observed in the ventricular CSF[66]. Elevated levels of both forms of oxidatively modified guanine have been demonstrated in the nDNA of AD

brains when compared to age-matched controls[21]. Interestingly, Aβ is an important factor in mitochondrial dysfunction and increases ROS production in AD[21]. Mitochondrial dysfunction and excessive levels of Aβ can activate the mitochondrial permeability transition pore leading to the destruction of neurons with defective mitochondria[34]. Furthermore, it has been demonstrated in the hippocampal neurons of AD patients that Aβ decreases the activity of essential ETC enzymes and alters mitochondrial dynamics[16,18,21]. These enzymes are highly susceptible to oxidative damage and the reduced activity of key enzymes involved in intermediate metabolism is a characteristic of abnormal cerebral glucose utilization[16]. Mitochondrial-induced oxidative stress may play an important role in the progression and pathophysiological changes in the brain of AD because neurons and mitochondria are sensitive to oxidative stress inducing mitochondrial dysfunction (**Figure 6**).

Previously, our lab investigated the role of mitochondria in T2D and cognitive impairment in MAs through analyzing blood-based features of mitochondrial abnormalities (i.e., mtDNA copy number and cell-free mtDNA)[2]. The data suggested mitochondrial dysfunction assessed by mtDNA copy number was closely related to both T2D and cognitive impairment[2]. Here, our objective was to determine if abnormal mitochondrial function, indicated by oxidative DNA damage, differs between population (MA vs NHW), as well as to evaluate the effects of sex, cognitive impairment, and T2D on AD risk. Using Illumina-based Next Generation Sequencing, we quantified oxidatively modified guanine residues in mtDNA. Our data show that 8oxoG mutational load is significantly higher in MAs than in NHWs and is associated with cognitive function, sex, and education. Particularly, the sex effect observed was moderated by population. Stratified analysis for 8oxoG mutational load in MAs suggests significant elevation when comparing MAs with Alzheimer's disease to normal controls.

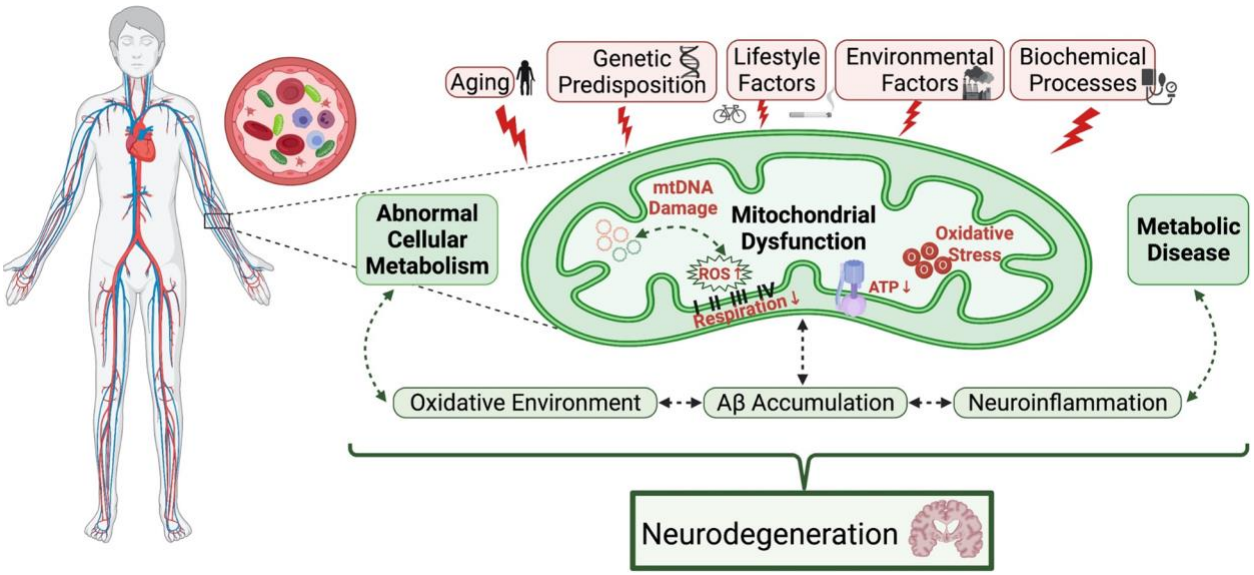


Figure 6. Graphical overview of global working hypothesis for risk factors and cellular/molecular processes that contribute to neurodegeneration. Modifiable and unmodifiable risk factors such as age, genetics, and lifestyle/environmental factors can induce elevated levels of ROS which could lead to mitochondrial and/or metabolic pathophysiology. This pathophysiology can contribute to and exacerbate an oxidative environment, neuroinflammation, and amyloid-beta accumulation that could ultimately promote neurodegeneration. This figure was created with BioRender.com. (Figure credit: Reid, D. M., Barber, R. C., Thorpe, R. J., Sun, J., Zhou, Z., & Phillips, N. R. (2022). *Npj Aging*)[44].

METHODS

Sample Acquisition and Description

Cohort

TARCC is the Texas Alzheimer's Research and Care Consortium, a longitudinal collaborative research initiative between ten Texas medical research institutions. The goal of TARCC is to investigate factors involved in the development and progression of AD in the MA population compared to NHWs.

Participants

This study was approved under the University of North Texas Health Science Center IRB #1330309-1; informed written consent was obtained from participants (or their legally authorized proxies) to take part in the study and allowing the publication of findings before data collection. Aging subjects enrolled in TARCC ($N = 559$; Table 1) who were diagnosed with AD ($n = 104$), MCI ($n = 127$), or normal cognition ($n = 328$) were selected to optimize matching with respect to age, sex and T2D distribution across MA and NHW fractions. An annual standardized assessment was conducted for each participant at one of the five original participating sites that

included a medical evaluation, neuropsychological testing, an interview, and a blood draw. Buffy coat samples from 261 NHWs and 299 MAs with the aforementioned cognitive phenotypes were analyzed in this work.

DNA Extraction, Amplification, and Sequencing

DNA extraction

DNA was extracted from 200 µL of buffy coat sample using the Mag-Bind® Blood & Tissue DNA HDQ 96 kit (Omega Bio-tek, Norcross, GA) using the Hamilton Microlab STARlet automated liquid handler (Hamilton Company, Reno, NV).

Whole mtDNA amplification

Whole mitochondrial genome for each sample was amplified using REPLI-g® Human Mitochondrial DNA kit (Qiagen, Venlo, Netherlands) following the manufacturer's protocol. This amplification approach follows a phi29 polymerase-based rolling circle and multiple displacement amplification. The purpose of mitochondrial genome amplification was to increase the amount of mtDNA relative to nuclear DNA to help with providing enough mtDNA for adequate coverage for whole-genome sequencing.

mtDNA sequencing

The Nextera XT™ DNA Library Preparation kit (Illumina, San Diego, CA) was used to prepare the library for sequencing following the manufacturer's protocol. The samples were sequenced on the NextSeq 550 Sequencer (Illumina) platform generating paired-end reads of 200 bp with an average read depth of 1855X.

Sequence mapping/alignment and variant calling

Raw mtDNA reads were aligned to the reference genome hg38 via BWA-MEM (v0.7.17) using the default parameter for mapping[67]. Generated SAM files were processed post-alignment with SAMtools (v.1.9) to produce BAM files that were sorted, indexed, and statistically assessed by coordinate[68]. All reads in the processed post-alignment BAM files were assigned to a single new read-group through the Picard tool AddOrReplaceReadGroups (http://broadinstitute.github.io/picard). Through GATK4 the Spark application of the Picard tool MarkDuplicates was employed on the single read-group BAM files to remove duplicate reads that may have resulted from sample preparation or the sequencing instrument[69]. BAM files with duplicate reads removed were indexed with SAMtools (v.1.9)[68] BAM files from the previous step

were used for calling somatic mutations with low allelic fractions for each sample excluding read orientation base qualities below 30 via a GATK4 tool variant caller named Mutect2 utilizing their mitochondria mode that automatically sets parameters for high-depth mitochondrial variant calling[69,70].

Oxidation artifact assessment

Oxidative somatic mutations have a low allelic fraction due to their prevalence which can also be affected by tissue heterogeneity (among other factors). CollectOxoGMetrics from Picard was utilized (http://broadinstitute.github.io/picard), a tool that calculates Phred-scaled probability scores based on low allelic frequency, sequence base context, and read orientation to distinguish alternative basecalls likely resulting from a true variant from those that may result from technical oxidative damage, specifically 8oxoG (**Figure 7**)[44]. Mutational oxidative damage results from 8oxoG base-pairing with cytosine or adenine during library preparation leading to G>T or C>A transversions during PCR amplification (https://support.illumina.com). See Costello et al. for a comprehensive analysis of next-generation sequencing 8oxoG artifact generation and detection[71]. The text file outputs from each file were subjected to manual review to exclude technical oxidative artifacts. Prior to the identification of total 8oxoG variant count, all detected somatic variants for each subject were assessed for any technical oxidative variants that may have been incorrectly identified as a true variant.

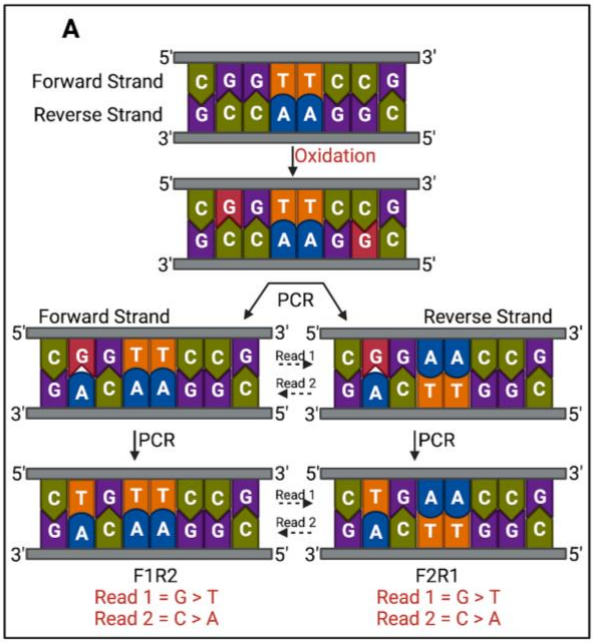


Figure 7. Next-generation sequencing read strand orientation bias and resulting artifacts. The presence of artifactual C>A/G>T transversions in sequencing data arises from the oxidation of guanine prior to PCR amplification. During PCR amplification adenine is misincorporated base-pairing with 8oxoG. Subsequent amplification steps produce the transversion mutations. Due to the nature of Illumina NGS chemistry, the G>T transversion will sequence on read 1 and the C>A transversion will sequence on read 2. This figure was created with BioRender.com. (Figure credit: Reid, D. M., Barber, R. C., Thorpe, R. J., Sun, J., Zhou, Z., & Phillips, N. R. (2022). *Npj Aging*)[44].

Identification of variants indicative of oxidative damage

From the variant call files (vcf), we aimed to identify the specific mutational events that would result from oxidative damage to the template DNA mtDNA. Samples vcf files were converted to tab delimited text files through vcflib, a library collection of tools to manipulate and describe sequence variation[72]. The variant call data were imported into Excel for manual data processing in order to remove indels, transitions, and non-oxidative transversions for the selection of oxidative variants. Oxidative variants were selected based on the mutagenic property of 8oxoG mispairing with adenine ultimately resulting in the signature oxidative transversion mutations (i.e., a G, T, C, or A alternative allele call where the reference allele call was a T, G, A, or C, respectively) shown in **Figure 4**[44]. Remaining variants indicative of oxidative damage were then further processed by removing variant calls with a read depth of less than 250 reads, removing individual SNPs (variants called in >90% reads), and removing variants where calls were limited to one orientation (forward or reverse; i.e., requiring coverage from both strands). Variants indicative of oxidative damage were summed for each sample and normalized for read depth (variant count per 1000 read depth) in both populations to test for group differences: cognitive function, sex, T2D, and comorbidity (T2D and cognitive impairment). Oxidative "hotspots" were identified as 8oxoG variant locations that occurred in at least 25 participants in the cohort.

Haplogroup assessment

In order to assess if background mitochondrial variants may be implicated in 8oxoG variant count, we used the NGS sequence data to derive haplogroups for statistical testing of group differences. Variant data were imported into Excel for manual processing in order to generate a list of individual SNPs for each sample (variants called in >90% reads). Each individual profile of mtDNA variants was imported into HaploGrep 2 (v.2.4.0), an online haplogroup

classification tool[73]. Haplogroups were defined in our statistical analyses based on the individual's identified macrohaplogroup or submacrohaplogroup. The sample size for this analysis was *n* = 560; one additional individual of unknown cognitive phenotype (specified as "other" and omitted from previously described analyses) was included here since this analysis is independent of cognitive phenotype.

Data Analysis

Statistical analyses were performed using Microsoft Excel, IBM SPSS software (v.24.0), and R software (v. 4.0.3). Welch's *t*-test (two-tailed) and two-way ANOVA were performed on 8oxoG mutational load to compare between both population groups and haplogroups. Multiple linear regression analysis was performed to evaluate the relationship between cognition, sex, age, education, and diabetes with 8oxoG variant count both within the whole study cohort and in stratified analyses of MAs and NHWs.

RESULTS

The descriptive statistics of the cohort are provided in **Table 1**. In both populations, MMSE, CDR sum, and years of education significantly differed between cognitive phenotypes as expected. Age was determined to significantly differ by cognitive diagnosis, and years of education was lower in the MA population. A Pearson correlation determined 8oxoG variant count did not significantly differ by age in the total cohort (**APPENDIX A**).

Table 1. Descriptive statistics of participants by population group and cognitive phenotype in the Texas Alzheimer's Research and Care Consortium.

	NC	MCI	AD	P-value[a]
Total Number of Subjects	328	127	104	
Non-Hispanic Whites	153	43	64	
Age [CI]	70.39 ± 1.178	71.35 ± 1.421	71.7 ± 1.056	0.338
Sex (F) [n, %]	78, 51.0%	21, 48.8%	29, 45.3%	0.749
Mini Mental State Exam (MMSE) [CI]	29.11 ± 0.1759	27.63 ± 0.6223	21.53 ± 1.413	0.000[b]
Clinical Dementia Rating (CDR) Sum [CI]	0.007 ± 0.009032	1.163 ± 0.2181	5.344 ± 0.8515	0.000[c]
Years of Education [CI]	16.07 ± 0.4063	14.56 ± 0.6597	15.11 ± 0.7524	0.001[d]
Diabetes (Y) [n, %]	59, 38.6%	18, 41.9%	22, 34.4%	0.726
Hyperlipidemia (Y) [n, %]	63, 41.2%	19, 44.2%	37, 57.8%	0.079
Obesity (Y) [n, %]	27, 17.6%	8, 18.6%	10, 15.6%	0.991
BMI kg/m^2 [CI]	27.331 ± 1.1778	27.272 ± 2.3284	27.394 ± 1.0624	0.996
	NC	**MCI**	**AD**	**P-value[a]**
Mexican Americans	175	84	40	
Age [CI]	67.62 ± 0.8156	69.88 ± 1.6912	73.38 ± 2.4848	0.000[e]
Sex (F) [n, %]	99, 56.6%	40, 47.6%	24, 60.0%	0.304
Mini Mental State Exam (MMSE) [CI]	28.14 ± 0.289	24.93 ± 0.787	19.88 ± 1.860	0.000[f]
Clinical Dementia Rating (CDR) Sum [CI]	0.006 ± 0.007897	1.113 ± 0.1560	5.738 ± 1.1827	0.000[g]
Years of Education [CI]	11.05 ± 0.6598	8.77 ± 1.1486	9.75 ± 1.5467	0.002[h]
Diabetes (Y) [n, %]	79, 45.1%	32, 38.1%	19, 47.5%	0.487
Hyperlipidemia (Y) [n, %]	101, 57.7%	48, 57.1%	20, 50.0%	0.670
Obesity (Y) [n, %]	84, 48.0%	38, 45.2%	8, 20.0%	0.005[i]
BMI kg/m^2 [CI]	30.917 ± 0.9992	31.295 ± 1.5374	28.718 ± 1.6508	0.116

a. The mean difference is significant at 0.05.

b. NC vs. MCI 0.016, NC vs. AD 0.000, MCI vs. AD 0.000

c. NC vs. MCI 0.000, NC vs. AD 0.000, MCI vs. AD 0.000

d. NC vs. MCI 0.003, NC vs. AD 0.042, MCI vs. AD 0.542

e. NC vs. MCI 0.028, NC vs. AD 0.000, MCI vs. AD 0.017

f. NC vs. MCI 0.000, NC vs. AD 0.000, MCI vs. AD 0.000

g. NC vs. MCI 0.000, NC vs. AD 0.000, MCI vs. AD 0.000

h. NC vs. MCI 0.001, NC vs. AD 0.273, MCI vs. AD 0.540

i. NC vs. MCI 0.905, NC vs. AD 0.003, MCI vs. AD 0.021

Total 8oxoG Variant Count is Significantly Higher in MA Females

Total 8oxoG variant count was significantly higher in the MA population compared to NHWs; mean = 7.46 and 5.96, respectively (**Figure 8**). In addition, female subjects had a higher 8oxoG variant count than males; mean = 7.06 and 6.43, respectively (**Figure 9**). The more comprehensive multiple linear regression model (**Table 2**) pointed to a significant interaction effect between population and sex related to 8oxoG variant count; p=0.01458, MA females being higher; p=0.0297 (**Figure 10**); additionally, years of education was identified as a significant factor (positive association). No other variables included in the multiple linear regression model demonstrated associated statistical significance (BMI, APOE, diabetes, cognition, age, population x education).

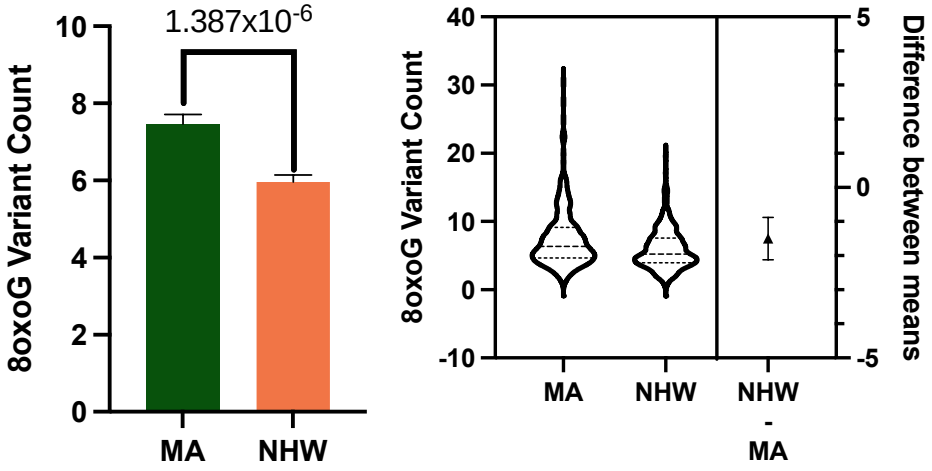


Figure 8. Cellular 8oxoG variant count is significantly higher in Mexican American population. **a** Total 8oxoG variant count was assessed by population using a two-tailed Welch's t-test (n = 559, t-statistic = 4.794, df = 558). Error bars represent standard error of the mean. **b** Violin plot showing the distribution of 8oxoG variant counts in Mexican American and non-Hispanic whites (n = 559) with effect size and confidence interval plotted on right y-axis. Dashed lines indicate the mean and dotted lines represent the 1st and 3rd quartile. The triangle represents the difference of the means, and the associated bar indicates the confidence interval.

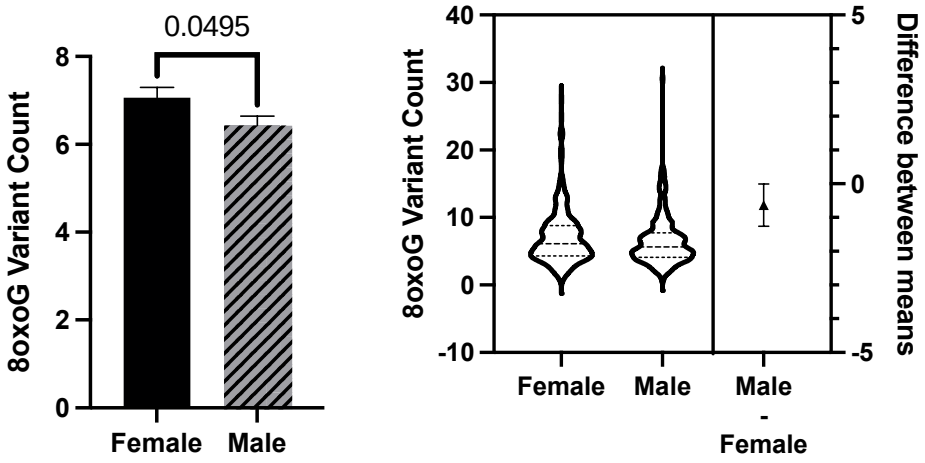


Figure 9. Cellular 8oxoG variant count is significantly higher in females. **a** Sex differences in 8oxoG variant count were determined using a two-tailed Welch's *t*-test (*n* = 559, *t*-statistic = 1.968, df = 558). Error bars represent standard error of the mean. **b** Violin plot showing the distribution of 8oxoG variant counts in females and males (*n* = 559) with effect size and confidence interval plotted on right *y*-axis. Dashed lines indicate the mean and dotted lines represent the 1st and 3rd quartile. The triangle represents the difference of the means, and the associated bar indicates the confidence interval.

Table 2. Cellular 8oxoG variant count and cognitive status (NC vs. MCI or AD) multiple linear regression model prediction considering population interaction effect with both sex and education.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	2.89096	2.21327	1.306	0.19204
Population (with respect to NHW)	-0.52522	1.46648	-0.358	0.72037
Cognitive Status (with respect to AD)	0.56007	0.45113	1.241	0.21497
Cognitive Status (with respect to MCI)	0.4515	0.40816	1.106	0.26914
Sex (with respect to Male)	-1.42142	0.44486	-3.195	***0.00148***
Diabetes (with respect to "Yes")	-0.35714	0.33836	-1.056	0.29166
Years of Education	0.14335	0.04556	3.146	***0.00174***
APOE ε2/ε2	-2.3329	2.6982	-0.865	0.38763
APOE ε2/ε3	0.64292	0.81025	0.793	0.42785
APOE ε2/ε4	0.47953	2.23482	0.215	0.83018
APOE ε3/ε3	0.61844	0.61248	1.01	0.31308
APOE ε3/ε4	0.36641	0.66434	0.552	0.58149
APOE ε4/ε4	0.65167	0.91108	0.715	0.47475
BMI	0.03357	0.02468	1.36	0.17424
Age	0.0308	0.0252	1.222	0.22207
Interaction: NHW x Male "Yes"	1.58842	0.64817	2.451	***0.01458***
Interaction: NHW x Years of Education	-0.15153	0.09808	-1.545	0.12291
R-squared	0.08275		*p*-value	5.87e-05
Adjusted R-squared	0.05567		df	16 and 542

F-statistic	3.056	Sample n	559

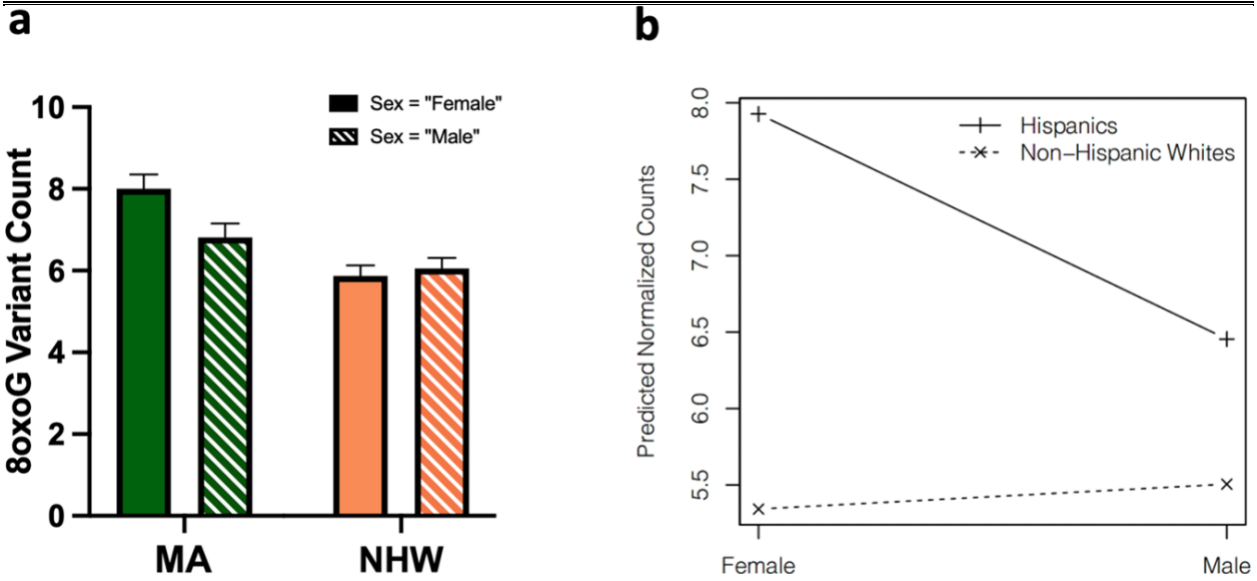


Figure 10: Population-by-sex interaction associated with total cellular 8oxoG variant count shows MA females have elevated 8oxoG counts. a Bar graph representing total 8oxoG variant count by population and sex as tested using a two-way ANOVA ($n = 559$, $p = 0.0297$, F-statistic = 4.75, df = 557) to determine if a population × sex interaction existed. **b** Interaction plot of predicted 8oxoG variant counts by sex in NHWs and MAs. Error bars represent standard error of the mean.

In a subsequent multiple linear regression analysis, we investigated the potential interaction between diabetes and cognitive status, in which we did not observe significant effects (**APPENDIX B**). We also derived the count of variants for each individual which corresponded to 8oxoG "hotspots" (i.e., frequently observed variants at certain locations within the mitochondrial genome) shown in **APPENDIX G**. In these "hotspot" analyses, we did not observe the same trends, and thus the metric proved to be generally less informative (**APPENDIX G-H and Tables J-Q**). Additionally, in the NHW population we observed associations between 8oxoG "hotspot" variant count and APOE status (**APPENDIX P-Q**), which was not observed in the MA population (**APPENDIX N-O**).

Population-specific Effects on 8oxoG Variant Count

As expected, based on the previous multiple linear regression analyses, 8oxoG variant count was significantly associated with sex (females higher) for MAs as shown in **Table 3**. However, interestingly, cognitive status of AD was in marginally significant association with

8oxoG variant count (shaded row, **Table 3**; bar graph provided in **Figure 11**), but this trend was not observed in NHWs (shaded row, **Table 4**). Two-way ANOVAs in NHWs did not show significance; however, in MAs there was significance for sex F(1,295)=5.8 and p=0.0166 (Figure 5). No other variables were associated with 8oxoG variant count in the MA population. BMI and age were marginally significant (both positive) in association with 8oxoG variant count in non-Hispanic Whites (**Table 4**); no other variables were associated with 8oxoG variant count. Another intriguing result is the significant positive association of education with 8oxoG variant count that is limited to the MA population (**Table 3**).

Table 3. Multiple linear regression results for cellular 8oxoG variant count within Mexican Americans.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	5.12074	3.63974	1.407	0.16054
Cognitive Status (with respect to AD)	1.4363	0.80455	1.785	*0.07528*
Cognitive Status (with respect to MCI)	0.75592	0.59613	1.268	0.20581
Sex (with respect to Male)	-1.39809	0.52709	-2.652	***0.00844***
Diabetes (with respect to "Yes")	-0.27268	0.52981	-0.515	0.60719
APOE ε2/ε2	-1.69728	4.72412	-0.359	0.71965
APOE ε2/ε3	0.2323	2.17461	0.107	0.91501
APOE ε3/ε3	-0.46508	1.9656	-0.237	0.81313
APOE ε3/ε4	-0.72004	2.00661	-0.359	0.71998
APOE ε4/ε4	-0.82949	2.92419	-0.284	0.77687
BMI	0.02671	0.03926	0.68	0.49687
Years of Education	0.14462	0.05415	2.671	***0.008***
Age	0.01253	0.03924	0.319	0.74983
R-squared	0.05857		p-value	0.1297
Adjusted R-squared	0.01907		df	12 and 286
F-statistic	1.483		Sample n	299

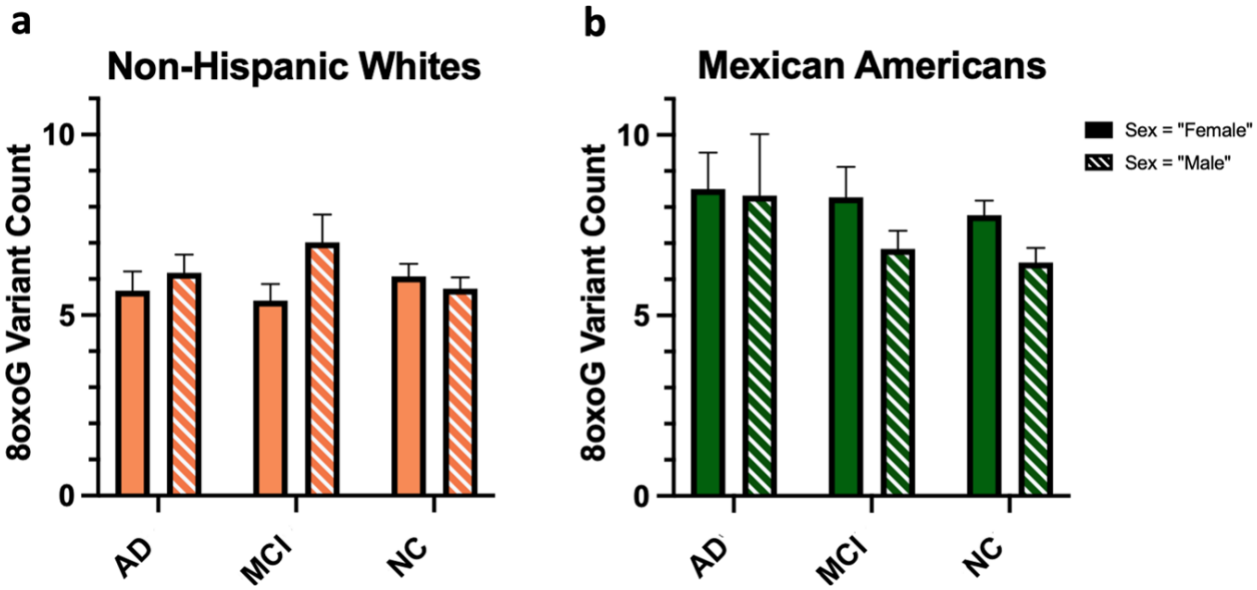


Figure 11. Cellular 8oxoG variant count by cognitive phenotype in each population. a Bar graph of 8oxoG count by cognition in NHWs tested using a two-way ANOVA (*n* = 260). **b** Bar graph of 8oxoG count by cognition in MAs tested using a two-way ANOVA (*n* = 299) to determine if a cognition × sex interaction existed in each population. Error bars represent standard error of the mean.

Table 4. Multiple linear regression results for cellular 8xoG variant count within non-Hispanic whites.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	0.59072	2.81035	0.21	0.8337
Cognitive Status with respect to AD	-0.23042	0.48975	-0.47	0.6384
Cognitive Status (with respect to MCI)	0.03883	0.54101	0.072	0.9428
Sex (with respect to Male)	0.18523	0.37372	0.496	0.6206
Diabetes (with respect to “Yes”)	-0.55416	0.40735	-1.36	0.1749
APOE ε2/ε2	-4.04738	2.99107	-1.353	0.1772
APOE ε2/ε3	0.30547	0.80613	0.379	0.7051
APOE ε2/ε4	0.29945	1.7921	0.167	0.8674
APOE ε3/ε3	0.89858	0.54737	1.642	0.1019
APOE ε3/ε4	0.76197	0.65579	1.162	0.2464
APOE ε4/ε4	1.28886	0.81301	1.585	0.1142
BMI	0.04855	0.02883	1.684	*0.0935*
Years of Education	-0.02771	0.07105	-0.39	0.6969
Age	0.05584	0.03101	1.801	*0.073*
R-squared	0.04846		*p*-value	0.488
Adjusted R-squared	-0.001829		df	13 and 246
F-statistic	0.9636		Sample n	260

Additional multiple linear regression analyses using cognition as a binary predictive variable (where MCI and AD are combined into cognitive impairment, CI, and NC is normal controls) were conducted (**APPENDIX C-F, K, M, O, and Q**); the higher 3-category resolution shown in **Table 3** (AD/MCI/NC) revealed a potential effect of AD on 8oxoG variant count in MAs, but the effect is not observable in the CI/NC regression analyses since it is diluted by the presence of MCI (**Table 4**).

Haplogroup-associated Elevation and Depression of 8oxoG Variant Count

Based on the Welch two-sided t-test, we observed haplogroup effects on 8oxoG variant burden within the combined cohort. Haplogroups A and C exhibited elevated 8oxoG variant counts (**Figure 12A; Table 5**). Conversely, haplogroups I and K exhibited lower 8oxoG variant counts (**Figure 12A; Table 5**). For population stratified inference, in the NHW population, using Welch's t-test, our results demonstrate haplogroup H displayed higher 8oxoG variant counts (**Figure 12B; Table 5**). Haplogroup I among NHWs showed reduced 8oxoG variant counts (**Figure 12B; Table 5**). In the MA population, the Welch's t-test reported haplogroup L had significantly reduced 8oxoG variant counts when compared to all other haplogroups observed in the MA population (**Figure 12C; Table 5**).

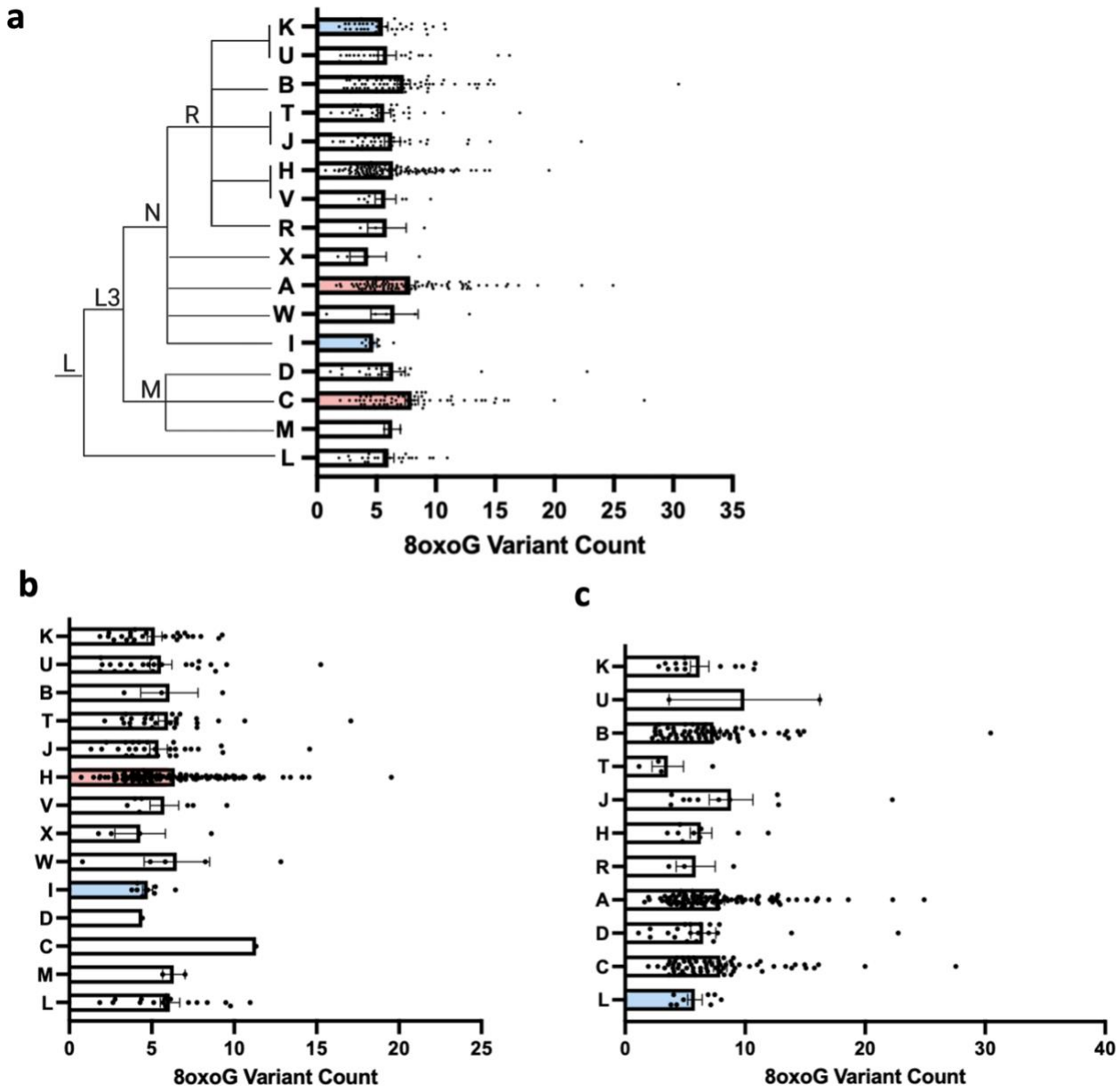


Figure 12. Cellular 8oxoG variant count by mitochondrial haplogroup. a Differences in total 8oxoG variant count by mitochondrial haplogroup of the cohort was assessed using Welch's *t*-test ($n = 560$). **b** Total 8oxoG variant count by mitochondrial haplogroup in NHW participants was assessed using Welch's *t*-test ($n = 261$). **c** Differences in 8oxoG variant count between mitochondrial groups in the MA population was determined by performing Welch's *t*-test ($n = 299$). Pink bars indicate significantly higher 8oxoG variant count and blue bars indicate significantly lower 8oxoG variant count. Error bars represent standard error of the mean. The mitochondrial haplogroup tree was illustrated based off the RSRS-oriented mtDNA tree build 17 from PhyloTree$_{mt}$ to include only macrohaplogroups and sub-macrohaplogroups represented in our cohort[74].

Table 5. Mitochondrial DNA haplogroup-associated cellular 8oxoG variant count mean within the combined cohort (NHW + MA; *n* = 560), MAs alone (*n* = 299), and NHWs alone (*n* = 261).

Combined Reference Haplogroup	Reference Haplogroup Mean	Non-Reference Haplogroup Mean	T-statistic	df	95% CI [LL,UL]	p-value
Haplogroup A	7.838462	6.51901	2.8867	132.94	[0.41535, 2.2236]	0.004545
Haplogroup B	7.303423	6.680043	1.1141	79.498	[-0.49026, 1.7370]	0.2686
Haplogroup C	7.9767	6.591353	2.4044	77.106	[0.23809, 2.5326]	0.0186
Haplogroup D	6.41393	6.7679	-0.34926	21.068	[-2.4612, 1.7533]	0.7304
Haplogroup H	6.364743	6.862089	-1.461	238.38	[-1.1679, 0.17326]	0.1453
Haplogroup I	4.752968	6.783636	-5.9126	11.613	[-2.7818, -1.2796]	8.127E-05
Haplogroup J	6.338881	6.784038	-0.65586	40.693	[-1.8162, 0.92590]	0.5156
Haplogroup K	5.555933	6.841888	-2.9575	51.28	[-2.1588, -0.41315]	0.004681
Haplogroup L	6.015808	6.790599	-1.635	32.557	[-1.7394, 0.18980]	0.1117
Haplogroup M	6.339475	6.756114	-0.59872	1.1166	[-7.3496, 6.5163]	0.6479
Haplogroup R	6.759379	5.872131	0.54395	2.0397	[-6.0017, 7.7762]	0.6401
Haplogroup T	6.81632	5.664707	1.9933	34.475	[-0.021895, 2.3251]	0.05419
Haplogroup U	6.794277	5.906105	1.1589	26.368	[-0.68615, 2.4625]	0.2569
Haplogroup V	6.767161	5.76442	1.1311	6.4231	[-1.1322, 3.1377]	0.2985
Haplogroup W	6.756741	6.519856	0.11912	4.0531	[-5.2561, 5.7298]	0.9109
Haplogroup X	6.772306	4.297114	1.6096	3.0671	[-2.3588, 7.3092]	0.2039
MA Reference Haplogroup	**Reference Haplogroup Mean**	**Non-Reference Haplogroup Mean**	**T-statistic**	**df**	**95% CI [LL,UL]**	**p-value**
Haplogroup A	7.838462	7.270255	1.0848	202.54	[-0.46460, 1.6010]	0.2793
Haplogroup B	7.361145	7.487292	-0.20331	97.394	[-1.3575, 1.1052]	0.8393
Haplogroup C	7.925182	7.331154	0.95278	98.058	[-0.64322, 1.8313]	0.343
Haplogroup D	6.513252	7.528179	-0.94198	21.341	[-3.2534, 1.2236]	0.3567
Haplogroup H	6.320519	7.495663	-1.2566	9.3373	[-3.2791, 0.92881]	0.2394
Haplogroup J	8.822136	7.413168	0.7684	9.3447	[-2.7158, 5.5338]	0.4612
Haplogroup K	6.207265	7.521843	-1.6302	16.132	[-3.0229, 0.39373]	0.1224
Haplogroup L	5.81628	7.505487	-2.5543	9.6501	[-3.1700, -0.20844]	0.0294
Haplogroup R	7.476387	5.872131	0.97675	2.0968	[-5.1588, 8.3674]	0.4276
Haplogroup T	7.513174	3.56018	2.9696	3.2244	[-0.12126, 8.0272]	0.05397
Haplogroup U	7.443624	9.935283	-0.39662	1.0031	[-81.727, 76.743]	0.7595

NHW Reference Haplogroup	Reference Haplogroup Mean	Non-Reference Haplogroup Mean	T-statistic	df	95% CI [LL,UL]	p-value
Haplogroup B	6.072026	5.944759	0.072626	2.0441	[-7.2587, 7.5133]	0.9486
Haplogroup H	6.368296	5.628957	1.9889	218.16	[0.0067015, 1.4720]	0.04796
Haplogroup I	4.752968	5.983953	-3.465	13.256	[-1.9970, -0.46450]	0.004076
Haplogroup J	5.419158	6.007037	-1.0462	33.322	[-1.7307, 0.55496]	0.303
Haplogroup K	5.17599	6.02422	-1.7433	32.388	[-1.8388, 0.14238]	0.09076
Haplogroup L	6.104487	5.934498	0.27424	20.689	[-1.1202, 1.4602]	0.7866
Haplogroup M	6.339475	5.943185	0.56521	1.1507	[-6.1779, 6.9705]	0.6622
Haplogroup T	5.941546	5.98848	-0.07583	30.492	[-1.3101, 1.2162]	0.94
Haplogroup U	5.983957	5.555742	0.62167	25.685	[-0.98851, 1.8449]	0.5396
Haplogroup V	5.951232	5.76442	0.20969	6.5517	[-1.9494, 2.3230]	0.8403
Haplogroup W	5.935018	6.519856	-0.29383	4.0673	[-6.0752, 4.9055]	0.7833
Haplogroup X	5.971888	4.297114	1.0874	3.0859	[-3.1504, 6.5000]	0.3544

DISCUSSION

Evaluating 8oxoG Variant Count by Population and Sex

Alzheimer's Disease was discovered over a century ago, and through research our understanding of the disease has exponentially grown. However, there are many gaps in our knowledge, particularly with respect to how this disease affects individuals from different ethnic/racial backgrounds. Our group investigated peripheral levels of mitochondrial 8oxoG, a characteristic of mitochondrial dysfunction, and its association with cognitive impairment, type-2 diabetes, and comorbidity (cognitive impairment and T2D) within the Mexican American population compared to non-Hispanic Whites. We hypothesized the MA population would demonstrate higher levels of mitochondrial oxidative damage due to the number of comorbid conditions burdening this population, such as cardiovascular disease, diabetes, and depression[27]. Overall, our results demonstrate that 8oxoG variant count was significantly higher in MAs compared to NHWs, and this effect was largely driven by MA females. In subsequent regression analyses, we observed that 8oxoG variant count is suggestively associated with AD cognitive

status (compared to control) particularly in MAs. Intriguingly, this analysis also revealed a positive association of 8oxoG variant count with education, warranting further investigation of biological and/or environmental influencers of 8oxoG.

The level of 8oxoG variant count in the mitochondrial genome was significantly higher in MAs compared to NHWs, which may be because MAs are at increased risk for metabolic disorders. Metabolic syndrome and obesity are associated with increased oxidative stress, which can lead to genomic instability such as increased levels of oxidative DNA damage[75,76]. Metabolic syndrome is a collection of conditions such as deficient glucose tolerance, fatty liver, and increased body weight, adiposity, and triglyceride levels[76]. Thus, metabolic health risk could account for the observed significant difference in levels of mitochondrial 8oxoG count. Furthermore, base excision repair (BER) is a predominant DNA repair pathway for oxidative DNA damage; failure of this system allows features of genomic instability to persist and accumulate[39]. Higher 8oxoG levels in MAs may be influenced by differences in DNA repair machinery expression due to the population's associated metabolic burden and/or population-specific variants that impact DNA repair efficiency.

Interestingly, recent evidence suggests that DNA damage repair is necessary for metabolic health, derived from observations demonstrating mtDNA repair glycosylase OGG1, an essential enzyme for BER, may influence metabolic phenotypes in high fat diet exposure[75–77]. Functional OGG1 prevents obesity and metabolic dysfunction[75,76] through altered *PGC-1α* expression and fatty acid oxidation[76]; reduced levels of PGC-1α has been reproducibly observed in T2D patients[15,78,79] and have been related to increased levels of ROS and decreased levels of β-oxidation enzymes[32]. The metabolic burden in MAs may be associated with metabolic

dysfunction which could alter OGG1 function causing elevated levels of 8oxoG. Interestingly, the genetic polymorphism in *OGG1* (rs1052133, Ser[326]Cys) has been associated with T2D risk in Mexican Americans[77] further suggesting that insufficient response to oxidative DNA damage may be implicated in metabolic disease in the Mexican American population.

There were significantly higher 8oxoG counts for Mexican American females compared to Mexican American males. In the literature, there is no clear consensus whether levels of DNA damage differ significantly based on biological sex, and this may be due to differing sample type, technique, and/or applied method of detection across studies[80,81]. In 2014, results of a meta-analysis indicated that there are no differences between sex and DNA damage[80]. Conversely, a recent review determined that men have higher levels when compared to women; however, inconsistency in reports indicate that other factors such as lifestyle may contribute to the sex effect on the prevalence of such lesions[81]. Further, most of the studies to date have not explicitly compared oxidative damage among different racial/ethnic groups in an aging population. Elevated levels of 8oxoG variant count in MA females may be partially explained by the fact that Mexican American women have a higher frequency of T2D[33]. Oxidative stress and mitochondrial dysfunction are well documented in T2D pathophysiology, and a restrictive diet reduces oxidative stress[15]. Additionally, there is accumulating evidence underlining sex differences in mitochondrial function and activity, and levels of oxidative stress in an age-dependent manner[82–84]. Silaidos, et al., observed that PBMCs from females exhibited significantly higher ATP levels, citrate synthase activity, uncoupled respiration, and ETC complex and system capacity when compared to PBMCs of men[82]. Recent evidence shows sex hormone status may be involved[85]. For example, mitochondrial function in female mice revealed that younger female mice display lower oxidative

stress levels compared to males and that subsequent ovariectomy limited the apparent protection against DNA damage; this protection was eliminated in aged female mice[84]. The lack of consistent data in the literature regarding sex differences in aging and age-related diseases emphasizes the need for further work to better understand sex-associated disease risk, especially in ethnic populations that are rapidly expanding.

Population-specific Associations with 8oxoG Variant Count

Our results from multiple linear regression analyses are suggestive of an AD-effect on 8oxoG variant count in the Mexican American population. In the literature there is accumulating evidence supporting the implication of mitochondrial dysfunction as a primary and/or secondary factor contributing to AD partially because of the significant levels of oxidative damage observed in various organs and tissues of individuals with cognitive impairment[18,21]. In particular, previous studies report significantly higher levels of 8oxoG and/or DNA damage in patients diagnosed with MCI or AD compared to controls, suggesting that (1) oxidative stress and subsequent DNA damage are features of AD pathophysiology, (2) accumulating oxidative DNA damage may be an early marker of AD, and (3) 8oxoG could potentially serve as a biomarker for MCI and/or AD[39,86–89]. However, there is little information regarding ethnic/racial differences in levels of oxidative DNA damage, and particularly peripheral levels of 8oxoG in the context of cognitive decline. Here we demonstrate population-specific variation in peripheral levels of mitochondrial oxidative DNA damage—the associations observed in the MA cohort were non-significant in the NHW cohort and had effect sizes in opposite directions; these findings emphasize the importance of future replication studies. As previously mentioned, it is possible that the Mexican American population has more pronounced effect due to their metabolic burden and potential genetic variation in

DNA repair machinery. Additionally, cognitive impairment has been well-documented in T2D, which increases the risk for AD by two-fold and has been associated with progression of more severe forms of cognitive impairment[16,32,61–63]. Moreover, oxidative stress is particularly related to amyloid and tau pathology through stimulating a vicious cycle of pathophysiology provoking mitochondrial dysfunction and metal toxicity, which would ultimately result in an increased mutational load and neurotoxic environment contributing to neuronal loss[18,39,89]. This gathering evidence may explain to an extent the observed suggestive association between 8oxoG variant count and AD in the Mexican American population. Correspondingly, the stronger association reported in MA females could be attributed to the extended lifespan of women and age-related decline in sex hormones diminishing the protective effects on antioxidant defenses and mitochondrial capacity[83,90–92]. Mitochondria are responsible for steroidogenesis and its interaction with sex steroids plays an important role in the brain[91]. Brain levels of sex hormones are known to decline with age, therefore, emphasizing lifestyle factors, metabolic, health, and age may be of particular importance in accounting for the vulnerability of Mexican American females to cognitive decline and associated pathophysiology[91].

Interestingly, the positive association of 8oxoG variant count in MAs extended to years of education. Fletcher, et al., reported associations between educational attainment and cognition in older age, after controlling for family background and genetic factors, and an interaction demonstrating those with an increased risk for AD mildly benefit from a higher educational background[93]. Educational attainment has been moderately studied in MAs with evidence indicating the disparity in cognitive impairment and dementia is due to genetic, behavioral, and socioeconomic factors[94]. Socioeconomic factors were found to be especially important in the

disparity, highlighting the inequity in educational attainment among underrepresented or immigrant populations which may contribute to their risk for cognitive decline[95]. Additionally, there are several reports indicating the protective effect of education on cognitive impairment does not entirely translate to MAs[96]. Data from a previous study suggests MAs may only benefit from cognitive-protective effects when years of education exceeds 12 years (i.e., education beyond high school)[96]. The reason for the observed positive association of 8oxoG with years of education in MAs is unclear; further studies investigating the effect of educational attainment on cognitive function and the paradoxical increase in 8oxoG in the Mexican American population are warranted.

Haplogroup-associated Elevated and Reduced 8oxoG Variant Count

In the whole cohort, we observed mitochondrial haplogroups A and C had significantly higher 8oxoG variant counts, while haplogroups K and I showed significantly reduced levels each independently compared to all other haplogroups. Previous data has shown haplogroup K to demonstrate a protective effect against AD in European populations[97]. The significantly lower levels of 8oxoG variant count in haplogroup K may be related to its apparent low risk for developing AD which is associated with increased oxidative damage. In the NHW population, haplogroup H was found to have significantly higher levels of 8oxoG variant counts compared to all other haplogroups observed in the population. Established features of European ancestry include mtDNA haplogroups associated with largest oxygen consumption, ineffective oxygen utilization, and slightly deficient DNA repair capacity causing elevated levels of ROS that could subsequently cause elevated levels of oxidative DNA damage[98]. Furthermore, a study recently demonstrated synergism between APOE e4 carrier status and mitochondrial haplogroup H—

when combined, individuals were at higher risk for AD[99]. Therefore, the elevated 8oxoG variant count exhibited by haplogroup H that we see here was not surprising, due to their associated altered mitochondrial capacity. Conversely, in the MA population haplogroup H did not demonstrate a significant elevation in 8oxoG variant count; however, as previously mentioned the APOE risk allele appears to have less of an effect in the MA population. This observation further suggests that MAs are differentially affected by established risk factors for cognitive impairment compared to their NHW counterparts. Nonetheless, studies investigating mitochondrial haplogroup risk in neurodegeneration is very limited, and thus, it is difficult to comment on whether there is evidence to confirm or refute our findings suggesting haplogroup-specific 8oxoG variant count differences (refer to review by Ienco et al., for a comprehensive assessment of the literature)[100]. Furthermore, due to the limited sample size (i.e., various haplotypes are observed less in one population compared to the other) our power to detect rare mitochondrial haplotype effects is limited, and thus presumably causing the lack of overlap between the mitochondrial haplogroups associated with 8oxoG variant count in both cohorts. Through the historic geographical migration of certain groups and maternal nature of mtDNA inheritance, there are observed variations in haplotype frequencies between societal-based ethnic/racial groups[101]. There is accumulating evidence indicating mitonuclear allelic interactions considerably alter the expression of important health-related phenotypes by influencing the quality of oxidative phosphorylation and metabolic function[101]. Gene flow of the mitochondrial genome differs from that of the nuclear genome and considering the generation of differing genetic variation throughout populations, it is hypothesized that the course of mito-nuclear coadaptation may be population specific[101]. This is likely relevant to the MA population as they

are considered an admixed population. Therefore, there will be limited overlap and significant results when comparing the two populations separately, especially in relation to the whole cohort.

CONCLUSION

While our results are potentially insightful, there are several limitations to note. First, it is important to acknowledge that the methods employed here are indirect measures/indicators of oxidative damage; however, this limitation is difficult to overcome since methods for detecting oxidative damage at the per-base resolution specific to mtDNA generally have (1) technical artifacts arise during library preparation, (2) low sequencing resolution, (3) higher detection limits, and/or (4) the requirement for specific and sensitive enzymes, proteins, or antibodies[102]. Another obvious limitation is the lack of data regarding metabolic disease in this cohort; our study was limited to self-described diabetes, which is likely an oversimplification given the highly heterogenous nature of metabolic syndrome in the MA population. Further, the inclusion of additional markers of metabolic health could have potentially helped with establishing an association. In general, it is challenging to interpret these results from a biological/mechanistic perspective, but importantly, they open the door for avenues of research that may prove highly relevant to addressing and resolving MA health disparities in age-related disease, namely, risk for AD.

Future studies will aim to increase the sample size and improve subject characterization of metabolic phenotypes to better resolve causal aspects of oxidative damage in MAs, specifically with respect to female vulnerability. We acknowledge that our data is suggestive in association to AD; however, future studies utilizing quantitative cognitive measures such as MMSE, CDR sum,

and other measures of neuropsychological testing and cognitive function, may improve our power to support the implication. Further, additional biochemical and genetic studies would solidify these results and aid in drawing conclusions. Such studies may include correlation analyses between 8oxoG variant load and expression of DNA repair machinery and ROS response systems, as genetic variant analysis of nuclear-encoded DNA repair genes and mito-nuclear epistatic effects. Ideally, the studies conducted and proposed here would be recapitulated in matched blood and brain tissue to validate the potential application of these peripheral phenotypes as biomarkers for brain pathology. Additionally, future studies will aim to include another population cohort and validate mitochondrial oxidative load using an alternative method such as liquid chromatography-tandem mass spectrometry (LC-MS/MS).

To conclude, the work we present here describes a differential effect of oxidative mitochondrial damage that is associated with cognitive decline among Mexican American females. We also describe a unique approach for sensitive quantification of putative oxidative damage in blood, a highly accessible tissue, and its potential relevance to cognitive aging in Mexican Americans. Further, we identify a potential role for mtDNA-based haplogroup risk in 8oxoG accumulation. The systemic elevation of 8oxoG load specifically in MA females may point to an underlying source of risk for cognitive decline in this vulnerable group, revealing avenues for more precise prevention, diagnosis, and treatment of cognitive dysfunction.

ACKNOWLEDGEMENTS

We thank Nicole Hales (currently at 10X Genomics) for library preparation and sequencing. Research reported in the publication was supported both by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number

R25GM125587 and by the National Institutes on Minority Health and Health Disparities of the National Institutes of Health under Award Number U54MD006882. Dr. Thorpe was supported by the National Institutes on Minority Health and Health Disparities of the National Institutes of Health under Award Number U54MD000214 and the National Institutes on Aging under Award number P30AG059298. This study was made possible by the Texas Alzheimer's Research and Care Consortium (TARCC) funded by the state of Texas through the Texas Council on Alzheimer's Disease and Related Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CHAPTER III: CIRCULATING CELL-FREE MITOCHONDRIAL DNA MUTATIONAL LOAD INDICATIVE OF 8OXOG OXIDATIVE DAMAGE IN COGNITIVELY IMPAIRED MEXICAN AMERICANS

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia that is characterized by symptoms of cognitive decline such as, having trouble remembering, problem-solving, communicating, along with additional cerebral incompetencies[12,103]. This heterogenous neurodegenerative disease is commonly known for its neurotoxic pathophysiological properties including the accumulation of amyloid beta (Aβ) peptides and hyperphosphorylated tau protein causing extracellular amyloid plaques and intracellular neurofibrillary tangles, respectively[12,18,103]. However, impaired mitochondrial function and chronic inflammation have been frequently reported and can be considered as contributors to the observed endophenotypic manifestations of cognitive impairment likely caused by Alzheimer's[18,31,103–107]. In particular, non-Hispanic Whites (NHWs) appear to exhibit an inflammatory endophenotype[31,104–106], while Mexican Americans (MAs) presented a metabolic endophenotype as demonstrated via donor blood serum-based protein biomarker profiles[31]. This evidence may point at biological and lifestyle factors influencing cognitive impairment that are distinct to a population group, which could in part explain endophenotypic differences. Identifying blood-based biomarkers capable of predicting and assessing disease progression is of great importance to elucidate factors that cause the pathophysiological heterogeneity of AD for the development of more precise therapeutics, as current established biomarkers are quite invasive and can be costly.

Type-2 diabetes (T2D) is a considerable risk factor for AD due to its comorbid association with cognitive impairment; however, the precise pathophysiological mechanisms connecting the complex diseases are unclear. Common pathology between AD and T2D include both metabolic and mitochondrial dysfunction, impaired glucose utilization, and reduced metabolic activity[2,16]. As the US aging population (i.e., individuals 65 years of age or older) continues to expand with the Hispanic/Latinx population is expected to exponentially increase compared to other

ethnic/racial groups, it is expected that there will be a prolific burden of age-related diseases affecting this population such as AD and diabetes[12,13]. Large amounts of evidence indicate the high disproportion of AD cases in Hispanics/Latinos compared to NHWs is attributed to the increased prevalence of metabolic syndrome, obesity, cardiovascular health risks, and diabetes[2,13,108]. This observation indicates that mitochondrial health and factors that affect it may hold greater biological importance in the development of age-related disease for individuals with Hispanic/Latinx descent, as the accumulating data indicates their mitochondrial capacity/function appears to be reduced compared to other racial/ethnic populations.

Currently in the US there are six drugs approved by the FDA for treatment of AD; however, five of the treatment options help temporarily relieve symptoms of the disease, while one drug (aducanumab) was approved in June 2021 under their accelerated pathway, which has been shown to alleviate the accumulating amyloid beta plaques in the brain, yet peer-reviewed clinical cognitive benefits have not been publicly available at this time[12,109]. The lack of available treatments for the AD continuum is one of the largest gaps within the Alzheimer's disease field among the comorbid associations and health disparity affecting underrepresented populations.

Mexican Americans are the largest population segment within the Hispanic/Latinx population and unfortunately, this group experiences earlier onset of cognitive impairment, late diagnosis, greater rate of cognitive decline, and more severe forms of dementia compared to their non-Hispanic White counterparts[12,28,30,31]. Additionally, the major risk allele for late-onset AD is the apolipoprotein E (APOE) e4 allele, which seems to be observed less frequently in the MA population and may not confer to a large effect size as for NHWs[12,30,31,110,111]. With little understanding of the genetic basis of AD in racial/ethnic populations, there is a crucial need to investigate the pathogenic mechanism of AD related to their molecular phenotypic presentation corresponding to lifestyle and metabolic health, since there are several mitochondrial-related diseases impacting this population to a greater degree compared to other populations in the absence of clear genetic contributors for cognitive decline.

Compared to the nuclear genome, the mitochondrial genome is especially susceptible to oxidative damage because of its location proximal to the oxidative phosphorylation machinery where majority of reactive oxygen species (ROS) are generated[16,35,112]. Additionally, mitochondria

lack protective histones and a robust DNA repair capacity which helps prevent genomic instability[16]. Elevated levels of ROS generate oxidative stress and an oxidative environment capable of damaging important biomolecules through oxidation reactions and may cause detrimental effects in genomic coding regions by altering function and/or the expression of mitochondrial genes[6,9,17]. The most common forms of oxidative DNA damage are: 8-oxo-7,8-dihydroguanine (8oxoG) and 8-oxo-7,8-dihydrodeoxyguanine (8oxodG) due to the low oxidation potential of the base guanine, thus causing it to be vulnerable to oxidative stress[9,10,17,21,43,49]. Oxidation to guanine encompasses unique mutagenic properties that when left unrepaired from the DNA damage response system, oxidized guanine can perturb cellular function through several different mechanisms and effect protein—DNA binding (e.g., transcription factors)[9,17,43,49]. Furthermore, it can develop substitution mutations which change the amino acid coded and may consequently modify the activity of encoded products, such as proteins[6,9,17,41,43]. Due to the nature of the mitochondrial genome within various cell types and tissues, the location, metabolic activity, and enzymatic processes will influence distinctive oxidative DNA damage to tissue with differing function[49].

Studies have demonstrated correlations between AD pathology and oxidative DNA damage, such as, Aβ was shown to induce oxidative stress in the CNS and peripheral tissues, decrease vital ETC enzymes activity, and modified mitochondrial dynamics[21,104]. Growing evidence implicates oxidative DNA damage as a primary and secondary contributor to pathology observed in the AD continuum[38,113–115]. Recent evidence from our lab analyzing blood-based indices of mtDNA copy number (CN) and cell-free mtDNA (cf-mtDNA) to investigate mitochondrial dysfunction in complex disease (T2D and cognitive impairment) among MAs showed that mtDNA CN was significantly associated with both T2D and cognitive impairment[2]. Also, cf-mtDNA was found to be higher in individuals with either disease, reaching significance in individuals with both diseases compared to healthy normal controls[2]. The quantification of mtDNA CN are indicators of mitochondrial biogenesis and cellular energetics[2,116], which can be used as a measurement of mitochondrial health since the amount of mtDNA released by cells is correlated[117]. Cf-mtDNA has been increasingly studied as a biomarker for systemic inflammation during cellular stress or apoptosis, as mtDNA fragments are released by cells into the

bloodstream as circulating cf-mtDNA (ccf-mtDNA), where it may act as a damage associated molecular pattern (DAMP) due to mitochondria's bacterial DNA origin and elicit an immune response by activating innate immune cells and inflammation[2,112,117,118]. There are numerous studies assessing ccf-mtDNA as a clinical diagnostic and predictive biomarker[112,117,119–122], and accumulating evidence indicates the extent of abnormal mitochondrial function and mtDNA damage may be attributed to correlations between disease severity and levels of ccf-mtDNA[112,121]. Furthermore, increased levels of ccf-mtDNA are shown in neurological diseases that exhibit inflammatory features, which could generate a vicious cycle of immune cell recruitment, enhancement of ROS generation and the oxidative environment, and mtDNA activation of systemic inflammation augmenting additional damage to mitochondria and its genome[2,112,119].

In a similar population-based cohort, we found that mtDNA variants indicative of 8oxoG in buffy coat PBMCs were significantly elevated in MAs compared to NHWs and were associated with cognitive function, sex, and education[44]. Correspondingly, a study reported by Miller, et al., revealed AD neurons compared to age-matched controls had significantly elevated levels of somatic single nucleotide variants (sSNVs) than anticipated when considering sSNVs are known to increase with age[38]. Variability in the increased levels of sSNVs among neurons corresponded with variability in AD pathology observed in neurons from affected brain regions effected, and the universal distribution of the variants are presumed to occur secondary to proceedings that develop disease pathology[38]. Supplementary analysis of sSNVs established potential mechanisms of oxidative DNA damage developed from 8oxoG (nonsynonymous mutations, e.g., C>A) that might contribute to the significant increase of sSNVs in AD[38]. Protein-coding genes exhibited increased substitution mutations, which could be increased ROS and oxidative stress commonly observed in AD brains, CNS, and periphery, which can contribute to inflammation and mitochondrial dysfunction—also frequently reported in AD[38,123]. This accumulating evidence may point at a mutational signature important to AD pathophysiology influencing the differing endophenotype reported in AD, particularly of those with different mitochondrial capacity, metabolic health, and comorbidities linked to ethnic/racial health disparities for developing cognitive decline.

Oxidative transversions may be associated with increased mitochondrial dysfunction and could contribute to the continuous progression of AD until a clinical endpoint, death. In this study, our objective was to use a blood-based measurement of 8oxoG sSNVs as an indicator of impaired mitochondrial function to investigate the role of mitochondria in pathophysiology of complex disease by (1) identifying influencers of 8oxoG in ccf-mtDNA such as population and sex, and (2) determining differences in 8oxoG buffy coat PBMCs and ccf-mtDNA from plasma based on AD/MCI diagnoses and/or related endophenotypes.

METHODS

Sample Acquisition and Description

Cohort

The Texas Alzheimer's Research and Care Consortium (TARCC) is a population-based collaborative longitudinal research initiative that has expanded between several Texas medical research institutions[124]. TARCC explores factors that may attribute to the development and progression of cognitive impairment due to AD in the MA population compared to their NHW counterparts.

Participants

The study received institutional review board approval under the University of North Texas Health Science Center IRB #1330309-1; informed written consent was obtained from participants and/or their legally authorized proxies to take part in the study and allow publication of findings before data collection. Volunteer aging participants enrolled in TARCC annually complete a medical evaluation, clinical interview, neuropsychological testing, and blood draw. Eligible participants obtained categorical clinical diagnoses of 'Alzheimer's disease', 'Mild Cognitive Impairment', and 'Normal Control' based on the criteria provided by the National Institute for Neurological Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association[125]. Additional information regarding the inclusion and exclusionary criteria of TARCC has been discussed elsewhere[126]. This study included NHW and MA subjects (N = 559; **Table 23**) diagnosed with Alzheimer's Disease (n = 104), Mild Cognitive Impairment (n = 127), or normal cognition (n = 328). Buffy coat PBMC samples obtained from NHWs (n = 261) and MAs (n = 299) and a subset of plasma samples from 62 NHWs and 57 MAs (N = 119; **Table 41**) collected at the same visit were selected to match the distribution of subjects with respect to age, sex, and

type-2 diabetes among both populations. The plasma subset did not include individuals diagnosed with MCI. These samples were analyzed to characterize cellular and circulating cell-free mtDNA (ccf-mtDNA) oxidative damage from blood.

DNA Extraction, Amplification, and Sequencing

DNA Extraction

DNA from both the buffy coat and plasma was extracted individually from 200 mL of each sample using the Mag-Bind® Blood & Tissue DNA HDQ 96 kit (Omega Bio-tek, Norcross, GA). Buffy coat extractions were conducted using the Hamilton Microlab STARlet automated liquid handler (Hamilton Company, Reno, NV) and plasma DNA was extracted manually using the same chemistry.

Whole mtDNA amplification:

The whole mitochondrial genome and large mtDNA fragments for each sample were amplified using the REPLI-g® Human Mitochondrial DNA kit (Qiagen, Venlo, Netherlands) following the manufacturer's protocol. This kit uses the high fidelity proofreading phi29 DNA polymerase capable of both rolling circle and multiple displacement amplification in combination of random hexamers[127]. Mitochondrial genome amplification was performed in order to increase mtDNA levels relative to nuclear DNA to enhance mtDNA coverage for whole genome sequencing. Amplified product was quantified via Qubit dsDNA BR assay on the Qubit 4 fluorometer (Invitrogen™, Thermo Fisher Scientific, Waltham, MA) for each sample and a small test size of approximately 12 samples were evaluated to determine the distribution of amplicon sizes using the 4200 TapeStation System (Agilent Technologies, Santa Clara, CA) following the manufacturer's protocol. The Genomic DNA ScreenTape and corresponding reagents were used to determine the presence of mtDNA fragments from 200 bp to the whole genome.

mtDNA Sequencing

The Nextera XT™ DNA Library Preparation kit (Illumina, San Diego, CA) was used to prepare the sample library for sequencing following the manufacturer's protocol. All buffy coat and plasma samples were sequenced on the NextSeq 550 Sequencer (Illumina) platform generating paired-end reads of 150bp.

Sequence Mapping/Alignment and Variant Calling

Raw mtDNA gzipped FASTQ pairs generated for each sample were aligned to the reference genome hg38 via BWA-MEM[67] (v0.7.17) using the default parameter for mapping to generate SAM files. Post-alignment SAM files were processed with SAMtools (v.1.9) to produce BAM files that were subsequently sorted, indexed, and statistically assessed by coordinate[68]. Resultant processed aligned sequence reads within the BAM files were assigned to a single new read-group through the Picard tool "AddOrReplaceReadGroups"[128]. Duplicate reads resulting from sample preparation, or the sequencing instrument were removed from each sample single new read-group BAM file with the GATK4 Spark application of the Picard tool "MarkDuplicates"[69]. The BAM files were then indexed with SAMtools (v.1.9)[68] and used for somatic variant calling including low allelic fractions and excluding read orientation base qualities (Phred score) under 30. High-depth mitochondrial somatic variants were called via the GATK4 variant caller, Mutect2, utilizing the mitochondria mode[69,70].

Oxidation Artifact Assessment

Picard tool, CollectOxoGMetrics, was used to calculate Phred-scaled probability scores for basecalls to differentiate biological alternative basecalls from technical oxidative damage due to 8oxoG (http://broadinstitute.github.io/picard). Readers are encouraged to review the study reported by Costello et al., for a comprehensive analysis of Next Generation Sequencing 8oxoG artifact generation and detection[71]. A text file was generated for each sample and were subjected to manual review to exclude technical oxidative artifacts with a Phred score below 30.

Identification of Variants Indicative of Oxidative Damage

The variant call files were manually assessed to identify 8oxoG transversions within the mitochondrial genome. The process of identifying these specific oxidative transversions has been previously described[44]. Variants indicative of 8oxoG damage for each subject from the buffy coat portion were summed and normalized by accounting for read depth (variant count per 1000 read depth) to evaluate group differences based on the following variables: population, cognition, sex, type-2 diabetes, comorbidity (cognitive impairment and disease), and lifestyle factors.

***APOE* and *OGG1* Genotyping Imputation**

Genome-wide SNP profiles were generated using the Illumina Infinium Multi-Ethnic Global Array which types 1.7million SNPs. Standard filtering based on SNP missingness, individual missingness, and minor allele frequency (5%) was conducted according to Anderson et al.,

2010[129]. Genetic imputation of APOE (rs7412 and rs429358 for individuals missing APOE genotypes) and OGG (rs1052133) was performed using Impute2 based on the 1000 Genomes Project Phase 3 data; probabilistic genotypes for were called at a threshold of 0.8.

DATA ANALYSIS

Statistical analyses were performed using Microsoft Excel, IBM SPSS software (v. 27.0), R software (v. 4.2.0), and GraphPad Prism software (v. 9.4.0). Welch's t-test (two-tailed) was performed on 8oxoG mutational load to compare between both population groups and haplogroups. Multiple linear regression analysis was performed to evaluate the relationship between cognition, sex, age, education, and diabetes with 8oxoG variant count both within the whole study cohort and in stratified analyses of MAs and NHWs.

RESULTS

Evaluation of 8oxoG Variant Count in the Buffy Coat PBMCs of MA and NHW TARCC Participants

Descriptive statistics for the cohort analyzed for cellular characterization of mitochondrial 8oxoG variants are displayed in **Table 23**. As anticipated the MMSE, CDR sum, and years of education in both populations used for buffy coat PBMC analysis had significantly different means when distributed across cognitive status groups. In the Mexican American population age was found to significantly differ between cognitive groups and overall years of education was observed at lower levels compared to non-Hispanic Whites.

Table 6. Descriptive statistics of MA and NHW participants categorized by population and cognitive phenotype in the Texas Alzheimer's Research and Care Consortium for buffy coat mitochondrial DNA oxidative mutational load.

	NC	MCI	AD	P-value[a]
Total Number of Subjects	328	127	104	
Non-Hispanic Whites	153	43	64	
Age [CI]	70.39 ± 1.178	71.35 ± 1.421	71.70 ± 1.056	0.338
Sex (F) [n, %]	78, 50.98%	21, 48.84%	29, 45.31%	
Mini Mental State Exam (MMSE) [CI]	29.11 ± 0.1759	27.63 ± 0.6223	21.53 ± 1.413	<0.000[b]
Clinical Dementia Rating (CDR) Sum [CI]	0.007 ± 0.009	1.163 ± 0.2181	5.344 ± 0.8515	<0.000[c]
Years of Education [CI]	16.07 ± 0.4063	14.56 ± 0.6597	15.11 ± 0.7524	0.001[d]
BMI kg/m^2 [CI]	27.331 ± 1.150	27.272 ± 2.328	27.394 ± 1.062	0.996
Diabetes (Y) [n, %]	59, 38.56%	18, 41.86%	22, 34.38%	
Hypercholesterolemia (Y) [n, %]	90, 58.82%	26, 60.47%	50, 78.13%	
Hyperlipidemia (Y) [n, %]	56, 36.60%	15, 34.88%	39, 60.94%	
Hypertension (Y) [n, %]	100, 65.36%	31, 72.09%	44, 68.75%	
Obesity (Y) [n, %]	27, 17.65%	8, 18.69%	10, 15.63%	
Depression (Y) [n, %]	12, 7.84%	6, 13.95%	17, 26.56%	
Tobacco Abuse (Y) [n, %]	51, 33.33%	18, 41.86%	28, 43.75%	
Alcohol Abuse (Y) [n, %]	3, 1.96%	5, 11.63%	3, 4.69%	
	NC	**MCI**	**AD**	**P-value[a]**
Mexican Americans	175	84	40	
Age [CI]	67.62 ± 0.8156	69.88 ± 1.691	73.37 ± 2.485	<0.000[e]
Sex (F) [n, %]	99, 56.57%	40, 47.62%	24, 60.00%	
Mini Mental State Exam (MMSE) [CI]	28.14 ± 0.2889	24.93 ± 1.140	19.87 ± 1.860	<0.000[f]
Clinical Dementia Rating (CDR) Sum [CI]	0.006 ± 7.897e-3	1.113 ± 0.2261	5.737 ± 1.183	<0.000[g]
Years of Education [CI]	11.05 ± 0.6598	8.77 ± 1.664	9.75 ± 1.547	0.002[h]
BMI kg/m^2 [CI]	30.917 ± 0.9992	31.295 ± 2.228	28.717 ± 1.651	0.116
Diabetes (Y) [n, %]	79, 45.14%	32, 38.10%	19, 47.50%	

Hypercholesterolemia (Y) [n, %]	103, 58.86%	52, 61.90%	23, 57.50%	
Hyperlipidemia (Y) [n, %]	89, 50.86%	35, 41.67%	12, 30.00%	
Hypertension (Y) [n, %]	120, 68.57%	63, 75.00%	28, 70.00%	
Obesity (Y) [n, %]	84, 48.00%	38, 45.24%	8, 20.00%	
Depression (Y) [n, %]	19, 10.86%	29, 34.52%	15, 37.50%	
Tobacco Abuse (Y) [n, %]	82, 46.86%	39, 46.43%	18, 45.00%	
Alcohol Abuse (Y) [n, %]	6, 3.43%	0, 0.00%	2, 5.00%	

a. The mean difference is significant at 0.05.

b. NC vs. MCI 0.016, NC vs. AD <0.001, MCI vs. AD <0.001

c. NC vs. MCI <0.001, NC vs. AD <0.001, MCI vs. AD <0.001

d. NC vs. MCI 0.003, NC vs. AD 0.042, MCI vs. AD 0.542

e. NC vs. MCI 0.028, NC vs. AD <0.001, MCI vs. AD 0.017

f. NC vs. MCI <0.001, NC vs. AD <0.001, MCI vs. AD <0.001

g. NC vs. MCI <0.001, NC vs. AD <0.001, MCI vs. AD <0.001

h. NC vs. MCI 0.001, NC vs. AD 0.273, MCI vs. AD 0.540

Genotype frequencies for *APOE* and *OGG1* are shown in **Table 7** distributed by cognitive status in each population.

Table 7. *APOE* and *OGG1* genotype frequencies in each population based on cognitive phenotype in TARCC participants assessed via buffy coat.

Non-Hispanic Whites	**N = 260**	**NC**	**MCI**	**AD**
APOE Genotype	e2/e2	1 (0.65%)	0 (0%)	0 (0%)
	e2/e3	15 (9.8%)	6 (14%)	1 (1.6%)
	e2/e4	3 (1.96%)	0 (0%)	0 (0%)
	e3/e3	85 (55.6%)	21 (48.8%)	19 (29.7%)
	e3/e4	24 (15.7%)	9 (20.9%)	30 (46.9%)
	e4/e4	4 (2.6%)	7 (16.3%)	13 (20.3%)
OGG1 Genotype	Ser326	53 (34.6%)	21 (48.8%)	30 (46.9%)
	Ser/Cys326	47 (30.7%)	8 (18.6%)	16 (25%)
	Cys326	5 (3.3%)	1 (2.3%)	3 (4.7%)
Mexican Americans	**N = 299**	**NC**	**MCI**	**AD**
APOE Genotype	e2/e2	1 (0.57%)	0 (0%)	0 (0%)
	e2/e3	17 (9.7%)	3 (3.6%)	0 (0%)
	e2/e4	0 (0%)	0 (0%)	0 (0%)
	e3/e3	121 (69.1%)	61 (72.6%)	25 (62.5%)
	e3/e4	33 (18.9%)	18 (21.4%)	15 (37.5%)
	e4/e4	2 (1.1%)	2 (2.3%)	0 (0%)
OGG1 Genotype	Ser326	55 (31.4%)	28 (33.3%)	18 (45%)
	Ser/Cys326	68 (38.9%)	31 (36.9%)	13 (32.5%)
	Cys326	17 (9.7%)	8 (9.5%)	4 (10%)

In the MA population 8oxoG variant count was significantly reduced for subjects reporting depression compared to those without depression; mean = 6.548 and 7.704, respectively (**Figure 13**). Tobacco abuse demonstrated an approach for significance in association with 8oxoG demonstrating a higher variant load compared to non-smokers in MAs; mean = 7.935 and 7.048, respectively (**Figure 14**). These trends were not observed in the NHW cohort.

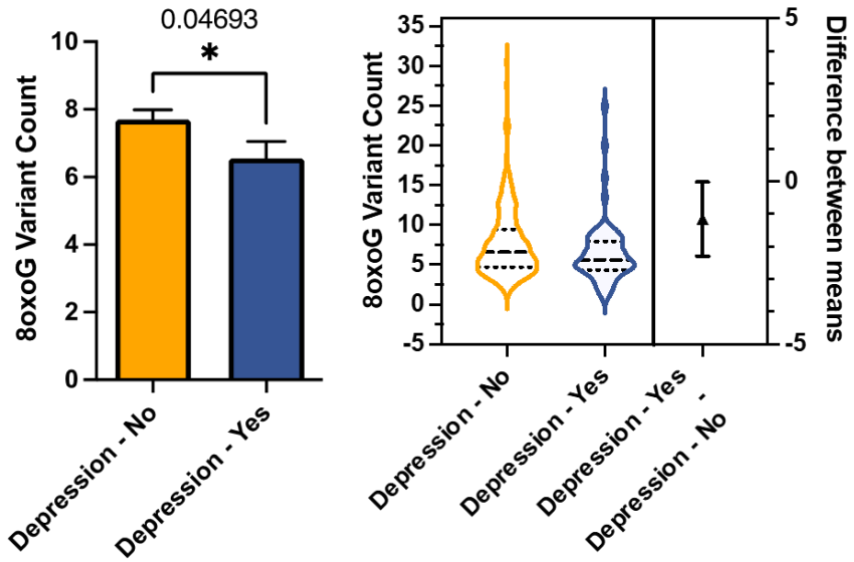


Figure 13. Cellular 8oxoG variant count is significantly higher in self-reported non-depressed Mexican Americans. a Total 8oxoG variant count was assessed by depression status using a two-tailed Welch's t-test (n = 299, t-statistic = 2.010, df = 105.8, p = 0.04693). Error bars represent standard error of the mean. **b** Violin plot displaying the distribution of 8oxoG variant counts in individuals with and without depression (n = 559) with effect size and confidence interval plotted on right y-axis. Dashed lines indicate the mean and dotted lines represent the 1st and 3rd quartile. The triangle represents the difference of the means, and the associated bar indicates the confidence interval.

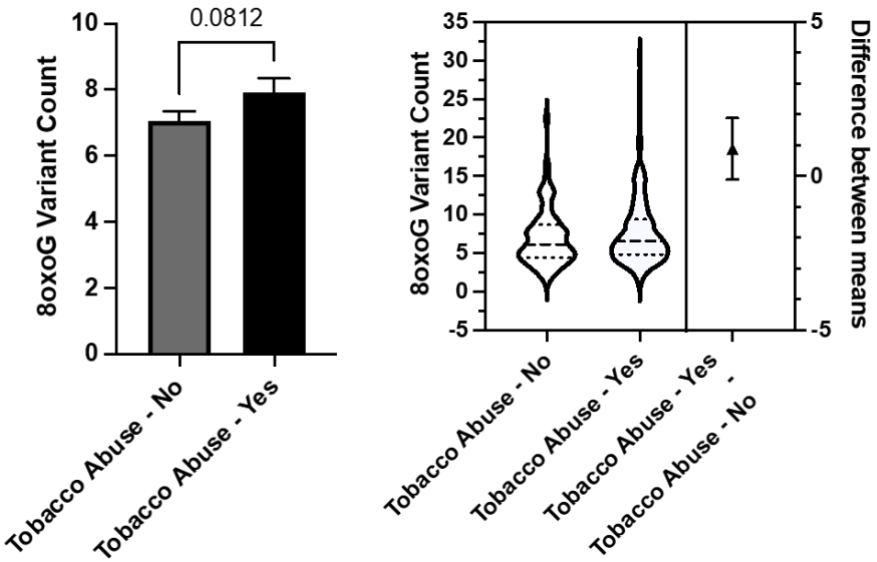


Figure 14. Total Cellular 8oxoG count approaches significance in MA individuals with a history of tobacco abuse. a Total 8oxoG variant count was evaluated by history of tobacco abuse using a two-tailed Welch's t-test (n = 299, t-statistic = 1.751, df = 257.2, p = 0.0812). Error bars represent standard error of the mean. **b** Violin plot displaying the distribution of 8oxoG variant counts in individuals with and without a history of tobacco abuse (n = 559) with effect size and confidence interval plotted on right y-axis. Dashed lines indicate the mean and dotted lines represent the 1st and 3rd quartile. The triangle represents the difference of the means, and the associated bar indicates the confidence interval.

Multiple linear regression model predictions in the whole cohort (MA and NHW combined) were performed to assess for associated factors to 8oxoG and to determine if there are predictive interactions. Sex (p = 0.000747), years of education (p = 0.005456), BMI (p = 0.028841), and tobacco abuse (p = 0.008646) were found to be significantly associated with 8oxoG variant count and the population-sex interaction demonstrated a significant interaction effect (p = 0.003762) (**Table 8**). The negative coefficient for sex indicates elevated 8oxoG for females. No other independent variables considered for the regression model showed associated statistical significance (population, cognitive status, age, diabetes, depression, APOE, OGG1, and population × education). Further analysis of 8oxoG variant count in both population and sex via two-way ANOVA indicates population is significantly associated (p = <0.0001), while sex was marginally significant (p = 0.0922).

Table 8. Cellular 8oxoG variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering population interaction effect with both sex and education.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	3.12218	2.50099	1.248	0.212613
Population (with respect to NHW)	-1.74971	1.79942	-0.972	0.331444
Cognitive Status (with respect to AD)	0.72882	0.54067	1.348	0.178412
Cognitive Status (with respect to MCI)	0.59658	0.49557	1.204	0.229364
Sex (with respect to Male)	-1.79283	0.52766	-3.398	***0.000747***
Age	0.01725	0.02955	0.584	0.559639
Years of Education	0.15024	0.05377	2.794	***0.005456***
BMI	0.0797	0.03633	2.193	***0.028841***
Diabetes (with respect to "Yes")	-0.63683	0.42547	-1.497	0.135232
Depression (with respect to "Yes")	-0.72191	0.52278	-1.381	0.168071
Tobacco Abuse (with respect to "Yes")	1.04508	0.39607	2.639	***0.008646***
APOE	0.16904	0.35467	0.477	0.63389
OGG1	-0.12608	0.29831	-0.423	0.672781
Interaction: NHW x Male "Yes"	2.26289	0.77646	2.914	***0.003762***
Interaction: NHW x Years of Education	-0.11933	0.11921	-1.001	0.31739

R-squared	0.1221	*p*-value	1.565e-06
Adjusted R-squared	0.09172	df	14 and 405
F-statistic	4.022	Sample n	420

In the subsequent multiple linear regression model, a diabetes × cognition interaction was evaluated and showed that 8oxoG variant count was significantly associated with population (p = 1.76e-06), sex (p = 0.0429), years of education (p = 0.0109), BMI (p = 0.0254), and tobacco abuse (p = 0.0155) (**Table 9**). An interaction effect between diabetes and cognition did not show significant association with 8oxoG variant count. Cognitive status with respect to AD displayed a suggestive association with AD compared to controls (p = 0.0556). No other variables included in the multiple linear regression model demonstrated significant association (cognitive status with respect to MCI, age, diabetes, depression, APOE, and OGG1.

Table 9. Cellular 8oxoG variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering diabetes interaction effect with cognitive status.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	2.16016	2.50269	0.863	0.3886
Population (with respect to NHW)	-2.35566	0.4857	-4.85	***1.76e-06***
Cognitive Status (with respect to AD)	1.24926	0.65088	1.919	*0.0556*
Cognitive Status (with respect to MCI)	0.54274	0.62565	0.867	0.3862
Sex (with respect to Male)	-0.81169	0.39958	-2.031	***0.0429***
Age	0.02659	0.0297	0.895	0.3711
Years of Education	0.12472	0.04877	2.557	***0.0109***
BMI	0.08239	0.03673	2.243	***0.0254***
Diabetes (with respect to "Yes")	-0.24581	0.56673	-0.434	0.6647
Depression (with respect to "Yes")	-0.62874	0.52599	-1.195	0.2326
Tobacco Abuse (with respect to "Yes")	0.96754	0.3979	2.432	***0.0155***
APOE	0.15846	0.35808	0.443	0.6584
OGG1	-0.20177	0.3006	-0.671	0.5025
Interaction: AD x Diabetes "Yes"	-1.38959	1.02584	-1.355	0.1763
Interaction: MCI x Diabetes "Yes"	-0.06118	0.98237	-0.062	0.9504
R-squared	0.107	*p*-value	2.31e-05	
Adjusted R-squared	0.07615	df	14 and 405	
F-statistic	3.467	Sample n	420	

Previously derived 8oxoG variant load for each subject corresponding to 8oxoG "hotspots" were analyzed via multiple linear regression prediction models to determine statistical associations. The predictive model assessing a population × sex and years of education interaction effect in association with 8oxoG "hotspots" did not demonstrate similar statistical significance (**Table Appendix R**). Similarly, the multiple linear regression model with a diabetes × cognition interaction effect did not display comparable statistical significance (**Table Appendix S**). Both models using "hotspot" as the dependent variable showed it was less informative than analyzing total 8oxoG variant count. Further analysis of 8oxoG hotspots via multiple linear regression modeling within each population lost significant statistical associations (**Tables Appendix T and U**) that were observed in **Appendix R and S**.

Based on stratification by population, multiple linear regression modelling indicated that total 8oxoG was significantly associated with cognitive status, sex, years of education, and tobacco abuse (**Table 10**). Surprisingly, within the MA population BMI did not show significant association as compared to the regression models investigating interactive effects with total

8oxoG variant count (**Tables 8-9**). Within the NHW population, the multiple linear regression model (**Table 11**) did not demonstrate any associations with the included predictive independent variables (cognition, sex, age, years of education, BMI, diabetes, depression, tobacco abuse, APOE, and OGG1).

Table 10. Multiple linear regression prediction model in the Mexican American population considering total cellular 8oxoG variant count.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	5.38624	3.66547	1.469	0.143088
Cognitive Status (with respect to AD)	1.88775	0.90381	2.089	***0.037847***
Cognitive Status (with respect to MCI)	1.11683	0.71666	1.558	0.120533
Sex (with respect to Male)	-2.23145	0.62169	-3.589	***0.000406***
Age	-0.01812	0.04432	-0.409	0.68306
Years of Education	0.15433	0.06272	2.461	***0.01461***
BMI	0.0733	0.05051	1.451	0.148079
Diabetes (with respect to "Yes")	-0.97291	0.62086	-1.567	0.118495
Depression (with respect to "Yes")	-1.4407	0.79867	-1.804	*0.072569*
Tobacco Abuse (with respect to "Yes")	1.97348	0.60289	3.273	***0.001228***
APOE	0.18362	0.65285	0.281	0.778769
OGG1	-0.03107	0.42738	-0.073	0.942111

R-squared	0.1199	*p*-value	0.001776
Adjusted R-squared	0.07743	df	11 and 228
F-statistic	2.824	Sample n	240

Table 11. Multiple linear regression prediction model in non-Hispanic Whites considering total cellular 8oxoG variant count.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	0.42632	3.39267	0.126	0.9
Cognitive Status (with respect to AD)	-0.1306	0.57357	-0.228	0.82
Cognitive Status (with respect to MCI)	0.23	0.64941	0.354	0.724
Sex (with respect to Male)	0.60026	0.45481	1.32	0.189
Age	0.04559	0.03686	1.237	0.218
Years of Education	0.0119	0.08375	0.142	0.887
BMI	0.07021	0.04965	1.414	0.159
Diabetes (with respect to "Yes")	-0.17897	0.53328	-0.336	0.738
Depression (with respect to "Yes")	0.01429	0.64799	0.022	0.982
Tobacco Abuse (with respect to "Yes")	-0.14434	0.46294	-0.312	0.756
APOE	0.17776	0.36646	0.485	0.628
OGG1	-0.28863	0.37952	-0.761	0.448
R-squared	0.04111	*p*-value	0.7794	
Adjusted R-squared	-0.02168	df	11 and 168	
F-statistic	0.6547	Sample n	180	

Additional prediction modelling used cognitive status as a binary variable to combine the effects of AD and MCI compared to NCs. Our results showed similar results to the models with greater resolution on cognitive status (**Appendix V-AC**).

Assessment of ccf-mtDNA 8oxoG Variant Count in MA and NHW TARCC Participants

The subset of participants included for the ccf-mtDNA 8oxoG variants are provided in **Table 12**. Sample size and participant selection used for the plasma analysis were selected from subjects included in the buffy coat portion of the study to compare the blood fractions collected from the same visit. Age, sex, and years of education were considered confounding variables for cognitive impairment and were utilized with the aim to pairwise match AD with normal control cases to help reduce the risk of confounders influencing false associations to AD due to the smaller sample size. Previously, our lab demonstrated a population-sex difference in mitochondrial 8oxoG variant count within the Mexican American population[44]. Genotype frequencies for *APOE* and *OGG1* are shown in **Table 13** distributed by cognitive status in each population.

Table 12. Descriptive statistics of participants classified by population and cognitive phenotype in TARCC for plasma (i.e., ccf) mitochondrial DNA oxidative mutational load

	NC	AD	P-value
Total Number of Subjects	63	59	
Non-Hispanic Whites	33	32	
Age [CI]	72.30 ± 3.112	72.19 ± 1.913	0.951
Sex (F) [n, %]	19, 57.58%	19, 58.62%	
Mini Mental State Exam (MMSE) [CI]	28.91 ± 0.4296	20.41 ± 2.214	< 0.001
Clinical Dementia Rating (CDR) Sum [CI]	0.000 ± 0.000	5.625 ± 1.353	< 0.001
Years of Education [CI]	14.12 ± 0.9178	14.13 ± 0.9549	0.996
BMI kg/m^2 [CI]	27.997 ± 2.291	27.709 ± 1.632	0.842
Diabetes (Y) [n, %]	15, 45.45%	16, 50.00%	
Hypercholesterolemia (Y) [n, %]	17, 51.52%	27, 84.38%	
Hyperlipidemia (Y) [n, %]	11, 33.33%	24, 75.00%	
Hypertension (Y) [n, %]	23, 69.70%	22, 68.75%	
Obesity (Y) [n, %]	25, 75.76%	26, 81.25%	
Depression (Y) [n, %]	6, 18.18%	12, 37.50%	
Tobacco Abuse (Y) [n, %]	13, 39.39%	16, 50.00%	
Alcohol Abuse (Y) [n, %]	1, 3.03%	2, 6.25%	
	NC	**AD**	**P-value**
Mexican Americans	30	27	
Age [CI]	73.23 ± 1.971	73.89 ± 3.184	0.733
Sex (F) [n, %]	17, 56.67%	14, 51.85%	
Mini Mental State Exam (MMSE) [CI]	27.83 ± 0.6846	19.63 ± 2.111	< 0.001
Clinical Dementia Rating (CDR) Sum [CI]	0.017 ± 0.03267	5.574 ± 1.282	< 0.001
Years of Education [CI]	9.93 ± 1.971	9.63 ± 1.794	0.824
BMI kg/m^2 [CI]	31.06 ± 2.166	28.319 ± 1.8647	0.065
Diabetes (Y) [n, %]	12, 40.00%	13, 48.15%	
Hypercholesterolemia (Y) [n, %]	17, 56.67%	17, 62.96%	
Hyperlipidemia (Y) [n, %]	13, 43.33%	8, 29.63%	
Hypertension (Y) [n, %]	21, 70.00%	18, 66.67%	
Obesity (Y) [n, %]	17, 56.67%	23, 85.19%	
Depression (Y) [n, %]	3, 10.00%	11, 40.74%	
Tobacco Abuse (Y) [n, %]	12, 40.00%	13, 48.15%	
Alcohol Abuse (Y) [n, %]	1, 3.33%	2, 7.41%	

Table 13. *APOE* and *OGG1* genotype frequencies in each population based on cognitive phenotype in TARCC participants assessed via plasma.

Non-Hispanic Whites	**N = 65**	**NC**	**AD**
APOE Genotype	e2/e2	0 (0%)	0 (0%)
	e2/e3	2 (6.1%)	0 (0%)
	e2/e4	1 (3%)	0 (0%)
	e3/e3	19 (57.6%)	8 (25%)
	e3/e4	5 (15.2%)	16 (50%)
	e4/e4	0 (0%)	7 (21.9%)
OGG1 Genotype	Ser326	10 (30.3%)	15 (46.9%)
	Ser/Cys326	8 (24.2%)	7 (21.9%)
	Cys326	2 (6.1%)	1 (3.1%)
Mexican Americans	**N = 57**	**NC**	**AD**
APOE Genotype	e2/e2	0 (0%)	0 (0%)
	e2/e3	3 (10%)	0 (0%)
	e2/e4	0 (0%)	0 (0%)
	e3/e3	22 (73.3%)	14 (51.2%)
	e3/e4	5 (16.7%)	12 (44.4%)
	e4/e4	0 (0%)	0 (0%)
OGG1 Genotype	Ser326	5 (16.7%)	9 (33.3%)
	Ser/Cys326	20 (66.7%)	9 (33.3%)
	Cys326	3 (10%)	4 (14.8%)

Although we attempted to match samples based on age, the final data set was not a complete match; in order to determine if age is a potential cofounder, we conducted a Pearson correlation to determine if age was associated to total 8oxoG variant count (**Figure 15**). The results demonstrate that age does not need to be considered a covariate in our dataset as it was not correlated with total ccf-mtDNA 8oxoG variant count.

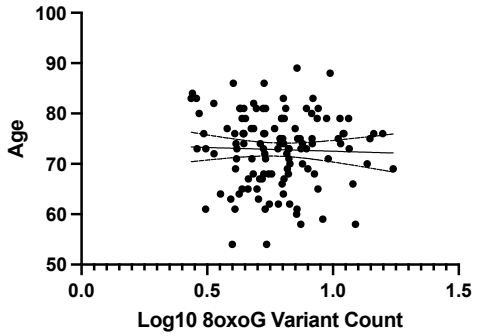


Figure 15. Scatter plot of plasma 8oxoG variant count by age. Sample means of total log transformed plasma 8oxoG variant count was assessed by age using a two-tailed Pearson correlation (n = 122). Dotted lines represent 95% confidence interval (-0.2113 to 0.1438), and the solid line indicates the best-fit line. Correlation statistics: r = -0.03489, R squared = 0.001217, p-value = 0.7028.

In the whole cohort ccf-8oxoG variant count was significantly elevated in the Mexican American population compared to non-Hispanic Whites; mean = 0.8500 and 0.7160, respectively (**Figure 16**). Interestingly, ccf-8oxoG variant count did not significantly differ based on cognitive status or sex (**Figure 17** and **Figure 18**).

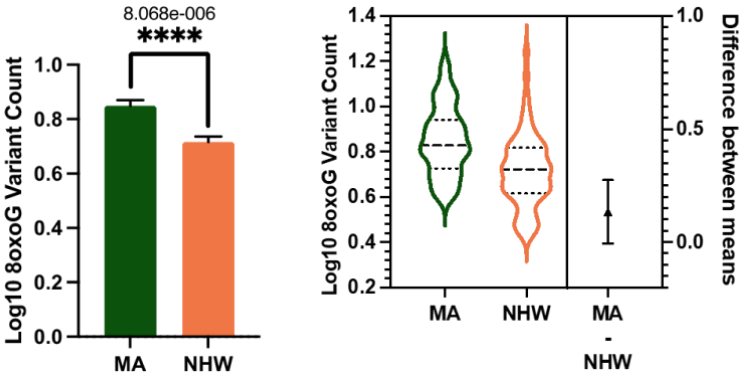


Figure 16. ccf-mtDNA 8oxoG variant count is significantly elevated in the Mexican American population. a Log transformed ccf-mtDNA 8oxoG variant count grouped by population using an unpaired, two-tailed t-test (*n* = 122, *t*-statistic = 4.666, df = 120, *p* = <0.0001). Error bars represent standard error of the mean. **b** Violin plot demonstrating distribution of 8oxoG variant count in MAs and NHWs (n= 122) with effect size and confidence interval plotted on right y-axis. Dashed lines indicate the mean and dotted lines represent the 1st and 3rd quartile. The triangle represents the difference of the means.

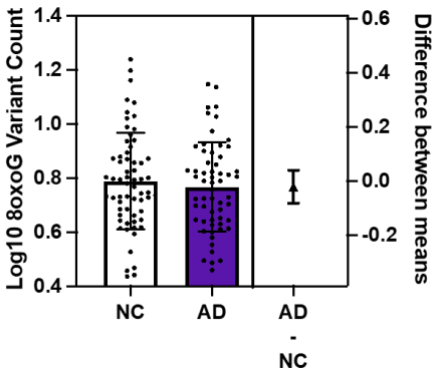


Figure 17. ccf-mtDNA 8oxoG variant count is not significantly associated with cognitive status in the whole cohort. Log transformed ccf-mtDNA 8oxoG variant count by cognitive phenotype was analyzed via an unpaired two-tailed t-test (*n* = 122, *t*-statistic = 0.6647, df = 120, *p* = 0.5075). Black filled circles represent individual data points and error bars represent standard error of the mean. Effect size and confident interval plotted on right y-axis. The triangle represents the difference of the means.

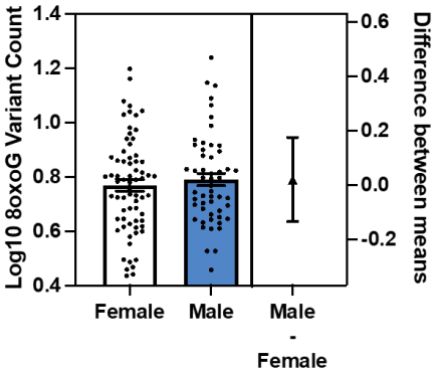


Figure 18. ccf-mtDNA 8oxoG variant count does not significantly differ by sex in the whole cohort. Log transformed ccf-mtDNA 8oxoG variant count grouped by sex via unpaired, two-tailed t-test (n = 122, t-statistic = 0.6705, df = 120, p = 0.5039). Closed circles indicate individual points and error bars are representative of the standard error of the mean. Effect size and confident interval plotted on right y-axis. The triangle represents the difference of the means.

Multiple linear regression modelling of ccf-mtDNA 8oxoG variant count in the whole cohort with respect to population- sex and education interactive effects demonstrated APOE status as a dosage effect (i.e., no e4 allele, one e4 allele, two e4 alleles) was indicated as a significant factor with a positive association (**Table 14**). Also, the population × sex interaction terms showed a significant positive association to plasma 8oxoG variant count while additional independent variables in the regression displayed significant association (population, AD, sex, age, years of education, BMI, diabetes, depression, tobacco abuse, APOE, OGG1, and population

× education. When assessing for a diabetes × cognition interaction, the multiple linear regression model revealed a significant association between ccf-8oxoG variant count and both population and APOE (**Table 15**). Other factors included in the regression model did not demonstrate significant association (AD, sex, age, years of education, diabetes, depression, tobacco abuse, and OGG1).

Table 14. Multiple linear regression predictive model for ccf-mtDNA 8oxoG variant count by cognitive status considering a population interaction with sex and education in the whole cohort.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	1.102547	0.228363	4.828	***6.73E-06***
Population with respect to NHW	-0.139851	0.137758	-1.015	0.3132
Cognitive Status with respect to AD	-0.065175	0.038484	-1.694	*0.0943*
Sex with respect to Male	-0.059352	0.046196	-1.285	0.2027
Age	-0.003315	0.002438	-1.36	0.1778
Years of Education	-0.001013	0.004628	-0.219	0.8274
BMI	0.002149	0.003673	0.585	0.5602
Diabetes with respect to "Yes"	-0.054443	0.036822	-1.479	0.1433
Depression (with respect to "Yes")	0.009989	0.039548	0.253	0.8013
Tobacco Abuse (with respect to "Yes")	-0.037633	0.034184	-1.101	0.2743
APOE	0.081716	0.032133	2.543	***0.013***
OGG1	0.006464	0.026095	0.248	0.805
Interaction: NHW x Male "Yes"	0.164836	0.065672	2.51	***0.0141***
Interaction: NHW x Years of Education	-0.006752	0.009877	-0.684	0.4962
R-squared	0.3607		*p*-value	0.0003819
Adjusted R-squared	0.2541		df	13 and 78
F-statistic	3.385		Sample n	92

Table 15. ccf-8oxoG variant count and cognitive status (NC vs AD) multiple linear regression model prediction considering diabetes interaction effect with AD.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	1.038734	0.23642	4.394	***3.43E-05***
Population with respect to NHW	-0.15482	0.041881	-3.697	***0.000401***
Cognitive Status with respect to AD	-0.045167	0.048985	-0.922	0.359305
Sex with respect to Male	0.019822	0.035072	0.565	0.573558
Age	-0.002561	0.002497	-1.026	0.308072
Years of Education	-0.003067	0.004381	-0.7	0.485864
BMI	0.002128	0.003805	0.559	0.577609
Diabetes with respect to "Yes"	-0.041097	0.054135	-0.759	0.450019
Depression (with respect to "Yes")	0.011365	0.040498	0.281	0.77973
Tobacco Abuse (with respect to "Yes")	-0.045777	0.035072	-1.305	0.195608
APOE	0.071074	0.032932	2.158	**0.033946**
OGG1	-0.001863	0.026801	-0.069	0.944769
Interaction: AD x Diabetes "Yes"	-0.03728	0.071914	-0.518	0.605626
R-squared	0.309		*p*-value	0.001998
Adjusted R-squared	0.204		df	12 and 79
F-statistic	2.943		Sample n	92

Stratified multiple linear regression models were conducted to evaluate the effect of population at a higher resolution. The MA regression model did not demonstrate any statistical significance between plasma 8oxoG variant count and included predictive variables (**Table 16**). Interestingly, multiple linear regression modelling in NHWs denoted ccf-mtDNA 8oxoG variant count was marginally significant with AD and diabetes (**Table 17**). Further, the model showed significant statistical association with sex and age, as well, but other variables included did not approach statistical significance.

Table 16. Multiple linear regression results for ccf-8oxoG variant count within Mexican Americans.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	0.9313922	0.3197653	2.913	***0.00597***
Cognitive Status with respect to AD	-0.0295081	0.056698	-0.52	0.60577
Sex with respect to Male	-0.0522656	0.0513703	-1.017	0.31538
Age	-0.0001923	0.0033756	-0.057	0.95487
Years of Education	-0.0007496	0.0052637	-0.142	0.88751
BMI	-0.0017547	0.0053372	-0.329	0.74414
Diabetes with respect to "Yes"	-0.0261234	0.056394	-0.463	0.64584
Depression (with respect to "Yes")	-0.003124	0.0668592	-0.047	0.96298
Tobacco Abuse (with respect to "Yes")	-0.0360425	0.0543473	-0.663	0.51121
APOE	0.0682865	0.0631564	1.081	0.28641
OGG1	0.0406068	0.0392238	1.035	0.30709
R-squared	0.153		*p*-value	0.7303
Adjusted R-squared	-0.06992		df	10 and 38
F-statistic	0.6863		Sample n	49

Table 17. Multiple linear regression results for ccf-8oxoG variant count within non-Hispanic Whites.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	1.115392	0.327981	3.401	***0.00182***
Cognitive Status with respect to AD	-0.107136	0.05317	-2.015	*0.05237*
Sex with respect to Male	0.107027	0.046919	2.281	***0.02934***
Age	-0.007998	0.003727	-2.146	***0.03954***
Years of Education	-0.002487	0.008295	-0.3	0.76624
BMI	0.009303	0.00555	1.676	0.10345
Diabetes with respect to "Yes"	-0.096071	0.051379	-1.87	*0.07068*
Depression (with respect to "Yes")	0.017801	0.048754	0.365	0.71742
Tobacco Abuse (with respect to "Yes")	-0.06144	0.044716	-1.374	0.17898
APOE	0.057165	0.039934	1.431	0.16199
OGG1	-0.024474	0.036719	-0.667	0.50986
R-squared	0.4417		*p*-value	0.2672
Adjusted R-squared	0.2672		df	10 and 32
F-statistic	2.532		Sample n	43

DISCUSSION

Ethnic/racial differences in the development of cognitive impairment are known to exist, yet there are a substantial number of reports investigating biological, behavioral, and lifestyle factors that lead to neurodegeneration in non-Hispanic Whites as opposed to other population

groups burdened by cognitive decline. Previously, our group demonstrated that peripheral levels of cellular mitochondrial 8oxoG transversion substitution mutations from buffy coat PBMCs were associated with population, sex, and years of education[44]. Multiple linear regression analyses approached statistical significance for association between 8oxoG variant count and AD compared to NC[44].

Here, our group expanded our previous investigation by including depression, tobacco abuse, and OGG1 genotype to determine if these known AD risk factors may be associated with buffy coat PBMC and circulating cell-free mitochondrial (ccf-mtDNA) 8oxoG variants. Additionally, we were interested in characterizing the predictability of 8oxoG variant count from buffy coat PBMCs versus ccf-mtDNA in assessing risk for cognitive decline in both populations. We hypothesized that accounting for depression, tobacco abuse, and OGG1 genotype would build better predictive models for assessing 8oxoG variant count in association with cognitive impairment, T2D, and comorbidity (i.e., cognitive impairment and T2D) within the Mexican American population compared to non-Hispanic Whites. We also hypothesize that evaluating mitochondrial dysfunction through ccf-mtDNA may be a better predictor for NHWs due to their inflammatory endophenotype compared to the metabolic endophenotype observed in MAs. Therefore, we hypothesize cellular mitochondrial 8oxoG variant load may serve as a better biomarker for MAs due to their observed endophenotype and metabolic burden. Altogether, our results confirm MAs, especially females, show greater oxidative damage to their mitochondrial genome compared to NHWs. Within the MA population tobacco abuse was closely related to increasing 8oxoG mutational load; however, interestingly, non-depressed individuals showed elevated 8oxoG load. Stratified regression analysis by population in buffy coat PBMCs demonstrated a suggestive association with AD cognitive status in the MA population for females, while this was not observed in NHWs. On the other hand, stratified regression analysis for NHWs in plasma showed a suggestive association with AD in younger aged males with diabetes.

Depression is a known risk factor for developing MCI and AD, and a previous study including MA participants from TARCC demonstrated a depressive endophenotype of MCI and AD[126]. Thus, it is not surprising that there are numerous studies reporting MAs experience more

depressive symptoms compared to other Hispanic/Latino subpopulations[130,131], as well as non-Hispanic Whites[65,96,132–135]. Furthermore, accumulating evidence indicates that individuals with depression have higher levels of 8oxoG and oxidative damage, as it is recognized that oxidative stress encompasses a critical role in depression pathophysiology through the activity of ROS[136,137]. Interestingly, our data indicated that MAs without depression experienced reduced levels of cellular 8oxoG variants. It is important to note that depression seemed to be underrepresented in our MA participants, since there is a substantial amount of evidence observing a higher prevalence of depression and depressive symptoms among MAs, yet our descriptive statistics demonstrate that more individuals without depression are included in the MCI and AD groups. Subsequent linear regression models in the whole cohort and MA population mostly show a negative association with 8oxoG variant count indicating MAs without depression are at an increased risk for elevated levels of oxidative damage compared to NHWs. It is possible that surveying for the presence or absence of depression may have poor resolution when investigating cognitive associations compared to assessing for a collection of depressive symptoms. Previous studies have reported distinct clustering of depressive symptoms is imperative when taking into account the connection between cognition and depression[138,139]. Our results for 8oxoG and depression among MAs warrants further investigation by implementing depressive symptoms and/or other indicators of depression.

Unsurprisingly, our results revealed increased levels of 8oxoG variants in buffy PMBCs of MAs with a history of tobacco abuse. Smoking tobacco and even exposure to tobacco smoke has been shown to cause elevated levels of 8oxoG compared to non-smokers[140–143] because of the various carcinogens contained within[144,145]. As previously mentioned, carcinogens are capable of forming DNA adducts and can lead to oxidative stress through the production of ROS. Additionally, cigarette smoke has been recognized to cause chronic inflammation leading to increasing oxidative stress which further results in accumulating oxidative damage, creating a vicious cycle[144,146]. The link between elevated 8oxoG levels in MA smokers could be attributed to the prevalence and frequency of smoking among this population, especially in MAs with lower educational attainment and income[147–149]. A recent study investigating the effect of smoking on cognitive function among aging Mexican Americans, concluded that smoking tobacco increased

risk for cognitive decline[150]. Following linear regression models that included MAs with 8oxoG mutational load as the outcome all demonstrated a significant link with tobacco abuse. These results seem to indicate tobacco use as a strong modifiable risk factor for increased mitochondrial oxidative damage in MAs and demonstrates the importance of addressing such behaviors in this population to prevent increased mtDNA 8oxoG damage that could lead to age-related diseases.

Generally, our higher resolution linear regression models evaluating several variables on predicting 8oxoG somatic variants in the *whole* cohort established that population $\times$ sex, tobacco abuse, BMI, years of education, and population were statistically associated. AD was suggestively associated with cellular 8oxoG somatic variants when assessing an interaction between diabetes and cognitive status. In the US aging population, it was reported that 25% of individuals are living with diabetes (whether they are aware or not) and approximately 50% of individuals are pre-diabetic[151]. Increasing evidence connects AD and T2D, showing a greater risk for cognitive decline due to T2D. Robust correlations have been previously reported between AD and high blood sugar, whereby high blood sugar was associated with the presence of Aβ plaques[151]. Moreover, brain dysfunction is frequently observed in earlier stages of T2D, and hemoglobin A1C (the established biomarker for T2D) has been related to decline in functional memory and reduced hippocampal size[151]. Links between T2D and AD implicate mitochondria dysfunction as a participating factor in the development and/or progression of neurodegeneration and may be of exceptional importance for ethnic/racial differences in disease severity and manifestation. In our data, population stratification limited to MAs evaluating the same independent variables indicated similar associations with AD. These results were not observed in the NHW population, and perhaps of interest is the fact that many of the coefficients were in the opposite direction (though not significant). This evidence further suggests that mitochondrial health could be the reason for the unexplained prevalence of cognitive impairment among Mexican Americans.

Results from evaluating ccf-mtDNA 8oxoG somatic variants demonstrated elevated levels in MAs; however, associations with both cognitive status and sex were not noteworthy. The elevated cell-free oxidative DNA damage in the MA population could be related to their mitochondrial health due to the various metabolic comorbidities affecting the population as previously mentioned. Linear regression modelling results show cell-free 8oxoG variant load was

notably associated with APOE when accounting for both the population × sex and diabetes × cognitive status interaction terms in the whole cohort. Knowing APOE status is of greater importance in the NHW population compared to MAs, it was not surprising the subsequent stratified regression in MAs did not demonstrate statistical significance compared to NHWs. Regression model stratification in NHWs showed significant associations between ccf-mtDNA 8oxoG count with sex and age. Cognitive status with respect to AD in NHWs was suggestively associated with 8oxoG (marginal significance). These accumulating results point to a differentiating role of mitochondrial dysfunction assessed by mitochondrial 8oxoG somatic variants in buffy coat PBMCs compared to ccf-mtDNA of plasma in relation to the metabolic and inflammatory endophenotypes, respectively. Biomarkers of cellular mitochondrial oxidative DNA damage may be more applicable in evaluating connections to cognitive decline in MAs, whereas cell-free could better explain the association to cognitive impairment in NHWs.

CONCLUSION

Altogether, the results of this study contribute to the current body of literature regarding oxidative damage in the context of aging and neurodegenerative disease. Uniquely, our data specifically point to novel, population-based effects in 8oxoG damage in cellular and cell-free mtDNA and investigate the relationship between 8oxoG damage and key comorbidities, modifiable risk factors, and AD. It is important to note that this study has its limitations including (1) the indirect measurement of 8oxoG lesions and oxidative damage, (2) a small sample size for the plasma dataset, (3) lacking biochemical, metabolic, and inflammatory phenotypes, (4) lack of assessment for nDNA variants, (5) only blood tissue is sampled, (6) evaluation was performed in one cohort, and (7) APOE in the NHW model could have reduced power due to a larger number of missingness compared to MAs.

To better understand the biological and mechanistic roles of mitochondrial dysfunction and oxidative DNA damage, future studies should incorporate a larger sample size, include more biological markers indicative of metabolic health and systemic inflammation, determine if utilizing MMSE, CDR sum, and/or other neuropsychological tests for cognitive function strengthens our power to further support/validate our results, and characterize the nuclear genetic background associated with the mitochondrial genome for each subject. Additionally,

future studies will aim to validate the applicability of peripheral cellular and cell-free pathophysiological phenotypes as biomarkers for assessing brain pathology, disease risk, and/or disease stage. These studies will also investigate expression of DNA repair machinery, the role of sex hormones, and validate mitochondrial oxidative DNA load using novel and established alternative methods.

ACKNOWLEDGEMENTS

We thank Nicole Hales (currently at 10X Genomics) for library preparation and sequencing of mtDNA from buffy coat PBMCs. We thank Viviana Mancilla (UNTHSC Genomic Core) for library preparation and sequencing of plasma mtDNA. Research reported in the publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R25GM125587 and by the National Institutes on Minority Health and Health Disparities of the National Institutes of Health under Award Number U54MD006882 and S21MD012472. R.J.T. was supported by the National Institutes on Minority Health and Health Disparities of the National Institutes of Health under Award Number U54MD000214 and the National Institutes on Aging under Award number P30AG059298. This study was made possible by the Texas Alzheimer's Research and Care Consortium (TARCC) funded by the state of Texas through the Texas Council on Alzheimer's Disease and Related Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CHAPTER IV: DISCUSSION AND FUTURE DIRECTIONS

INTEGRATING BLOOD-BASED SIGNATURES OF OXIDATIVE DAMAGE IN MITOCHONDRIAL DNA FOR PHENOTYPES OF COGNITIVE IMPAIRMENT

There is particular interest in identifying and validating blood-based biomarkers capable of evaluating risk for cognitive impairment in addition to aiding in diagnosis and prognosis in the clinical setting because typically disease tissue biopsy may not be available[152], especially in Alzheimer's disease. It is important to note that biomarkers are not diagnostic tests and are used to determine underlying brain changes causing cognitive impairment due to dementia by distinguishing the presence or absence of disease or risk for disease[12]. Blood tissue is a great alternative to utilizing brain tissue, as it is less invasive in nature and represents the whole body since these cells are in the periphery. It is recognized that the brain is particularly susceptible to OS attributed to the high energy demand, high oxygen consumption, abundance of easily peroxidizable polyunsaturated fatty acids, high levels of catalyst iron, etc., in addition to engaging approximately 25% of inhaled oxygen[107]. Furthermore, brain neuronal cells are remarkably defenseless against elevated levels of ROS and brain astrocytes are known for their neuroprotective functionality, as they help alleviate the oxidative environment by neutralizing ROS via glutathione[107]. Astrocytes aid in the removal cellular debris from dead/dying cells and hallmark AD pathology observed in neurodegeneration[18,107]. The oxidative environment, neuroinflammation, and mitochondrial dysfunction such as reduced energy metabolism are often observed in combination with AD pathophysiology[18,107,153] and indices of these features may serve as biomarkers for cognitive decline related to Alzheimer's disease, especially in populations disproportionately burdened by the disease or other metabolic comorbid conditions that may propagate the vicious cycle of these underlying pathophysiological conditions.

Oxidized mtDNA is a prominent pro-inflammatory initiator due to its recognition as a DAMP triggering an inflammatory response. There is accumulating evidence and urge supporting its application in serving as an inflammatory biomarker since it is frequently observed at elevated levels in inflammatory diseases. Further, evidence from a recent lupus study suggests oxidized mtDNA in the form of 8oxoG is more inflammatory compared to nDNA[154]. We have discussed in length mechanisms contributing to oxidized mtDNA and logic implicates OS, impaired

mitochondrial dysfunction, inflammation, and mitochondrial capacity related to mitochondrial inheritance, therefore our integrative blood-based assessment of oxidative mtDNA in MAs compared to NHWs could help better understand the cause of ethnic/racial differences in cognitive decline. In the literature, there are a limited number of studies investigating blood-based biomarkers specifically in cognitively impaired Mexican Americans. These studies have demonstrated that mitochondrial dysfunction is a clear prominent feature of MAs with cognitive impairment and blood-based biomarkers are capable of distinguishing the presence of disease and/or comorbidity[2,31,96,155,156]. In general, blood-based biomarkers of mitochondrial dysfunction investigated in various diseases indicate distinct differences between normal healthy controls and those with disease[2,81,86,121,122,154]. Studies characterizing mitochondrial blood-based markers in disease usually focus on one particular blood fraction or use whole blood[157]. Cellular mtDNA are encompassed in mitochondria for encoding mitochondrial encoded genes essential for OXPHOS and protein synthesis, whereas ccf-mtDNA are mitochondrial genome fragments that potentiate inflammatory responses that could lead to disease pathology[157]. As a consequence, the functional differences in cellular and cell-free mtDNA in peripheral blood from different blood fractions is important to describe in combination and separately for elucidation of the role of mtDNA in related pathology[158]. Here we characterized mtDNA 8oxoG variants in two different blood fractions: buffy coat PBMCs and plasma, to capture the degree of mitochondrial dysfunction in cognitively impaired individuals from different ethnic/racial groups with distinct risk and comorbid factors affecting health outcome.

Overall, our results indicate both cellular and ccf- 8oxoG variants mtDNA are significantly elevated in the MA population compared to the NHW population (**Figure 19**). Also, sex-differences in 8oxoG variant levels are observed from both. Notably, our results from evaluating oxidized cellular mtDNA compared to ccf-mtDNA ultimately signifies strong predictive capability in MAs, especially MA females, compared to NHWs due to the observed statistical significance of assessing various independent variables in the MA population that also demonstrated to have opposite effects in NHWs. This evidence implies mitochondrial dysfunction in cellular mtDNA may be distinctly related to disease pathology in MAs with cognitive decline. In contrast, our ccf-mtDNA results within each population separately revealed distinct differences in associations

between 8oxoG variants and variables known to influence risk for cognitive impairment compared to results from cellular mtDNA. 8oxoG variant count in ccf-mtDNA displayed poor associations in the MA population compared to NHWs. This evidence altogether supports the notion that (1) blood-based signatures of mitochondrial dysfunction differ between ethnic populations, (2) cellular and ccf-mtDNA possess different functionality in potentially developing pathophysiological condition, and (3) ethnic/racial differences exist in the manifestation of neurodegeneration through the assessment of mitochondrial oxidative DNA damage from different blood fractions.

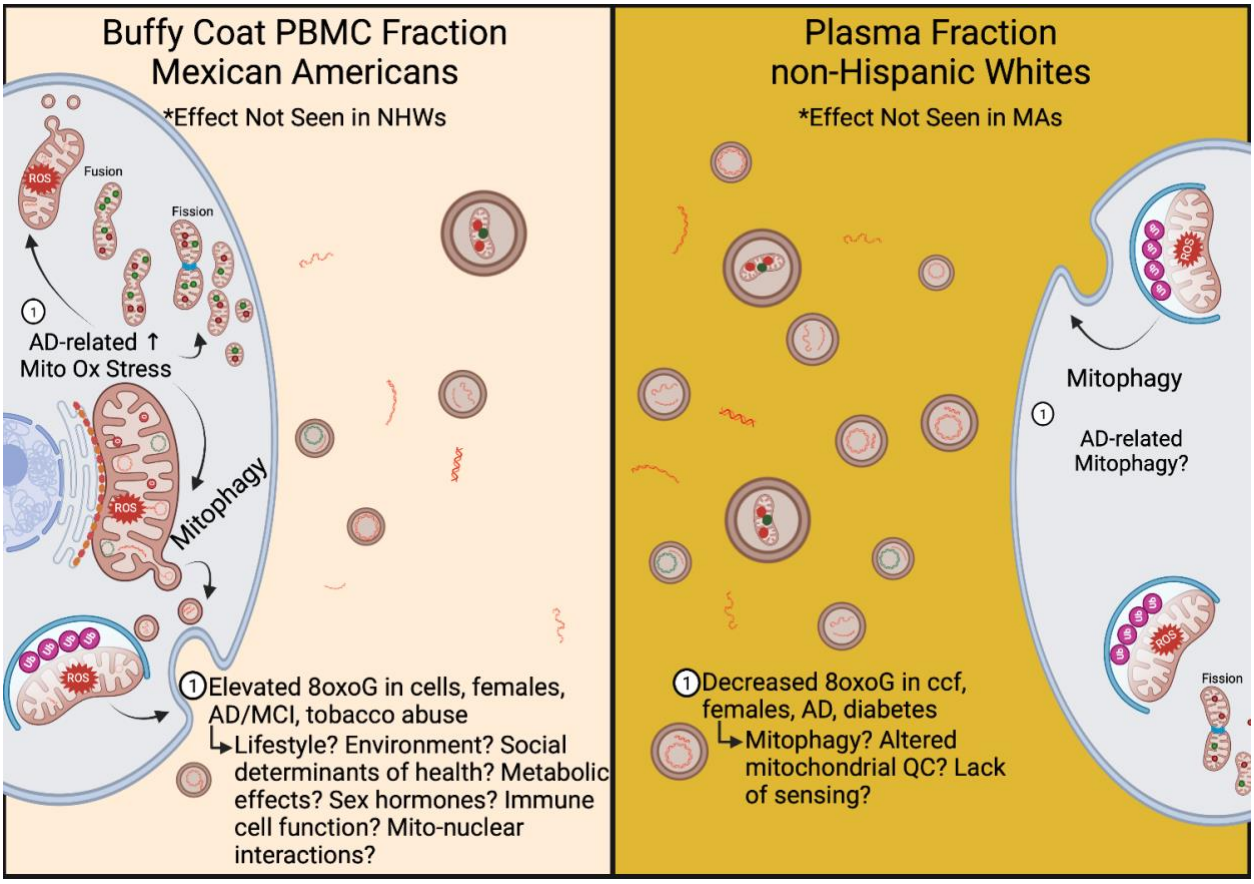


Figure 19. Overall results from cellular and ccf-mtDNA 8oxoG variants in MA vs NHW participants of TARCC. In the buffy coat PBMC blood fraction (left) observations in cellular 8oxoG variant count were found to be significant in MAs and the effects were not observed in NHWs. In buffy coat PBMCs, cellular 8oxoG variant count was significantly elevated in MAs, especially females, and was associated with CI and tobacco abuse. Modifiable risk factors, biological processes, and genetics such as lifestyle, environment, social determinants of health, immune cell function, and mito-nuclear interactions are theorized to contribute to elevated oxidative damage to mtDNA in MAs from TARCC. In the blood plasma fraction (right) observations of ccf-8oxoG variant count were significantly reduced in NHW females and were associated with AD

and diabetes. These results suggest possible alterations in mitochondrial quality control and/or lack of sensing in NHW females. Overall results from cellular and ccf-mtDNA 8oxoG variant loads suggest mitophagy may play a role in ethnic/racial differences in AD etiology.

Future directions of this approach for characterizing oxidative mtDNA damage in the context of neurodegeneration will investigate mtDNA deletions from both blood fractions to determine mechanistic differences in oxidized mtDNA from mitochondria compared to circulating cell-free. Genomic deletions are representative of DNA damage and are associated with oxidative DNA damage[159]. Mitochondrial genomic deletions are frequently observed in Alzheimer's disease— the 5kb "common deletion" that primarily causes Kearns-Sayre syndrome and accumulates in specific tissues with age is a predominant deletion observed in AD patients[159,160]. We will have information on mitochondrial genomic deletions for participants involved in our studies (**Appendix AD and AE**) to evaluate severity of damage by population, disease, sex, etc.

NOVEL NGS METHOD FOR 8OXOG DETECTION

With the technological advances in sequencing chemistry researchers are able to gain insight on mitochondrial health by obtaining information on mitochondrial variants, mutational load, nuclear-encoded variants impacting mitochondrial bioenergetics, and mtDNA CN[152]. Two single-molecule long-read sequencing platforms are currently available through Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT)[50,58,59]. Interestingly, these platforms are not actually generating long-reads, but are based on library preparation approaches using barcodes to computationally generate long-reads[59]. While both PacBio and ONT platforms are able to produce long reads and detect modifications in native strands of DNA, each utilizes a distinct detection mechanism for base determination, and thus present unique advantages and disadvantages[50,51].

PacBio's Single Molecule Real-Time sequencing (SMRT-seq) platform was introduced as one of the first single molecule sequencing technologies[50,58,59]. The SMRT-seq platform collects data during replication of target DNA[51,59]. Attachment of two hairpin adapters during library preparation produces a SMRTbell template that facilitates continuous circular sequencing[40,50,51,59]. Reads obtained from SMRT-seq can be as long as 60kb and has reduced PCR

amplification bias compared to previous sequencing generations[50,51,59]. Despite the numerous advantages offered by this third-generation sequencing platform, SMRT-seq possesses several limitations including relatively low throughput, higher cost, PCR product bias when compared to ONT, and high error rate typically observed as single base pair insertions and deletions[40,51]. The cost of sequencing runs can be exceptionally high due to the need for high coverage ($1000 per Gb) and pose a challenge for accessibility for small laboratories[59]. SMRT-seq is able to detect modified bases such as C5-methylcytosine (m5C), N6-methylaadenosine (m6A), and 5-hydroxymethylcytosine (hm5C) due to shift in kinetics of the DNA polymerase. Accurate characterization, however, requires deep coverage because of the dynamic nature of the DNA polymerase incorporating modified bases at a slower rate[58].

In contrast to the detection of fluorescently labeled nucleotides used in PacBio SMRT-seq, ONT sequencing platforms rely on a nanoscopic hole formed in an electrically resistant synthetic polymer membrane occupied by a biological pore protein to directly read strands in a given library[40,50,161]. Hundreds to thousands of bespoke nanopore proteins are contained within the sensor array of a given flowcell, allowing for many strands of DNA to be sequenced simultaneously[50,59]. The MinION flowcell contains up to 2048 pores that is monitored in groups of 512 via the MinKNOW software program[50,59,161]. Nanopore sequencing is facilitated by a molecular motor protein attached during library preparation[59]. This enzyme not only directs strands to an available nanopore for sequencing, but also unwinds the DNA and facilitates the strands passage through at a specific speed. As bases are ratcheted through (3-6 k-mer length of bases or base pairs within a given sequence) the electrical conductance of the pore will change due to voltage across the membrane[59,161]. Unique disruptions caused by the motif presence within the pore at any given time is detected and recorded by the ASIC sensor[50,59]. The raw data is represented by squiggle tracings from changes in voltage and can then be converted into the traditional four nucleotides (A, G, C, and T) by the Guppy basecaller integrated into MinKNOW[50,59]. There are several workflows that can be utilized depending on the type of analysis desired and real-time sequencing allows users to analyze data during sequencing runs[50]. Although this device has a higher error rate than competitors (i.e., PacBio), particularly in regions of low complexity such as the homopolymeric stretches characteristic of mitochondrial genome,

great efforts have been taken to improve the accuracy and data generated[50]. Performance and reproducibility of the device was assessed by the MinION Analysis and Reference Consortium and the investigators reported low variations in performance, as well as consistency with base error rate, throughput, and read length[50].

The MinION device allows for real-time, direct sequencing of long DNA/RNA fragments, has a low startup fee (~$1000 for instrument and $750 per Gb), uses minimal technology, and is small in size; the smallest sequencing device currently available[40,50,51,59]. While the capabilities and limitations of nanopore-based sequencing platforms are similar to those of SMRT-seq long read lengths, relatively high error rates, and capabilities in base modification detection there are numerous advantages that set ONT technology apart from other commercially available platforms[40,58,162]. First and foremost, nanopore-based sequencing is scalable, offering a range of devices that can meet the needs of any size laboratory. The device is approximately 3 cm x 10 cm and uses a USB port to run off computers making the device able to accommodate laboratories of any size with any throughput need[59]. Other advantages of nanopore-based sequencing include: (1) users have flexibility in their desired read length, (2) can obtain reads covering repetitive regions, (3) simple/fast workflows, (4) less costly materials, (5) is easily accessible to labs, (7) portable, (8) PCR-free and chemical labeling free library preparation, and (9) achieves high yields of data up to 30 Gb[40,51,161]. The size of the device and its ability to detect oxidative modifications is ideal for clinical/biomedical settings[51,161].

Several studies have recently demonstrated the ability to detect and accurately characterize modified bases using the MinION device[162]. Nanopore sequencing chemistry monitors the electrical conductance to allow for more than 1,000 signals to be detected depending on the k-mer[59,162]. Thus, the presence of a modified base within the pore will cause a unique disruption from that of a modified base presented in the squiggle tracings[59,161]. Currently, only m5C and m6A have been evaluated and applied using nanopore sequencing methods[58]. Successful detection of methylated bases provides compelling support for the proposed success of detecting both forms of oxidatively modified guanine. Machine learning and statistical testing tools/packages using nanopore sequencing exist to detect modified bases, however, there are no validated algorithms to detect oxidative modifications. Currently, there is an assortment of

tools available to use for detection of additional modifications[58]. In particular, the open-source research basecaller Bonito (https://github.com/nanoporetech/bonito) is presently the most accurate basecaller for nanopore sequencing data compared to their current recommended basecaller Guppy[163]. Bonito can be utilized to develop algorithms for modified base detection. Therefore, Oxford nanopore sequencing technology may perform as an improved alternative approach to detect and quantify oxidative damage by training basecalling models in Bonito.

SCIENTIFIC IMPACT

Mitochondrial genetics has been shown great importance in various diseases and has continually gained recognition for its involvement in disease pathophysiology. Results reported here are insightful and may shape our understanding in the role of mitochondria in neurodegeneration, such as AD. It has been recognized that ethnic/racial differences exist for AD risk, development, and manifestation. The growing MA aging population poses a great threat to our healthcare system, yet research efforts are low in number investigating the AD continuum in MAs. Our studies confirm previous reports of enhanced mitochondrial dysfunction and oxidative DNA damage in AD. Further we demonstrate population-based differences in mitochondrial DNA oxidative damage of cognitively impaired individuals. Our results show incorporating biological, molecular, and behavioral phenotypes allow for a more comprehensive understanding of their associations to population and sex differences observed in AD. Also, this work described here evaluated and partially revealed the potential use of mitochondrial indices of mitochondrial dysfunction in peripheral blood. Our results provide new information regarding blood-based signatures of mitochondrial dysfunction in the MA population compared to NHWs in relation to factors influencing risk for cognitive decline which could prove valuable in improving the prevention, diagnosis, and/or identification of therapeutic targets for cognitive decline.

REFERENCES

1. Finkel, T. Signal transduction by reactive oxygen species. *Journal of Cell Biology* **194**, 7–15 (2011).
2. Silzer, T. *et al.* Circulating mitochondrial DNA: New indices of type 2 diabetes-related cognitive impairment in Mexican Americans. *PLOS ONE* **14**, e0213527 (2019).
3. Harman, D. Aging: a theory based on free radical and radiation chemistry. *J Geritol* **11**, 298–300 (1956).
4. Go, Y.-M. & Jones, D. P. Redox theory of aging: implications for health and disease. *Clinical Science* **131**, 1669–1688 (2017).
5. Sun, N., Youle, R. J. & Finkel, T. The Mitochondrial Basis of Aging. *Molecular Cell* **61**, 654–666 (2016).
6. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal* **17**, 1195–1214 (2003).
7. Thannickal, V. J. & Fanburg, B. L. Reactive oxygen species in cell signaling. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **279**, L1005–L1028 (2000).
8. Ravanat, J.-L., Turesky, R. J., Gremaud, E., Trudel, L. J. & Stadler, R. H. Determination of 8-Oxoguanine in DNA by Gas Chromatography-Mass Spectrometry and HPLC-Electrochemical Detection: Overestimation of the Background Level of the Oxidized Base by the Gas Chromatography-Mass Spectrometry Assay. *Chemical Research in Toxicology* **8**, 1039–1045 (1995).
9. Wu, L. L., Chiou, C.-C., Chang, P.-Y. & Wu, J. T. Urinary 8-OHdG: a marker of oxidative stress to DNA and a risk factor for cancer, atherosclerosis and diabetics. *Clinica Chimica Acta* **339**, 1–9 (2004).
10. Cadet, J., Douki, T. & Ravanat, J.-L. Measurement of oxidatively generated base damage in cellular DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **711**, 3–12 (2011).
11. Swerdlow, R. H. Brain aging, Alzheimer's disease, and mitochondria. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1812**, 1630–1639 (2011).
12. 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **18**, 700–789 (2022).
13. Wu, S., Vega, W. A., Resendez, J. & Jin, H. *Latinos and Alzheimer's Disease: New Numbers Behind the Crisis*. https://www.usagainstalzheimers.org/sites/default/files/Latinos-and-AD_USC_UsA2-Impact-Report.pdf (2016).
14. 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **15**, 321–387 (2019).
15. Wallace, D. C. A Mitochondrial Paradigm of Metabolic and Degenerative Diseases, Aging, and Cancer: A Dawn for Evolutionary Medicine. *Annual Review of Genetics* **39**, 359–407 (2005).
16. Moreira, P. I., Santos, M. S., Seiça, R. & Oliveira, C. R. Brain mitochondrial dysfunction as a link between Alzheimer's disease and diabetes. *Journal of the Neurological Sciences* **257**, 206–214 (2007).
17. Liguori, I. *et al.* Oxidative stress, aging, and diseases. *Clinical Interventions in Aging* **Volume 13**, 757–772 (2018).
18. Chen, Z. & Zhong, C. Oxidative stress in Alzheimer's disease. *Neuroscience Bulletin* **30**, 271–281 (2014).

19. Giri, M., Lü, Y. & Zhang, M. Genes associated with Alzheimer's disease: an overview and current status. *Clinical Interventions in Aging* 665 (2016) doi:10.2147/CIA.S105769.
20. Gatz, M. *et al.* Role of Genes and Environments for Explaining Alzheimer Disease. *Archives of General Psychiatry* **63**, 168 (2006).
21. Wang, X. *et al.* Oxidative stress and mitochondrial dysfunction in Alzheimer's disease. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1842**, 1240–1247 (2014).
22. How Is Alzheimer's Disease Diagnosed? *NIH National Institute on Aging (NIA)* https://www.nia.nih.gov/health/how-alzheimers-disease-diagnosed (2021).
23. McKhann, G. M. *et al.* The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7**, 263–269 (2011).
24. Jack, C. R. *et al.* Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7**, 257–262 (2011).
25. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **7**, 270–279 (2011).
26. Babulal, G. M. *et al.* Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: Update and areas of immediate need. *Alzheimer's & Dementia* **15**, 292–312 (2019).
27. 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **17**, 327–406 (2021).
28. Johnson, L. A., Large, S. E., Izurieta Munoz, H., Hall, J. R. & O'Bryant, S. E. Vascular Depression and Cognition in Mexican Americans. *Dementia and Geriatric Cognitive Disorders* **47**, 68–78 (2019).
29. Bertoni, B., Budowle, B., Sans, M., Barton, S. A. & Chakraborty, R. Admixture in Hispanics: Distribution of Ancestral Population Contributions in the United States. *Human Biology* **75**, 1–11 (2003).
30. Rose, K. M. Mild Cognitive Impairment in Hispanic Americans: An Overview of the State of the Science. *Archives of Psychiatric Nursing* **19**, 205–209 (2005).
31. O'Bryant, S. E. *et al.* Biomarkers of Alzheimer's Disease Among Mexican Americans. *Journal of Alzheimer's Disease* **34**, 841–849 (2013).
32. Zilliox, L. A., Chadrasekaran, K., Kwan, J. Y. & Russell, J. W. Diabetes and Cognitive Impairment. *Current Diabetes Reports* **16**, 87 (2016).
33. Martorell, R. Diabetes and Mexicans: why the two are linked. *Prev Chronic Dis* **2**, A04 (2005).
34. Wallace, D. C. A mitochondrial bioenergetic etiology of disease. *Journal of Clinical Investigation* **123**, 1405–1412 (2013).
35. Stewart, J. B. & Chinnery, P. F. Extreme heterogeneity of human mitochondrial DNA from organelles to populations. *Nature Reviews Genetics* **22**, 106–118 (2021).
36. van den Ameele, J., Li, A. Y. Z., Ma, H. & Chinnery, P. F. Mitochondrial heteroplasmy beyond the oocyte bottleneck. *Seminars in Cell & Developmental Biology* **97**, 156–166 (2020).

37. Chinnery, P. F. & Samuels, D. C. Relaxed Replication of mtDNA: A Model with Implications for the Expression of Disease. *The American Journal of Human Genetics* **64**, 1158–1165 (1999).
38. Miller, M. B. *et al.* Somatic genomic changes in single Alzheimer's disease neurons. *Nature* **604**, 714–722 (2022).
39. Kwiatkowski, D. *et al.* Associations between DNA Damage, DNA Base Excision Repair Gene Variability and Alzheimer's Disease Risk. *Dementia and Geriatric Cognitive Disorders* **41**, 152–171 (2016).
40. Wang, B., Kumar, V., Olson, A. & Ware, D. Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. *Frontiers in Genetics* **10**, (2019).
41. Loft, S. *et al.* Biomarkers of oxidative damage to DNA and repair. *Biochemical Society Transactions* **36**, 1071–1076 (2008).
42. Gordan, R. *et al.* Involvement of cytosolic and mitochondrial iron in iron overload cardiomyopathy: an update. *Heart Failure Reviews* **23**, 801–816 (2018).
43. Kino, K., Hirao-Suzuki, M., Morikawa, M., Sakaga, A. & Miyazawa, H. Generation, repair and replication of guanine oxidation products. *Genes and Environment* **39**, 21 (2017).
44. Reid, D. M. *et al.* Mitochondrial DNA oxidative mutations are elevated in Mexican American women potentially implicating Alzheimer's disease. *npj Aging* **8**, (2022).
45. Gottesman, I. I. & Gould, T. D. The Endophenotype Concept in Psychiatry: Etymology and Strategic Intentions. *American Journal of Psychiatry* **160**, 636–645 (2003).
46. Cannon, T. D. & Keller, M. C. Endophenotypes in the Genetic Analyses of Mental Disorders. *Annual Review of Clinical Psychology* **2**, 267–290 (2006).
47. Meyer-Lindenberg, A. & Weinberger, D. R. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nature Reviews Neuroscience* **7**, 818–827 (2006).
48. Amente, S. *et al.* Genome-wide mapping of genomic DNA damage: methods and implications. *Cellular and Molecular Life Sciences* **78**, 6745–6762 (2021).
49. Poetsch, A. R. The genomics of oxidative DNA damage, repair, and resulting mutagenesis. *Computational and Structural Biotechnology Journal* **18**, 207–219 (2020).
50. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics* **14**, 265–279 (2016).
51. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics* **13**, 278–289 (2015).
52. Ding, Y., Fleming, A. M. & Burrows, C. J. Sequencing the Mouse Genome for the Oxidatively Modified Base 8-Oxo-7,8-dihydroguanine by OG-Seq. *J Am Chem Soc* **139**, 2569–2572 (2017).
53. Poetsch, A. R., Boulton, S. J. & Luscombe, N. M. Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome Biology* **19**, 215 (2018).
54. Amente, S. *et al.* Genome-wide mapping of 8-oxo-7,8-dihydro-2′-deoxyguanosine reveals accumulation of oxidatively-generated damage at DNA replication origins within transcribed long genes of mammalian cells. *Nucleic Acids Research* **47**, 221–236 (2019).
55. Wu, J., McKeague, M. & Sturla, S. J. Nucleotide-Resolution Genome-Wide Mapping of Oxidative DNA Damage by Click-Code-Seq. *J Am Chem Soc* **140**, 9783–9787 (2018).

56. Liu, Z. J., Martínez Cuesta, S., van Delft, P. & Balasubramanian, S. Sequencing abasic sites in DNA at single-nucleotide resolution. *Nature Chemistry* **11**, 629–637 (2019).
57. Fang, Y. & Zou, P. Genome-Wide Mapping of Oxidative DNA Damage via Engineering of 8-Oxoguanine DNA Glycosylase. *Biochemistry* **59**, 85–89 (2020).
58. Hofer, A., Liu, Z. J. & Balasubramanian, S. Detection, Structure and Function of Modified DNA Bases. *J Am Chem Soc* **141**, 6420–6429 (2019).
59. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* **17**, 333–351 (2016).
60. Kauppila, J. H. K. *et al.* Base-excision repair deficiency alone or combined with increased oxidative stress does not increase mtDNA point mutations in mice. *Nucleic Acids Research* **46**, 6642–6669 (2018).
61. Kianpour Rad, S. *et al.* Mechanism involved in insulin resistance via accumulation of beta-amyloid and neurofibrillary tangles: link between type 2 diabetes and Alzheimer's disease. *Drug Design, Development and Therapy* **Volume 12**, 3999–4021 (2018).
62. Mullins, R. J., Diehl, T. C., Chia, C. W. & Kapogiannis, D. Insulin Resistance as a Link between Amyloid-Beta and Tau Pathologies in Alzheimer's Disease. *Frontiers in Aging Neuroscience* **9**, (2017).
63. Gonçalves, R. A., Wijesekara, N., Fraser, P. E. & de Felice, F. G. The Link Between Tau and Insulin Signaling: Implications for Alzheimer's Disease and Other Tauopathies. *Frontiers in Cellular Neuroscience* **13**, (2019).
64. Aguilar-Navarro, S. G. *et al.* Association between ApoE ε4 Carrier Status and Cardiovascular Risk Factors on Mild Cognitive Impairment among Mexican Older Adults. *Brain Sciences* **11**, 68 (2021).
65. Johnson, L. A. *et al.* Depression, inflammation, and memory loss among Mexican Americans: analysis of the HABLE cohort. *International Psychogeriatrics* **29**, 1693–1699 (2017).
66. Cooke, M. S., Olinski, R. & Evans, M. D. Does measurement of oxidative damage to DNA have clinical significance? *Clinica Chimica Acta* **365**, 30–49 (2006).
67. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).
68. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
69. van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud*. (O'Reilly Media, Inc., 2020).
70. Benjamin, D. *et al.* Calling Somatic SNVs and Indels with Mutect2. *bioRxiv 861054* (2019) doi:https://doi.org/10.1101/861054.
71. Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Research* **41**, e67–e67 (2013).
72. Garrison, E., Kronenberg, Z. N., Dawson, E. T., Pederson, B. S. & Prins, P. *Vcflib and tools for processing the VCF variant call format*. (2021) doi:https://doi.org/10.1101/2021.05.21.445151.
73. Weissensteiner, H. *et al.* HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research* **44**, W58–W63 (2016).
74. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation* **30**, E386–E394 (2009).

75. Komakula, S. S. B. *et al.* The DNA Repair Protein OGG1 Protects Against Obesity by Altering Mitochondrial Energetics in White Adipose Tissue. *Scientific Reports* **8**, 14886 (2018).
76. Sampath, H. *et al.* 8-Oxoguanine DNA Glycosylase (OGG1) Deficiency Increases Susceptibility to Obesity and Metabolic Dysfunction. *PLoS ONE* **7**, e51697 (2012).
77. Thameem, F. *et al.* The Ser(326)Cys Polymorphism of 8-Oxoguanine Glycosylase 1 (OGG1) Is Associated with Type 2 Diabetes in Mexican Americans. *Human Heredity* **70**, 97–101 (2010).
78. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**, 267–273 (2003).
79. Patti, M. E. *et al.* Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proceedings of the National Academy of Sciences* **100**, 8466–8471 (2003).
80. Soares, J. P. *et al.* Aging and DNA damage in humans: a meta-analysis study. *Aging* **6**, 432–439 (2014).
81. Møller, P. Effect of age and sex on the level of DNA strand breaks and oxidatively damaged DNA in human blood cells. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **838**, 16–21 (2019).
82. Silaidos, C. *et al.* Sex-associated differences in mitochondrial function in human peripheral blood mononuclear cells (PBMCs) and brain. *Biology of Sex Differences* **9**, 34 (2018).
83. Ventura-Clapier, R. *et al.* Mitochondria: a central target for sex differences in pathologies. *Clinical Science* **131**, 803–822 (2017).
84. Gaignard, P. *et al.* Effect of Sex Differences on Brain Mitochondrial Function and Its Suppression by Ovariectomy and in Aged Mice. *Endocrinology* **156**, 2893–2904 (2015).
85. Duong, P. *et al.* Neuroprotective and neurotoxic outcomes of androgens and estrogens in an oxidative stress environment. *Biology of Sex Differences* **11**, 12 (2020).
86. Migliore, L. *et al.* Oxidative DNA damage in peripheral leukocytes of mild cognitive impairment and AD patients. *Neurobiology of Aging* **26**, 567–573 (2005).
87. Wang, J., Markesbery, W. R. & Lovell, M. A. Increased oxidative damage in nuclear and mitochondrial DNA in mild cognitive impairment. *Journal of Neurochemistry* **96**, 825–832 (2006).
88. Torres, L. L. *et al.* Peripheral Oxidative Stress Biomarkers in Mild Cognitive Impairment and Alzheimer's Disease. *Journal of Alzheimer's Disease* **26**, 59–68 (2011).
89. Sliwinska, A. *et al.* The levels of 7,8-dihydrodeoxyguanosine (8-oxoG) and 8-oxoguanine DNA glycosylase 1 (OGG1) – A potential diagnostic biomarkers of Alzheimer's disease. *Journal of the Neurological Sciences* **368**, 155–159 (2016).
90. Andrew, M. K. & Tierney, M. C. The puzzle of sex, gender and Alzheimer's disease: Why are women more often affected than men? *Women's Health* **14**, 174550651881799 (2018).
91. Gaignard, P. *et al.* Role of Sex Hormones on Brain Mitochondrial Function, with Special Reference to Aging and Neurodegenerative Diseases. *Frontiers in Aging Neuroscience* **9**, (2017).
92. Garcia, M. A. *et al.* Age of Migration Differentials in Life Expectancy With Cognitive Impairment: 20-Year Findings From the Hispanic-EPESE. *Gerontologist* **58**, 894–903 (2018).

93. Fletcher, J., Topping, M., Zheng, F. & Lu, Q. The effects of education on cognition in older age: Evidence from genotyped Siblings. *Social Science and Medicine* **280**, (2021).
94. Rote, S. M. & Angel, J. L. Gender-Based Pathways to Cognitive Aging in the Mexican-Origin Population in the United States: The Significance of Work and Family. *The Journals of Gerontology: Series B* **76**, e165–e175 (2021).
95. Garcia, M., Saenz, J., Downer, B. & Wong, R. The role of education in the association between race/ethnicity/nativity, cognitive impairment, and dementia among older adults in the United States. *Demographic Research* **38**, 155–168 (2018).
96. O'Bryant, S. E. *et al.* Risk factors for mild cognitive impairment among Mexican Americans. *Alzheimer's & Dementia* **9**, 622 (2013).
97. Carrieri, G. *et al.* Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. *Human Genetics* **108**, 194–198 (2001).
98. Nayyar, A. & Chakalova, L. APOE4, oxidative stress and decreased repair capacity - a no-brainer. Faulty lipid metabolism and increased levels of oxidative damage may be risk factors in the pathogenesis of late-onset dementia. *Biodiscovery* 1 (2015) doi:10.7750/BioDiscovery.2015.17.1.
99. Wang, Y. & Brinton, R. D. Triad of Risk for Late Onset Alzheimer's: Mitochondrial Haplotype, APOE Genotype and Chromosomal Sex. *Frontiers in Aging Neuroscience* **8**, (2016).
100. Ienco, E. C. *et al.* May "Mitochondrial Eve" and Mitochondrial Haplogroups Play a Role in Neurodegeneration and Alzheimer's Disease? *International Journal of Alzheimer's Disease* **2011**, 1–11 (2011).
101. Wolff, J. N., Ladoukakis, E. D., Enríquez, J. A. & Dowling, D. K. Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130443 (2014).
102. Li, W. & Sancar, A. Methodologies for detecting environmentally induced DNA damage and repair. *Environmental and Molecular Mutagenesis* **61**, 664–679 (2020).
103. Verri, M. *et al.* Mitochondrial Alterations, Oxidative Stress and Neuroinflammation in Alzheimer's Disease. *International Journal of Immunopathology and Pharmacology* **25**, 345–353 (2012).
104. Jun, G. Meta-analysis Confirms CR1, CLU, and PICALM as Alzheimer Disease Risk Loci and Reveals Interactions With APOE Genotypes. *Archives of Neurology* **67**, 1473 (2010).
105. O'Bryant, S. E. *et al.* Decreased C-reactive protein levels in Alzheimer disease. *J Geriatr Psychiatry Neurol* **23**, 49–53 (2010).
106. O'Bryant, S. E. *et al.* A serum protein-based algorithm for the detection of Alzheimer disease. *Arch Neurol* **67**, 1077–81 (2010).
107. Gamba, P. *et al.* Oxidized cholesterol as the driving force behind the development of Alzheimer's disease. *Frontiers in Aging Neuroscience* **7**, (2015).
108. Mayeda, E. R., Haan, M. N., Kanaya, A. M., Yaffe, K. & Neuhaus, J. Type 2 Diabetes and 10-Year Risk of Dementia and Cognitive Impairment Among Older Mexican Americans. *Diabetes Care* **36**, 2600–2606 (2013).
109. Tampi, R. R., Forester, B. P. & Agronin, M. Aducanumab: evidence from clinical trial data and controversies. *Drugs Context* **10**, (2021).
110. Tang, M.-X. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA* **279**, 751 (1998).

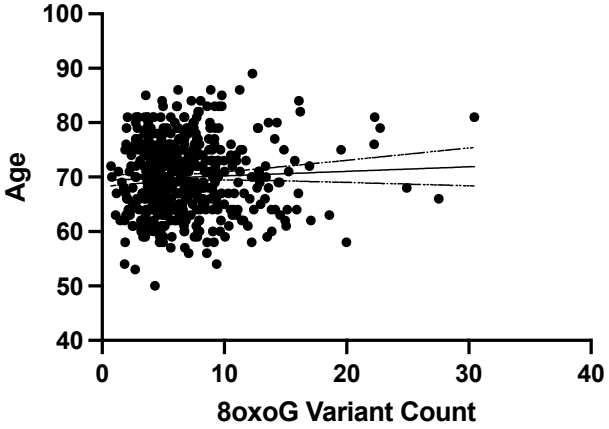
111. Campos, M., Edland, S. D. & Peavy, G. M. Exploratory Study of Apolipoprotein E ε4 Genotype and Risk of Alzheimer's Disease in Mexican Hispanics. *J Am Geriatr Soc* **61**, 1038–1040 (2013).
112. Gambardella, S. *et al.* ccf-mtDNA as a Potential Link Between the Brain and Immune System in Neuro-Immunological Disorders. *Frontiers in Immunology* **10**, (2019).
113. Lovell, M. A. & Markesbery, W. R. Oxidative DNA damage in mild cognitive impairment and late-stage Alzheimer's disease. *Nucleic Acids Research* **35**, 7497–7504 (2007).
114. Mao, P. & Reddy, P. H. Aging and amyloid beta-induced oxidative DNA damage and mitochondrial dysfunction in Alzheimer's disease: Implications for early intervention and therapeutics. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1812**, 1359–1370 (2011).
115. Peña-Bautista, C. *et al.* Oxidative Damage of DNA as Early Marker of Alzheimer's Disease. *International Journal of Molecular Sciences* **20**, 6136 (2019).
116. Clay Montier, L. L., Deng, J. J. & Bai, Y. Number matters: control of mammalian mitochondrial DNA copy number. *Journal of Genetics and Genomics* **36**, 125–131 (2009).
117. Jeong, H. *et al.* Peripheral biomarkers of mitochondrial dysfunction in adolescents with bipolar disorder. *Journal of Psychiatric Research* **123**, 187–193 (2020).
118. Boyapati, R. K., Tamborska, A., Dorward, D. A. & Ho, G.-T. Advances in the understanding of mitochondrial DNA as a pathogenic factor in inflammatory diseases. *F1000Res* **6**, 169 (2017).
119. Trifunov, S. *et al.* Circulating Cell-Free Mitochondrial DNA in Cerebrospinal Fluid as a Biomarker for Mitochondrial Diseases. *Clinical Chemistry* **67**, 1113–1121 (2021).
120. Duvvuri, B. & Lood, C. Cell-Free DNA as a Biomarker in Autoimmune Rheumatic Diseases. *Frontiers in Immunology* **10**, (2019).
121. Polina, I. A., Ilatovskaya, D. v. & DeLeon-Pennell, K. Y. Cell free DNA as a diagnostic and prognostic marker for cardiovascular diseases. *Clinica Chimica Acta* **503**, 145–150 (2020).
122. Szilágyi, M. *et al.* Circulating Cell-Free Nucleic Acids: Main Characteristics and Clinical Application. *International Journal of Molecular Sciences* **21**, 6827 (2020).
123. Kaur, U. *et al.* Reactive oxygen species, redox signaling and neuroinflammation in Alzheimer's disease: the NF-κB connection. *Current Topics in Medicinal Chemistry* **15**, 446–457 (2015).
124. Waring, S. *et al.* The Texas Alzheimer's Research Consortium longitudinal research cohort: study design and baseline characteristics. *Texas Public Health Journal* **60**, 9–13 (2008).
125. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–939 (1984).
126. Johnson, L. A., Hall, J. R. & O'Bryant, S. E. A Depressive Endophenotype of Mild Cognitive Impairment and Alzheimer's Disease. *PLoS ONE* **8**, e68848 (2013).
127. Qiagen. REPLI-g Mitochondrial DNA Kit. https://www.qiagen.com/us/products/discovery-and-translational-research/pcr-qpcr-dpcr/preamplification/repli-g/repli-g-mitochondrial-dna-kit/.
128. Picard Toolkit. *Broad Institute* https://broadinstitute.github.io/picard/ (2019).
129. Anderson, C. A. *et al.* Data quality control in genetic case-control association studies. *Nature Protocols* **5**, 1564–1573 (2010).

130. Camacho, Á. *et al.* Anxious-depression among Hispanic/Latinos from different backgrounds: results from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Social Psychiatry and Psychiatric Epidemiology* **50**, 1669–1677 (2015).
131. Hooker, K. *et al.* Depression Among Older Adults in the United States by Disaggregated Race and Ethnicity. *Gerontologist* **59**, 886–891 (2019).
132. Gonzalez, H. M., Haan, M. N. & Hinton, L. Acculturation and the Prevalence of Depression in Older Mexican Americans: Baseline Results of the Sacramento Area Latino Study on Aging. *J Am Geriatr Soc* **49**, 948–953 (2001).
133. de Oliveira, G., Cianelli, R., Gattamorta, K., Kowalski, N. & Peragallo, N. Social Determinants of Depression Among Hispanic Women. *J Am Psychiatr Nurses Assoc* **23**, 28–36 (2017).
134. Diniz, B. S., Fisher-Hoch, S. & McCormick, J. The association between insulin resistance, metabolic variables, and depressive symptoms in Mexican-American elderly: A population-based study. *International Journal of Geriatric Psychiatry* **33**, e294–e299 (2018).
135. O'Bryant, S. E., Humphreys, J. D., Schiffer, R. B. & Sutker, P. B. Presentation of Mexican Americans to a Memory Disorder Clinic. *Journal of Psychopathology and Behavioral Assessment* **29**, 137–140 (2007).
136. Ceylan, D. *et al.* Alterations in levels of 8-Oxo-2'-deoxyguanosine and 8-Oxoguanine DNA glycosylase 1 during a current episode and after remission in unipolar and bipolar depression. *Psychoneuroendocrinology* **114**, 104600 (2020).
137. Vaváková, M., Ďuračková, Z. & Trebatická, J. Markers of Oxidative Stress and Neuroprogression in Depression Disorder. *Oxidative Medicine and Cellular Longevity* **2015**, 1–12 (2015).
138. Hall, J. *et al.* Higher Groundwater Selenium Exposure Is Associated with Better Memory: A Project FRONTIER Study. *Neuroscience and Medicine* **03**, 18–25 (2012).
139. Ivnik, R. J. *et al.* Mayo's older americans normative studies: WAIS-R norms for ages 56 to 97. *Clinical Neuropsychologist* **6**, 1–30 (1992).
140. Asami, S. Cigarette smoking induces an increase in oxidative DNA damage, 8-hydroxydeoxyguanosine, in a central site of the human lung. *Carcinogenesis* **18**, 1763–1766 (1997).
141. Romano, G. *et al.* Evaluation of 8-hydroxydeoxyguanosine in human oral cells: the importance of tobacco smoke and urban environment. *Anticancer Res* **20**, 3801–5 (2000).
142. Mesaros, C., Arora, J. S., Wholer, A., Vachani, A. & Blair, I. A. 8-Oxo-2′-deoxyguanosine as a biomarker of tobacco-smoking-induced oxidative stress. *Free Radical Biology and Medicine* **53**, 610–617 (2012).
143. Lodovici, M. *et al.* Active and Passive Smoking and Lifestyle Determinants of 8-Oxo-7,8-Dihydro-2′-Deoxyguanosine Levels in Human Leukocyte DNA. *Cancer Epidemiology, Biomarkers & Prevention* **14**, 2975–2977 (2005).
144. Caliri, A. W., Tommasi, S. & Besaratinia, A. Relationships among smoking, oxidative stress, inflammation, macromolecular damage, and cancer. *Mutat Res Rev Mutat Res* **787**, 108365 (2021).
145. Wiencke, J. K. DNA adduct burden and tobacco carcinogenesis. *Oncogene* **21**, 7376–7391 (2002).
146. Saetta, M. Airway Inflammation in Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine* **160**, S17–S20 (1999).

147. Markides, K. S., Coreil, J. & Ray, L. A. Smoking among Mexican Americans: a three-generation study. *Am J Public Health* **77**, 708–11 (1987).
148. Kaplan, R. C. *et al.* Smoking among U.S. Hispanic/Latino adults: the Hispanic community health study/study of Latinos. *Am J Prev Med* **46**, 496–506 (2014).
149. Rodriquez, E. J., Fernández, A., Livaudais-Toman, J. C. & Pérez-Stable, E. J. How Does Acculturation Influence Smoking Behavior Among Latinos? The Role of Education and National Background. *Ethn Dis* **29**, 227–238 (2019).
150. Collins, N., Sachs-Ericsson, N., Preacher, K. J., Sheffield, K. M. & Markides, K. Smoking Increases Risk for Cognitive Decline Among Community-Dwelling Older Mexican Americans. *The American Journal of Geriatric Psychiatry* **17**, 934–942 (2009).
151. Alzheimer's Association. *Diabetes and cognitive decline*. (2021).
152. Hubens, W. H. G. *et al.* Blood biomarkers for assessment of mitochondrial dysfunction: An expert review. *Mitochondrion* **62**, 187–204 (2022).
153. Liu, C.-C., Liu, C.-C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106–18 (2013).
154. Giaglis, S. *et al.* Circulating mitochondrial DNA copy numbers represent a sensitive marker for diagnosis and monitoring of disease activity in systemic lupus erythematosus. *RMD Open* **7**, e002010 (2021).
155. O'Bryant, S. E. *et al.* A blood screening tool for detecting mild cognitive impairment and Alzheimer's disease among community-dwelling Mexican Americans and non-Hispanic Whites: A method for increasing representation of diverse populations in clinical research. *Alzheimer's & Dementia* **18**, 77–87 (2022).
156. Li, A. *et al.* Periodontitis and cognitive impairment in older adults: The mediating role of mitochondrial dysfunction. *Journal of Periodontology* (2022) doi:10.1002/JPER.21-0620.
157. Rosa, H. & Malik, A. N. Accurate Measurement of Cellular and Cell-Free Circulating Mitochondrial DNA Content from Human Blood Samples Using Real-Time Quantitative PCR. in 247–268 (2021). doi:10.1007/978-1-0716-1270-5_15.
158. Rosa, H. S., Ajaz, S., Gnudi, L. & Malik, A. N. A case for measuring both cellular and cell-free mitochondrial DNA as a disease biomarker in human blood. *The FASEB Journal* **34**, 12278–12288 (2020).
159. Phillips, N. R., Simpkins, J. W. & Roby, R. K. Mitochondrial DNA deletions in Alzheimer's brains: A review. *Alzheimer's & Dementia* **10**, 393–400 (2014).
160. Wei, Y.-H. Mitochondrial DNA alterations as ageing-associated molecular events. *Mutation Research/DNAging* **275**, 145–155 (1992).
161. Xu, L. & Seki, M. Recent advances in the detection of base modifications using the Nanopore sequencer. *Journal of Human Genetics* **65**, 25–33 (2020).
162. Stoiber, M. *et al. De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing*. (2017) doi:https://doi.org/10.1101/094672.
163. Xu, Z. *et al.* Fast-bonito: A faster deep learning based basecaller for nanopore sequencing. *Artificial Intelligence in the Life Sciences* **1**, 100011 (2021).

APPENDIX

APPENDIX A. Scatter plot of cellular 8oxoG variant count by age.



Sample means of total 8oxoG variant count was assessed by age using a two-tailed Pearson correlation (n = 559). Dotted lines represent 95% confidence interval (-0.03757 to 0.1333), and the solid line indicates best-fit line. Correlation statistics: r = 0.04824, R squared = 0.002327, p-value = 0.2704.

APPENDIX B. Cellular 8oxoG variant count multiple linear regression model prediction considering cognitive status (NC vs MCI or AD) and diabetes interaction effect.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	2.065866	2.198627	0.94	0.34783
Population (with respect to NHW)	-1.959649	0.406268	-4.824	***1.84E-06***
Cognitive Status (with respect to AD)	1.086334	0.570781	1.903	*0.05754*
Cognitive Status (with respect to MCI)	0.50188	0.525202	0.956	0.3397
Sex (with respect to Male)	-0.68253	0.322023	-2.12	***0.0345***
Diabetes (with respect to "Yes")	-0.009736	0.444694	-0.022	0.98254
APOE ε2/ε2	-1.464225	2.70427	-0.541	0.58842
APOE ε2/ε3	0.84649	0.813543	1.04	0.29857
APOE ε2/ε4	0.138294	2.235666	0.062	0.9507
APOE ε3/ε3	0.716064	0.620893	1.153	0.2493
APOE ε3/ε4	0.380755	0.670123	0.568	0.57014
APOE ε4/ε4	0.989753	0.918605	1.077	0.28176
BMI	0.035386	0.02488	1.422	0.15552
Years of Education	0.107933	0.041059	2.629	***0.00881***
Age	0.039031	0.025168	1.551	0.12154
Interaction: AD x Diabetes "Yes"	-1.170123	0.864129	-1.354	0.17627
Interaction: MCI x Diabetes "Yes"	-0.224683	0.807738	-0.278	0.78099
R-squared	0.07337		*p*-value	4.185e-04
Adjusted R-squared	0.04601		df	16 and 542
F-statistic	2.682		Sample n	559

APPENDIX C. Cellular 8oxoG variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD) and diabetes interaction effect.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	2.058131	2.195539	0.937	0.34896
Population (with respect to NHW)	-1.927984	0.404489	-4.766	***2.41E-06***
Cognitive Impairment	0.76542	0.439084	1.743	*0.08186*
Sex (with respect to Male)	-0.685558	0.321389	-2.133	***0.03336***
Diabetes (with respect to "Yes")	-0.003315	0.444255	-0.007	0.99405
APOE ε2/ε2	-1.469991	2.701135	-0.544	0.58652
APOE ε2/ε3	0.805548	0.808028	0.997	0.31924
APOE ε2/ε4	0.111867	2.231898	0.05	0.96004
APOE ε3/ε3	0.71578	0.616826	1.16	0.24638
APOE ε3/ε4	0.408495	0.668919	0.611	0.54167
APOE ε4/ε4	0.95647	0.916872	1.043	0.29732
BMI	0.034368	0.024827	1.384	0.16683
Years of Education	0.109596	0.040924	2.678	***0.00763***
Age	0.039027	0.025094	1.555	0.12048
Interaction: Cognitive Impairment x Diabetes "Yes"	-0.653041	0.663536	-0.984	0.32546

R-squared	0.07159	*p*-value	1.875e-04
Adjusted R-squared	0.0477	df	14 and 544
F-statistic	2.997	Sample n	559

APPENDIX D. Cellular 8oxoG variant count and cognitive impairment (NC vs MCI + AD) multiple linear regression model prediction considering population interaction effect with both sex and education.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	2.88116	2.21084	1.303	0.19306
Population (with respect to NHW)	-0.51694	1.46467	-0.353	0.72427
Cognitive Impairment	0.49811	0.34392	1.448	0.1481
Sex (with respect to Male)	-1.42915	0.44298	-3.226	***0.00133***
Diabetes (with respect to "Yes")	-0.35695	0.33806	-1.056	0.2915
Years of Education	0.1441	0.04539	3.175	***0.00158***
APOE ε2/ε2	-2.34923	2.69473	-0.872	0.38371
APOE ε2/ε3	0.62346	0.80435	0.775	0.43861
APOE ε2/ε4	0.46293	2.23149	0.207	0.83573
APOE ε3/ε3	0.6054	0.60887	0.994	0.32051
APOE ε3/ε4	0.36706	0.66375	0.553	0.58049
APOE ε4/ε4	0.65263	0.91027	0.717	0.4737
BMI	0.03332	0.02463	1.353	0.17657
Age	0.03107	0.02514	1.236	0.2171
Interaction: NHW x Male "Yes"	1.59823	0.64596	2.474	***0.01366***
Interaction: NHW x Years of Education	-0.15208	0.09796	-1.553	0.12111
R-squared	0.08267		*p*-value	3.163e-05
Adjusted R-squared	0.05733		df	15 and 543
F-statistic	3.262		Sample n	559

APPENDIX E. Cellular 8oxoG variant count stratification in the Mexican American population by cognitive impairment (NC vs MCI + AD).

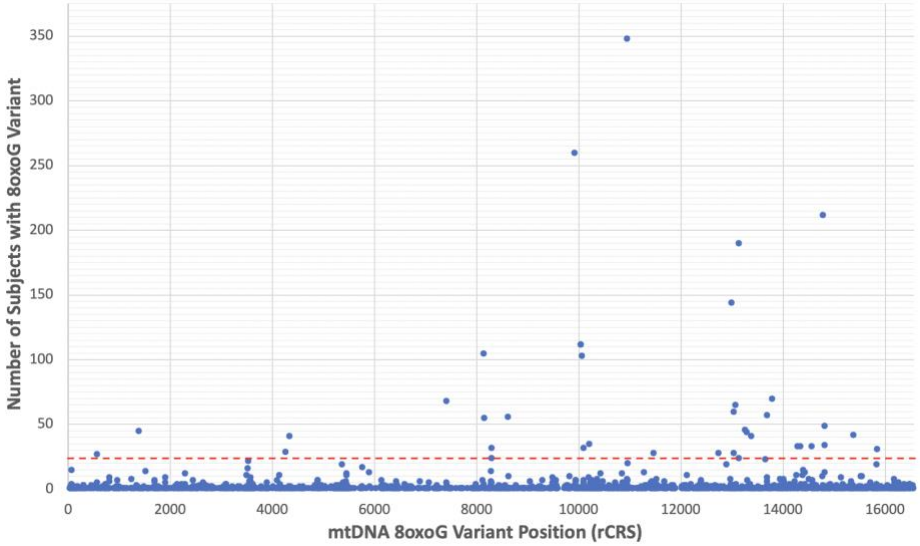
Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	5.09069	3.63725	1.4	0.16271
Cognitive Impairment	0.95919	0.53876	1.78	*0.07607*
Sex (with respect to Male)	-1.43742	0.52445	-2.741	***0.00651***
Diabetes (with respect to "Yes")	-0.24172	0.52806	-0.458	0.64747
APOE ε2/ε2	-1.85575	4.71698	-0.393	0.6943
APOE ε2/ε3	0.06738	2.16343	0.031	0.97517
APOE ε3/ε3	-0.60564	1.95648	-0.31	0.75712
APOE ε3/ε4	-0.80412	2.00258	-0.402	0.68832
APOE ε4/ε4	-1.07929	2.90559	-0.371	0.71057
BMI	0.02298	0.03896	0.59	0.55584
Years of Education	0.14618	0.05408	2.703	***0.00728***
Age	0.01642	0.03892	0.422	0.67335
R-squared	0.05647		*p*-value	0.1096
Adjusted R-squared	0.02031		df	11 and 287
F-statistic	1.562		Sample n	299

APPENDIX F. Stratification analysis in NHWs for cellular 8oxoG variant count by cognitive impairment (NC vs MCI or AD).

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	0.61747	2.80517	0.22	0.826
Cognitive Impairment	-0.11555	0.41685	-0.277	0.7819
Sex (with respect to Male)	0.18141	0.37302	0.486	0.6272
Diabetes (with respect to "Yes")	-0.54057	0.40557	-1.333	0.1838
APOE ε2/ε2	-4.01421	2.98532	-1.345	0.18
APOE ε2/ε3	0.37297	0.79069	0.472	0.6376
APOE ε2/ε4	0.34083	1.78683	0.191	0.8489
APOE ε3/ε3	0.93935	0.5389	1.743	*0.0826*
APOE ε3/ε4	0.76551	0.65468	1.169	0.2434
APOE ε4/ε4	1.29813	0.81143	1.6	0.1109
BMI	0.04823	0.02878	1.676	*0.095*
Years of Education	-0.02983	0.07078	-0.421	0.6738
Age	0.0556	0.03095	1.796	*0.0736*

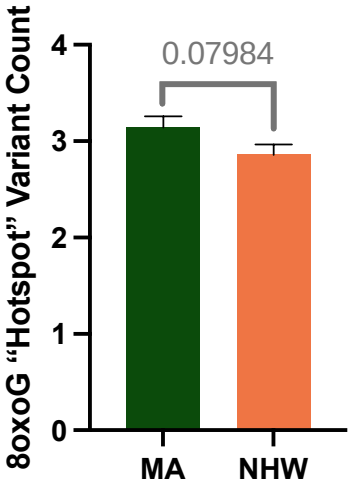
R-squared	0.04768	*p*-value	0.4212
Adjusted R-squared	0.001411	df	12 and 247
F-statistic	1.03	Sample n	260

APPENDIX G. Distribution of cellular 8oxoG variant in mitochondrial genome by number of individuals with 8oxoG variant- 8oxoG "hotspot" variant selection.



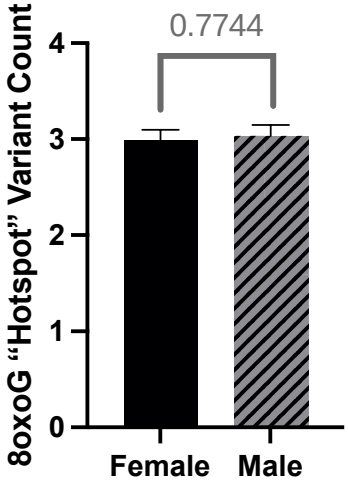
The red dashed line intercepts at 25 subjects with cellular 8oxoG variant. 8oxoG variants above the intercept were conveniently selected as "hotspots".

APPENDIX H. Cellular 8oxoG “hotspot” variant count does not differ between populations.



Welch’s t-test was performed to determine statistical difference in cellular 8oxoG “hotspot” variant count between MAs and NHWs (n = 559). Error bars represent standard error of the mean.

APPENDIX I. Cellular 8oxoG “hotspot” variant count does not differ between sexes.



Welch’s t-test was performed on cellular 8oxoG “hotspot” variant count by sex (n =559). Error bars represent standard error of the mean.

APPENDIX J. Cellular 8oxoG “hotspot” variant count multiple linear regression model prediction considering cognitive status (NC vs MCI or AD) and diabetes interaction effect.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	1.031513	1.106376	0.932	0.3516
Population with respect to NHW	-0.236656	0.204439	-1.158	0.2475
Cognitive Status with respect to AD	0.034295	0.287224	0.119	0.905
Cognitive Status with respect to MCI	0.09451	0.264288	0.358	0.7208
Sex with respect to Male	0.077441	0.162046	0.478	0.6329
Diabetes with respect to "Yes"	0.089844	0.223776	0.401	0.6882
APOE ε2/ε2	0.462665	1.360822	0.34	0.734
APOE ε2/ε3	0.662786	0.409385	1.619	0.106
APOE ε2/ε4	1.064559	1.125014	0.946	0.3444
APOE ε3/ε3	0.623034	0.312441	1.994	***0.0466***
APOE ε3/ε4	0.606919	0.337214	1.8	*0.0724*
APOE ε4/ε4	0.963228	0.462253	2.084	***0.0376***
BMI	0.016726	0.01252	1.336	0.1821
Years of Education	0.007475	0.020661	0.362	0.7177
Age	0.012951	0.012665	1.023	0.307
Interaction: AD x Diabetes "Yes"	-0.182847	0.43484	-0.42	0.6743
Interaction: MCI x Diabetes "Yes"	-0.767987	0.406463	-1.889	*0.0594*
R-squared	0.02788		*p*-value	0.4869
Adjusted R-squared	-0.0008199		df	16 and 542
F-statistic	0.9714		Sample n	559

APPENDIX K. Cellular 8oxoG “hotspot” variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD) and diabetes interaction effect.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	0.996468	1.105562	0.901	0.3678
Population with respect to NHW	-0.241563	0.20368	-1.186	0.2361
Cognitive Impairment	0.060094	0.221101	0.272	0.7859
Sex with respect to Male	0.070685	0.161835	0.437	0.6624
Diabetes with respect to "Yes"	0.086946	0.223705	0.389	0.6977
APOE ε2/ε2	0.430834	1.360154	0.317	0.7516
APOE ε2/ε3	0.636282	0.406882	1.564	0.1184
APOE ε2/ε4	1.028611	1.123871	0.915	0.3605
APOE ε3/ε3	0.587519	0.310602	1.892	*0.0591*
APOE ε3/ε4	0.591401	0.336834	1.756	*0.0797*
APOE ε4/ε4	0.989615	0.46169	2.143	***0.0325***
BMI	0.016723	0.012502	1.338	0.1816
Years of Education	0.008087	0.020607	0.392	0.6949
Age	0.013824	0.012636	1.094	0.2744
Interaction: Cognitive Impairment x Diabetes "Yes"	-0.501576	0.334123	-1.501	0.1339

R-squared	0.02471	*p*-value	0.4677
Adjusted R-squared	-0.0003888	df	14 and 544
F-statistic	0.9845	Sample n	559

APPENDIX L. Cellular 8oxoG "hotspot" variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering population interaction effect with both sex and education.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	2.246331	0.974249	2.306	***0.0215***
Population with respect to NHW	-0.753823	0.743259	-1.014	0.3109
Cognitive Status with respect to AD	-0.055	0.228654	-0.241	0.81
Cognitive Status with respect to MCI	-0.185378	0.206759	-0.897	0.3703
Sex with respect to Male	-0.163209	0.22485	-0.726	0.4682
Diabetes with respect to "Yes"	-0.089799	0.167338	-0.537	0.5917
Years of Education	0.002068	0.023077	0.09	0.9286
APOE ε2/ε2	0.223413	1.367797	0.163	0.8703
APOE ε2/ε3	0.578424	0.410699	1.408	0.1596
APOE ε2/ε4	0.949603	1.132505	0.838	0.4021
APOE ε3/ε3	0.514769	0.310391	1.658	*0.0978*
APOE ε3/ε4	0.545231	0.33675	1.619	0.106
APOE ε4/ε4	0.853611	0.461773	1.849	*0.0651*
Age	0.007646	0.012544	0.61	0.5424
Interaction: NHW x Male "Yes"	0.471257	0.328227	1.436	0.1516
Interaction: NHW x Years of Education	0.016038	0.049696	0.323	0.747
R-squared	0.02163		*p*-value	0.6776
Adjusted R-squared	-0.005392		df	15 and 543
F-statistic	0.8005		Sample n	559

APPENDIX M. Cellular 8oxoG “hotspot” variant count and cognitive impairment (NC vs MCI + AD) multiple linear regression model prediction considering population interaction effect with both sex and education.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	2.221287	0.972313	2.285	***0.0227***
Population with respect to NHW	-0.743506	0.742467	-1.001	0.3171
Cognitive Impairment	-0.129473	0.174356	-0.743	0.4581
Sex with respect to Male	-0.172106	0.224002	-0.768	0.4426
Diabetes with respect to "Yes"	-0.09046	0.167219	-0.541	0.5888
Years of Education	0.002985	0.02299	0.13	0.8967
APOE ε2/ε2	0.203655	1.366297	0.149	0.8816
APOE ε2/ε3	0.555142	0.407812	1.361	0.174
APOE ε2/ε4	0.928906	1.130984	0.821	0.4118
APOE ε3/ε3	0.499258	0.30865	1.618	0.1063
APOE ε3/ε4	0.546107	0.336514	1.623	0.1052
APOE ε4/ε4	0.854969	0.461449	1.853	*0.0645*
Age	0.008026	0.012513	0.641	0.5215
Interaction: NHW x Male "Yes"	0.482702	0.327216	1.475	0.1407
Interaction: NHW x Years of Education	0.015416	0.049646	0.311	0.7563
R-squared	0.02118		*p*-value	0.6246
Adjusted R-squared	-0.004013		df	14 and 544
F-statistic	0.8407		Sample n	559

APPENDIX N. Multiple linear regression results for cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) within Mexican Americans.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	3.119633	1.728499	1.805	*0.0722*
Cognitive Status (with respect to AD)	-0.18077	0.382078	-0.473	0.6365
Cognitive Status (with respect to MCI)	-0.427275	0.2831	-1.509	0.1323
Sex (with respect to Male)	-0.099865	0.250312	-0.399	0.6902
Diabetes (with respect to "Yes")	-0.04695	0.251607	-0.187	0.8521
APOE ε2/ε2	0.721106	2.243469	0.321	0.7481
APOE ε2/ε3	-0.64859	1.032714	-0.628	0.5305
APOE ε3/ε3	-0.711241	0.933457	-0.762	0.4467
APOE ε3/ε4	-0.583709	0.95293	-0.613	0.5407
APOE ε4/ε4	-0.060057	1.388688	-0.043	0.9655
BMI	0.003853	0.018646	0.207	0.8364
Years of Education	-0.007986	0.025716	-0.311	0.7564
Age	0.012322	0.018637	0.661	0.509
R-squared	0.01516		*p*-value	0.974
Adjusted R-squared	-0.02616		df	12 and 286
F-statistic	0.3669		Sample n	299

APPENDIX O. Multiple linear regression results for cellular 8oxoG "hotspot" variant count and cognitive impairment (NC vs MCI + AD) within Mexican Americans.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	3.108745	1.726513	1.801	*0.0728*
Cognitive Impairment	-0.353629	0.255736	-1.383	0.1678
Sex (with respect to Male)	-0.114114	0.248945	-0.458	0.647
Diabetes (with respect to "Yes")	-0.035737	0.250658	-0.143	0.8867
APOE ε2/ε2	0.663692	2.239034	0.296	0.7671
APOE ε2/ε3	-0.708339	1.026924	-0.69	0.4909
APOE ε3/ε3	-0.762167	0.928691	-0.821	0.4125
APOE ε3/ε4	-0.61417	0.950576	-0.646	0.5187
APOE ε4/ε4	-0.15056	1.379213	-0.109	0.9131
BMI	0.0025	0.018493	0.135	0.8926
Years of Education	-0.007422	0.025671	-0.289	0.7727
Age	0.013734	0.018472	0.743	0.4578
R-squared	0.01388		*p*-value	0.9677
Adjusted R-squared	-0.02392		df	11 and 287
F-statistic	0.3672		Sample n	299

APPENDIX P. Multiple linear regression results for cellular 8oxoG "hotspot" variant count and cognitive status (NC vs MCI or AD) within non-Hispanic Whites.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	-0.37463	1.60903	-0.233	0.8161
Cognitive Status (with respect to AD)	0.05397	0.2804	0.192	0.8475
Cognitive Status (with respect to MCI)	0.19302	0.30975	0.623	0.5338
Sex (with respect to Male)	0.30995	0.21397	1.449	0.1487
Diabetes (with respect to "Yes")	-0.31488	0.23322	-1.35	0.1782
APOE ε2/ε2	-1.33584	1.71249	-0.78	0.4361
APOE ε2/ε3	0.63187	0.46154	1.369	0.1722
APOE ε2/ε4	0.93667	1.02604	0.913	0.3622
APOE ε3/ε3	0.6952	0.31339	2.218	***0.0274***
APOE ε3/ε4	0.52071	0.37546	1.387	0.1667
APOE ε4/ε4	0.80166	0.46547	1.722	*0.0863*
BMI	0.03831	0.01651	2.321	***0.0211***
Years of Education	0.03061	0.04068	0.753	0.4524
Age	0.01545	0.01775	0.87	0.3849
R-squared	0.06899		*p*-value	0.1585
Adjusted R-squared	0.01979		df	13 and 246
F-statistic	1.402		Sample n	260

APPENDIX Q. Multiple linear regression results for cellular 8oxoG "hotspot" variant count and cognitive impairment (NC vs MCI + AD) within non-Hispanic Whites.

Variable	Coefficient	Std. Error	t-statistic	p-value
Constant	-0.36081	1.60594	-0.225	0.8224
Cognitive Impairment	0.11329	0.23864	0.475	0.6354
Sex (with respect to Male)	0.30798	0.21355	1.442	0.1505
Diabetes (with respect to "Yes")	-0.30786	0.23218	-1.326	0.1861
APOE ε2/ε2	-1.31871	1.70907	-0.772	0.4411
APOE ε2/ε3	0.66673	0.45266	1.473	0.1421
APOE ε2/ε4	0.95804	1.02294	0.937	0.3499
APOE ε3/ε3	0.71626	0.30852	2.322	***0.0211***
APOE ε3/ε4	0.52254	0.3748	1.394	0.1645
APOE ε4/ε4	0.80645	0.46454	1.736	*0.0838*
BMI	0.03815	0.01647	2.316	***0.0214***
Years of Education	0.02952	0.04052	0.728	0.467
Age	0.01533	0.01772	0.865	0.3877

R-squared	0.06837	*p*-value	0.1206
Adjusted R-squared	0.02311	df	12 and 247
F-statistic	1.511	Sample n	260

APPENDIX R. Cellular 8oxoG "hotspot" variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype and population interaction effect with both sex and education.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	2.262487	1.248584	1.812	***0.0707***
Population (with respect to NHW)	-1.267441	0.898335	-1.411	0.159
Cognitive Status (with respect to AD)	-0.051627	0.26992	-0.191	0.8484
Cognitive Status (with respect to MCI)	-0.038292	0.247408	-0.155	0.8771
Sex (with respect to Male)	-0.213908	0.263429	-0.812	0.4173
Age	0.004708	0.014752	0.319	0.7498
Years of Education	0.007887	0.026846	0.294	0.7691
BMI	0.025208	0.018139	1.39	0.1654
Diabetes (with respect to "Yes")	-0.160855	0.212408	-0.757	0.4493
Depression (with respect to "Yes")	-0.096772	0.260993	-0.371	0.711
Tobacco Abuse (with respect to "Yes")	0.205513	0.197733	1.039	0.2993
APOE	0.086777	0.177063	0.49	0.6243
OGG1	-0.154641	0.148929	-1.038	0.2997
Interaction: NHW x Male "Yes"	0.41861	0.387636	1.08	0.2808
Interaction: NHW x Years of Education	0.044539	0.059512	0.748	0.4547

R-squared	0.2393	*p*-value	0.7656
Adjusted R-squared	-0.009812	df	14 and 405
F-statistic	0.7092	Sample n	420

APPENDIX S. Cellular 8oxoG "hotspot" variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype as well as diabetes and cognitive status interaction.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	2.036356	1.236478	1.647	0.1004
Population (with respect to NHW)	-0.37884	0.239963	-1.579	0.1152
Cognitive Status (with respect to AD)	-0.06664	0.321575	-0.207	0.8359
Cognitive Status (with respect to MCI)	0.266971	0.309107	0.864	0.3883
Sex (with respect to Male)	0.003812	0.197416	0.019	***9.85e-01***
Age	0.005012	0.014672	0.342	0.7328
Years of Education	0.015568	0.024098	0.646	0.5186
BMI	0.023945	0.018147	1.319	0.1878
Diabetes (with respect to "Yes")	0.053767	0.279997	0.192	0.8478
Depression (with respect to "Yes")	-0.10751	0.259868	-0.414	0.6793
Tobacco Abuse (with respect to "Yes")	0.161687	0.196587	0.822	0.4113
APOE	0.061665	0.176915	0.349	0.7276
OGG1	-0.13653	0.148512	-0.919	0.3585
Interaction: AD x Diabetes "Yes"	0.044303	0.506826	0.087	0.9304
Interaction: MCI x Diabetes "Yes"	-0.86358	0.48535	-1.779	***0.0759***
R-squared	0.02768	*p*-value	0.6436	
Adjusted R-squared	-0.005934	df	14 and 405	
F-statistic	0.8235	Sample n	420	

APPENDIX T. Cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype within the Mexican American population.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	3.186749	1.725239	1.847	*0.066*
Cognitive Status (with respect to AD)	0.082055	0.425397	0.193	0.847
Cognitive Status (with respect to MCI)	-0.086177	0.337315	-0.255	0.799
Sex (with respect to Male)	-0.329255	0.292614	-1.125	0.262
Age	-0.004077	0.020858	-0.195	0.845
Years of Education	0.003534	0.02952	0.12	0.905
BMI	0.014829	0.023773	0.624	0.533
Diabetes (with respect to "Yes")	-0.254014	0.292223	-0.869	0.386
Depression (with respect to "Yes")	-0.375336	0.375911	-0.998	0.319
Tobacco Abuse (with respect to "Yes")	0.454095	0.283764	1.6	0.111
APOE	0.206517	0.307277	0.672	0.502
OGG1	-0.088752	0.201158	-0.441	0.659
R-squared	0.02522	*p*-value	0.8775	
Adjusted R-squared	-0.02181	df	11 and 228	
F-statistic	0.5363	Sample n	240	

APPENDIX U. Cellular 8oxoG “hotspot” variant count and cognitive status (NC vs MCI or AD) multiple linear regression model prediction considering *OGG1* genotype within the non-Hispanic White population.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	0.240313	1.997209	0.12	0.904
Cognitive Status (with respect to AD)	-0.053166	0.33765	-0.157	0.875
Cognitive Status (with respect to MCI)	0.276141	0.382296	0.722	0.471
Sex (with respect to Male)	0.188589	0.267741	0.704	0.482
Age	0.009629	0.021702	0.444	0.658
Years of Education	0.06132	0.049302	1.244	0.215
BMI	0.040103	0.029228	1.372	0.172
Diabetes (with respect to "Yes")	-0.097021	0.313935	-0.309	0.758
Depression (with respect to "Yes")	0.35329	0.381463	0.926	0.356
Tobacco Abuse (with respect to "Yes")	-0.119536	0.272527	-0.439	0.661
APOE	-0.079565	0.215729	-0.369	0.713
OGG1	-0.287485	0.223417	-1.287	0.2
R-squared	0.04725	*p*-value	0.6819	
Adjusted R-squared	-0.01513	df	11 and 168	
F-statistic	0.7574	Sample n	180	

APPENDIX V. Total cellular 8oxoG variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD), *OGG1* genotype, and population interaction effect with both sex and years of education.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	3.09706	2.4955	1.241	0.215301
Population (with respect to NHW)	-1.74819	1.7973	-0.973	0.331292
Cognitive Impairment	0.65399	0.42223	1.549	0.122185
Sex (with respect to Male)	-1.80365	0.52479	-3.437	***0.000649***
Age	0.01777	0.02942	0.604	0.546137
Years of Education	0.15093	0.05362	2.815	***0.005119***
BMI	0.0789	0.03611	2.185	***0.029476***
Diabetes (with respect to "Yes")	-0.63093	0.42414	-1.488	0.137643
Depression (with respect to "Yes")	-0.71806	0.52188	-1.376	0.169612
Tobacco Abuse (with respect to "Yes")	1.04785	0.39541	2.65	***0.008363***
APOE	0.18073	0.35033	0.516	0.606218
OGG1	-0.12746	0.2979	-0.428	0.668967
Interaction: NHW x Male "Yes"	2.27335	0.77412	2.937	***0.003506***
Interaction: NHW x Years of Education	-0.11929	0.11907	-1.002	0.316985

R-squared	0.122	*p*-value	7.353e-07
Adjusted R-squared	0.09385	df	13 and 406
F-statistic	4.338	Sample n	420

APPENDIX W. Total cellular 8oxoG variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD), *OGG1* genotype, and diabetes interaction effect.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	2.20767	2.4953	0.885	0.3768
Population (with respect to NHW)	-2.31379	0.48336	-4.787	***2.37e-06***
Cognitive Impairment	0.8922	0.51652	1.727	*0.0849*
Sex (with respect to Male)	-0.79321	0.39745	-1.996	***0.0466***
Age	0.02638	0.02955	0.893	0.3725
Years of Education	0.12382	0.0486	2.548	***0.0112***
BMI	0.08066	0.03652	2.208	***0.0278***
Diabetes (with respect to "Yes")	-0.22149	0.56487	-0.392	0.6952
Depression (with respect to "Yes")	-0.69608	0.5208	-1.337	0.1821
Tobacco Abuse (with respect to "Yes")	0.94259	0.39659	2.377	***0.0179***
APOE	0.15661	0.35343	0.443	0.6579
OGG1	-0.18157	0.29956	-0.606	0.5448
Interaction: Cognitive Impairment x Diabetes "Yes"	-0.69414	0.79965	-0.868	0.3859

R-squared	0.1041	*p*-value	9.263e-06
Adjusted R-squared	0.07766	df	12 and 407
F-statistic	3.94	Sample n	420

APPENDIX X. Cellular 8oxoG “hotspot” variant count multiple linear regression model prediction considering cognitive impairment (NC vs MCI + AD), *OGG1* genotype, and population interaction effect with both sex and years of education within the whole cohort.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	2.265021	1.24577	1.818	*0.0698*
Population (with respect to NHW)	-1.267594	0.897224	-1.413	0.1585
Cognitive Impairment	-0.044081	0.21078	-0.209	0.8344
Sex (with respect to Male)	-0.212817	0.261981	-0.812	0.4171
Age	0.004656	0.014688	0.317	0.7514
Years of Education	0.007817	0.026768	0.292	0.7704
BMI	0.025288	0.018028	1.403	0.1615
Diabetes (with respect to "Yes")	-0.161449	0.211734	-0.763	0.4462
Depression (with respect to "Yes")	-0.097161	0.260528	-0.373	0.7094
Tobacco Abuse (with respect to "Yes")	0.205233	0.197392	1.04	0.2991
APOE	0.085599	0.174886	0.489	0.6248
OGG1	-0.154501	0.148714	-1.039	0.2995
Interaction: NHW x Male "Yes"	0.417555	0.386446	1.081	0.2806
Interaction: NHW x Years of Education	0.044535	0.059439	0.749	0.4541

R-squared	0.02392	*p*-value	0.6969
Adjusted R-squared	-0.00733	df	13 and 406
F-statistic	0.7655	Sample n	420

APPENDIX Y. Cellular 8oxoG "hotspot" variant count multiple linear regression model considering cognitive impairment (NC vs. MCI + AD), *OGG1* genotype, and diabetes interaction with cognition.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	1.969724	1.234291	1.596	0.111
Population (with respect to NHW)	-0.39896	0.239093	-1.669	***0.096***
Cognitive Impairment	0.094742	0.255492	0.371	0.711
Sex (with respect to Male)	-0.01567	0.196599	-0.08	0.937
Age	0.005782	0.014616	0.396	***6.93e-01***
Years of Education	0.016958	0.02404	0.705	0.481
BMI	0.024231	0.018065	1.341	0.181
Diabetes (with respect to "Yes")	0.046774	0.27941	0.167	0.867
Depression (with respect to "Yes")	-0.05664	0.257613	-0.22	0.826
Tobacco Abuse (with respect to "Yes")	0.181266	0.196172	0.924	0.356
APOE	0.075946	0.174825	0.434	0.664
OGG1	-0.15207	0.148175	-1.026	0.305
Interaction: Cognitive Impairment x Diabetes "Yes"	-0.43665	0.395542	-1.104	0.27
R-squared	0.02215	*p*-value	0.6835	
Adjusted R-squared	-0.006685	df	12 and 407	
F-statistic	0.7681	Sample n	420	

APPENDIX Z. Cellular 8oxoG variant count multiple linear regression prediction model considering cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in the Mexican American population.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	5.22225	3.65763	1.428	0.154721
Cognitive Impairment	1.36712	0.64952	2.105	***0.036397***
Sex (with respect to Male)	-2.28525	0.61787	-3.699	***0.000271***
Age	-0.01356	0.04394	-0.308	0.757991
Years of Education	0.15573	0.06265	2.486	***0.013646***
BMI	0.06814	0.05009	1.36	0.17506
Diabetes (with respect to "Yes")	-0.92301	0.61752	-1.495	0.136364
Depression (with respect to "Yes")	-1.4394	0.79812	-1.803	*0.072626*
Tobacco Abuse (with respect to "Yes")	1.99263	0.60204	3.31	***0.001084***
APOE	0.21207	0.6515	0.326	0.745087
OGG1	-0.05146	0.42638	-0.121	0.904049
R-squared	0.1172	*p*-value	0.001224	
Adjusted R-squared	0.07869	df	10 and 229	
F-statistic	3.041	Sample n	240	

APPENDIX AA. Cellular 8oxoG "hotspot" variant count multiple linear regression considering cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in the Mexican American population.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	3.150962	1.719516	1.832	*0.0682*
Cognitive Impairment	-0.031558	0.30535	-0.103	0.9178
Sex (with respect to Male)	-0.340996	0.290473	-1.174	0.2416
Age	-0.003081	0.020658	-0.149	0.8816
Years of Education	0.00384	0.029454	0.13	0.8964
BMI	0.013702	0.023548	0.582	0.5612
Diabetes (with respect to "Yes")	-0.243125	0.290305	-0.837	0.4032
Depression (with respect to "Yes")	-0.375051	0.37521	-1	0.3186
Tobacco Abuse (with respect to "Yes")	0.458274	0.283027	1.619	0.1068
APOE	0.212727	0.306281	0.695	0.488
OGG1	-0.093201	0.200451	-0.465	0.6424
R-squared	0.02459	*p*-value	0.8318	
Adjusted R-squared	-0.01801	df	10 and 229	
F-statistic	0.5773	Sample n	201	

APPENDIX AB. Cellular 8oxoG variant count multiple linear regression predictive model considering cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in non-Hispanic Whites.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	0.513269	3.380988	0.152	0.88
Cognitive Impairment	0.016123	0.495422	0.033	0.974
Sex (with respect to Male)	0.604656	0.453736	1.333	0.184
Age	0.044461	0.036718	1.211	0.228
Years of Education	0.008396	0.083286	0.101	0.92
BMI	0.07249	0.049341	1.469	0.144
Diabetes (with respect to "Yes")	-0.175279	0.532067	-0.329	0.742
Depression (with respect to "Yes")	-0.02399	0.642238	-0.037	0.97
Tobacco Abuse (with respect to "Yes")	-0.14622	0.461915	-0.317	0.752
APOE	0.151415	0.362023	0.418	0.676
OGG1	-0.29547	0.378453	-0.781	0.436
R-squared	0.03962	*p*-value	0.7262	
Adjusted R-squared	-0.01721	df	10 and 169	
F-statistic	0.6971	Sample n	180	

APPENDIX AC. Multiple linear regression predictive model assessing cellular 8oxoG "hotspot" variant count with cognitive impairment (NC vs. MCI + AD) and *OGG1* genotype in non-Hispanic Whites.

Variable	Coefficient	Std. Error	t-statistic	*p*-value
Constant	0.319717	1.9925	0.16	0.873
Cognitive Impairment	0.080826	0.291964	0.277	0.782
Sex (with respect to Male)	0.192608	0.267398	0.72	0.472
Age	0.008598	0.021639	0.397	0.692
Years of Education	0.05812	0.049082	1.184	0.238
BMI	0.042181	0.029078	1.451	0.149
Diabetes (with respect to "Yes")	-0.093654	0.31356	-0.299	0.766
Depression (with respect to "Yes")	0.318328	0.378487	0.841	0.402
Tobacco Abuse (with respect to "Yes")	-0.121255	0.272218	-0.445	0.657
APOE	-0.103619	0.213349	-0.486	0.628
OGG1	-0.293728	0.223032	-1.317	0.19

R-squared	0.04369	*p*-value	0.6556
Adjusted R-squared	-0.0129	df	10 and 169
F-statistic	0.772	Sample n	180

APPENDIX AD. Example TSV of deletion profile for one sample.

sample	cluster.id	alt.reads	ref.reads	heteroplasmy	del.start.range	del.end.range	del.size	final.event	final.start	final.end	final.size	seq1	seq2	seq
indel/Phillips	cluster11622	5	3919	0.1274	4860 - 4860	14140 - 14140	9280	dup	14141	4860	7289	CTAACCCTACTCCTAA*TCACATAACCTATTC	TCCGGCCTGCTTCTTC*TCACATGACAAAAAC	TCACAT
indel/Phillips	cluster15282	5	3480	0.1435	6330 - 6330	13993 - 13993	7663	del	6331	13993	7663	CACCCTGGAGCCTCCG*TAGACCTAACCATCT	CTGCCCCTACTCCTCC*TAGACCTAACCTGAC	TAGACCTAACC
indel/Phillips	cluster16421	5	5212	0.0958	6841 - 6841	11136 - 11136	4295	del	6842	11136	4295	CTACCATAATCATCGC*TATCCCCACCGGCGT	TCTTCGAAACCACACT*TATCCCCACCTTGGC	TATCCCCACC
indel/Phillips	cluster20071	6	4424	0.1354	8616 - 8616	11137 - 11137	2521	del	8617	11137	2521	ATTTCCCCCTCTATTG*ATCCCCACCTCCAAA	CTTCGAAACCACACTT*ATCCCCACCTTGGCT	ATCCCCACCT
indel/Phillips	cluster21801	5	6121	0.0816	9006 - 9006	12732 - 12732	3726	del	9007	12732	3726	CCTGGCCGTACGCCTA*ACCGCTAACATTACT	CATACTAATCTTAGTT*ACCGCTAACAACCTA	ACCGCTAACA
indel/Phillips	cluster26361	9	6384	0.1408	10668 - 10668	12733 - 12733	2065	del	10669	12733	2065	GCCATACTAGTCTTTG*CCGCCTGCGAAGCAG	ATACTAATCTTAGTTA*CCGCTAACAACCTAT	CCGC
indel/Phillips	cluster27241	5	4950	0.1009	10868 - 10868	11977 - 11977	1109	del	10869	11977	1109	CTAATTATTAGCATCA*TCCCTCTACTATTTT	AGTCACAGCCCTATAC*TCCCTCTACATATTT	TCCCTCTAC
indel/Phillips	cluster27411	8	4404	0.1813	10897 - 10897	12738 - 12738	1841	del	10898	12738	1841	TTTTAACCAAATCAAC*AACAACCTATTTAGC	AATCTTAGTTACCGCT*AACAACCTATTCCAA	AACAACCTATT
indel/Phillips	cluster27621	5	3957	0.1262	10927 - 10927	12092 - 12092	1165	del	10928	12092	1165	CTGTTCCCCAACCTTT*TCCTCCGACCCCCTA	CACCTATCCCCCATTC*TCCTCCTATCCCTCA	TCCTCC
indel/Phillips	cluster28061	7	5257	0.133	11031 - 11031	12418 - 12418	1387	del	11032	12418	1387	GTGAACCACTATCACG*AAAAAAACTCTACCT	CTCGTTAACCCTAACA*AAAAAAACTCATACC	AAAAAAACTC
indel/Phillips	cluster29272	5	6234	0.0801	11320 - 11320	12915 - 12915	1595	del	11321	12915	1595	TGCCCAAGAACTATCA*AACTCCTGAGCCAAC	ATTTATCCTACACTCC*AACTCATGAGACCCA	AACTC
indel/Phillips	cluster29461	11	7137	0.1539	11358 - 11358	11820 - 11820	462	del	11359	11820	462	TATGACTAGCTTACAC*AATAGCTTTTATAGT	AAACTCTACTCCCACT*AATAGCTTTTTGATG	AATAGCTTTT
indel/Phillips	cluster29881	6	7659	0.0783	11509 - 11509	11670 - 11670	161	del	11510	11670	161	CCTCACACTCATTCTC*AACCCCCTGACAAAA	CAGCCATTCTCATCCA*AACCCCCTGAAGCTT	AACCCCCTGA
indel/Phillips	cluster30142	5	5826	0.0857	11559 - 11559	14078 - 14078	2519	del	11560	14078	2519	TTGTACTATCCCTATG*AGGCATAATTATAAC	TCACCTCAACCCAAAA*AGGCATAATTAAACT	AGGCATAATTA
indel/Phillips	cluster30511	11	7064	0.1555	11632 - 11632	12539 - 12539	907	del	11633	12539	907	TGCATACTCTTCAATC*AGCCACATAGCCCTC	TCTCGAACTGACACTG*AGCCACAACCCAAAC	AGCCACA
indel/Phillips	cluster31632	6	7105	0.0844	11841 - 11841	12845 - 12845	1004	del	11842	12845	1004	CTTTTTGATGACTTCT*AGCAAGCCTCGCTAA	ACACAGCAGCCATTCA*AGCAATCCTATACAA	AGCAA
indel/Phillips	cluster31722	6	7244	0.0828	11856 - 11856	12884 - 12884	1028	del	11857	12884	1028	TAGCAAGCCTCGCTAA*CCTCGCCTTACCCCC	GCGATATCGGTTTCAT*CCTCGCCTTAGCATG	CCTCGCCTTA
indel/Phillips	cluster32082	5	4474	0.1116	11966 - 11966	13066 - 13066	1100	del	11967	13066	1100	CTCAACATACTAGTCA*CAGCCCTATACTCCC	GGCCCCACCCCAGTCT*CAGCCCTACTCCACT	CAGCCCTA
indel/Phillips	cluster32633	6	5940	0.1009	12062 - 12062	16185 - 16185	4123	dup	16186	12062	12446	AATCCACATCAAAACC*CCCTCCCCATGCTTA	TTCACACGAGAAAACA*CCCTCATGTTCATAC	AAAAC
indel/Phillips	cluster32842	5	5647	0.0885	12102 - 12102	14065 - 14065	1963	del	12103	14065	1963	CCATTCTCCTCCTATC*CCTCAACCCCGACAT	TCCACCTCCATCATCA*CCTCAACCCAAAAAG	CCTCAACCC
indel/Phillips	cluster32892	5	5816	0.0859	12121 - 12121	14064 - 14064	1943	del	12122	14064	1943	CAACCCCGACATCATT*ACCGGGTTTTCCTCT	CTCCACCTCCATCATC*ACCTCAACCCAAAAA	CATCAT
indel/Phillips	cluster32963	5	5382	0.0928	12183 - 12183	16193 - 16193	4010	dup	16194	12183	12559	TCAAAACCCCCTCCCC*ATGCTTACAAGCAAG	GTGAATCTGACAACAG*AGGCTTACGACCCCT	CA
indel/Phillips	cluster34222	7	5586	0.1252	12541 - 12541	13376 - 13376	835	del	12542	13376	835	TCGAACTGACACTGAG*CCACAACCCAAACAA	GCTCCGGGTCCATCAT*CCACAACCTTAACAA	CCACAACC
indel/Phillips	cluster35152	11	4688	0.2341	12934 - 12934	13139 - 13139	205	del	12935	13139	205	TCATGAGACCCACAAC*AAATAGCCCTTCTAA	TCCACCCCCTAGCAGA*AAATAGCCCACTAAT	AAATAGCCC
indel/Phillips	cluster35282	9	4640	0.1936	12952 - 12952	13149 - 13149	197	del	12953	13149	197	ATAGCCCTTCTAAACG*CTAATCCAAGCCTCA	AGCAGAAAATAGCCCA*CTAATCCAAACTCTA	CTAATCCAA
indel/Phillips	cluster35662	5	3790	0.1318	13019 - 13019	14047 - 14047	1028	del	13020	14047	1028	AATCAGCCCAATTAGG*TCTCCACCCCTGACT	ATTTCACAGCACCAAA*TCTCCACCTCCATCA	TCTCCACC
indel/Phillips	cluster35712	5	2890	0.1727	13024 - 13024	14416 - 14416	1392	del	13025	14416	1392	GCCCAATTAGGTCTCC*ACCCCTGACTCCCCT	ACTCACCAAGACCTCA*ACCCCTGACCCCCAT	ACCCCTGAC
indel/Phillips	cluster36482	5	4928	0.1014	13292 - 13292	14330 - 14330	1038	del	13293	14330	1038	TAGTTACAATCGGCAT*CAACCAACCACACCT	CTATTAAAGTTTACCA*CAACCACCACCCCAT	CAACCA
indel/Phillips	cluster36912	5	4492	0.1112	13434 - 13434	14282 - 14282	848	del	13435	14282	848	ACTACTCAAAACCATA*CCTCTCACTTCAACC	CGAATCAACCCTGACC*CCTCTCCTTCATAAA	CCTCTC
indel/Phillips	cluster37012	5	5003	0.0998	13450 - 13450	14074 - 14074	624	del	13451	14074	624	CCTCTCACTTCAACCT*CCCTCACCATTGGCA	ATCATCACCTCAACCC*AAAAAGGCATAATTA	TCAACC
indel/Phillips	cluster37402	5	3580	0.1395	13527 - 13527	14565 - 14565	1038	del	13528	14565	1038	AGACCACATCATCGAA*ACCGCAAACATATCA	TAACACACCCGACCAC*ACCGCTAACAATCAA	ACCGC
indel/Phillips	cluster37462	9	2684	0.3342	13547 - 13547	13698 - 13698	151	del	13548	13698	151	CAAACATATCATACAC*AAACGCCTGAGCCCT	CCTACTAAACCCCATT*AAACGCCTGGCAGCC	AAACGCCTG
indel/Phillips	cluster37642	6	3444	0.1739	13586 - 13586	14002 - 14002	416	del	13587	14002	416	CTCTCATCGCTACCTC*CCTGACAAGCGCCTA	CTCCTCCTAGACCTAA*CCTGACTAGAAAAGC	CCTGAC
indel/Phillips	cluster38762	6	3000	0.1996	13933 - 13933	14562 - 14562	629	del	13934	14562	629	TTCTACCCTAGCATCA*CACACCGCACAATCC	TAATAACACACCCGAC*CACACCGCTAACAAT	CACACCGC

APPENDIX AE. Example mtDNA plot of deletions (blue) and duplications (red) with varying heteroplasmy for one sample.