Assembly of 809 whole mitochondrial genomes with clinical, imaging, and fluid biomarker phenotyping

Perry G. Ridge[a], **Mark E. Wadsworth**[a], **Justin B. Miller**[a], **Andrew J. Saykin**[b], **Robert C. Green**[c], **the Alzheimer's Disease Neuroimaging Initiative**[1], and **John S. K. Kauwe**[a,d,*]

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

[b]Radiology and Imaging Sciences, Medical and Molecular Genetics and the Indiana Alzheimer's Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA

[c]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Partners HealthCare Personalized Medicine, The Broad Institute and Harvard Medical School, Boston, MA, USA

[d]Department of Neuroscience, Brigham Young University, Provo, UT, USA

Abstract

Introduction—Mitochondrial genetics are an important but largely neglected area of research in Alzheimer's disease. A major impediment is the lack of data sets.

Methods—We used an innovative, rigorous approach, combining several existing tools with our own, to accurately assemble and call variants in 809 whole mitochondrial genomes.

Results—To help address this impediment, we prepared a data set that consists of 809 complete and annotated mitochondrial genomes with samples from the Alzheimer's Disease Neuroimaging Initiative. These whole mitochondrial genomes include rich phenotyping, such as clinical, fluid biomarker, and imaging data, all of which is available through the Alzheimer's Disease Neuroimaging Initiative website. Genomes are cleaned, annotated, and prepared for analysis.

Discussion—These data provide an important resource for investigating the impact of mitochondrial genetic variation on risk for Alzheimer's disease and other phenotypes that have been measured in the Alzheimer's Disease Neuroimaging Initiative samples.

Keywords

Alzheimer's disease; ADNI; Mitochondrial genetics; Whole mitochondrial genomes; Next-generation sequencing

---

*Corresponding author. Tel.: 11-801-422-2993; Fax: 11-801-422-0004. kauwe@byu.edu.

1. Introduction

Alzheimer's disease (AD), the most common form of dementia, affects >20 million people worldwide and is the only one of the top 10 causes of death that has no effective treatments [1–3]. Full-time care is required as AD progresses, further impacting patients and their loved ones and stressing the health-care system. With incidence expected to increase to 1 in 85 people by 2050 [2], it is essential to achieve early diagnosis, effective treatments, and a better understanding of the underlying etiology.

Understanding the underlying mechanisms of risk for AD is a key for both diagnosis and treatment. Swerdlow et al. [4] proposed the Mitochondrial Cascade Hypothesis of AD. Briefly, an individual's genetics determine the baseline mitochondrial function and how mitochondria change as a person ages and declining mitochondrial function causes AD-specific pathologies.

In addition to the evidence provided by Swerdlow et al. [4], several lines of evidence support a role for mitochondria in AD. First, mitochondria fundamentally change in a number of ways in AD and contribute to its progression and onset [5]: metabolism decreases [6], mitochondrial fusion/fission are disrupted [7], mitochondrial concentration (i.e. the ratio of mitochondrial genomes to nuclear genomes) decreases in cerebrospinal fluid [8,9], mitochondrial morphology changes [4,10], mitochondrial-encoded enzymes in the electron transport chain are altered [5,11], amyloid plaques aggregate in mitochondria [12,13], and many of these changes take place near plaques [14].

Second, individuals with a maternal family history of AD have as high as 9 times the risk of AD compared with individuals with a paternal family history of AD [15,16], or no family history. Furthermore, individuals with a maternal family history of AD also score lower on cognitive tests [17], have a lower age of onset of AD [15,18], and have more pronounced brain abnormalities consistent with AD (e.g. cerebral metabolic changes [19], higher amyloid β burden [20], reduction in gray matter volume [21,22], and increased global Pittsburgh Compound-B uptake Pittsburgh Compound-B-positron emission tomography [23]). Moreover, we found that some of these brain abnormalities are associated with mitochondrial haplotypes [24].

This mitochondrial impact on AD risk could be influenced by several factors, including differential responses to the oxidative stress, variation in nuclear-encoded mitochondrial genes, and variation in the mitochondrial genome. In this article, we focus on an important resource for investigating mitochondrial genomic variation and others [25]. Several groups have reported a relationship between mitochondrial genetics and risk for AD (summarized in Ridge et al. [3], Table 2). Twelve different haplotypes have been implicated in mitochondrial genetic studies, but the majority of these were reported only once and not replicated [26–33], and six different studies reported no association between mitochondrial genetic variants and AD [34–39]. Among reported associations, there is no consensus, and sometimes, results appear to be contradictory. For example, Haplogroup U has been reported as both a risk and protective haplogroup [28,31,32]. However, potentially explaining the confounding nature of discoveries to date, the majority of studies used incomplete sequence data and/or had very

small sample sizes [26–39], thus most were underpowered and lacked the resolution to identify correlations for all but the most common haplogroups. Only a single study used whole mitochondrial data [30], whereas most genotyped only a handful of single nucleotide polymorphisms (SNPs). Furthermore, only one study used a large data set, but in this particular data set, the authors only genotyped 138 SNPs [39]. In summary, there is strong evidence to suggest a relationship between the mitochondrial genome and AD, yet the relationship remains undefined.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) recently sequenced the whole genomes, including mitochondrial genomes, of 809 individuals. Each of the genomes was analyzed using tools and pipelines developed for diploid genomes. However, these analysis pipelines, particularly variant identification that relies on a likelihood model expecting diploid sequences, are inaccurate for use on the mitochondrial genome, which is haploid. Here, we report not only an AD data set of 809 annotated whole mitochondrial genomes with extensive phenotypes (Table 1) but also an appropriate pipeline to analyze mitochondrial genomes. We hope to facilitate research in this important area by providing a data set and analysis pipeline for future researchers to augment this initial data set.

2. Methods

2.1. Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership and is an ongoing, longitudinal, highly collaborative study. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and used for the early diagnosis of AD. ADNI has undergone several phases (ADNI1, ADNI GO, and ADNI 2), with each phase adding additional samples. In 2012, 818 ADNI samples were selected for whole genome sequencing to further the goals of ADNI. DNA sequence data were collected from DNA derived from whole blood. All subjects in our analyses had self-reported ancestry of non-Hispanic European American. All the data (whole genome sequence, phenotype, and newly assembled and annotated whole mitochondrial genomes) are publically available through ADNI (http://adni.loni.usc.edu/data-samples/).

2.2. Genome sequencing, assembly, and variant detection

ADNI genomes were sequenced on an Illumina HiSeq. Reads were paired-end, 100 base-pair reads. Before read mapping, adapters were removed. ADNI mapped the whole genome sequences and called variants using default settings in the Burrows-Wheeler Aligner [40] for mapping and standard best practices from the Genome Analysis Toolkit [41,42] for variant detection (for details see http://adni.loni.usc.edu/data-samples/genetic-data/wgs/). However, these steps needed to be redone for two reasons: First, original mappings were to Hg19. Historically, determining a reference mitochondrial genome has been a bit confusing. The standard mitochondrial genome, NC_012920, was released first in 1981 (the Cambridge Reference Sequence, or CRS) [43] and corrected in 1999 (the new Cambridge Reference

Sequence, or rCRS) [44]. NC_012920 is a European haplogroup and thus is a leaf on the mitochondrial haplotype phylogenetic tree. A number of other mitochondrial genomes have been suggested as the "correct" mitochondrial genome reference, including a reconstructed hypothetical mitochondrial genome corresponding to Mitochondrial Eve [45]. Nevertheless, NC_012920 remains the standard version of the mitochondrial genome used in mitochondrial genetics studies. Hg19, which was used for read mapping by ADNI, has a version of the mitochondrial genome, represented as chrM, but sometimes corresponds to NC_001807 and sometimes NC_012920. Therefore, to be consistent with mitochondrial genetic standards, reads needed to be remapped. Second, the standard Genome Analysis Toolkit pipeline includes many steps. One of these steps is the HaplotypeCaller. The HaplotypeCaller builds a likelihood model based on possible reconstructed haplotypes in a genomic region, but it assumes sequences are diploid. Consequently, mitochondrial genomic variants identified using the HaplotypeCaller are likely to include many inaccuracies.

Since chrMand NC_012920 only differ by a few bases, we were able to extract only those reads that mapped to chrM (with SAMtools [46]), rather than all reads corresponding to the whole nuclear and mitochondrial genomes. Extracted reads were remapped to NC_012920 using Burrows-Wheeler Aligner. Next, we performed local realignments around indels and base recalibration, which are not affected by ploidy, with Genome Analysis Toolkit to refine the new mappings. Finally, we used FreeBayes (-p 1 –F 0.6, and removed variants with quality less than 20) [47] to joint-call variants and converted the resulting variant call format (VCF) file to fasta with vcf2fasta (vcflib, https://github.com/vcflib/vcflib). An overview of the whole process is outlined in Supplementary Figure 1.

2.3. Genome annotation

We annotated mitochondrial variants and haplotypes for each sample. We downloaded 9228 mitochondrial DNA coding and RNA sequence variants and 2792 control region variants from MITOMAP [48]. For each variant present in the data sets downloaded from MITOMAP, we added complete information (i.e. frequency, source, locus names, etc.) to the "Info" column in the VCF file and added the corresponding annotation information to the header lines. For each variant that was reported by multiple studies in MITOMAP, we included all studies in the annotation.

Next, we annotated mitochondrial haplotypes with Phy-Mer [49]. Phy-Mer reports the five most likely mitochondrial haplotypes and a score, where 1 is a perfect score. For each of the samples, we selected the top hit. All samples had scores >0.99 except for one that had a score of 0.988.

2.4. Variant validation

SNP data from the Illumina 2.5 M Array were collected from the same DNA extraction that was used for whole genome sequence (WGS) data collected. We compared 256 mitochondrial variants genotyped on that array to the variant calls from our WGS analysis pipeline. Validation was evaluated by simply looking at the concordance of calls on the individual level for the two sets of genotype data.

3. Results

Mitochondrial genomes were sequenced to an average depth of 2986 reads and ranged from an average depth of 1515 reads to 7831 reads. We identified 1649 total mitochondrial genetic variants from these genomes, of which 1336 have been previously reported. Samples had an average of about 27 variants, with a range of 1 to 96 variants. We validated our variant calls using 256 mitochondrial genetic genotypes from an SNP array performed on the same samples. Overall, 98.18% of WGS variant calls matched SNP genotypes acquired from the array.

We identified 506 different mitochondrial haplotypes in the data set, all of which have been previously observed. The majority, 350, only appear in a single sample in the data set. The most common haplotype, K1A1B1A, is shared by 15 individuals. When considering only the major mitochondrial haplogroup (e.g. H, V, U, etc.), the majority of individuals had haplotypes in the H and U groups, 338 and 105, respectively. We report number of individuals in each major mitochondrial haplogroup in Table 2 and counts of all haplotypes in Supplementary Table 1. These frequencies are similar to those observed in other cohorts of non-Hispanic European American ancestry [50]. Several individuals have haplotypes reported to be associated with risk for AD: five have haplotype H5 (a risk haplogroup [29]).

4. Discussion

We have presented the application of a novel approach to accurately assemble and genotype a data set of whole mitochondrial genomes from the ADNI study. Several accurate algorithms exist for calling genotypes in diploid NGS data, but use models that are inappropriate for haploid samples. We used existing software, but tested and refined parameter settings to achieve high levels of genotype accuracy in our mitochondrial genome samples. We validated the identified genotypes by repurposing existing publically available data. Our genotypes are >98% concordant with genotypes from Illumina SNP chips in the same samples. This high level of accuracy is approaching the expected error rate of SNP chips and makes it difficult to definitively determine whether our genotypes, or the SNP chip genotypes, are correct.

In addition, we have made these data publically available through the ADNI website. A multisample VCF file with mitochondrial genomic variants for each subject is available for download from the Download—Genetic Data section of the ADNI repository (http://adni.loni.usc.edu/data-samples/access-data/). Also included is an .XLSX file with mitochondrial haplotype information for each subject. The new data sets are named "ADNI WGS Whole mitochondrial genome variants" and "ADNI WGS Whole mitochondrial genome—Haplotypes". This data set is now ready to be applied in AD studies and to help elucidate the relationship between mitochondria and AD, which has thus far alluded researchers.

Complete mitochondrial genomic data result in high resolution of haplotype definition, including large numbers of singleton haplotypes. Methods that group haplotypes in evolutionarily meaningful ways are necessary to fully leverage these data. In our previous

analyses of complete mitochondrial genomics data, we incorporated TreeScanning to concentrate statistical power on evolutionarily meaningful groups of haplotypes [24,30,51,52]. This approach has been applied successfully in the study of both mitochondrial and nuclear genomic contributions to AD risk and related phenotypes [24,30,51,53,54].

The ADNI subjects include AD cases, subjects with mild cognitive impairment, and cognitively normal controls. Nearly all of these subjects are also associated with extensive clinical, imaging, and fluid biomarker data, including some longitudinal data. As such, they provide great value in evaluating the multitude of factors that are associated with AD, dementia, dementia progression, and conversion from mild cognitive impairment to dementia. While this study is underpowered to identify haplotypes associated with AD risk, we anticipate it will serve as the foundation for additional data collection and an expanded study in the future. These data will also prove valuable for validation and discovery analyses related to the multitude of other phenotypes that are available for these subjects. Association analyses using imaging and fluid biomarker data are outside the scope of this study, but we anticipate that these data will be leveraged by several groups, including ours, for these kinds of analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Querfurth HW, LaFerla FM. Alzheimer's disease. N Engl J Med. 2010; 362:329–44. [PubMed: 20107219]

2. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement. 2007; 3:186–91. [PubMed: 19595937]

3. Ridge PG, Ebbert MT, Kauwe JS. Genetics of Alzheimer's disease. Biomed Res Int. 2013; 2013:254954. [PubMed: 23984328]

4. Swerdlow RH, Burns JM, Khan SM. The Alzheimer's disease mitochondrial cascade hypothesis. J Alzheimers Dis. 2010; 20(Suppl 2):S265–79. [PubMed: 20442494]

5. Santos RX, Correia SC, Wang X, Perry G, Smith MA, Moreira PI, et al. Alzheimer's disease: diverse aspects of mitochondrial malfunctioning. Int J Clin Exp Pathol. 2010; 3:570–81. [PubMed: 20661404]

6. Cunnane S, Nugent S, Roy M, Courchesne-Loyer A, Croteau E, Tremblay S, et al. Brain fuel metabolism, aging, and Alzheimer's disease. Nutrition. 2011; 27:3–20. [PubMed: 21035308]

7. Wang X, Su B, Lee HG, Li X, Perry G, Smith MA, et al. Impaired balance of mitochondrial fission and fusion in Alzheimer's disease. J Neurosci. 2009; 29:9090–103. [PubMed: 19605646]

8. Podlesniy P, Llorens F, Golanska E, Sikorska B, Liberski P, Zerr I, et al. Mitochondrial DNA differentiates Alzheimer's disease from Creutzfeldt-Jakob disease. Alzheimers Dement. 2016; 12:546–55. [PubMed: 26806388]

9. Podlesniy P, Figueiro-Silva J, Llado A, Antonell A, Sanchez-Valle R, Alcolea D, et al. Low cerebrospinal fluid concentration of mitochondrial DNA in preclinical Alzheimer disease. Ann Neurol. 2013; 74:655–68. [PubMed: 23794434]

10. Chan DC. Mitochondria: dynamic organelles in disease, aging, and development. Cell. 2006; 125:1241–52. [PubMed: 16814712]

11. Du H, Guo L, Yan S, Sosunov AA, McKhann GM, Yan SS. Early deficits in synaptic mitochondria in an Alzheimer's disease mouse model. Proc Natl Acad Sci U S A. 2010; 107:18670–5. [PubMed: 20937894]

12. Anandatheerthavarada HK, Devi L. Amyloid precursor protein and mitochondrial dysfunction in Alzheimer's disease. Neuroscientist. 2007; 13:626–38. [PubMed: 17911214]

13. Devi L, Prabhu BM, Galati DF, Avadhani NG, Anandatheerthavarada HK. Accumulation of amyloid precursor protein in the mitochondrial import channels of human Alzheimer's disease brain is associated with mitochondrial dysfunction. J Neurosci. 2006; 26:9057–68. [PubMed: 16943564]

14. Xie H, Guan J, Borrelli LA, Xu J, Serrano-Pozo A, Bacskai BJ. Mitochondrial alterations near amyloid plaques in an Alzheimer's disease mouse model. J Neurosci. 2013; 33:17042–51. [PubMed: 24155308]

15. Gomez-Tortosa E, Barquero MS, Baron M, Sainz MJ, Manzano S, Payno M, et al. Variability of age at onset in siblings with familial Alzheimer disease. Arch Neurol. 2007; 64:1743–8. [PubMed: 18071037]

16. Edland SD, Silverman JM, Peskind ER, Tsuang D, Wijsman E, Morris JC. Increased risk of dementia in mothers of Alzheimer's disease cases: evidence for maternal inheritance. Neurology. 1996; 47:254–6. [PubMed: 8710088]

17. Seshadri S, Wolf PA, Beiser A, Au R, McNulty K, White R, et al. Lifetime risk of dementia and Alzheimer's disease. The impact of mortality on risk estimates in the Framingham Study. Neurology. 1997; 49:1498–504. [PubMed: 9409336]

18. Duara R, Barker WW, Lopez-Alberola R, Loewenstein DA, Grau LB, Gilchrist D, et al. Alzheimer's disease: interaction of apolipoprotein E genotype, family history of dementia, gender, education, ethnicity, and age of onset. Neurology. 1996; 46:1575–9. [PubMed: 8649551]

19. Mosconi L, Mistur R, Switalski R, Brys M, Glodzik L, Rich K, et al. Declining brain glucose metabolism in normal individuals with a maternal history of Alzheimer disease. Neurology. 2009; 72:513–20. [PubMed: 19005175]

20. Mosconi L, Rinne JO, Tsui WH, Berti V, Li Y, Wang H, et al. Increased fibrillar amyloid-{beta} burden in normal individuals with a family history of late-onset Alzheimer's. Proc Natl Acad Sci U S A. 2010; 107:5949–54. [PubMed: 20231448]

21. Honea RA, Swerdlow RH, Vidoni ED, Goodwin J, Burns JM. Reduced gray matter volume in normal adults with a maternal family history of Alzheimer disease. Neurology. 2010; 74:113–20. [PubMed: 20065246]

22. Honea RA, Swerdlow RH, Vidoni ED, Burns JM. Progressive regional atrophy in normal adults with a maternal history of Alzheimer disease. Neurology. 2011; 76:822–9. [PubMed: 21357834]

23. Honea RA, Vidoni ED, Swerdlow RH, Burns JM. Maternal family history is associated with Alzheimer's disease biomarkers. J Alzheimers Dis. 2012; 31:659–68. [PubMed: 22669011]

24. Ridge PG, Koop A, Maxwell TJ, Bailey MH, Swerdlow RH, Kauwe JS, et al. Mitochondrial haplotypes associated with biomarkers for Alzheimer's disease. PLoS One. 2013; 8:e74158. [PubMed: 24040196]

25. Swerdlow, RH. Mitochondria and Mitochondrial Cascades in Alzheimer's Disease. J Alzheimers Dis. [published online ahead of print October 7, 2017]. https://doi.org/10.3233/JAD-170585

26. Takasaki S. Mitochondrial haplogroups associated with Japanese Alzheimer's patients. J Bioenerg Biomembr. 2009; 41:407–10. [PubMed: 19795196]

27. Maruszak A, Canter JA, Styczynska M, Zekanowski C, Barcikowska M. Mitochondrial haplogroup H and Alzheimer's disease–is there a connection? Neurobiol Aging. 2009; 30:1749–55. [PubMed: 18308428]

28. Fesahat F, Houshmand M, Panahi MS, Gharagozli K, Mirzajani F. Do haplogroups H and U act to increase the penetrance of Alzheimer's disease? Cell Mol Neurobiol. 2007; 27:329–34. [PubMed: 17186363]

29. Santoro A, Balbi V, Balducci E, Pirazzini C, Rosini F, Tavano F, et al. Evidence for sub-haplogroup h5 of mitochondrial DNA as a risk factor for late onset Alzheimer's disease. PLoS One. 2010; 5:e12037. [PubMed: 20700462]

30. Ridge PG, Maxwell T, Corcoran C, Norton M, Tschanz J, O'Brien E, et al. Mitochondrial genomic analysis of late onset Alzheimer's disease reveals protective haplogroups H6A1A/H6A1B: the Cache County Study on Memory in Aging. PLoS One. 2012; 7:e45134. [PubMed: 23028804]

31. Carrieri G, Bonafe M, De Luca M, Rose G, Varcasia O, Bruni A, et al. Mitochondrial DNA haplogroups and APOE4 allele are nonindependent variables in sporadic Alzheimer's disease. Hum Genet. 2001; 108:194–8. [PubMed: 11354629]

32. van der Walt JM, Dementieva YA, Martin ER, Scott WK, Nicodemus KK, Kroner CC, et al. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. Neurosci Lett. 2004; 365:28–32. [PubMed: 15234467]

33. Lakatos A, Derbeneva O, Younes D, Keator D, Bakken T, Lvova M, et al. Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort. Neurobiol Aging. 2010; 31:1355–63. [PubMed: 20538375]

34. Zsurka G, Kalman J, Csaszar A, Rasko I, Janka Z, Venetianer P. No mitochondrial haplotype was found to increase risk for Alzheimer's disease. Biol Psychiatry. 1998; 44:371–3. [PubMed: 9755361]

35. Chinnery PF, Taylor GA, Howell N, Andrews RM, Morris CM, Taylor RW, et al. Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. Neurology. 2000; 55:302–4. [PubMed: 10908912]

36. Pyle A, Foltynie T, Tiangyou W, Lambert C, Keers SM, Allcock LM, et al. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann Neurol. 2005; 57:564–7. [PubMed: 15786469]

37. Mancuso M, Nardini M, Micheli D, Rocchi A, Nesti C, Giglioli NJ, et al. Lack of association between mtDNA haplogroups and Alzheimer's disease in Tuscany. Neurol Sci. 2007; 28:142–7. [PubMed: 17603766]

38. Kruger J, Hinttala R, Majamaa K, Remes AM. Mitochondrial DNA haplogroups in early-onset Alzheimer's disease and frontotemporal lobar degeneration. Mol Neurodegener. 2010; 5:8. [PubMed: 20181062]

39. Hudson G, Sims R, Harold D, Chapman J, Hollingworth P, Gerrish A, et al. No consistent evidence for association between mtDNA variants and Alzheimer disease. Neurology. 2012; 78:1038–42. [PubMed: 22442439]

40. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–60. [PubMed: 19451168]

41. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–8. [PubMed: 21478889]

42. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303. [PubMed: 20644199]

43. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981; 290:457–65. [PubMed: 7219534]

44. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999; 23:147. [PubMed: 10508508]

45. Behar DM, van Oven M, Rosset S, Metspalu M, Loogvali EL, Silva NM, et al. A "Copernican" reassessment of the human mitochondrial DNA tree from its root. Am J Hum Genet. 2012; 90:675–84. [PubMed: 22482806]

46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–9. [PubMed: 19505943]

47. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907. 2012

48. Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, et al. mtDNA variation and analysis using Mitomap and Mitomaster. Curr Protoc Bioinformatics. 2013; 44:1.23.1–1.23.26. [PubMed: 25489354]

49. Navarro-Gomez D, Leipzig J, Shen L, Lott M, Stassen AP, Wallace DC, et al. Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier. Bioinformatics. 2015; 31:1310–2. [PubMed: 25505086]

50. Mitchell SL, Goodloe R, Brown-Gentry K, Pendergrass SA, Murdock DG, Crawford DC. Characterization of mitochondrial haplogroups in a large population-based sample from the United States. Hum Genet. 2014; 133:861–8. [PubMed: 24488180]

51. Ridge PG, Maxwell TJ, Foutz SJ, Bailey MH, Corcoran CD, Tschanz JT, et al. Mitochondrial genomic variation associated with higher mitochondrial copy number: the Cache County Study on Memory Health and Aging. BMC Bioinformatics. 2014; 15(Suppl 7):S6.

52. Templeton AR, Maxwell T, Posada D, Stengard JH, Boerwinkle E, Sing CF. Tree scanning: a method for using haplotype trees in phenotype/genotype association studies. Genetics. 2005; 169:441–53. [PubMed: 15371364]

53. Li Y, Grupe A, Rowland C, Nowotny P, Kauwe JS, Smemo S, et al. DAPK1 variants are associated with Alzheimer's disease and allele-specific expression. Hum Mol Genet. 2006; 15:2560–8. [PubMed: 16847012]

54. Grupe A, Li Y, Rowland C, Nowotny P, Hinrichs AL, Smemo S, et al. A scan of chromosome 10 identifies a novel locus showing strong association with late-onset Alzheimer disease. Am J Hum Genet. 2006; 78:78–88. [PubMed: 16385451]

RESEARCH IN CONTEXT

1. Systematic review: We reviewed existing literature using PubMed and Google to identify data sets with whole mitochondrial genome sequences. Although several whole genome data sets have been generated for Alzheimer's disease, none of these specifically addressed the whole mitochondrial genomes, which require specialized analytic approaches because they are haploid.

2. Interpretation: Mitochondria have a role in Alzheimer's disease, but the precise role has thus far eluded researchers. A primary impediment is a lack sufficient data. Our goals were to (1) develop an approach for analyzing whole mitochondrial genomes; and (2) develop a data set of whole mitochondrial genomes and make them publicly available to encourage research in this important area.

3. Future directions: We anticipate that making this data set and analytical approach will facilitate additional studies of the effect the mitochondrial genome has on risk for Alzheimer's disease.

Table 1

Demographics

	Count	Sex (male/female)	Average age	APOE status ε2/ε2,ε2/ε3,ε3/ε3, ε3/ε4,ε4/ε4,ε2/ε4
Cases	191	126/65	74.42	0/8/74/80/25/4
Controls	279	135/144	74.51	0/35/167/68/7/2
MCI	333	183/149 *	71.57 *	1/26/162/110/25/9
Total	803	444/358 *	73.17 *	1/69/403/258/57/15

APOE, apolipoprotein E; MCI, mild cognitive impairment.

Demographic and phenotype information is available for 803 of the 809 mitochondrial genomes in the data set. APOE status refers to APOE genotype.

*
Missing data for one sample.

Table 2

Counts and frequencies of major mitochondrial haplogroups

Major mitochondrial haplogroup	Count	Frequency (%)	# Cases/controls/ MCI/unknown
A	5	0.62	3/1/1/0
B	7	0.87	3/1/3/0
C	3	0.37	0/0/3/0
F	1	0.12	1/0/0/0
H	338	41.78	73/128/133/4
I	27	3.34	4/12/11/0
J	69	8.53	24/18/26/1
K	68	8.41	19/24/25/0
L	33	4.08	5/16/12/0
M	11	1.36	2/4/5/0
N	7	0.87	2/2/3/0
R	5	0.62	1/3/1/0
T	78	9.64	16/25/37/0
U	105	12.98	27/31/46/1
V	28	3.46	4/10/14/0
W	15	1.85	3/2/10/0
X	9	1.11	4/2/3/0

MCI, mild cognitive impairment.

Numbers refer to individuals who have the listed major mitochondrial haplogroup or a subgroup (e.g. individuals with H5 are counted as part of the H group in this table).