Documentation for the Mitochondrial DNA Heteroplasmy Dataset

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

1. **General description and dataset structure 2**
2. **Methods 2**
3. **References 3**

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 inherited and acquired sequence variants related to resting metabolic rate, total energy expenditure, and activity energy expenditure. Platelets were collected during year 2.

Mitochondrial DNA (mtDNA) heteroplasmy is a mixture of normal and mutant mtDNA molecules in a cell. High levels of heteroplasmy at specific mtDNA sites leads to inherited mitochondrial diseases with neurological, sensory, and movement impairments. Lower heteroplasmy levels are also common and often increase with age. Here we describe the quantification of heteroplasmy at 20 disease-causing mtDNA sites from the 138 sequenced mtDNAs. The mtDNA sites include: seven mtDNA mutations leading to complex I deficiency and brain MRI abnormalities [1] (m.10158T>C, m.10191T>C, m.10197G>A, m.13091T>C, m.13513G>A, m.13514A>G, m.14487T>C), three primary Leber's hereditary optic neuropathy mutations (m.3460G>A, m.11778G>A, and m.14484T>C), two nonsyndromic deafness mutations (m.7445A>G and m.7511T>C), and eight mitochondrial myopathy (m.3243A>G, m.3302A>G, m.4308G>A, m.5650G>A, m.5703G>A, m.7497G>A, m.12315G>A, and m.14709T>C) [2].

The data file is a csv file containing columns of percent heteroplasmy for each mtDNA site and HABCID. Each mutation is coded with a number indicating mtDNA site and major>minor nucleotides detected at that site. For example, values for m.10158T>C indicate the percent of the minor C allele at mtDNA nucleotide 10158 for each sample. Missing values for the nucleotide calls are left blank.

The data file is 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:

HetMitochondria.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

DNA extraction and Mitochondrial DNA sequencing methods are described in *Documentation for the Sequence of Mitochondrial DNA Dataset, Dr. Greg Tranah, Health ABC Ancillary Study #AS06-84*.

Heteroplasmy calculation

The MitoChip provides a quantitative output of signal intensity data for each nucleotide, allowing quantitative estimates of allelic contribution. We calculated quantitative estimates of heteroplasmy with established methods [5]. Heteroplasmy was derived for each mtDNA nucleotide by first calculating (a) the minor allele signal (MAS) defined as the raw signal intensity of the highest ‘‘non-expected’’ or minor allele minus the raw signal intensity value of the smallest contributing or background allele, and (b) the background subtracted expected allele signal (EAS) defined as the raw signal intensity of the expected allele (revised Cambridge Reference Sequence; NC_012920.1) minus the raw signal intensity of the smallest contributing or background allele. Percent heteroplasmy was then defined as MAS/(MAS + EAS). The MitoChip has been shown to have a detection limit of 2% heteroplasmy [6] and we excluded from analysis all loci with a calculated heteroplasmy value <2%. Twenty samples were repeated for concordance testing which yielded >98% sequence concordance of nucleotide calls. Within-chip error rate of 0.0028%, which is comparable to previously published rates of 0.0025% and 0.0021% [6]. The coefficient of variation (CV) values for the 20 repeated samples ranged from 0.09-0.27 for the 20 examined heteroplasmic mtDNA mutations.

3. References

1. Lebre AS, *et al.* A common pattern of brain MRI imaging in mitochondrial diseases with complex I deficiency. *J Med Genet* **48**, 16-23 (2013).

2. MITOMAP: A Human Mitochondrial Genome Database. http://www.mitomap.org.

3. Tranah GJ, *et al.* Mitochondrial DNA sequence variation is associated with free-living activity energy expenditure in the elderly. *Biochimica et biophysica acta*, (2012).

4. Symons S, Weber K, Bonin M, Nieselt K. ResqMi - a Versatile Algorithm and Software for Resequencing Microarrays. In: *Proceedings of the German Conference on Bioinformatics* (ed^(eds Beyer A, Schroeder M) (2008).

5. Coon KD, *et al.* Quantitation of heteroplasmy of mtDNA sequence variants identified in a population of AD patients and controls by array-based resequencing. *Mitochondrion* **6**, 194-210 (2006).

6. Maitra A, *et al.* The Human MitoChip: a high-throughput sequencing microarray for mitochondrial mutation detection. *Genome Res* **14**, 812-819 (2004).