Supplementary Information

Selective constraints and pathogenicity of mitochondrial DNA variants inferred from a novel database of 196,554 unrelated individuals.

Alexandre Bolze[1,*,#], Fernando Mendez[1,*], Simon White[1], Francisco Tanudjaja[1], Magnus Isaksson[1], Misha Rashkin[1], Johnathan Bowes[1], Elizabeth T. Cirulli[1], William J. Metcalf[2,3], Joseph J. Grzymski[2,3], William Lee[1], James T. Lu[1], Nicole L. Washington[1,#]

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

There are 1 supplementary figure and 9 supplementary tables. The tables are in a separate Excel document.

Figure S1. Comparison between HelixMTdb and MITOMAP

Comparison of allele frequencies (AF) of variants present in both HelixMTdb and MITOMAP. The left panels are scatter plots where each variant is represented by one dot. The x-axis represents AF in HelixMTdb, and the y-axis represented AF in MITOMAP. The spearman rho coefficient rho is indicated on the upper left of the plot. The right panels represent similar data. The variants were first grouped by bins. Seven categories of bins were defined based on counts and AF of a variant: n=0, n=1, n=2 to 0.01%, 0.01%-0.1%, 0.1%-1%, 1%-10%, >10%. Each square therefore represents the variants that fall in a particular HelixMTdb bin and a particular MITOMAP bin. For example, the square at the lower left represents variants that have a count of zero in each database. The square on the upper right represents variants with an AF >10% in each database. The color of the square represents the number of variants in this square. The darker the square, the more variants are in each. **(A)** AF of all homoplasmic variants in HelixMTdb, and all variants in MITOMAP are represented. **(B)** AF of homoplasmic SNVs in HelixMTdb, and SNVs in MITOMAP are represented. **(C)** AF of homoplasmic insertions in HelixMTdb, and insertions in MITOMAP are represented. **(D)** AF of homoplasmic deletions in HelixMTdb, and deletions in MITOMAP are represented.

Figure S1

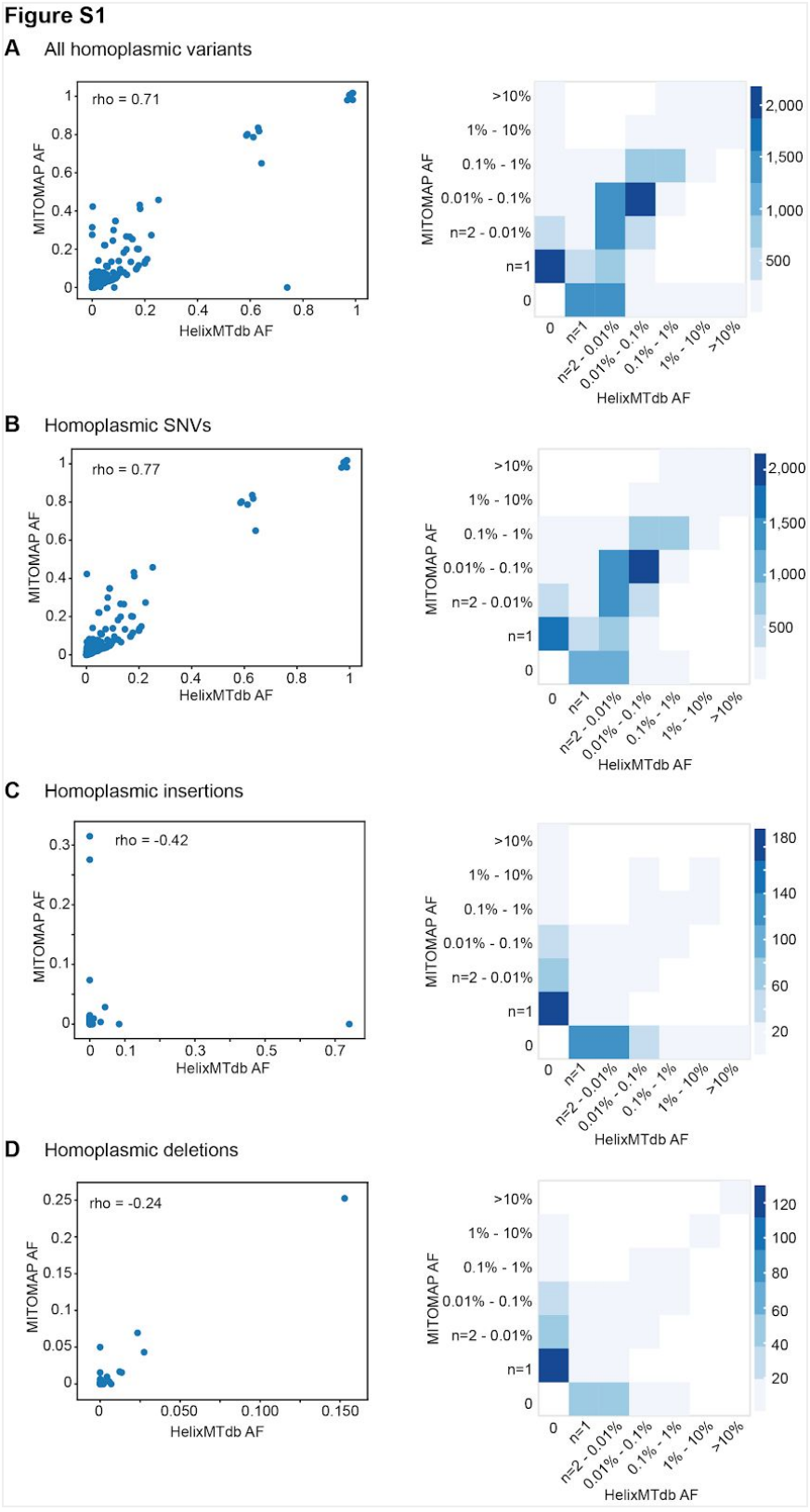
A All homoplasmic variants
rho = 0.71
MITOMAP AF
HelixMTdb AF
>10%
1% - 10%
0.1% - 1%
0.01% - 0.1%
n=2 - 0.01%
n=1
0
2,000
1,500
1,000
500
B Homoplasmic SNVs
rho = 0.77
C Homoplasmic insertions
rho = -0.42
180
140
100
60
20
D Homoplasmic deletions
rho = -0.24
120
100
80
60
40
20

Table S1: Distribution of mitochondrial haplogroups in HelixMTdb

Table S2: List of all constrained mitochondrial regions inferred from homoplasmic calls
Regions / intervals of 1bp were not included in this analysis.

Table S3: List of all constrained mitochondrial regions inferred from homoplasmic calls and heteroplasmic calls with a RAF >=0.5
Regions / intervals of 1bp were not included in this analysis.

Table S4: List of all constrained mitochondrial regions inferred from all homoplasmic calls and all heteroplasmic calls
Regions / intervals of 1bp were not included in this analysis.

Table S5: Variant attributes by genomic features
Mean allele frequency = ratio of n_non_ref / n_samples. So the % of individuals either with a homoplasmic or heteroplasmic variant vs number of total samples.
% conservation = hl.agg.mean(phastcons100v)

Table S6: Details of all putative loss-of-function variants in HelixMTdb

Table S7: Counts of all LHON variants reported in HelixMTdb and the UK Biobank

Table S8: Location of genomic features in the mitochondrial genome

Table S9: ICD10 codes in Healthy Nevada Project and UK Biobank