Published in final edited form as:

Hum Genet. 2014 July ; 133(7): 861–868. doi:10.1007/s00439-014-1421-9.

Characterization of mitochondrial haplogroups in a large population-based sample from the United States

Sabrina L. Mitchell¹, Robert Goodloe¹, Kristin Brown-Gentry¹, Sarah A. Pendergrass², Deborah G. Murdock¹, and Dana C. Crawford¹

¹Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN

²Center for Systems Genomics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, Eberly College of Science, The Huck Institutes of the Life Sciences, University Park, Pennsylvania

Abstract

Mitochondrial DNA (mtDNA) haplogroups are valuable for investigations in forensic science, molecular anthropology, and human genetics. In this study, we developed a custom panel of 61 mtDNA markers for high-throughput classification of European, African, and Native American/Asian mitochondrial haplogroup lineages. Using these mtDNA markers we constructed a mitochondrial haplogroup classification tree and classified 18,832 participants from the National Health and Nutrition Examination Surveys (NHANES). To our knowledge, this is the largest study to date characterizing mitochondrial haplogroups in a population-based sample from the United States, and the first study characterizing mitochondrial haplogroup distributions in self-identified Mexican Americans separately from Hispanic Americans of other descent. We observed clear differences in the distribution of maternal genetic ancestry consistent with proposed admixture models for these subpopulations, underscoring the genetic heterogeneity of the United States Hispanic population. The mitochondrial haplogroup distributions in the other self-identified racial/ethnic groups within NHANES were largely comparable to previous studies. Mitochondrial haplogroup classification was highly concordant with self-identified race/ethnicity (SIRE) in non-Hispanic whites (94.8%), but was considerably lower in admixed populations including non-Hispanic blacks (88.3%), Mexican Americans (81.8%), and other Hispanics (61.6%), suggesting SIRE does not accurately reflect maternal genetic ancestry, particularly in populations with greater proportions of admixture. Thus, it is important to consider inconsistencies between SIRE and genetic ancestry when performing genetic association studies. The mitochondrial haplogroup data that we have generated, coupled with the epidemiologic variables in NHANES, is a valuable resource for future studies investigating the contribution of mtDNA variation to human health and disease.

Keywords

mitochondrial haplogroups; NHANES; mitochondrial genetic variation; Sequenom; multiplex genotyping

Introduction

Mitochondria, the primary energy producers of the cell, are unique organelles that maintain their own DNA (mtDNA) separate from the nucleus. The human mitochondrial genome, a double stranded, circular chromosome spanning approximately 16.6 kilobases, encodes 22 tRNAs, 2 rRNAs, and 13 proteins that are part of the respiratory chain complexes. This compact genome is maternally inherited, has a high mutation rate, and does not undergo recombination. Mitochondria are essential to cellular metabolism, and mtDNA variation has been associated with multiple complex diseases including cancer (Canter et al. 2005; Wallace 2012; Zheng et al. 2012), type 2 diabetes (Wang et al. 2001; Crispim et al. 2006; Tang et al. 2006; Soini et al. 2012), and neurological disorders such as Parkinson's and Alzheimer's (van der Walt et al. 2003; van der Walt et al. 2004). Additionally, epidemiologic studies indicate that some common complex diseases, such as diabetes (Alcolado and Alcolado 1991; Thomas et al. 1994; Young et al. 1995) and preterm birth (Alleman et al. 2012), exhibit strong maternal inheritance supporting the notion that mitochondrial genome variation could contribute to underlying disease etiology.

Mitochondrial haplogroups are collections of similar haplotypes defined by combinations of single nuclear polymorphisms (SNPs) in mtDNA inherited from a common ancestor. These haplogroups are formed as a result of the sequential accumulation of mutations through maternal lineages (Wallace 2013). Common complex diseases have been associated with specific mitochondrial haplogroups, including Alzheimer's disease with haplogroup U (van der Walt et al. 2004) and age-related macular degeneration with the J1U haplogroup cluster (Kenney et al. 2013). Due to population migration, distinct lineages of mtDNA are associated with major global groups including African, European, Native American/Asian, & Oceanic/Aboriginal Australian (Forster 2004; Wallace 2013) raising the possibility that mtDNA variation could contribute to the differences in disease prevalence observed among racial/ethnic groups.

Despite strong evidence that mtDNA variation plays a role in the development and progression of complex human diseases, mitochondrial genetic variation has been largely ignored in the context of genome-wide association studies (GWAS), highlighting the need for studies investigating the contribution of mtDNA variation to human health and disease. Single mtDNA variants can be used in association studies; however the use of mitochondrial haplogroups is also an important way to characterize the effect of mtDNA variation on normal trait variance or disease phenotypes. A consistent lack of adequate mitochondrial markers on GWAS fixed content arrays, along with poor characterization of mitochondrial haplogroups contribute to the general paucity of mitochondrial genetic association discovery. In the present study, we identify a set of mitochondrial markers for high-throughput classification of mitochondrial haplogroups and use these markers to

characterize the mitochondrial haplogroups in one of the largest collections of population-based samples from the United States creating a valuable resource for future genotype-phenotype correlations.

Materials & Methods

Study Population

The National Health and Nutrition Examination Surveys (NHANES) are large, population-based surveys conducted in the United States by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). Each survey collects demographic, lifestyle, and health data from participants using questionnaires, standardized medical examinations, and clinical laboratory measures such as levels of cholesterol, triglycerides, hormones, and other biomarkers. Additionally, DNA is collected for a subset of survey participants for Genetic NHANES. All study procedures were approved by the CDC's Ethics Review Board and written informed consent was obtained from all study participants.

We accessed DNA samples from NHANES III, NHANES 1999–2002, and NHANES 2007–2008, for a total of 19,608 samples. Each of these surveys employed a complex study design that oversampled from minority populations. NHANES III was conducted in two three-year phases initiating in 1988 and 1991; DNA sample acquisition began in phase II (1991–1994) and a total of 7,159 DNA samples from participants greater than or equal to 12 years of age are available for study. Beginning in 1999, NHANES became a continuous survey conducted annually. A total of 7,838 and 4,611 DNA samples from participants aged 20 years or older are available for NHANES 1999–2002 and NHANES 2007–2008, respectively. No identifying information is available to study investigators; therefore studies accessing NHANES are considered non-human subjects research by Vanderbilt University's Institutional Review Board (IRB #061062; HS2 committee).

The self-identified racial/ethnic groups included in NHANES are non-Hispanic white, non-Hispanic black, Mexican American, other Hispanic, and other race. The “other Hispanic” category was created in the transition from NHANES III to continuous NHANES and includes participants of Hispanic descent who did not self-identify as Mexican American, including Hispanic blacks. As there was not a separate category for other Hispanics in NHANES III, participants of Hispanic descent who did not self-identify as Mexican American were included in the “other race” category which also included those participants that self-identified with racial groups other than “white” or “black”. As a result, the “other race” category is not comparable between NHANES III and continuous NHANES. Therefore in the present study, data for those self-identified as “other race” in NHANES III were analyzed separately.

Each survey employs a complex sampling design that oversamples minorities, therefore, racial/ethnic distribution differs across the NHANES III, NHANES 1999–2002, and 2007–2008 data sets (Table 1). The percent of other Hispanic samples in NHANES 2007–2008 (11%) is more than twice that in NHANES 1999–2002 (5%). This increase, along with the decrease in the percent of samples self-identified as Mexican American between these

surveys, reflects the change in the sampling scheme that occurred in the 2007–2008 survey where all Hispanics were oversampled, not just Mexican Americans (http://www.cdc.gov/nchs/nhanes/nhanes2007-2008/generaldoc_e.htm).

SNP Selection and Genotyping

Using Phylotree, a comprehensive phylogenetic tree of global mitochondrial DNA variation (van Oven and Kayser 2009), as well as an extensive literature search on PubMed for prior studies with haplogroup classification (Herrnstadt et al. 2002; van der Walt et al. 2003; Poole et al. 2010; Paneto et al. 2011), a total of 63 mitochondrial SNPs (mtSNPs) were selected for genotyping to classify the major self-identified racial/ethnic groups within NHANES (non-Hispanic white, non-Hispanic black, and Mexican American) into haplogroups (Table S1). European (H, I, J, K, T, U, V, W, & X), African (L0, L1, L2, L3, & L4), and Native American/Asian (A, B, C, & D) mitochondrial haplogroup-defining SNPs were chosen. Those SNPs appearing only once in Phylotree, and thus defining a single haplogroup, were preferentially selected.

DNA for NHANES III is distributed as crude cell lysates, while purified DNA aliquots are distributed for NHANES 1999–2002 and 2007–2008 samples. SNPs were genotyped using the Sequenom iPLEX[®] Gold MassArray platform by the Center for Human Genetics Research Open Wet Lab and DNA Resources Core at Vanderbilt University. Three SNP pool assays were designed using Sequenom MassARRAY software, multiplex primer extension was performed and the extension products were analyzed via MALDI-TOF mass spectrometry (Tang et al. 2004).

All genotype data generated on NHANES DNA samples were returned to the CDC and deposited into the Genetic NHANES database where it is available for secondary analyses via NCHS application. In addition to experimental NHANES DNA, we genotyped blinded duplicates provided by the CDC for concordance checks and quality control. SNPs that are greater than 5% discordant among duplicate samples are rejected by the CDC and not deposited into the Genetic NHANES database. One SNP in NHANES III and five SNPs in NHANES 1999–2002 did not meet the duplicate concordance required by the CDC, and therefore were excluded. We also performed internal quality control separately for each NHANES data set. Two SNPs were excluded in all three data sets and an additional three SNPs were excluded in the NHANES III and 1999–2002 data sets due to low call rate (less than 95%). Overall, 57 SNPs for NHANES III, 53 SNPs for NHANES 1999–2002, and 61 SNPs for NHANES 2007–2008 were available for haplogroup classification (Supplemental table 1).

Mitochondrial Haplogroup classification

A simplified mitochondrial classification tree, rooted at the most recent common ancestor (MRCA), was created based on the mitochondrial SNPs genotyped (Figure 1). We developed an algorithm to automate mitochondrial haplogroup classification and all resultant classifications were manually reviewed for accuracy. In this study, we do not present the data for individual haplogroup subclades, but collapse participants classified in a subclade into the parent haplogroup. For example, participants in haplogroup A include

those classified as 'A', 'A4', and 'A2'. Overall, approximately 4% of samples were not classified due to missing genotype data. NHANES III had the highest proportion of unclassified samples (6.7%) likely due to lower quality of DNA compared with the continuous NHANES samples of which fewer than 3% were unclassified.

Results

Mitochondrial haplogroup classifications

Mitochondrial haplogroups were determined for 96% of participants in the three NHANES data sets. As expected, the distribution of haplogroup ancestries was similar to the distribution of self-identified race/ethnicities within each survey (Figure 2). Among the total self-identified non-Hispanic whites classified ($n=8,537$) across the three surveys, 94.8% of mitochondrial haplogroups were of European origin (Figure 3). Not unexpectedly, the most common haplogroup observed among non-Hispanic whites was haplogroup H with an average frequency of 43.0% across the surveys (Table 2). Haplogroup U was the second most common haplogroup observed among non-Hispanic whites with an average frequency of 13.6% across the surveys. The average frequencies of haplogroups T, J, and K were 9.6%, 9.2%, and 8.4%, respectively. Each of the remaining European haplogroups (I, V, W, X, HV, and JT) were observed among fewer than 5% of non-Hispanic whites. The Native American/Asian haplogroup lineages and African haplogroups were observed at combined average frequencies of 1.1% and 1.0%, respectively, among self-identified non-Hispanic whites.

Among the total self-identified non-Hispanic blacks classified ($n = 4,078$) across the three surveys, 88.3% of mitochondrial haplogroups observed were of African lineages (Figure 3). The most frequently observed haplogroup among non-Hispanic blacks was L3 with an average frequency of 33.9% across the three surveys, followed by L2 (29.9%), L1 (19.0%), L0 (4.3%), and L4 (1.2%). The European and Native American/Asian lineage haplogroups were observed at combined average frequencies of 6.3% and 1.7%, respectively, among self-identified non-Hispanic blacks.

Among the total self-identified Mexican Americans classified ($n=4,623$) across the three NHANES surveys, 81.8% of mitochondrial haplogroups observed were of Native American/Asian origin (Figure 3). Haplogroup A was the most commonly observed haplogroup among Mexican Americans, with a combined average frequency of 34.2% across the surveys, followed by haplogroups B (20.1%), C (20.0%), and D (7.5%). Haplogroups of European and African origin were observed among self-identified Mexican Americans at combined average frequencies of 13.2% and 4.4%, respectively.

Among the total classified from those self-identified as other Hispanic in continuous NHANES ($n=921$), 61.1% of haplogroups were of Native American/Asian origin, while 18.0% and 19.3% were of European and African descent, respectively (Figure 3). While the distribution of ancestries differed between the Mexican Americans and other Hispanics the relative frequencies of Native American haplogroups were similar between the two groups, with haplogroup A being the most common.

Among the total classified from those self-identified as “other race” (n=311) in NHANES III, approximately one-third (33.8%) of mitochondrial haplogroups were of Native American/Asian origin, 27.3% were of African origin, and 16.1% were of European origin (Figure 3). Contrastingly, among those self-identified as “other race” (n=362) in the combined continuous NHANES data sets only 19.3% of mitochondrial haplogroups were of Native American/Asian origin, while 22.4% and 9.4% were of European and African origin, respectively (Figure 3).

Approximately half (48.9%), of those who self-identified as “other race” in continuous NHANES, could only be classified at the resolution of Eurasian macrohaplogroup (M, N, or R). This is not unexpected given that SNPs were selected to define mitochondrial haplogroups of European, African, and Native American origin. It is likely these participants have mtDNA lineages not captured by the SNPs genotyped, such as East Asian haplogroups F, G, and subclades of M.

Discussion

This is the largest study to date characterizing mitochondrial haplogroups in a population-based sample from the U.S. In this study we developed a custom mitochondrial SNP panel for high throughput classification of European, African, and Native American/Asian mitochondrial haplogroup lineages for eventual genotype-phenotype association studies. Using these markers, a simplified mitochondrial phylogenetic tree was constructed for haplogroup classification. Applying this method we determined the mitochondrial haplogroups for 18,832 participants from the NHANES III, 1999–2002, and 2007–2008 data sets.

The distributions of mitochondrial haplogroup ancestries among non-Hispanic whites and non-Hispanic blacks within NHANES are comparable to previous reports, including those from the Scientific Working Group on DNA Analysis Methods (SWGDM) for Caucasians (n = 1402) and African Americans (n = 1148) (Allard et al. 2002; Allard et al. 2005). One notable exception was the identification of haplogroups L0 and L4 among the NHANES non-Hispanic blacks, as these haplogroups were not observed among the SWGDAM African American samples. This difference in mitochondrial haplogroup distributions between NHANES and SWGDAM data sets is most likely due to differences in sample size and possibly geographic origin of samples.

The U.S. Hispanic population is a genetically heterogeneous group that includes individuals originating from both Latin America and the Iberian Peninsula. Based on data from the 2011 American Community Survey conducted by the U.S. Census Bureau, Hispanics make up nearly 17% of the U.S. population, and approximately two-thirds of U.S. Hispanics originated from Mexico (http://www.pewhispanic.org/2013/02/15/hispanic-population-trends/ph_13-01-23_ss_hispanics1/). In general, previous studies characterizing mitochondrial haplogroup distributions in U.S. Hispanics have not differentiated between individuals of Mexican descent and Hispanics originating from other countries (Merriwether et al. 1997; Allard et al. 2006; Bertoni et al. 2011; Levran et al. 2012). In this study, we show there is a difference in the distribution of mitochondrial haplogroup ancestries between

self-identified Mexican Americans and other Hispanics in NHANES. The distribution of haplogroup ancestries in the NHANES Mexican Americans is very similar to those reported in a recent study by Kumar et al, in which the mitochondrial genomes of 384 Mexican Americans were sequenced (Kumar et al. 2011). In contrast, the distribution of mitochondrial haplogroup ancestries in the SWGDAM U.S. Hispanics (Allard et al. 2006) differs from both the NHANES Mexican Americans and other Hispanics. The SWGDAM Hispanic Americans (n = 686) have a lower proportion of Native American ancestral haplogroups compared with the NHANES Mexican Americans and a higher proportion compared with the NHANES other Hispanics. However, when the southwestern U.S. Hispanics (n = 245) in the SWGDAM are examined separately, the mitochondrial haplogroup distributions are comparable to the NHANES Mexican Americans. Likewise, the distribution of mitochondrial haplogroup ancestries in the remaining SWGDAM U.S. Hispanics (n=441) is very similar to the NHANES other Hispanics. This makes sense given that the SWGDAM southwestern U.S. Hispanics were from Texas and New Mexico and therefore most likely of Mexican descent. The observed differences are consistent with proposed admixture models for these populations in which Mexican Americans are primarily the result of a two way admixture event between Native Americans and Europeans, and to a lesser degree Africans, while Hispanics of other origins exhibit a tri-hybrid ancestry including Native American, European, and African contributions (Hanis et al. 1991; Bertoni et al. 2011; Kumar et al. 2011).

Overall, we observed high concordance between self-identified race/ethnicity and mitochondrial ancestral lineage for the non-Hispanic white group. As expected, concordance was lower among non-Hispanic blacks, Mexican Americans, and the other Hispanics, consistent with the higher percentage of population admixture typically observed in these racial/ethnic groups. Thus, while relatively few non-Hispanic whites (<3%) would be reclassified based on mitochondrial haplogroup classification, a larger percentage of non-Hispanic blacks (8.0%), Mexican Americans (17.5%), and other Hispanics (37.3%) would be reclassified (Table S2). This suggests that self-identified race/ethnicity may not accurately reflect maternal genetic ancestry particularly in populations with greater proportions of admixture, and thus is not a reliable surrogate for genetic ancestry.

While there are a number of studies which describe the distribution of mitochondrial haplogroups for U.S. Hispanics, to our knowledge this is the first study characterizing the distribution of mitochondrial haplogroups in self-identified Mexican Americans separately from self-identified other Hispanics in the U.S. The differences observed in maternal ancestry between these two groups highlight the importance of including mitochondrial variation in genetic studies to adjust for population stratification. The mitochondrial SNP panels designed in this study can be used to capture the maternal contribution to admixture and would complement the current practice of characterizing admixture using only nuclear ancestry informative markers.

One limitation of this study is the resolution of mitochondrial haplogroup classification that can be achieved with the SNPs genotyped. In this study we performed targeted genotyping of mitochondrial variants that allows samples to be broadly classified into haplogroups, with relatively few SNPs that define subclades of these major haplogroups. While sequencing the

entire mitochondrial genome would provide much finer resolution, it is expensive and therefore is generally not feasible for large-scale studies such as this one.

The unique characteristics of mtDNA, including high copy number, high mutation rate, and maternal inheritance, make this genome useful for a variety of applications including molecular anthropology and forensic science, as well as human and medical genetics. The inclusion of mitochondrial variants in genetic studies is essential for understanding the role mitochondrial genetic variation plays in human health and disease. We have generated a valuable resource that, coupled with the epidemiologic and clinical lab variables available in NHANES, can be used for future studies investigating genotype-phenotype correlations. These mtDNA variants can be used in conjunction with nuclear SNPs to adjust for population stratification in association studies involving the nuclear genome as well as in direct investigations of the effect of mitochondrial genetic variation on specific phenotypes, and analysis of nuclear-mitochondrial genome interactions. Such studies will help uncover the contribution of mitochondrial genetic variation to normal trait variance as well as to complex disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Genotyping in NHANES was supported in part by the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study (U01HG004798 and its ARRA supplements) as part of the Population Architecture using Genomics and Epidemiology (PAGE) study established by the National Human Genome Research Institute (NHGRI). We, at EAGLE, would like to thank Melissa Allen and Paxton Baker for their work in genotyping the samples through the Center for Human Genetic Research Open Wet Lab (OWL) and DNA Resources Core, respectively. We would also like to thank Dr. Geraldine McQuillan and Jody McLean for their help in accessing the Genetic NHANES data. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institutes for Health or the Centers for Disease Control and Prevention.

References

- Alcolado JC, Alcolado R. Importance of maternal history of non-insulin dependent diabetic patients. *BMJ*. 1991; 302:1178–1180. [PubMed: 2043812]
- Allard MW, Miller K, Wilson M, Monson K, Budowle B. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. Scientific Working Group on DNA Analysis Methods. *J Forensic Sci*. 2002; 47:1215–1223. [PubMed: 12455642]
- Allard MW, Polanskey D, Miller K, Wilson MR, Monson KL, Budowle B. Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. *Forensic Sci Int*. 2005; 148:169–179. [PubMed: 15639612]
- Allard MW, Polanskey D, Wilson MR, Monson KL, Budowle B. Evaluation of Variation in Control Region Sequences for Hispanic Individuals in the SWGDAM mtDNA Data Set. *J Forensic Sci*. 2006; 51:566–573. [PubMed: 16696703]
- Alleman BW, Myking S, Ryckman KK, Myhre R, Feingold E, Feenstra B, Geller F, Boyd HA, Shaffer JR, Zhang Q, Begum F, Crosslin D, et al. No observed association for mitochondrial SNPs with preterm delivery and related outcomes. *Pediatr Res*. 2012; 72:539–544. [PubMed: 22902432]

- Bertoni B, Budowle B, Sans M, Barton S, Chakraborty R. Admixture in Hispanics: Distribution of Ancestral Population Contributions in the Continental United States. *Hum Biol.* 2011; 75
- Canter JA, Kallianpur AR, Parl FF, Millikan RC. Mitochondrial DNA G10398A Polymorphism and Invasive Breast Cancer in African-American Women. *Cancer Res.* 2005; 65:8028–8033. [PubMed: 16140977]
- Crispim D, Canani LH, Gross JL, Tschiedel B, Souto KEP, Roisenberg I. The European-Specific Mitochondrial Cluster J/T Could Confer an Increased Risk of Insulin-Resistance and Type 2 Diabetes: An Analysis of the m.4216T > C and m.4917A > G Variants. *Ann Hum Genet.* 2006; 70:488–495. [PubMed: 16759180]
- Forster P. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc B Biol Sci.* 2004; 359:255–264.
- Hanis CL, Hewett-Emmett D, Bertin TK, Schull WJ. Origins of U.S. Hispanics. Implications for diabetes. *Diabetes Care.* 1991; 14:618–627. [PubMed: 1914811]
- Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N. Reduced-median-network analysis of complete mitochondrial DNA coding region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet.* 2002; 70:1152–1171. [PubMed: 11938495]
- Kenney MC, Hertzog D, Chak G, Atilano SR, Khatibi N, Soe K, Nobe A, Yang E, Chwa M, Zhu F, Memarzadeh M, King J, et al. Mitochondrial DNA haplogroups confer differences in risk for age-related macular degeneration: a case control study. *BMC Med Genet.* 2013; 14:4. [PubMed: 23302509]
- Kumar S, Bellis C, Zlojutro M, Melton PE, Blangero J, Curran JE. Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC Evol Biol.* 2011; 11:293. [PubMed: 21978175]
- Levrin O, Awolesi O, Shen P-H, Adelson M, Kreek MJ. Estimating ancestral proportions in a multi-ethnic US sample: implications for studies of admixed populations. *Hum Genomics.* 2012; 6:2. [PubMed: 23244743]
- Merriwether DA, Huston S, Iyengar S, Hamman R, Norris JM, Shetterly SM, Kamboh MI, Ferrell RE. Mitochondrial versus nuclear admixture estimates demonstrate a past history of directional mating. *Am J Phys Anthropol.* 1997; 102:153–159. [PubMed: 9066897]
- Paneto GG, Köhnemann S, Martins JA, Cicarelli RMB, Pfeiffer H. A single multiplex PCR and SNaPshot minisequencing reaction of 42 SNPs to classify admixture populations into mitochondrial DNA haplogroups. *Mitochondrion.* 2011; 11:296–302. [PubMed: 21172459]
- Poole JC, Procaccio V, Brandon MC, Merrick G, Wallace DC. Multiplex analysis of mitochondrial DNA pathogenic and polymorphic sequence variants. *Biol Chem.* 2010; 391:1115–1130. [PubMed: 20707610]
- Soini HK, Moilanen JS, Finnilä S, Majamaa K. Mitochondrial DNA sequence variation in Finnish patients with matrilineal diabetes mellitus. *BMC Res Notes.* 2012; 5:350. [PubMed: 22780954]
- Tang D-L, Zhou X, Li X, Zhao L, Liu F. Variation of mitochondrial gene and the association with type 2 diabetes mellitus in a Chinese population. *Diabetes Res Clin Pract.* 2006; 73:77–82. [PubMed: 16414144]
- Tang K, Oeth P, Kammerer S, Denissenko MF, Ekblom J, Jurinke C, van den Boom D, Braun A, Cantor CR. Mining disease susceptibility genes through SNP analyses and expression profiling using MALDI-TOF mass spectrometry. *J Proteome Res.* 2004; 3:218–227. [PubMed: 15113097]
- Thomas F, Balkau B, Vauzelle-Kervroedan F, Papoz L. Maternal effect and familial aggregation in NIDDM. The CODIAB Study. CODIAB-INSERM-ZENECA Study Group. *Diabetes.* 1994; 43:63–67. [PubMed: 8262318]
- van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009; 30:E386–E394. [PubMed: 18853457]
- van der Walt JM, Nicodemus KK, Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Haines JL, Koller WC, Lyons K, Pahwa R, Stern MB, et al. Mitochondrial Polymorphisms Significantly Reduce the Risk of Parkinson Disease. *Am J Hum Genet.* 2003; 72:804–811. [PubMed: 12618962]
- van der Walt JM, Dementieva YA, Martin ER, Scott WK, Nicodemus KK, Kroner CC, Welsh-Bohmer KA, Saunders AM, Roses AD, Small GW, Schmechel DE, Murali Doraiswamy P, et al. Analysis

- of European mitochondrial haplogroups with Alzheimer disease risk. *Neurosci Lett*. 2004; 365:28–32. [PubMed: 15234467]
- Wallace DC. Mitochondria and cancer. *Nat Rev Cancer*. 2012; 12:685–698. [PubMed: 23001348]
- Wallace DC. Bioenergetics in human evolution and disease: implications for the origins of biological complexity and the missing genetic variation of common diseases. *Philos Trans R Soc B Biol Sci*. 2013; 368(1622) 20120267.
- Wang D, Taniyama M, Suzuki Y, Katagiri T, Ban Y. Association of the mitochondrial DNA 5178A/C polymorphism with maternal inheritance and onset of type 2 diabetes in Japanese patients. *Exp Clin Endocrinol Diabetes Off J Ger Soc Endocrinol Ger Diabetes Assoc*. 2001; 109:361–364.
- Young CA, Kumar S, Young MJ, Boulton AJM. Excess maternal history of diabetes in Caucasian and Afro-origin non-insulin-dependent diabetic patients suggests dominant maternal factors in disease transmission. *Diabetes Res Clin Pract*. 1995; 28:47–49. [PubMed: 7587912]
- Zheng S, Qian P, Li F, Qian G, Wang C, Wu G, Li Q, Chen Y, Li J, Li H, He B, Ji F. Association of Mitochondrial DNA Variations with Lung Cancer Risk in a Han Chinese Population from Southwestern China. *PLoS One*. 2012; 7:e31322. [PubMed: 22363619]

Figure 1. Mitochondrial haplogroup classification tree

Simplified mitochondrial phylogenetic tree for classification of European (blue), African (red), and Native American/Asian (green) mitochondrial haplogroup lineages; Eurasian macrohaplogroups are in orange and the checkered boxes indicate haplogroups that were included in counts of the corresponding parent Eurasian macrohaplogroup. mtSNPs in bold are found in a single haplogroup and those followed by (!) indicate reversions.

Figure 2. Distribution of mitochondrial haplogroup lineages by survey

Data include 19,608 samples genotyped from NHANES III, NHANES 1999–2002, and NHANES 2007–2008.

Figure 3. Distribution of mitochondrial haplogroup lineages by self-identified race/ethnicity

Data include only the total classified within each self-identified race/ethnicity.

Abbreviations: Cont. NH (Continuous NHANES; includes 1999–2002 and 2007–2008 data sets), NHIII (NHANES III), NHW (non-Hispanic white), NHB (non-Hispanic black), MA (Mexican American)

Table 1
Distribution of self-identified race/ethnicity by survey

Counts and percentages are given for each Genetic NHANES, unweighted by survey design.

	NHANES III	NHANES 1999–2002	NHANES 2007–2008
Non-Hispanic white	2,631 (37%)	4,003 (51%)	2,222 (48%)
Non-Hispanic black	2,108 (29%)	1,350 (17%)	867 (19%)
Mexican American	2,073 (29%)	1,876 (24%)	817 (18%)
Other Hispanic	-	418 (5%)	526 (11%)
Other race	347 (5%)	191 (2%)	179 (4%)
Total	7,159	7,838	4,611

Table 2

Distribution of mitochondrial haplogroups by self-identified race/ethnicity

Percentages were calculated based on the total number classified for each self-identified racial/ethnic group.

Mitochondrial Haplogroup	Non-Hispanic White	Non-Hispanic Black	Mexican American	Other Hispanic	Other race (NHII)	Other race (Cont. NH)
A	31 (0.4%)	31 (0.8%)	1,581 (34.2%)	282 (30.6%)	52 (16.7%)	16 (4.4%)
B	21 (0.2%)	17 (0.4%)	927 (20.1%)	123 (13.4%)	17 (5.5%)	16 (4.4%)
C	38 (0.4%)	14 (0.3%)	926 (20.0%)	118 (12.8%)	21 (6.8%)	10 (2.8%)
D	11 (0.1%)	8 (0.2%)	346 (7.5%)	40 (4.3%)	15 (4.8%)	28 (7.7%)
H	3,668 (43.0%)	112 (2.7%)	244 (5.3%)	62 (6.7%)	21 (6.8%)	33 (9.1%)
I	197 (2.3%)	8 (0.2%)	28 (0.6%)	*	*	*
J	789 (9.2%)	18 (0.4%)	64 (1.4%)	21 (2.3%)	*	*
K	715 (8.4%)	17 (0.4%)	106 (2.3%)	13 (1.4%)	6 (1.9%)	7 (1.9%)
T	823 (9.6%)	16 (0.4%)	42 (0.9%)	21 (2.3%)	*	10 (2.8%)
U	1,160 (13.6%)	57 (1.4%)	74 (1.6%)	35 (3.8%)	8 (2.6%)	17 (4.7%)
V	223 (2.6%)	10 (0.2%)	17 (0.4%)	*	*	*
W	161 (1.9%)	*	6 (0.1%)	*	*	*
X	143 (1.7%)	*	6 (0.1%)	5 (0.5%)	*	*
HV	133 (1.6%)	9 (0.2%)	15 (0.3%)	*	5 (1.6%)	*
JT	83 (1.0%)	*	*	*	*	*
L0	*	174 (4.3%)	9 (0.2%)	6 (0.7%)	5 (1.6%)	*
L1	8 (0.1%)	775 (19.0%)	34 (0.7%)	56 (6.1%)	22 (7.1%)	*
L2	28 (0.3%)	1,221 (29.9%)	82 (1.8%)	44 (4.8%)	24 (7.7%)	9 (2.5%)
L3	50 (0.6%)	1,382 (33.9%)	77 (1.7%)	72 (7.8%)	34 (10.9%)	22 (6.1%)
L4	*	47 (1.2%)	*	*	*	*
M	20 (0.2%)	30 (0.7%)	9 (0.2%)	*	29 (9.3%)	100 (27.6%)
N	200 (2.3%)	100 (2.5%)	15 (0.3%)	8 (0.9%)	8 (2.6%)	50 (13.8%)
R	29 (0.3%)	21 (0.5%)	8 (0.2%)	*	34 (10.9%)	27 (7.5%)
Total	8,537	4,078	4,623	921	311	362

* indicates fewer than 5 observations; per our data use agreement with CDC, data are suppressed to maintain study participant confidentiality.

Abbreviations: NHIII (NHANES III), Cont. NH (Continuous NHANES; includes 1999–2002 and 2007–2008 data sets)