NIAGADS

RESEARCH USE STATEMENT AND NON-TECHNICAL SUMMARY

Research Use Statement for Application for Genomic Data from NIAGADS

Mitochondrial Peptide DNA Variation in Alzheimer's Disease

Objective:

Our objective is to verify preliminary observations that a particular mitochondrial single nucleotide polymorphism (mtSNP) associates with neurodegeneration phenotypes including Alzheimer's disease (AD). Importantly, our lab has identified a mtSNP that is enriched in AD patients of the ADNI cohort. This same mtSNP predicts longitudinal cognitive decline in the Health and Retirement Study, associates with multi-phenotypic brain imaging markers in the UK Biobank, and differentiates the brain temporal cortex transcriptome in a Mayo Clinic cohort. We've further functionalized this mtSNP in the context of a novel peptide (which we call SHMOOSE, and which we separately identified by mass-spec analysis of CSF) that is derived from a mitochondrial small open reading frame affected by this SNP. This neurodegeneration-associated-mtSNP is non-synonymous to the SHMOOSE peptide and leads to reduced protection against amyloid beta toxicity *in vitro*. Our goal is to further study the association of this SHMOOSE mtSNP with the diagnosis and phenotype of AD in additional cohorts.

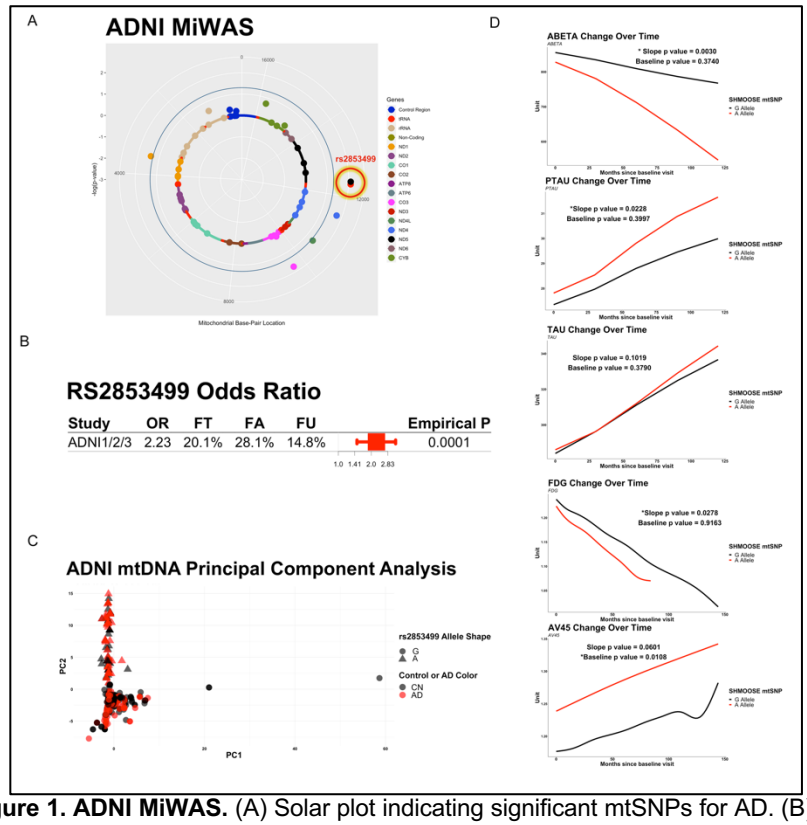

Figure 1. ADNI MiWAS. (A) Solar plot indicating significant mtSNPs for AD. (B) Odds ratio for most significant AD-associative mtSNP. (C) Principal component analysis on mtDNA. (D) AD-related biomarkers assessed as a function of time and RS2853449 (SHMOOSE mtSNP).

Study Design: We have developed an analytic plan to identify mtSNPs that associate with any desired phenotype. Many genome wide association studies do not include the mitochondrial genome due to its non-nuclear genetic behavior. Studies that do examine the mitochondrial genome focus on the known 37 genes and haplotypes. We have named our analytic plan MiWAS (**M**itochondrial **W**ide **A**ssociation **S**tudy) that focuses on both the known 37 genes and mitochondrial smORFs that encode for MDPs. We have published mtSNPs that associate with cataract in a Hispanic population using MiWAS, shown the utility of mtSNPs for assessing ethnic-specific traits, and identified a mtSNP in the humanin open reading frame that associated with circulating humanin levels.[1-3] In our unpublished data included here, we have identified a mtSNP that associates with AD in the ADNI cohort (Figure 1) and also differentiates the temporal cortex transcriptome in patient RNASeq data collected by Mayo Clinic (included in this proposal is a request for the Mayo Clinic GWAS data).

We will estimate the effects of the SHMOOSE mtSNP on AD while considering genetic ancestry, biological sex, and age. These data were considered priority characteristics, although mtSNPs were likely filtered from such standard GWAS pipelines. These mtSNPs are included on the Illumina-based genotype arrays. All data are available in the following data sets that we are requesting:

1. NG00042 - Miami, Vanderbilt, and Medical School of Mount Sinai (UMVUMSSM) GWAS
2. NG00031 - MIRAGE Caucasian GWAS

3. NG00026 - University of Pittsburgh GWAS
4. NG00022 - ADC1- Alzheimer Disease Center Dataset 1
5. NG00023 - ADC2- Alzheimer Disease Center Dataset 2
6. NG00020 - NIA-LOAD GWAS

Analysis plan

Our analysis plan will estimate the effects of the SHMOOSE mtSNP on AD diagnosis. The phenotype of AD diagnosis is consistent with NIAGADS date use. We will examine the effect of the SHMOOSE mtSNP on AD separately by each cohort. Analyses will be conducted solely on individuals of European ancestry because the SHMOOSE mtSNP is the European haplogroup U determining mtSNP. We will ensure European ancestry of the studied samples by using the 1000 Genomes as a reference population. That is, we will conduct principal component analysis on the requested data sets and the 1000 Genomes reference population to ensure accurate ethnic determination. See Figure 2 for our examples of PCA derived from genotype data for both the nuclear and mitochondrial genome. The expected SHMOOSE mtSNP frequency is 20-25%. Due to the high frequency nature of the SHMOOSE mtSNP, we will have ample power to estimate effects for AD diagnosis in each cohort.

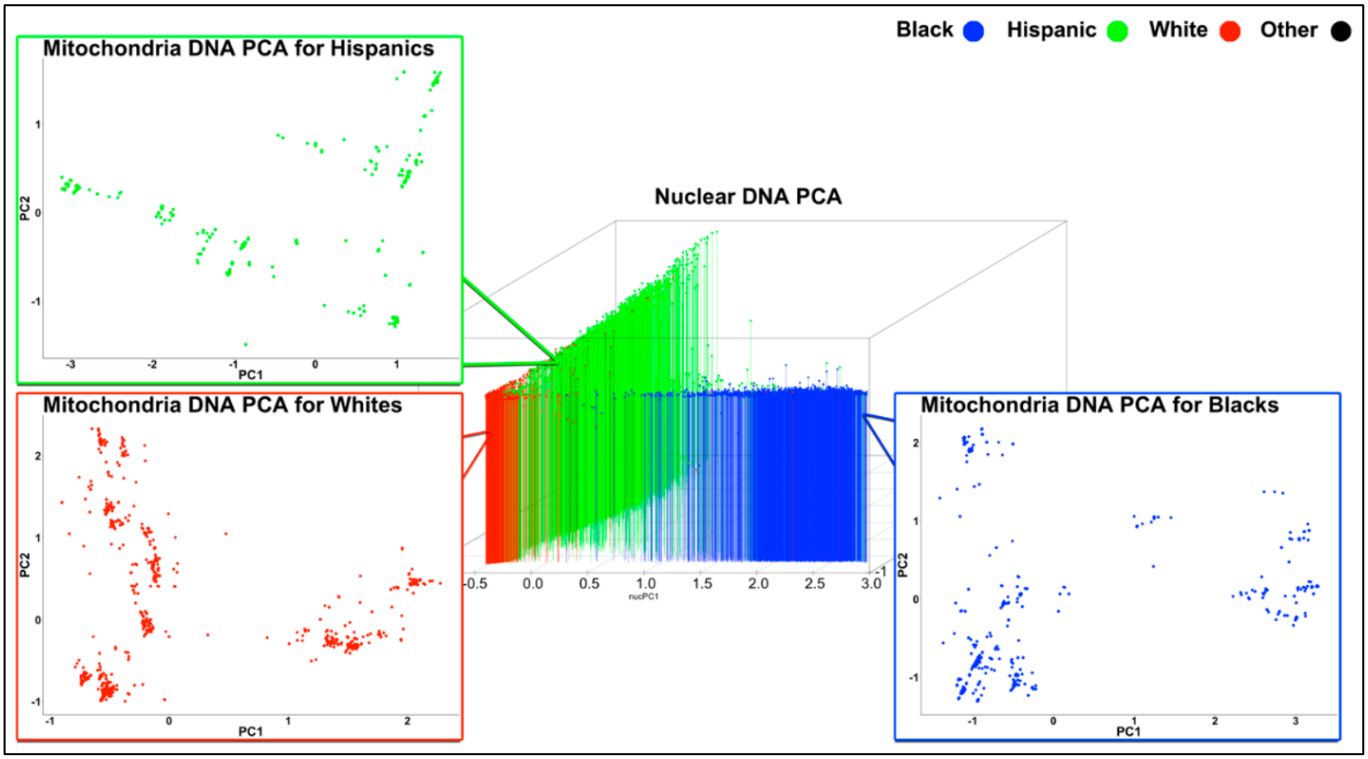

Figure 2. PCA. Principal component analysis on HRS data. Background image is a PCA on nuclear genotypes. Sub-boxes are PCAs on mitochondrial DNA genotypes.

Logistic regression analyses on AD will be conducted with the independent variables of SHMOOSE mtSNP, nuclear genetic ancestry, mitochondrial genetic ancestry, biological sex, and age. SHMOOSE mtSNP data will be extracted from genotype files (e.g., VCF or PLINK files). Nuclear genetic ancestry will be captured via principal component analysis on the entire genotype data set. PLINK will be used to generate eigenvalues and eigenvectors; PLINK uses EIGENSOFT for principal component analysis. Eigenvectors will be derived from LD-pruned data with PLINK input values for SNP window of 50, window shift of 5, and threshold of 2. Mitochondrial genetic ancestry will be conducted using the R *prcomp*

command, as previously described.[1] Analyses on AD diagnosis will be conducted in R using custom scripts.

Overall, we will utilize NIAGADS cohort to determine the effect of the SHMOOSE mtSNP on AD, results of which are relevant to novel microprotein characterization and future therapeutic development.

References

1 Miller, B. *et al.* Comparing the Utility of Mitochondrial and Nuclear DNA to Adjust for Genetic Ancestry in Association Studies. *Cells* **8**, 306 (2019).
2 Miller, B. *et al.* A Mitochondrial Genome-Wide Association Study of Cataract in a Latino Population. *Translational Vision Science & Technology* **9**, 25-25, doi:10.1167/tvst.9.6.25 (2020).
3 Yen, K. *et al.* Humanin Prevents Age-Related Cognitive Decline in Mice and is Associated with Improved Cognitive Age in Humans. *Sci Rep* **8**, 14212, doi:10.1038/s41598-018-32616-7 (2018).

Non-Technical Summary for Application for Genomic Data from NIAGADS

The objective of our research is to estimate the effects of mitochondrial DNA genotypes on Alzheimer's disease (AD). Many genome wide association techniques do not include the mitochondrial genome due to its non-nuclear genetic behavior. Studies that do examine the mitochondrial genome focus on the known 37 genes and haplotypes. Our group has developed a novel strategy called MiWAS that focuses on both the known 37 genes and mitochondrial smORFs that encode for peptides, which we call mitochondrial-derived peptides (MDPs).

Our preliminary analyses have identified one such mitochondrial DNA genotype that associates with many markers of neurodegeneration (i.e., AD diagnosis, brain imaging, and gene expression). Our goal is to further estimate the effects of this mitochondrial DNA genotype on AD in the cohorts of NIAGADS.