Analysis of mitochondrial DNA variants in 200,000 individuals

9th Human Genetics in NYC Conference

20 July 2021

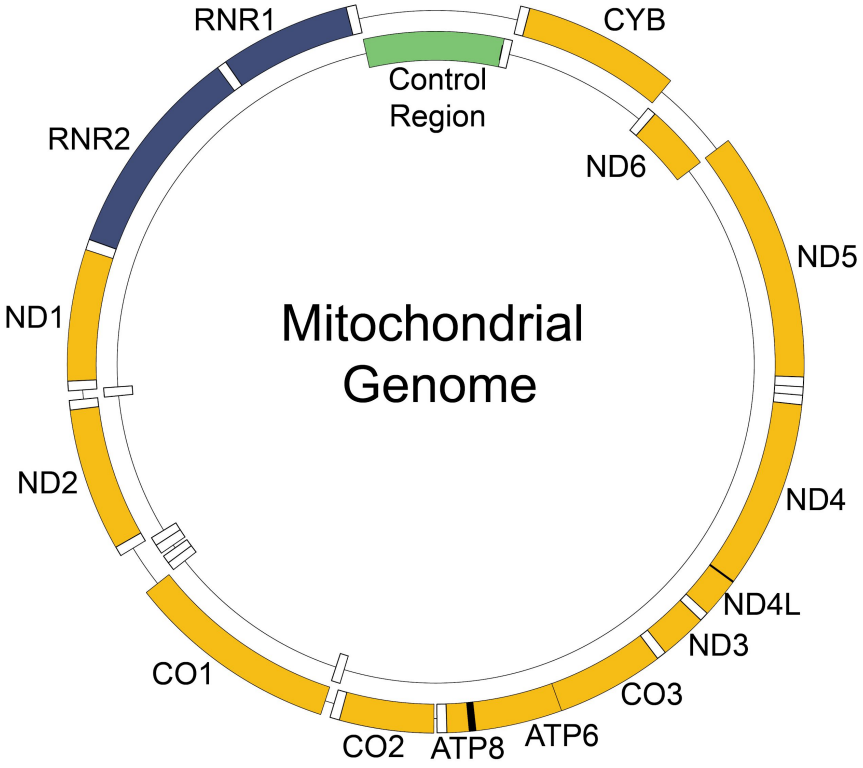
Alexandre Bolze

Helix

Disclosures

I am a full time employee of Helix

Human mitochondrial genome is 16,569 bases



Control region

rRNA (2 genes)

tRNA (22 genes)

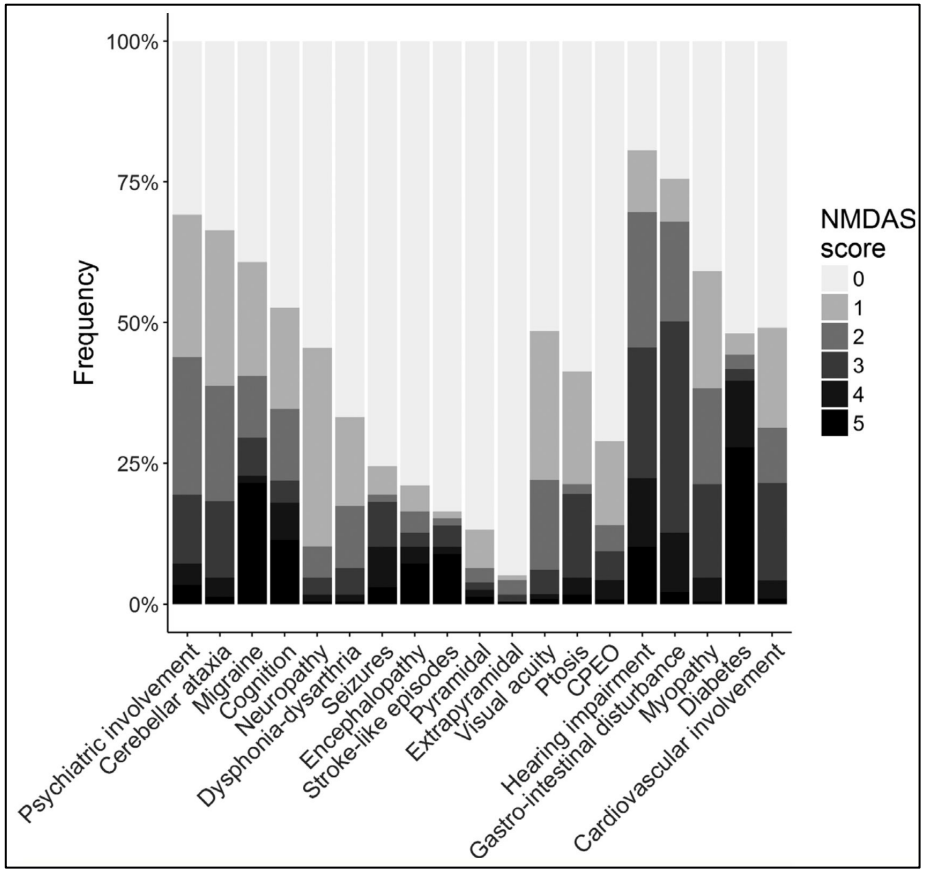
Protein-coding (13 genes)

mtDNA variants can cause disease, example: m.3243A>G

- Heteroplasmy for m.3243A>G in *MT-TL1* gene

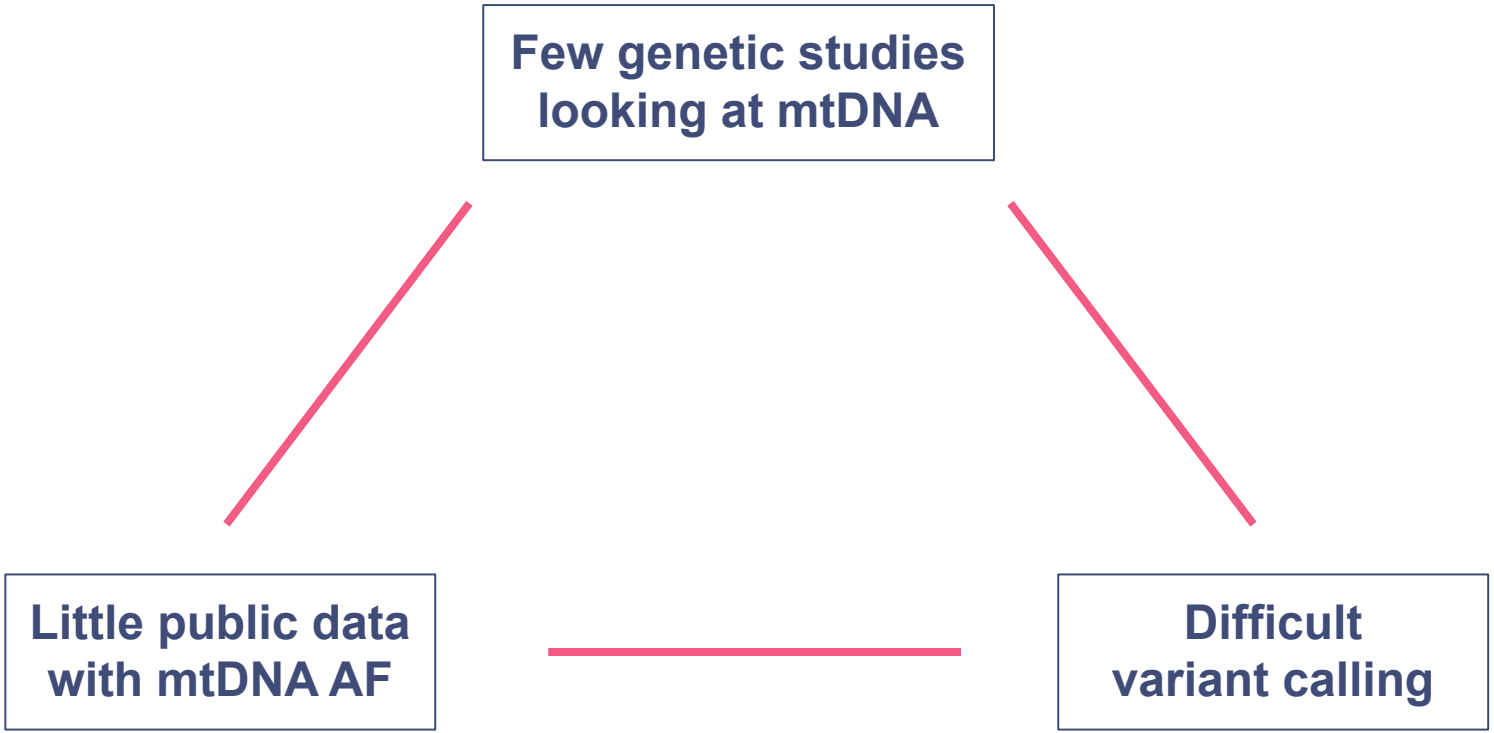
  is linked to MELAS **M**itochondrial **E**ncephalopathy, **L**actic **A**cidosis, and **S**troke-like episodes.

- Broad and heterogeneous spectrum of phenotypes (based on 238 carriers)



Ref: Pickett et al., Annals of Clinical and Translational Neurology, 2020

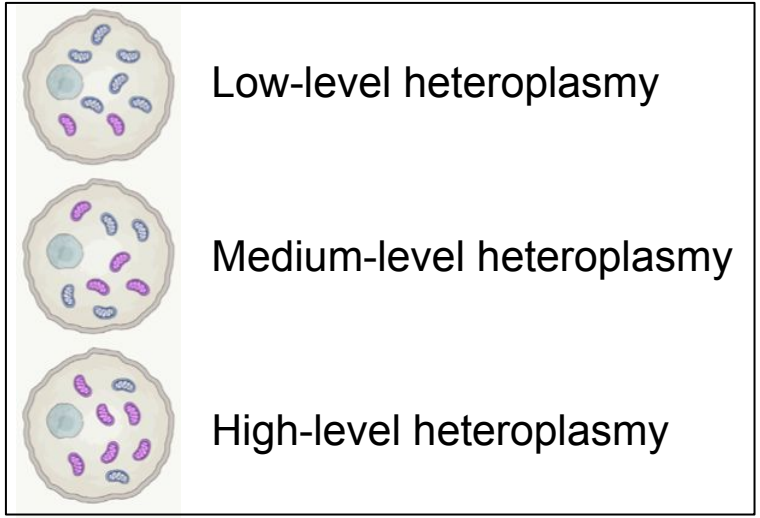
Interpreting mtDNA variants is challenging and mtDNA is often forgotten in human genetic studies



Challenges for mtDNA variant calling

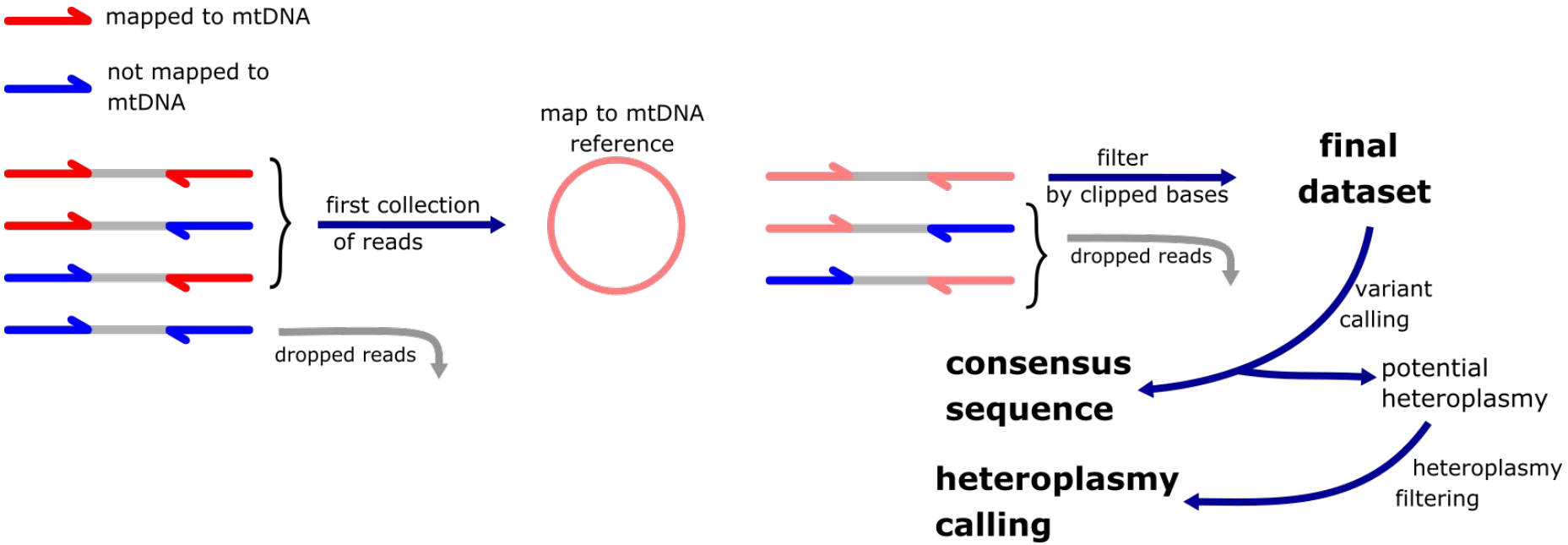
- NUMTs
  NUclear-encoded MiTochondrial DNA segments

- Heteroplasmy, especially low levels of heteroplasmy
  - 1,000 to 10,000s copies per cell
  - Levels change between tissues, over time, between mother and child



Adapted from Russell et al., Cell, 2020

Specific read alignment to deal with NUMTs



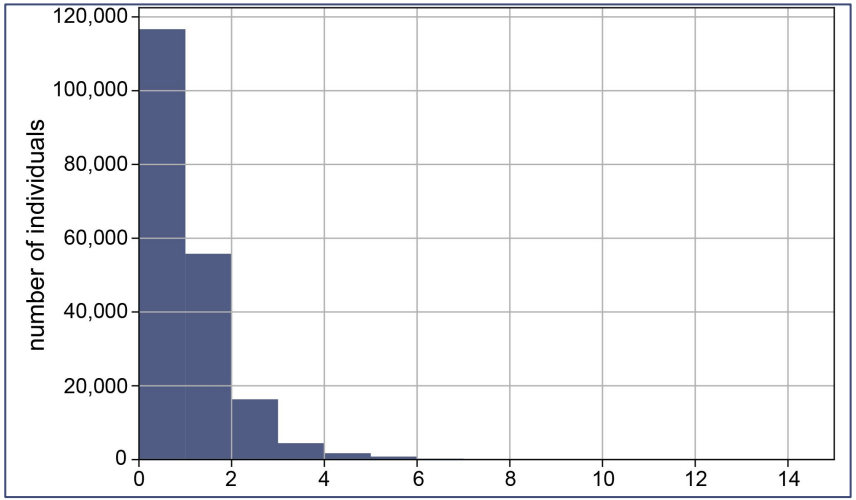
Thresholds and quality control for heteroplasmic calls

Based on overall coverage

- WES with mean DP ~200x,
  we called heteroplasmy levels >5%

- For WGS, can probably call
  heteroplasmy as low as 1%

N of heteroplasmic variants per individual



After QC, individual had on average 1 heteroplasmic call (range: 0 to 13)

Few genetic studies looking at mtDNA

Little public data with mtDNA AF

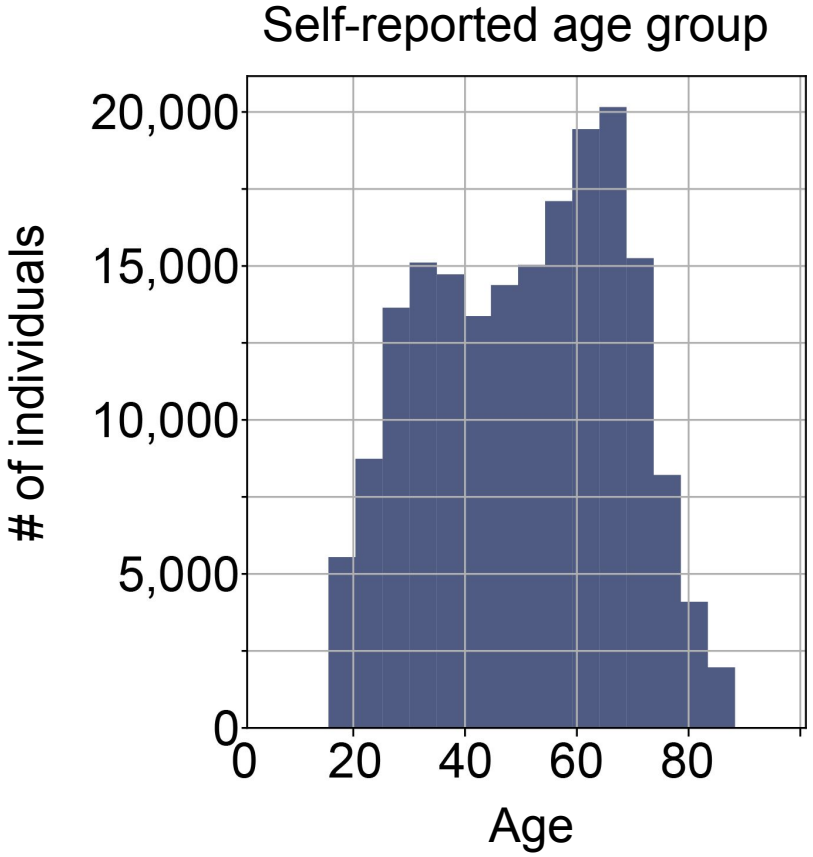
Difficult variant calling

- NUMTs
- Heteroplasmy
- Mutect2 software by GATK

HelixMTdb is a public resource based on 195,983 participants with no ‘mito’ bias

- Unrelated adults living in the U.S., no criteria based on mito

- **www.helix.com/mito**



New Results

A catalog of homoplasmic and heteroplasmic mitochondrial DNA variants in humans

Alexandre Bolze, Fernando Mendez, Simon White, Francisco Tanudjaja, Magnus Isaksson, Ruomu Jiang, Andrew Dei Rossi, Elizabeth T. Cirulli, Misha Rashkin, William J. Metcalf, Joseph J. Grzymski, William Lee, James T. Lu, Nicole L. Washington

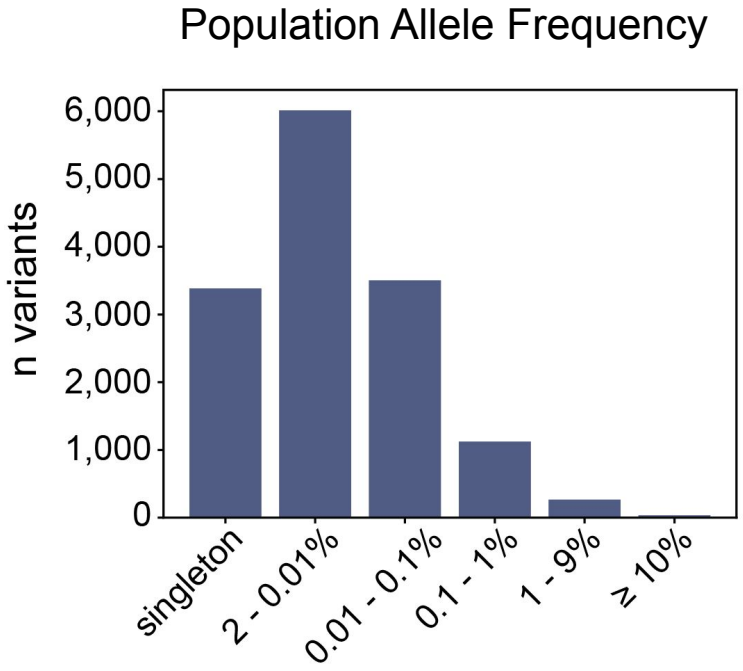
doi: https://doi.org/10.1101/798264

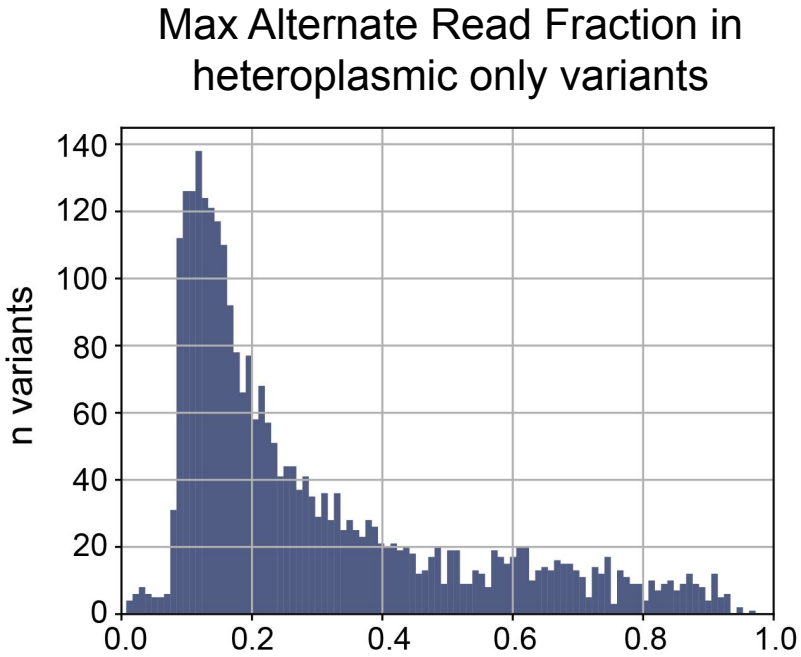
Cold Spring Harbor Laboratory

bioRχiv

THE PREPRINT SERVER FOR BIOLOGY

Majority of 14,323 mtDNA variants are very rare

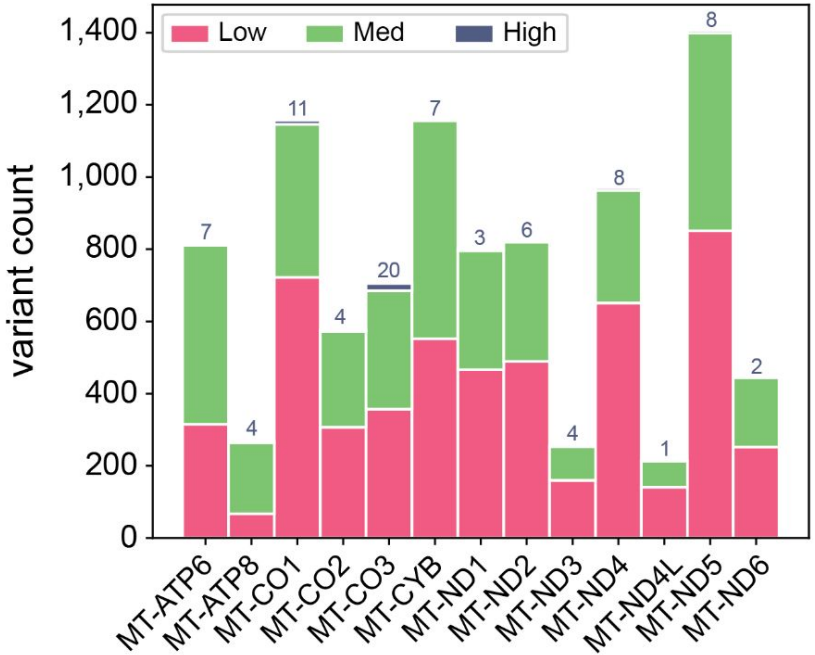




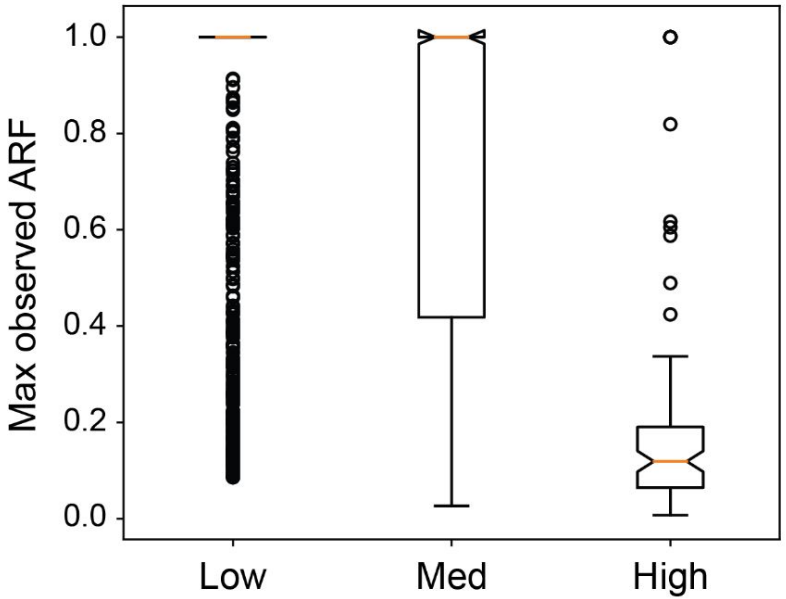
20% of variants were only seen at heteroplasmic levels, most of them at low levels of heteroplasmy

Not tolerant to truncating variants in protein-coding genes

Distribution of coding variants by severity



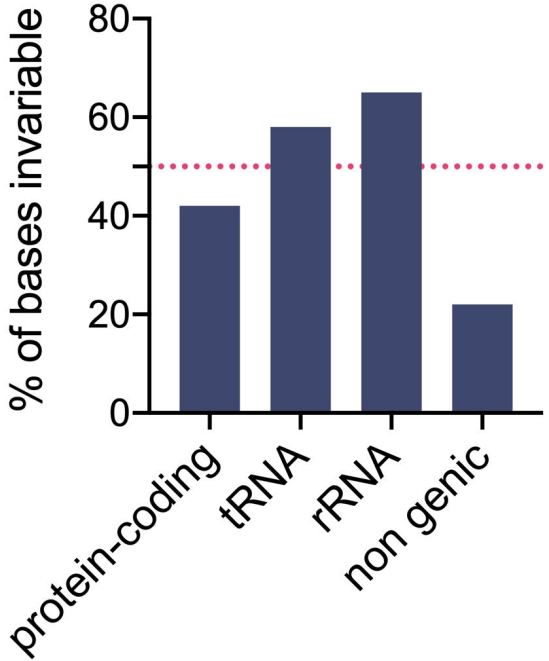
Maximum Alternate Read Fraction by severity

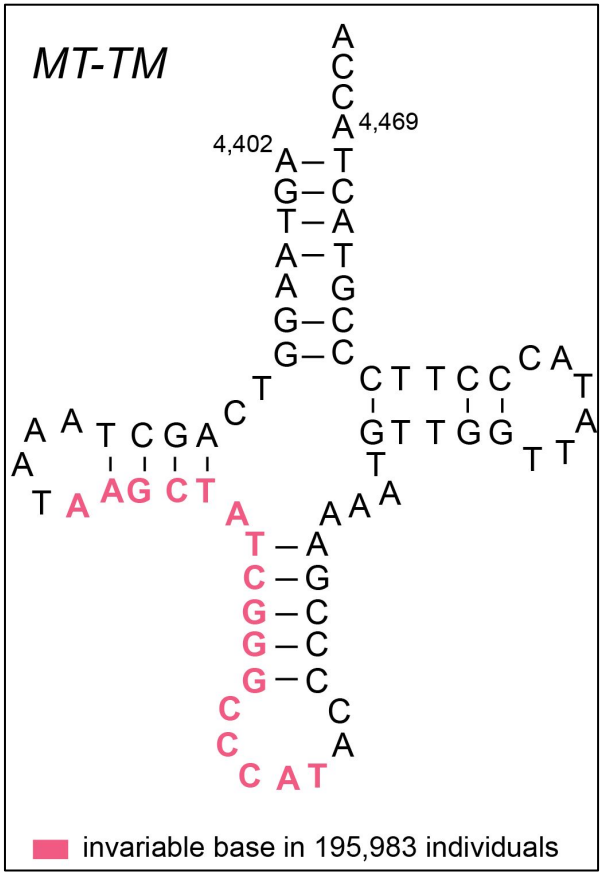


Only nonsense variant observed at homoplasmic levels was p.M1* in *MT-ND1* gene

mtDNA regions under high constraint

% of bases without any homoplasmic or heteroplasmic (>50% level) variant





Ref: Bolze*, Mendez* et al., BioRxiv, 2020

Few genetic studies looking at mtDNA

Little public data with mtDNA AF

- HelixMTdb
- gnomAD v3.1 (November 2020)
- on mitomap (March 2021)

Difficult variant calling

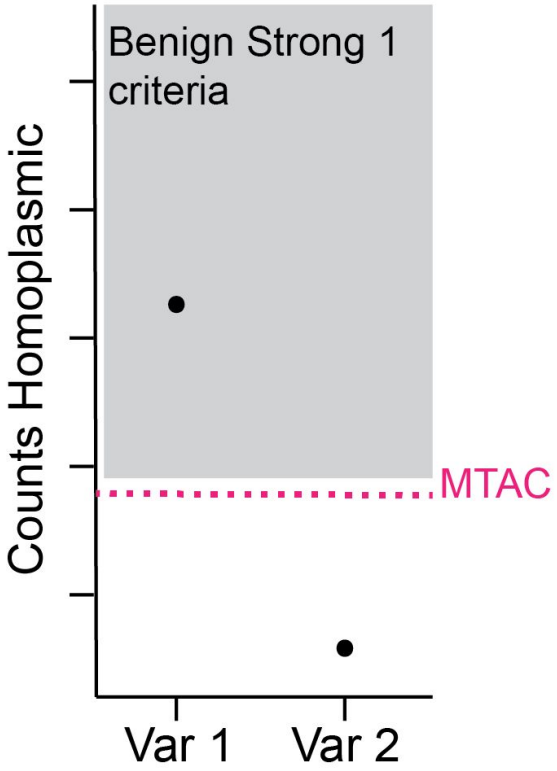
- NUMTs
- Heteroplasmy
- Mutect2 software by GATK

Use case: re-classification of LHON variants

Leber Hereditary Optic Neuropathy

- Early adult onset
- Prevalence: ~ 1 in 30,000
- Mostly caused by 3 mtDNA mutations (homoplasmic levels)
- Other risk factors: male, smoking

Use case: re-classification of LHON variants



$$AF_{max}(LHON) \sim prevalence \times \frac{genetic}{homogeneity} \times \frac{1}{penetrance}$$

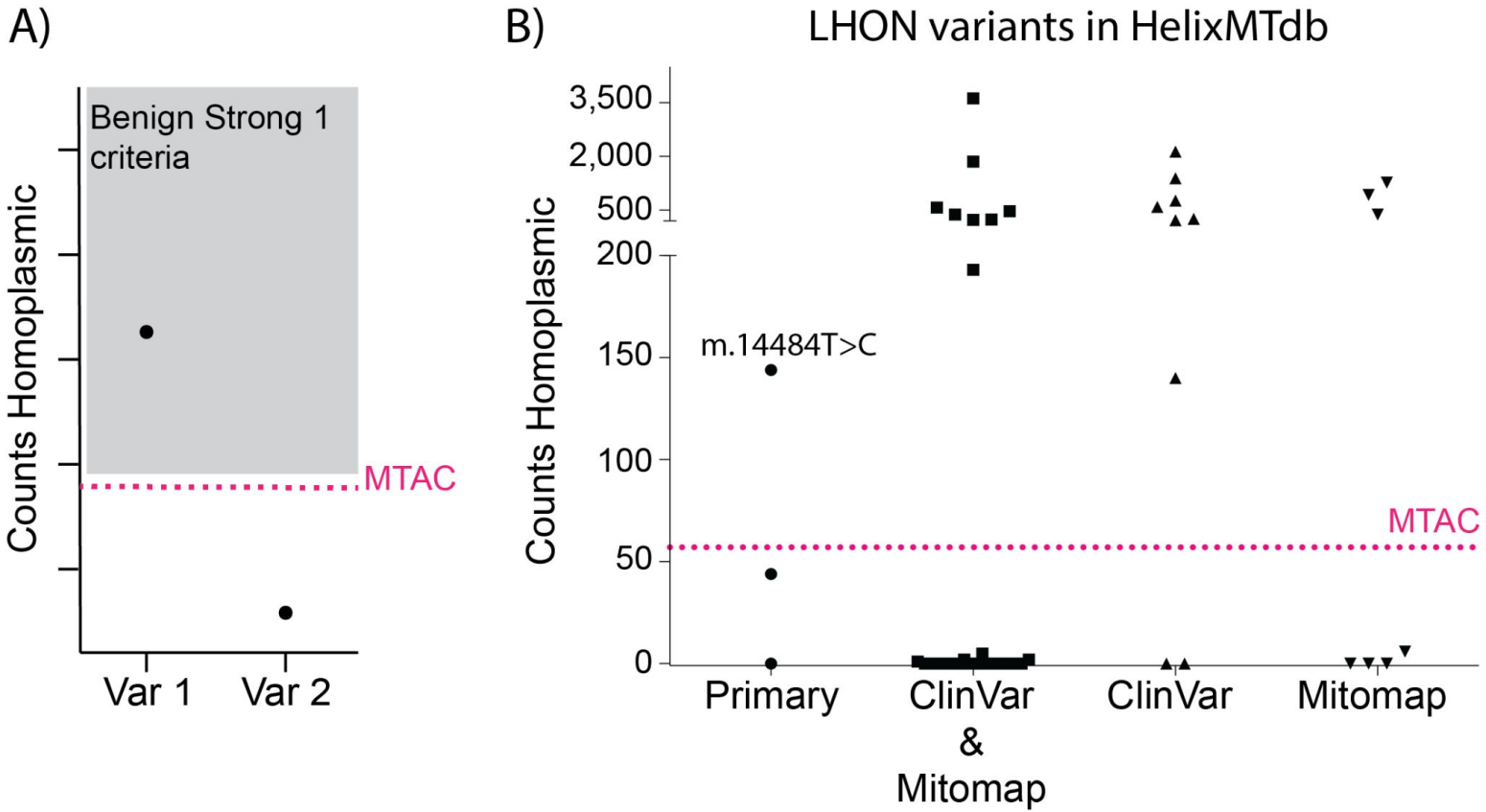
$$AF_{max}(LHON) \sim \frac{2}{10{,}000}$$

Max Tolerated Allele Count in 195,983 mito genomes

$$MTAC_{LHON} = 57$$

Ref: Whiffin et al., GIM, 2017

Use case: re-classification of LHON variants



ClinVar & Mitomap accessed Oct. 2019

Use case: re-classification of LHON variants

Table: phenotype of individuals carrying the m.14484T>C variant

	HelixMTdb all individuals	HelixMTdb homoplasmic m.14484T>C	UK Biobank all individuals	UK Biobank homoplasmic m.14484T>C
Individuals with EHR & mito available	28,503	27	413,647	318
Individuals with ICD10 code H47.2 (optic atrophy)	20 (0.07%)	0	97 (0.02%)	0

Use case: re-classification of LHON variants

Penetrance of the m.14484T>C variant is likely less than 10%,
even when restricted to men.

Overall, numbers in HelixMTdb suggest that up to **40% of variants** reported to be Pathogenic for LHON **could be reclassified**, based on population allele frequency information.

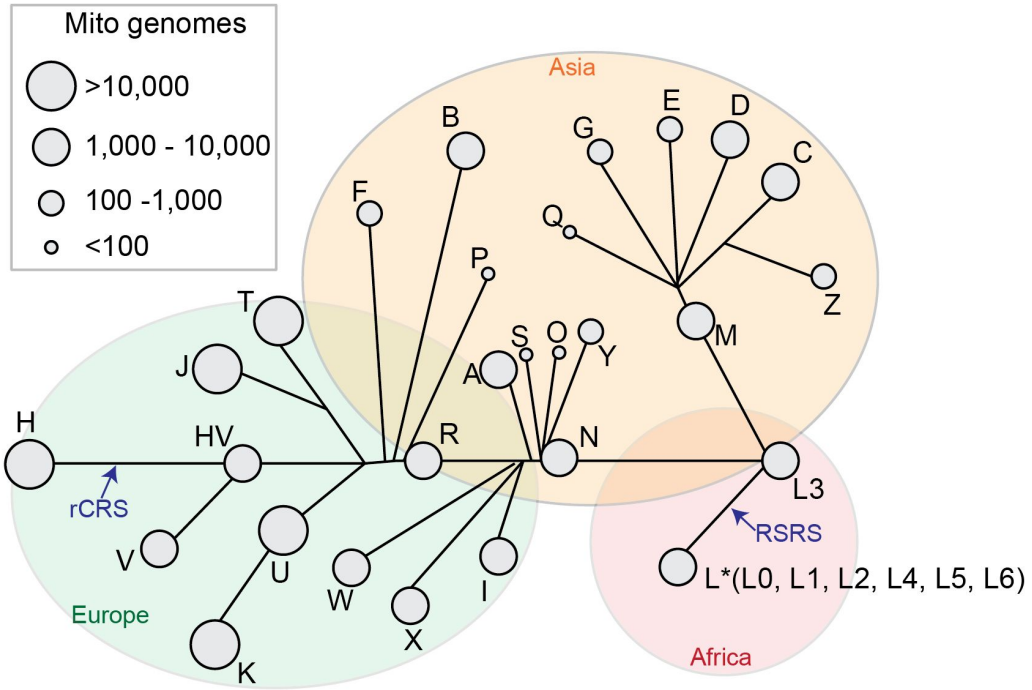
Limitations of HelixMTdb

- 91% from N lineages
  5% from M lineages
  4% from L lineages

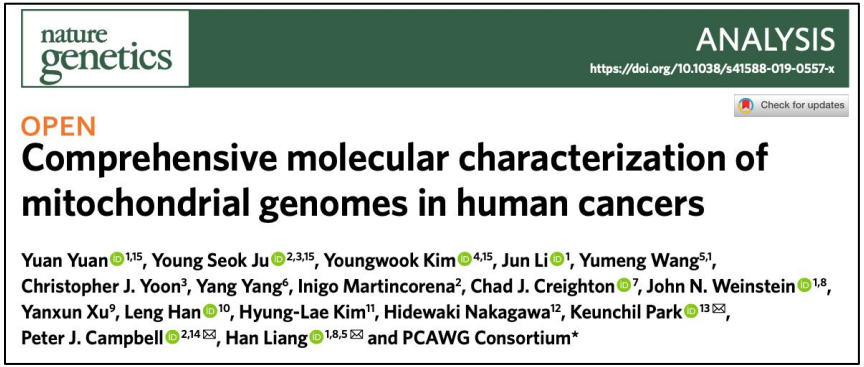
- Low sensitivity for very low levels of heteroplasmy (<10%)

- From Saliva.
  Not always the tissue you want

- What is actionable?



Recent examples from the literature



nature genetics

ANALYSIS

https://doi.org/10.1038/s41588-019-0557-x

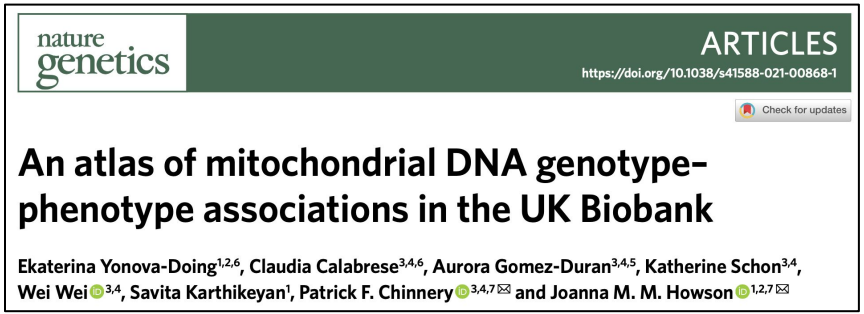
Check for updates

OPEN

Comprehensive molecular characterization of mitochondrial genomes in human cancers

Yuan Yuan[1,15], Young Seok Ju[2,3,15], Youngwook Kim[4,15], Jun Li[1], Yumeng Wang[5,1], Christopher J. Yoon[3], Yang Yang[6], Inigo Martincorena[2], Chad J. Creighton[7], John N. Weinstein[1,8], Yanxun Xu[9], Leng Han[10], Hyung-Lae Kim[11], Hidewaki Nakagawa[12], Keunchil Park[13] ✉, Peter J. Campbell[2,14] ✉, Han Liang[1,8,5] ✉ and PCAWG Consortium*

Truncating mutations in mtDNA were enriched in kidney, colorectal and thyroid cancers



nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00868-1

Check for updates

An atlas of mitochondrial DNA genotype–phenotype associations in the UK Biobank

Ekaterina Yonova-Doing[1,2,6], Claudia Calabrese[3,4,6], Aurora Gomez-Duran[3,4,5], Katherine Schon[3,4], Wei Wei[3,4], Savita Karthikeyan[1], Patrick F. Chinnery[3,4,7] ✉ and Joanna M. M. Howson[1,2,7] ✉

Top hit:
m.11719G>A (synonymous in *MT-ND4*) with Mean Corpuscular Hemoglobin, $p=4.5 \times 10^{-20}$

Take home

Few genetic studies looking at mtDNA

- Interpretation of variants
- GWAS for common homoplasmic variants
- Gene burden tests

Little public data with mtDNA AF

- HelixMTdb
- gnomAD v3.1 (November 2020)
- on mitomap (March 2021)

Difficult variant calling

- NUMTs
- Heteroplasmy
- Mutect2 software by GATK

Acknowledgements

All participants in Helix DNA Discovery Project, Healthy Nevada Project and UK Biobank

Helix & collaborators

Nicole Washington
Fernando Mendez
Andrew Dei Rossi
William Lee
James Lu
William Metcalf (DRI)
Joe Grzymski (DRI)

Early users of the database

Marie Lott & Mitomap team
Ed Reznik (MSKCC)
Cory Dunn (University of Helsinki)
Linda Mathisen (Oslo University Hospital)
Emma Watson (Newcastle Upon Tyne)

Database available: www.helix.com/mito

Preprint: Bolze*, Mendez* et al., BioRxiv, 2020

Email: alexandre.bolze@helix.com

Twitter: @alexbolze