The spectrum of mitochondrial genomic variation across 250,000 individuals

Nicole L. Washington[1], Alexandre Bolze[1], Fernando Mendez[1], Elizabeth T. Cirulli[1], Simon White[1], M. Rashkin[1], William J Metcalf[2,3], Joseph J. Grzymski[2,3], James Lu[1]

[1] Helix, San Carlos, California
[2] Desert Research Institute, Reno, Nevada
[3] Renown Institute of Health Innovation, Reno, Nevada

Mitochondrial variants are causative for rare mitochondrial diseases (1 in 5,000 clinically affected adults), and are also linked to common health conditions such as obesity and Alzheimer's disease. For those with deleterious variants, differences in heteroplasmic levels can lead to varying phenotypic presentations of the same disease in a population. Understanding the genetic contribution of mitochondrial variation to human health is challenging due to the unique nature of mitochondrial biology and its coexistent, complementary, yet distinct inheritance from that of the nuclear genome.

Smaller mitochondrial-specific databases are biased due to their composition; our current understanding of human mitochondrial variation is largely informed by studies that recruit for patients with inherited mitochondrial disease, and by studies of human evolution and migration patterns. In both cases, there is often bias in recruitment, and baseline rates of variation may be skewed. Information on the frequency of a homoplasmic or heteroplasmic variant in the population would help develop, test, and refine or disprove hypotheses for causality.

Here we characterize the mitochondrial genomes of more than 250,000 individuals sequenced in the Helix clinical laboratory. This collection is unbiased towards individuals with a mitochondrial disorder. We also provide the community with a research resource of all mtDNA variants identified and their frequencies. Our database includes more than 10,000 mtDNA variants, with ~20% present as singletons, and ~54% mtDNA bases invariant in the population. While our variant count is similar to MitoMAP, our singleton rate is lower and invariant base fraction higher as is expected in an unbiased cohort with improved mtDNA variant frequency estimates. We find that heteroplasmic-only variants are enriched at sites of high conservation, supporting the presence of heteroplasmy as an indicator of pathogenicity. We show that this large population-based approach improves our ability to interpret disease-associated variants by analyzing those reported to be pathogenic for Leber's Hereditary Optic Neuropathy (LHON). We find that many reportedly pathogenic variants have a frequency far above the maximum plausible allele frequency for a LHON-causing variant, including the primary variant m.14484T>C. We further confirm these findings with matched medical records for participants in the UK Biobank and Healthy Nevada Project.