

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No specialized software or code was used for the collection of primary data; most of which was downloaded from public databases and or published datasets.
Data analysis	Code used for data analysis was written and implemented in Python (v3.10), R (v3.6.1, v4.0.4 or v4.4.0), and ChimeraX (v1.3). All code is available through Github: https://github.com/leklab/mitochondrial_constraint . Variant calling on genome sequencing data from the UK Biobank and SPARK cohorts was done using MitoHPC version 20230418 (https://github.com/ArkingLab/MitoHPC) and GATK Mutect2 v4.1.7 (per https://github.com/leklab/sfari_mito_calling) respectively. Sequence alignment of the human and chimpanzee reference sequences was done using msa R package v1.36.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data analyzed or generated during this study are included in this article and its Supplementary files, and available via https://github.com/lekclab/mitochondrial_constraint. Constraint metrics are provided in the Supplementary Datasets, and will also be available via <http://gnomad.broadinstitute.org>. Publicly-available datasets used in this study are available from the following sources: ClinVar, <https://www.ncbi.nlm.nih.gov/clinvar/>; DECIPHER, <https://www.deciphergenomics.org/ddd/ddgenes> (developmental disorder genes); gnomAD, <http://gnomad.broadinstitute.org>; HelixMTdb, <https://www.helix.com/mitochondrial-variant-database>; HmtVar, <https://www.hmtvar.uniba.it/>; IMPC, <https://www.ebi.ac.uk/mi/imp/essential-genes-search/> (essential genes); MitImpact, <https://mitimpact.css-mendel.it/> (APOGEE predictions); MITOMAP, <https://www.mitomap.org/MITOMAP>; NCBI Genome, <https://www.ncbi.nlm.nih.gov/datasets/genome/>; Phylotree, <https://www.phylotree.org/> (haplogroup variants); Protein Data Bank, <https://www.rcsb.org/>; UCSC, <https://genome.ucsc.edu/> (phyloP scores); UniProt, <https://www.uniprot.org/>. A detailed description of these datasets and their application is also provided at https://github.com/lekclab/mitochondrial_constraint/tree/main/required_files.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size Mitochondrial DNA variation from 56,434 individuals in gnomAD was used for all of our constraint analyses. We did not use any methods to predetermine sample size as we were using a publicly-accessible dataset, which represented the largest publicly-accessible population dataset of mtDNA variation from genome sequencing data. Our study relied on existing population-based datasets, and did not generate new population datasets, so increasing the size of the dataset was out of scope. Due to restrictions on sample size, confidence intervals and error associated with results generated were reported.

Data exclusions For the de novo mutations used to build the mutational model, checks were performed to confirm that mutations across sources had highly similar mutation characteristics, and one outlier group was excluded (see Supplementary Methods). Variants in gnomAD that did not pass quality filters were excluded from constraint analysis. Variants at bases which were not called by gnomAD (m.301, m.302, m.310, m.316, m.16182) or in the replication dataset HelixMTdb (m.300-316, m.513-525, m.16182-16194) were excluded from observed:expected calculations. Disease-associated variants from ClinVar with a conflicting interpretation, or listed only to be associated with cancer, were excluded. Samples or variants which did not pass quality filters were excluded from the UK Biobank and SPARK variant calls used, as described in the Methods.

Replication The majority of the study involved the analysis of existing datasets, for which number of replicates associated with experimental data generation is not applicable. Assessment of constraint for functional classes of mtDNA variation was replicated using another large, publicly-accessible mitochondrial population database with heteroplasmy data, HelixMTdb. This was a successful replication of our initial results that used the gnomAD dataset. All technical and biological replicates associated with the mitochondrial base editing experiments are mentioned in the methods and figure legends. Three or five technical replicates were used in each cell growth and seahorse biological replicate, respectively. Four or five biological replicates were performed for the cell growth assay for each target site. Three or four biological replicates were performed for the seahorse experiment for each target site. These replicates were used to report an average or range.

Randomization

Randomization was not relevant to our study design. The majority of the study involved the analysis of existing datasets for a population-based study, hence why randomization was not relevant. Randomization was not relevant for experimental work associated with mitochondrial base editing as there were no experimental groups for samples to be randomly assigned to.

Blinding

Blinding was not relevant to our study design. The majority of the study involved the analysis of existing datasets for a population-based study, hence why blinding was not relevant. The experimental results and conclusions associated with mitochondrial base editing are based upon machine measurements that would not be impacted by blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

HEK293T cells obtained from American Type Culture Collection (ATCC CRL-3216) were used.

Authentication

The cell line was confirmed to be hypotriploid, which is a signature of HEK cell lines. No other authentication was used after purchase from ATCC.

Mycoplasma contamination

The HEK293T cell line tested negative for mycoplasma.

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified cell lines were used in this study.