

Documentation for the Sequence of Mitochondrial DNA Dataset

Dr. Greg Tranah, Health ABC Ancillary Study #AS06-84

1. General description and dataset structure 2

2. Methods..... 3

3. mtDNA Sequencing Details 3

4. References 4

1. General description and dataset structure

138 European Ancestry participants who had been part of the Energy Expenditure substudy were identified for Dr. Greg Tranah's Health ABC Ancillary Study #AS06-84. mtDNA was isolated from platelets and the entire mtDNA was sequenced from these participants to identify sequence variants related to resting metabolic rate, total energy expenditure, and activity energy expenditure. Platelets were collected during year 2. Energy Expenditure measurements were performed during years 2, 3, and 10.

Two data files have been released. The larger file (SeqMitochondria_full.csv) contains the sequence data as it was delivered by the laboratory. Use of this dataset requires special knowledge of how mtDNA works. Specifically, all mtDNA sequence positions are assigned as they align to a universal reference sequence (rCRS; NC_012920.1). Unfortunately this universal reference sequence has two extremely rare insertions/deletions (311c-315c and 3107 deletion). When analyzing mtDNA sequences it is important to account for these insertions/deletions when assigning nucleotide positions. So, anyone analyzing this dataset will have to take these shifts into account as well. The second smaller file (SeqMitochondria_abridged.csv) is the one that is most useful for analysis. This file only contains data for the variable (polymorphic sites) and takes the insertions/deletions into account. This is most likely the file to be used for any future analyses. All of Dr. Tranah's Health ABC-mtDNA papers have been (and will be) based on the data in this file.

The data files are csv files containing rows of nucleotide calls for each mitochondria polymorphic site, and columns for each HABCID. For example see below a small sample of the dataset structure

mtDNA	1001	1041	1046	1078	1084	1088	1093	1110	1111	1139	1142	1153
114	c	c	c	c	c	c	c	c	c	c	c	c
119	t	t	t	t	t	t	t	t	t	t	t	t
146	t	t	t	t	t	t	t	t	t	t	t	t

Where mtDNA 114, 119, 146, etc. are the polymorphic sites in the Mitochondria DNA and 1001, 1041, 1046, etc. are the HABCIDs. Nucleotide calls include a, g, t, c or n which refers to a "failed call" for each polymorphic site.

Missing values for the nucleotide calls are coded as "n."

The two data files are in a self-extracting zip file that can be downloaded from the "Current Datasets" listing on the Health ABC website under the "Datasets/Documentation" link. This zip file is called:

SeqMitochondria.exe

Key variables:

HABCID HABC Enrollment ID without the 2-letter prefix

To use the data, please contact the PI at your site.

2. Methods

mtDNA extraction protocol performed in the lab of Pamela Paris, University of California, San Francisco.

1. Thaw frozen platelets at 37°C for ~3 min.
2. Spin to acquire platelet pellet @ 16,100g for 3 min (cryo tube).
 - Transfer supernatant into clean cryo tubes (with original barcode sticker added)
 - Store @ -80°C
3. Add 200uL of Cell Lysis Solution (Qiagen Puregene Tissue Kit) to platelet pellet (in original cryo tube).
4. Vortex for 20 seconds.
5. Add 1.0 µl Proteinase K Solution (Qiagen Puregene Tissue Kit) and incubate at 55°C for 1 hour.
6. Add 1.5uL of RNase A Solution (Qiagen Puregene Tissue Kit) to cell lysate.
7. Mix sample by inverting tube 25 times and incubate at 37°C for 15 minutes.
8. Cool sample to room temperature for 5 min and add 100uL of Protein Precipitation Solution (Qiagen Puregene Tissue Kit) to cell lysate.
9. Vortex vigorously at high speed for 20 seconds. Put on ice for 5-10 min.
10. Centrifuge at 13,200 rpm for 5 minutes @ 20°C.
11. Pipette the supernatant containing the DNA (leaving behind the precipitated protein pellet) into clean Eppendorf tube. Pipette 300uL of 100% isopropanol into a 1.5 mL tube and add 0.5uL of glycogen solution.
12. Mix the sample by inverting gently 50 times.
13. Centrifuge at 13,200 rpm for 3 minute @ 20°C. The DNA should be visible as a white pellet.
14. Remove supernatant, leave ~50uL. Add 500uL of COLD 70% ethanol and invert the tube several times to wash the DNA pellet.
15. Centrifuge at 13,200 rpm for 3 minute. Carefully remove all the ethanol.
16. Air dry for 5 min - 15 min.
17. Add 20uL of DNA Hydration Solution (Qiagen Puregene Tissue Kit).
18. Heat at 65°C for 15 min.
19. Store at 4°C.

3. mtDNA Sequencing Details

A total of 138 Health ABC participants of European Ancestry that were part of an Energy Expenditure substudy were sequenced. This project was funded through the R03 mechanism and the budget would only cover the cost of sequencing 138 participants. mtDNA was extracted from platelets and sequenced with the Affymetrix Mitochondrial Resequencing Array 2.0 (MitoChip, Affymetrix, Santa Clara, CA) as previously

described. The MitoChip interrogates the forward and reverse strands of the 16.5 kb mitochondrial genome for a total of ~30 kb sequence, enables the detection of known and novel mutations and has redundant probe tiling for detecting the major human mitochondrial haplotypes and known disease-related mutations. Built-in redundancy via independent probe sets also allows a test of within-chip reproducibility. Briefly, the entire mitochondrial genome was first amplified in two long-range PCR reactions using LA PCR Kit (Takara Bio U.S.A., Madison, WI) for each sample using two sets of overlapping primers. Mitochondrial fragments were amplified and prepared for array hybridization according to the Affymetrix protocol for GeneChip CustomSeq Resequencing Array. The resulting PCR products were assessed qualitatively by 1% agarose gel electrophoresis and purified using a Clonotech Clean-Up plate (Clonotech, Mountain View, CA). The purified DNA was quantified by PicoGreen and for selected samples, confirmed by NanoDrop measurements. The amplicons were pooled at equimolar concentrations. Chemical fragmentation was performed and products were confirmed to be in the size range of 20-200 bp by 20% polyacrylamide gel electrophoresis with SYBR Gold staining. The IQ-EX control template, a 7.5 kb plasmid DNA, was used as a positive control. The samples were labeled with TdT and hybridized to the array in a 49°C rotating hybridization oven for 16 hours. Finally, streptavidin phycoerythrin (SAPE), and then antibody staining was performed. The microarrays were processed in the GeneChip Fluidic Station and the GeneChip Scanner. Signal intensity data was output for all four nucleotides, permitting quantitative estimates of allelic contribution. The allelic contribution was assessed using the raw data from the individual signal intensities by deriving the ratio of expected allele (REA), which is the log ratio of the raw signal intensity of the expected allele at any site (as defined by the mtDNA reference sequence) to the average raw signal intensity of the other three alleles, at each site for every individual. DAT files with raw pixel data were generated and used as input for grid alignment. CEL files generated from DAT files were analyzed in batches using GSEQ. Samples with call rates of less than 95% were discarded. Ten samples were repeated for concordance testing. For samples passing initial filtering, ResqMi 1.2 [100] was used for re-analysis of bases originally called as “N” by GSEQ. Analysis was performed using custom Perl scripts. Data was extracted from gene regions as defined by NCBI annotations for the revised Cambridge Reference Sequence (rCRS; NC_012920.1).

4. References

1. [Tranah](#), G. J., M. A. Nalls, S. M. Katzman, et al. **2012**. Mitochondrial DNA sequence variation associated with dementia and cognitive function in the elderly. *Journal of Alzheimer’s Disease*. 32(2). PMID: 22785396.
2. [Tranah](#), G. J., E. T. Lam, S. M. Katzman, et al. **2012**. Mitochondrial DNA sequence variation is associated with free-living activity energy expenditure in the elderly. *Biochimica et Biophysica Acta – Bioenergetics*. 1817(9):1691-1700. PMID: 22659402.