Investigating mitonuclear interactions in human admixed populations

Arslan A. Zaidi and Kateryna D. Makova

To function properly, mitochondria utilize products of 37 mitochondrial and >1,000 nuclear genes, which should be compatible with each other. Discordance between mitochondrial and nuclear genetic ancestry could contribute to phenotypic variation in admixed populations. Here, we explored potential mitonuclear incompatibility in six admixed human populations from the Americas: African Americans, African Caribbeans, Colombians, Mexicans, Peruvians and Puerto Ricans. By comparing nuclear versus mitochondrial ancestry in these populations, we first show that mitochondrial DNA (mtDNA) copy number decreases with increasing discordance between nuclear and mtDNA ancestry. The direction of this effect is consistent across mtDNA haplogroups of different geographic origins. This observation indicates suboptimal regulation of mtDNA replication when its components are encoded by nuclear and mtDNA genes with different ancestry. Second, while most populations analysed exhibit no such trend, in African Americans and Puerto Ricans, we find a significant enrichment of ancestry at nuclear-encoded mitochondrial genes towards the source populations contributing the most prevalent mtDNA haplogroups (African and Native American, respectively). This possibly reflects compensatory effects of selection in recovering mitonuclear interactions optimized in the source populations. Our results provide evidence of mitonuclear interactions in human admixed populations and we discuss their implications for human health and disease.

Mitochondria participate in vital functions of eukaryotic cells, such as generation of adenosine tri-phosphate (ATP) via oxidative phosphorylation (OXPHOS), regulation of calcium uptake, apoptosis and metabolism of essential nutrients. Even though mitochondria harbour their own genome (mtDNA) with 37 genes, >1,000 genes involved in mitochondrial functions are encoded by the nuclear genome. Thus, mitochondrial functions rely on fine-tuned mitonuclear interactions, resulting in mitonuclear coevolution over time. Mitonuclear coevolution has been demonstrated in non-human organisms, in which it manifests as mitonuclear incompatibilities arising in inter-specific and interpopulation hybrids. Whether these incompatibilities exist and contribute to phenotypic variation in humans is currently unknown. In this study, we explore mitonuclear incompatibility in recently admixed human populations.

Evidence of mitonuclear coevolution comes primarily from laboratory crosses of model organisms, such as fruit flies1–5, wasps6–10, seed beetles11, marine copepods12–15 and yeast16–18. Inter-population hybrids in these organisms frequently exhibit reduced viability and fecundity16,17, which can be explained by altered expression of OXPHOS genes19, reduced OXPHOS activity, decreased ATP production1, altered mtDNA copy number11,19 and elevated oxidative damage4. Fitness can often be restored if the hybrids are backcrossed with the maternal line but not with the paternal line11, suggesting that their reduced fitness was caused by differences in ancestry between mitochondrial and nuclear genomes, hereafter called mitonuclear DNA discordance. The phenotypic consequences of mitonuclear discordance are known as mitonuclear incompatibility. Mitonuclear incompatibility has also been observed in naturally occurring populations18–22. For example, two populations of the eastern yellow robin carrying different mtDNA haplotypes exhibit high differentiation at nuclear-encoded mitochondrial genes, suggesting that mitonuclear incompatibility may have maintained the divergence between the two populations in spite of continued gene flow22. In another example, the admixture fraction at nuclear-encoded mitochondrial loci in hybrids of two killifish populations, who also have divergent mtDNA haplogroups, appears to be correlated with OXPHOS activity20.

Currently, we know very little about the extent of mitonuclear incompatibility and its contribution to phenotypic variation in humans. It is established that altered interactions between mtDNA- and nuclear-encoded factors can modulate disease phenotypes for cardiomyopathy, predisposition to type 2 diabetes and possibly hearing loss and Huntington’s disease20,21. However, no elevated mitonuclear linkage disequilibrium was found across a set of 51 human populations from the Human Genome Diversity Project25–27. Mitonuclear linkage disequilibrium is expected if selection had favoured certain allelic combinations between mtDNA and nuclear-encoded mitochondrial genes across diverging human populations28. A recent study has also shown that nuclear-encoded mitochondrial genes are significantly underrepresented in Neanderthal, but not Denisovan, introgressed regions in the human genome29. Such a depletion is expected because there is no evidence of introgression of Neanderthal and Denisovan mtDNA into modern humans29,30. Therefore, if admixture between Neanderthals/Denisovans and humans results in mitonuclear incompatibility, selection should favour human-specific alleles at nuclear-encoded mitochondrial genes.

In this manuscript, we explored signatures of mitonuclear incompatibility and coevolution in six admixed human populations from the Americas studied as part of the 1000 Genomes Project31,32: (1) African Americans from the Southwest; (2) African Caribbeans from Barbados; (3) Colombians from Medellín, Colombia; (4) Mexicans from Los Angeles; (5) Peruvians from Lima, Peru and (6) Puerto Ricans from Puerto Rico (see Methods and Table 1). These populations are characterized by admixture among Africans, Europeans and Native Americans. Our hypothesis is that increasing discordance between nuclear and mitochondrial ancestry in
Table 1 | The number of individuals from each population used in the study

| Ancestry group | Population | Number of individuals with sequence alignments | Number of individuals with genotype data | Data source |
|----------------|------------|-----------------------------------------------|------------------------------------------|-------------|
| Admixed        | African Americans from the Southwest (ASW) | 60 | 61 | 1000 Genomes Project Data② |
|                | African Caribbeans from Barbados (ACB) | 96 | 96 | 1000 Genomes Project Data② |
|                | Mexicans from Los Angeles (MXL) | 67 | 64 | 1000 Genomes Project Data② |
|                | Peruvians (PEL) | 86 | 85 | 1000 Genomes Project Data② |
|                | Puerto Ricans (PUR) | 105 | 104 | 1000 Genomes Project Data② |
|                | Colombians (CLM) | 94 | 94 | 1000 Genomes Project Data② |
| African        | Yoruba from Nigeria (YRI) | 109 | 108 | 1000 Genomes Project Data② |
| European       | Utah residents of Northern and Western European ancestry (CEU) | 99 | 99 | 1000 Genomes Project Data② |
| Native American | Aymara | ND | 25 | Mao et al.⑥ |
|                | Nahuan | ND | 14 | Mao et al.⑥ |
|                | Quechua | ND | 24 | Mao et al.⑥ |
|                | Maya | ND | 25 | Mao et al.⑥ |
| Total          | | 717 | 799 | |

ND: data not determined, available or used in this study.

admixed individuals will lead to an increase in the degree of mitonuclear incompatibility (Fig. 1a). For instance, incompatibility might arise between mtDNA origins of replication and nuclear-encoded mtDNA replication machinery③④⑤ and therefore, might lead to a decrease in mtDNA replication efficiency. If our hypothesis is correct, then mtDNA copy number should decrease with increasing degree of mitonuclear DNA discordance in admixed individuals (Fig. 1a). To test this hypothesis, we determined the mtDNA copy number based on the ratio between the depth of reads aligning to the mtDNA and the nuclear genome from publicly available sequence alignments of DNA extracted from lymphoblastoid cell lines (LCLs; see Methods). To compute the degree of mitonuclear DNA discordance, we first determined mtDNA haplogroups and assigned their origins to African, European or Native American ancestry (see Methods and Fig. 2c,d). Next, nuclear ancestry proportions were estimated using unsupervised clustering with ADMIXTURE⑥⑦ from single nucleotide polymorphism (SNP) genotype data, which included individuals representing ancestry from three source populations: (1) Europeans (represented by Utah residents of Northern and Western European ancestry), (2) Africans (represented by Yoruba from Nigeria) and (3) Native Americans (represented by the Aymara, Nahuan, Quechua and Maya; see Methods for more details and Fig. 2a,b). The degree of mitonuclear discordance was computed as the proportion of nuclear ancestry of a different origin from the mtDNA. For example, for individuals carrying the African mtDNA haplogroup (L), mitonuclear discordance is one minus the African ancestry fraction in the nuclear genome.

We found a significant negative correlation between mtDNA copy number and mitonuclear DNA discordance in admixed individuals (Fig. 3a; β = −0.193, one-sided P value = 1.14×10⁻⁴⁹), suggesting that mtDNA copy number declines as nuclear ancestry becomes increasingly dissimilar to mtDNA ancestry. The y intercept, that is, when mitonuclear DNA discordance is zero, is similar to the median mtDNA copy number in the individuals from source populations, Europeans (represented by CEU) and Africans (represented by YRI), who are not admixed (Fig. 3b). The negative correlation between mtDNA copy number and mitonuclear DNA discordance in admixed individuals is consistent across mtDNA haplogroups from three different geographic origins—Native American (β = −1.06; P = 1.53×10⁻⁴⁹), African (β = −0.63; P = 0.026) and European (β = −0.30; P = 0.370)—even though it is not statistically significant in every case (Supplementary Fig. 1). Moreover, copy number for mtDNA of one geographic origin decreases with increasing nuclear ancestry from each of the other two geographic origins (panels outside of the top-left to bottom-right diagonal in Supplementary Fig. 1). For instance, the mtDNA copy number decreases with increase in both African and European ancestry in individuals with Native American mtDNA haplogroups. Overall, our results show that mtDNA copy number is negatively correlated with mitonuclear DNA discordance, consistent with mitonuclear incompatibility in admixed individuals (Fig. 1a). The lack of statistical power, especially in individuals carrying European haplogroups, is probably due to the sex-biased nature of admixture in Americans.

It is established that European gene flow in the Americas was male-biased①②③④⑤⑥⑦⑧, and we replicate this observation in the studied admixed populations. We show that African mtDNA haplogroups are more frequent in the African Caribbeans (frequency = 0.94) and African Americans (0.93), and Native American mtDNA haplogroups are more frequent in Colombians (0.87), Mexicans (0.87), Peruvians (0.95) and Puerto Ricans (0.67), consistent with predominantly non-European female contribution to admixture (Fig. 4). We validate this by showing that the observed mtDNA haplogroup frequencies conform with expectations based on the proportion of females from each of the three source populations derived by comparing the ancestry fractions on the autosome and X chromosome (see Fig. 4 and Methods). Two exceptions to this result are Colombians and Puerto Ricans, in whom the frequency of Native American mtDNA haplogroups is much higher than expected (Fig. 4; African Caribbeans also exhibit a departure from expectation, but it
This is an interesting observation that cannot be explained simply by post-admixture drift (see Methods and Supplementary Fig. 2) and, therefore, requires further exploration. Nevertheless, it is clear that the mtDNA haplogroups in American populations are predominantly non-European.

We leveraged this finding to test whether selection has acted against mitonuclear incompatibilities arising in admixed populations in the Americas. Our reasoning is that, if the mtDNA haplogroups in these populations are predominantly non-European, co-adapted mitonuclear combinations might be disrupted in individuals carrying European alleles at nuclear-encoded mitochondrial genes, resulting in a reduction in fitness. Hence, post-admixture selection would favour, and lead to an enrichment of, non-European ancestry at these genes compared to the rest of the genome (Fig. 1b).

To evaluate this prediction, we analysed local, that is SNP-specific, ancestry inferred from the genotypes of the studied admixed populations. Admixed population might contain varying levels of ancestry from multiple ancestral populations, whereas mtDNA derives its ancestry from only one population. Increasing discordance (concordance) between mitochondrial and nuclear ancestry in individuals can lead to increasing (decreasing) levels of mitonuclear incompatibility, which can have phenotypic consequences. Sex-biased admixture can lead to varying levels of ancestry proportions on different chromosome types on the basis of their modes of inheritance. In extreme cases, the mtDNA ancestry can be entirely from one population even though the nuclear ancestry is highly admixed. This discordance in ancestry between the nuclear and mitochondrial genomes might result in selective pressure for 'matching' ancestry at nuclear-encoded mitochondrial genes.
Next, we computed the deviations in African, European and Native American ancestry at each SNP by subtracting them from the mean global nuclear ancestry proportion separately for each admixed group (see Methods). For example, if the frequency of African and European ancestry at a SNP is 0.3 and 0.7, respectively, in a population where both source populations contributed equally (that is, mean admixture fraction is 0.5), the deviations are $-0.2$ and $0.2$, respectively.

The expected deviation in ancestry across neutral loci is zero and its variance due to drift is a function of the time since admixture and effective population size. For nuclear-encoded mitochondrial genes, we expect local ancestry to deviate, on average, from neutral expectations in favour of the source population contributing the highest proportion of mtDNA. To assess whether the data support this expectation, we downloaded a list of nuclear genes from MitoCarta 2.0 and split them into mitochondrial ($N = 960$) versus non-mitochondrial ($N = 17,456$, see Supplementary Table 5). The mitochondrial genes encode proteins with experimental evidence of mitochondrial localization, whereas the non-mitochondrial genes have no such evidence. We next followed a published approach and further split the 960 mitochondrial genes into two subsets—167 'high-mt' genes (known genes encoding proteins that are part of the mtDNA replication and transcription machinery, and of ribosomal and OXPHOS complexes) and remaining 793 'low-mt' genes (Supplementary Table 5).

---

**Fig. 2** | Ternary plots showing the distribution of African, European and Native American ancestry in the samples analysed (samples sizes are shown in parentheses). a. ADMIXTURE components 1, 2 and 3 correspond to European (EUR), African (AFR) and Native American (NAT) ancestry, respectively. The grey points are the admixed samples from the 1000 Genomes dataset and the coloured points are samples serving as proxies for source populations. b. The ancestry structure of admixed populations. c. Distribution of mtDNA haplogroups among admixed individuals. d. mtDNA haplogroups grouped by region where they are thought to have been most commonly found prior to admixture. ACB, African Caribbeans from Barbados; ASW, African Americans from South Western USA; CLM, Colombians from Colombia; MXL, Mexicans from Los Angeles, USA; PEL, Peruvians from Lima, Peru; PUR, Puerto Ricans from Puerto Rico.

**Fig. 3** | mtDNA copy number in admixed and non-admixed populations. a. mtDNA copy number is negatively correlated with the discordance between mitochondrial and nuclear DNA ancestry. Standardized $\beta$ coefficient, $t$-statistic, one-sided $P$ value and correlation coefficient are shown. The discordance score is one minus the ancestry proportion from the same source population as the mtDNA haplogroup. The red dashed line is the median mtDNA copy number calculated across the individuals from European (CEU) and African (YRI) source populations, plotted separately in b.
We calculated gene-specific ancestry deviation by averaging across all SNPs in 5-Mb windows flanking the midpoint of the coding sequence. Subsequently, we tested whether the mean ancestry at mitochondrial genes (high-mt or low-mt) deviates significantly from that at non-mitochondrial genes, which we used as a proxy for the amount of drift experienced by neutrally evolving loci. We used this approach, as opposed to a genome-wide scan because our ability to detect such a signature at individual loci is limited due to: (1) the large number of genes involved in mitochondrial function (>1,000)\(^{16}\), (2) the low amount of genetic differentiation among human populations (\(F_{ST} \approx 0.1\))\(^{16}\), (3) only a relatively recent admixture history in the populations analysed\(^{16,37,38}\) and (4) the large number of tests involved in an agnostic genome-wide search. We used a bootstrap approach to generate a distribution of the mean deviation in ancestry for each functional gene category (for 167 high-mt, 793 low-mt and 17,456 non-mitochondrial genes, see Methods for more details). We show using simulations (see Methods) that our approach has adequate statistical power to detect selection with the given sample sizes at these time scales (Supplementary Fig. 3).

We found a significant enrichment in Native American ancestry at high-mt genes in Puerto Ricans, and a significant enrichment in African ancestry at high-mt genes in African Americans (Fig. 5). Because Native American mtDNA haplogroups are more frequent in Puerto Ricans and African mtDNA haplogroups are more frequent in African Americans (Fig. 4), these results are consistent with our prediction (Fig. 1b). However, in three out of six admixed populations analysed, the mean ancestry of nuclear-encoded mitochondrial genes does not significantly deviate from expectation (the 95% bootstrapped confidence interval spans the zero line, see Fig. 5), consistent with no evidence of selection for local ancestry at such genes. Surprisingly, we also observe a significant enrichment in European ancestry at high-mt genes in Mexicans (Fig. 5), which contradicts our hypothesis. This lack of a consistent pattern of enrichment across admixed populations paints a complex picture of selection acting on nuclear-encoded mitochondrial genes, requiring further exploration.

**Discussion**

We report that mtDNA copy number decreases as the nuclear and mtDNA become increasingly dissimilar in ancestry among human admixed individuals. Because mitonuclear compatibility is important for regulation of mtDNA copy number, its reduction due to mitonuclear DNA discordance might reflect the incompatibility between the origins of mtDNA replication and nuclear-encoded proteins involved in mtDNA replication\(^{46}\). Future functional experiments should validate effects of mtDNA haplogroup and nuclear DNA combinations on mtDNA copy number directly. This can be performed, for instance, in cybrids—cellular hybrid lines carrying the same nuclear genetic background but different mtDNA haplotypes\(^{47}\). Our results and the suggested future experiments are important because there is a growing body of evidence implicating mtDNA copy number as a biomarker for different diseases including diabetes\(^{54}\), Parkinson's disease\(^{46}\), male infertility\(^{54}\), depression\(^{51}\), ageing\(^{52}\) and many types of cancer\(^{53–55}\), even though it may not always correlate with mitochondrial content and activity\(^{56}\).

Multiple biological and environmental sources can contribute to variation in mtDNA copy number. However, the design of our study minimizes these. For instance, even though mtDNA copy number is known to vary among tissues\(^{56}\), this could not have contributed to our estimates because we made our measurements in cell lines of the same type—LCLs. mtDNA copy number is also known to vary depending on the physiological conditions of an individual, which may be driven by the environment (for example, stress\(^{57}\)). However, the use of cell lines that are cultured under standard laboratory conditions should largely remove mtDNA copy number variation originally present among samples due to these factors. Having said that, the observed correlation between mitonuclear DNA discordance and mtDNA copy number that we found should be replicated.
and explored in future studies, ideally with mtDNA copy number measured across biological and technical replicates, to minimize biological variation and measurement error. It would also be interesting to test the effect of mitonuclear DNA discordance on other mitochondrial phenotypes such as mitochondrial morphology and rate of ATP production in this and other cell types. However, the caveat of using cell lines is that it is difficult to say whether or not the effects of mitonuclear ancestry discordance exist in vivo in the individuals from whom they are derived. Therefore, future in vivo studies will also be of great value.

Corroborating previous studies\textsuperscript{36--39}, we found a significant sex bias in admixture for all six admixed populations studied. In most admixed populations we found a high level of concordance between the female contribution from each source population estimated from nuclear genetic markers and that estimated from mtDNA. However, in Colombians and Puerto Ricans, the frequency of

---

**Fig. 5 | Systematic deviations in local ancestry for different functional categories of genes.** The y axis shows local ancestry deviation and the x axis lists the functional categories. High-mt (167 genes) are nuclear genes that encode important subunits of mitochondrial replication, transcription and the OXPHOS complexes. Low-mt (793 genes) are nuclear genes that were inferred to have mitochondrial function by Mitocarta 2.0\textsuperscript{42}, but are not part of the high-mt gene set. Non-mt (17,456 genes) are genes that do not have known or inferred mitochondrial function based on Mitocarta 2.0\textsuperscript{42}. A block bootstrap approach (see Methods) was used to generate the distributions. Briefly, we sampled 167 windows of 5 Mb spanning the genes, with replacement, within each category and calculated the mean local ancestry deviation across these windows. This was repeated for 1,000 bootstraps to generate the distributions. The horizontal bars indicate the empirical 95% confidence interval of the mean ancestry deviation.
Native American ancestry in mtDNA is significantly higher than expected based on the ancestry information in the nuclear genome. This deviation cannot be explained by genetic drift experienced by mtDNA since admixture due to its smaller effective population size, and could be due to selection for the Native American haplogroups in Colombians and Puerto Ricans, if the Native American mtDNA was better adapted to the environment. Indeed, climate adaptation may have played a role in mtDNA diversification across human populations. This observation, which has not been reported previously, requires further investigation.

In line with the hypothesis that selection against mitonuclear incompatibility should disfavor European ancestry at nuclear-encoded mitochondrial genes, given that the mtDNA haplogroups in admixed American populations are predominantly African or Native American, we also tested for an enrichment for non-European ancestry at nuclear-encoded mitochondrial genes in admixed populations. We found that the ancestry profiles at nuclear-encoded mitochondrial genes in three out of the six admixed populations studied (African Caribbeans, Colombians and Peruvians) do not deviate from neutral expectation; that is, they are indistinguishable from the genomic background (Fig. 5). This could mean that there has been no selection on mitonuclear interactions in these populations. We observed an enrichment of Native American ancestry in Puerto Ricans, and an enrichment of African ancestry in African Americans, at nuclear-encoded mitochondrial genes. Because Puerto Ricans predominantly carry Native American mtDNA haplogroups, whereas African Americans primarily carry African haplogroups, this observation is consistent with our expectation of selection against mitonuclear incompatibility. However, we also found a significant enrichment of European ancestry in Mexicans—another population with predominantly Native American mtDNA haplogroups. This observation is inconsistent with selection against mitonuclear incompatibility in admixed populations. This complex picture among different admixed populations suggests that other competing selective pressures on mitochondrial genes might be involved and it may not be straightforward to detect signatures of mitonuclear incompatibility from deviations in local ancestry alone.

Our ability to detect a signature of ancestry enrichment driven by mitonuclear incompatibility is influenced by several factors including degree of sex bias, time since admixture, effective population size, selection strength and the number of loci under selection. One potential limitation of our analysis is that we assume that all nuclear-encoded mitochondrial genes have equal effect sizes for mitochondrial function. Since mitochondrial function is a highly complex trait, this assumption is probably incorrect. A more powerful way of testing for ancestry enrichment across these genes would be to weigh the contribution of each locus by its effect size, as we demonstrate using simulations (Supplementary Fig. 3). Unfortunately, we do not know what these effect sizes are because genome-wide association studies of mitochondrial phenotypes, such as mtDNA copy number and rate of ATP production, are yet to be conducted in humans. While systematic collection of such data for large cohorts of individuals is pending, it would also be highly informative to explore the effects of mitonuclear interactions on various health-related phenotypes in large-scale datasets such as the UK Biobank.

In conclusion, our results demonstrate that discordance between mtDNA and nuclear ancestry for mitochondrial genes might contribute to phenotypic variation in admixed individuals. Mitonuclear DNA discordance contributes to disease phenotypes in non-admixed individuals (reviewed in). Therefore, we expect this phenomenon to contribute even more to disease variation in admixed individuals, which needs to be evaluated in future studies. Such evaluation is also critical for making advances in mitochondrial replacement therapy (MRT), a technique in which the mtDNA carrying disease-associated mutations in a patient’s oocyte is replaced with mtDNA from a healthy donor oocyte. Despite the success of MRT, many human and non-human primate embryos created via mitochondrial replacement do not develop normally. Mitochondrial replacement can also lead to detrimental effects on growth, development, respiration, metabolism, ageing, fertility and survival in non-primate animals. Human cybrid lines also show variation in mtDNA copy number, ATP turnover rates, reactive oxygen species production and expression of OXPHOS genes. Despite these observations, the degree to which the mitochondrial haplogroup of a donor should match the genetic background of the ‘nuclear’ parents in MRT in humans remains unanswered. The answer to this question is even less clear for admixed individuals, or for individuals whose parents belong to different ethnic groups, as their nuclear genomes originate from different populations. Our results highlight the potential of studying admixed individuals to better understand phenotypic effects of mitonuclear DNA discordance, which will be useful in elucidating MRT-associated risks and evaluating disease susceptibility in contemporary admixed and non-admixed populations.

Methods

Global ancestry and mtDNA haplogroup. We downloaded the 1000 Genomes phase 3 vcf files and retained individuals who belonged to one of the following admixed populations: (1) African Americans from the Southwestern US; (2) African Caribbeans from Barbados; (3) Colombians from Medellin, Colombia; (4) Mexicans from Los Angeles; (5) Peruvians from Lima, Peru and (6) Puerto Ricans from Puerto Rico. We also used data from Utah residents of Northern and Western European ancestry and from Yorubans from Ibadan, Nigeria, who serve as proxies for the European and African source populations, respectively. To represent the Native American ancestry component, we analysed previously published genotype data from the following four groups: (1) Aymara; (2) Nahuan; (3) Quechua and (4) Maya. The data is summarized in Table 1. We determined global ancestry—the overall contribution of African, European and Native American ancestry to the nuclear genome of each individual using ADMIXTURE (Fig. 2). For this purpose, we merged the 1000 Genomes genotype data with the genotype data from Native American populations published by Mao and colleagues. We included SNPs that overlapped across both datasets (a total of 691,435 SNPs) and converted the genotypes to binary format for use with PLINK. We further removed palindromic (A/T, G/C) SNPs to ensure strand consistency across both datasets. Subsequently, the two datasets were merged and SNPs were pruned for linkage disequilibrium (PLINK’s –indep-pairwise function with window size of 50, step size of five and r threshold of 0.16970), which resulted in 88,442 SNPs. We ran unsupervised ADMIXTURE analysis on this genotype dataset for k = 1, 2, 3, 4 and 5, and used the ancestry proportions for k = 3 for all downstream analyses, since it had the lowest cross-validation error (Supplementary Figs. 4 and 5). In agreement with previous studies, the individuals from admixed populations derive their genetic ancestry from three primary source populations: Native American, European and African. The proportion of ancestry from each source population varies among the admixed populations (Fig. 2, Supplementary Fig. 4 and Supplementary Tables 2 and 3) because of differences in admixture histories.

We used Haploreg version 2.02 to determine the mtDNA haplogroup, for individuals from the 1000 Genomes Project only, as the individuals from Mao et al. were not genotyped for mitochondrial variants. All mtDNA haplogroups were called with high accuracy (minimum posterior probability of 0.78). To increase statistical power, we grouped together haplogroups belonging to the same major haplogroup (for example, L1b1a3 was grouped with L3d1b1 under the L major haplogroup Supplementary Table 4). We further grouped major haplogroups into regional groups, corresponding to pre-colonization origins (L: African; A, B, C, D: Native American; H, J, K, T, U, V, W, European) (Fig. 2c and Supplementary Table 4). We excluded two individuals (HG01272 and NA19982), whose mtDNA was predicted to belong to the H haplogroup, which is most frequently found in South Asia. This way, both nuclear and mitochondrial ancestries were categorized into only three regional groups: Native American, European and African (Fig. 2d and Supplementary Table 4). Among the admixed individuals, African (L) and Native American (A, B, C and D) mtDNA haplogroups are more frequent than European haplogroups (H, J, K, T, U, V and W, see Supplementary Fig. 6), consistent with female bias in the non-European contribution.

We calculated mitonuclear discordance in ancestry as the fraction of the nuclear genome from a different geographical origin than that of the mtDNA haplogroup. For instance, for individuals with Native American mtDNA haplogroups, mitonuclear DNA discordance is the proportion of nuclear ancestry that is not Native American (that is, African plus European).
mtDNA copy number estimation. Given the average sequencing depth of the autosomes and mtDNA, and the fact that there are two autosomal chromosome copies per cell, we can compute the number of copies of mtDNA per cell:

\[
\text{mtDNA copy number} = \frac{\text{Average mtDNA depth}}{\text{Average autosomal depth}} \times 2
\]

(1)

In equation (2) for average depth, \(N\) is the total number of reads aligning to the chromosome, \(L\) is the average length of a read, and \(G\) is the size of the chromosome in base pairs. Similar approaches have been used previously to accurately estimate mtDNA copy number per cell\(^{25}\). We first calculated mtDNA copy number for each autosomal chromosome separately and then calculated the mean across all chromosomes.

A subset of samples (a total of 24) in the 1000 Genomes Project Data were sequenced at both low (\(\times 2\)–4) and high coverage (\(\times 20\)–40). We used these to validate whether mtDNA copy numbers calculated from the low-versus-high-coverage alignments agree with each other. As shown in Supplementary Fig. 7, the copy numbers generally agree, with a few exceptions. Some samples show appreciable mtDNA copy numbers when calculated using the high-coverage alignments but low copy numbers when calculated using the low-coverage alignments (Supplementary Fig. 7). The source annotation of these samples indicates that some of them (samples for which this information is available) were sequenced from peripheral blood mononuclear cells (PBMCs), instead of LCLs, the only two types of cell line used to carry out sequencing for the 1000 Genomes Project (Supplementary Fig. 7 and Supplementary Table 1). The difference in mtDNA copy number seen between the two cell lines is consistent with previous observations that LCLs are known to carry significantly higher mtDNA copy numbers than PBMCs\(^{25,26}\). Because annotation for the source DNA is not available for all samples (Supplementary Table 1), we plotted the density of mtDNA copy number calculated from the low-coverage alignments and observed a clear separation between samples sequenced from PBMCs and LCLs (Supplementary Fig. 8). Based on this separation, we inferred that samples with more than 250 mtDNA copies per cell are LCLs (Supplementary Table 1) and only retained these for analysis with mtDNA copy number to exclude samples that were sequenced from PBMCs in an effort to limit variation due to DNA source. After removing such samples, the correlation coefficient between mtDNA copy number from low-coverage and high-coverage alignments is 0.71, as opposed to 0.66, before removing them.

An advantage of using LCLs to study mtDNA copy number variation is that they exhibit high mtDNA content and elevated expression of genes involved in mtDNA replication and transcription, as well as of respiratory genes, consistent with elevated mitochondrial biogenesis\(^3\). Furthermore, because they are maintained following standard protocols in a laboratory, variation due to differences in individuals’ environments, from whom the LCLs are derived, is unlikely to systematically confound our analysis of mtDNA copy number.

Local ancestry. Local ancestry for autosomes was generated using RFMix\(^7\) by the 1000 Genomes Project admixture working group as described by Martin and colleagues\(^{25}\) (https://personal.broadinstitute.org/armartin/tgp_admixture/tnp_pop/). For downstream analyses, we masked out regions of the genome where local ancestry was called with less than 0.9 maximum posterior probability. Global ancestry calculated from these filtered local ancestry calls is highly correlated with global ancestry estimated using ADMIXTURE\(^7\) (Supplementary Fig. 9).

Sex-biased admixture. Males and females may contribute different proportions of ancestry from the source populations during admixture. This sex bias can lead to a situation where different chromosomes give different estimates of ancestry proportions (Fig. 1b). For example, for every three X chromosomes, two are contributed by females while one is contributed by males. This is in contrast to autosomes, which are contributed by both sexes equally. Therefore, any sex bias in the admixture composition of the ancestral population will influence the ancestry fraction of the X chromosome more than that of autosomes (Fig. 1b). We took advantage of the difference in admixture fraction between the X chromosome and autosomes to estimate the relative contribution of males and females from each of the three relevant ancestral groups (African, European and Native American) using the approach described in\(^1\). If \(f_{\text{female}}\) is the proportion of ancestry of the admixed group, which were female and from population \(i\) and \(f_{\text{male}}\) who were male and from population \(i\), we assume that for each admixed group (for example, Puerto Ricans, Colombians and so on), \(f_{\text{female}} + f_{\text{male}} = 1\), the mean autosomal ancestry fraction from population \(i\), where \(f_{\text{female}}\) and \(f_{\text{male}}\) the expected ancestry fraction for the X chromosome in the population, \(f_{\text{X}}\) is\(^1\):

\[
f_{\text{female}} = \frac{f_{\text{male}} + 2f_{\text{female}}}{1.5}
\]

(3)

We performed a grid search for the values of \(f_{\text{female}}\) and \(f_{\text{male}}\) that equal \(f_{\text{X}}\) and minimize the squared deviation between \(f_{\text{X}}\); predicted using equation (3) and the mean plausibly adjusted X-chromosomal ancestry fraction inferred from genotype data. Estimated values of \(f_{\text{female}}\) and \(f_{\text{male}}\) and confidence intervals around these estimates, generated by bootstrapping (sampling individuals with replacement), are shown in Supplementary Fig. 10.

Simulation of drift in mtDNA since admixture. Our finding that the frequency of Native American mtDNA in Puerto Ricans is very similar to that of Colombians suggests that Native American ancestry in African Americans can be explained by at least two factors. First, since mtDNA represents a single genealogical history, it yields a ‘noisier’ estimate of the proportion of females from each parental population than the estimate based on autosomal and X-chromosomal loci, which represent multiple genealogical histories because of recombination and independent assortment. Second, we expect larger fluctuations in mtDNA ancestry as a result of drift because of its smaller effective population size compared to that of autosomal or X-chromosomal loci\(^{25}\). To test whether genetic drift since the occurrence of admixture can account for the deviation in mtDNA frequency in Puerto Ricans and Colombians, we simulated the expected amount of drift in local ancestry since admixture in Puerto Ricans and Colombians using a simple hybrid-isolation demographic model\(^8\). In this model, all three parental groups mixed at some time in the past in proportions equal to the mean global ancestry averaged across individuals in the population. In subsequent generations, 2\(N\) autosomes and 2\(N\) copies of mtDNA were drawn from a multinomial distribution with the probability of drawing European, African or Native American ancestry determined by the relative ancestry proportions in the previous generation. After \(G\) generations, the final ancestry frequency at each locus, averaged across individuals in the population, was recorded. This process was repeated 10,000 times to simulate the amount of drift experienced by 10,000 independent loci. We used 17 generations for \(G\) in Puerto Ricans and 14 generations for \(G\) in Colombians, similar to values estimated for these populations in ref. \(^9\) and 1,250 for \(N\) following ref. \(^9\). We show in Supplementary Fig. 11 that the simulated distribution of local ancestry using these parameters matches the observed distribution of local ancestry well.

The assumed model of admixture dynamics has several limitations, which could lead to biased estimates of the female and male contributions. Specifically, we assume a hybrid-isolation model with equal reproductive variance and similar generation times between males and females. First, models that incorporate continuous gene flow are likely to yield slightly different estimates, especially if admixture occurred recently; that is, within the last five generations\(^4\). Because admixture for the populations used in this study started much earlier (>10 generations ago\(^{25,26}\)), this is not a major concern in our case. Second, shorter generation times in females than in males would result in a smaller effective population size for the mtDNA compared to the nuclear genome\(^25\). This would lead to stronger drift in mtDNA, which might explain the observed discrepancy in ancestry proportions between mtDNA and the nuclear genome in Colombians and Puerto Ricans without invoking selection. Third, men typically tend to have higher variance in reproductive success relative to women\(^{27,28}\), which would increase the effective population size of the mtDNA compared to the nuclear genome.

A detailed discourse of how these competing processes affect the inference of sex bias in admixture dynamics is beyond the scope of this paper, but needs to be explored in future studies.

Local ancestry enrichment in nuclear-encoded mitochondrial genes. Strong deviations in local ancestry frequency in a population can be indicative of post-admixture selection (for example, Tang et al.\(^4\)). Our ability to detect such a signature at individual loci is limited, as we explain in the main text of the manuscript. A more feasible approach is to test for significant deviations in ancestry across a number of loci relative to the genomic background. This is the underlying basis for the approach used in ref. \(^1\), as well as our approach to detecting selection on nuclear-encoded mitochondrial genes. For each admixed population, we begin by calculating the frequency of Native American, European and African ancestry at every SNP in the genome by averaging across all individuals in the population. These were subtracted from the mean ancestry frequency across all SNPs (the expectation) to calculate the deviation in ancestry at each SNP. We used the gene annotation provided in MitoCarta 2.0 and generated a list of mitochondrial and non-mitochondrial genes\(^26\). We further split the mitochondrial genes into ‘high-mt’ (\(N = 167\) and ‘low-mt’ (\(N = 793\) subsets, as curated in\(^19\)). An unweighted block bootstrap approach was used to generate the distribution of average deviation in local ancestry, for each gene category. We generated windows of 5\(\text{Mb}\) spanning each gene (4.5\(\text{Mb}\) on either side of a gene’s midpoint) to take into account linkage disequilibrium among SNPs. Subsequently, we used bedtools\(^{88}\) to intersect SNPs, at which local ancestry deviation was minimized the squared deviation between \(f_{\text{X}}\); predicted using equation (3) and the mean plausibly adjusted X-chromosomal ancestry fraction inferred from genotype data. Estimated values of \(f_{\text{female}}\) and \(f_{\text{male}}\) and confidence intervals around these estimates, generated by bootstrapping (sampling individuals with replacement), are shown in Supplementary Fig. 10.
(that is, high-nt) and calculated the mean ancestry deviation first for each window, and then across windows. This process was repeated 1,000 times to generate a distribution of mean deviation in local ancestry for each gene category.

Statistical power to detect selection. We carried out forward simulations to test whether we have adequate statistical power to detect selection with our approach. We started by simulating 167 independent biallelic loci, where, at each locus, the favoured allele has a starting frequency $f_i$ and an effect size, $β$, randomly drawn from a standard normal distribution. At each locus, the relative fitness of the favoured allele is $w = 1 + βfi$, where $s$ is the selection coefficient. In every generation, the allele frequency after a round of genetic selection on survival, $fi$, was calculated as follows:

$fi = \frac{fi - w}{fi - w + (1 - f)}$

In the next generation, we sampled $2N(2x1,250=2,500)$ chromosomes at random from a binomial distribution where the probability of drawing a chromosome with a favoured allele is equal to $fi$. We repeated this until the 6th generation, at which point we sampled 2 chromosomes at random, where $N$ is the sample size (that is, the number of diploid individuals). This last step was carried out to take into account sampling error.

We chose 0.13 for $fi$, equal to the mean Native American ancestry proportion in Puerto Ricans and used a range of values for $s$ ($0.001, 0.002, 0.003, 0.004, 0.005, 0.01$), the selection coefficient. We set $s$ at 17 generations and $N$ at 1,250, to match the degree of drift experienced by Puerto Ricans, supported by the comparison of simulated versus observed distribution of local ancestry (Supplementary Fig. 11). We carried out 1,000 simulations and, for each simulation, generated the distribution of the mean deviation in ancestry using an unweighted bootstrap approach similar to that used for the real data. Briefly, for every bootstrap, 167 loci were sampled with replacement and the deviation for each locus was calculated by subtracting the expected allele frequency, equal to the frequency at the start of the simulation, from the observed frequency, equal to the frequency at the 6th generation. The deviation was averaged across loci and the sampling was repeated 1,000 times to generate a distribution around the unweighted mean. To test the idea that knowledge of effect sizes might provide more statistical power to detect selection as mentioned in the discussion section, we also calculated weighted $d$-values by multiplying the deviation at each locus by its effect size prior to summation.

We show in Supplementary Fig. 3 that, even at relatively low selection pressures ($s = 0.005$), we have more than 80% power to detect selection at the 0.05 level of significance using the unweighted approach with a sample size of 100, which is comparable to the sample sizes used in this study (Table 1). We also show that the weighted approach provides more statistical power to detect selection than using an unweighted approach. Last, we show that in the absence of selection ($s=0$), the false positive rate is equal to the level of significance (–0.05).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All analyses were conducted using publicly available data. The 1000 Genomes Project data are available on its FTP site. Intermediate files and code have been made publicly available on github (https://github.com/makovalab-psu/Mito_nuclear_incompatibility).
38. Ruiz-Linares, A. et al. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet.* **10**, e1004572 (2014).
39. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
40. Long, J. C. The genetic structure of admixed populations. *Genetics** **127**, 417–428 (1991).
41. Hartl, D. L., Clark, A. G. & Clark, A. G. Principles of Population Genetics (Sinauer Associates, Sunderland, 1997).
42. Calvo, S. E., Clausner, K. R. & Mootha, V. K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.* **44**, D1251–D1257 (2015).
43. Pagliarini, D. J. et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell** **134**, 112–123 (2008).
44. Schatz, G. The protein import machinery of mitochondria. *Protein Sci.* **2**, 141–146 (2008).
45. Bhattacharya, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
46. Holt, I. J. & Eyre, W. A. Human mitochondrial DNA replication. *Cold Spring Harb. Perspect. Biol.* **4**, a012971 (2012).
47. Trounce, I., Neill, S. & Wallace, D. C. Cytoplasmic transfer of the mitochondrial genome as a means for evolutionary adaptation and implications for population susceptibility to diseases. *Biochim. Biophys. Acta** **1842**, 1787–1796 (2015).
48. Lee, H. K. et al. Decreased mitochondrial DNA content in peripheral blood precedes the development of non-insulin-dependent diabetes mellitus. *Diabetes Res. Clin. Pract.* **42**, 161–167 (1998).
49. Pyle, A. et al. Reduced mitochondrial DNA copy number is a biomarker of Parkinson's disease. *Neurology** **38**, 216.216–216 (2016).
50. Jiang, M. et al. Increased total mtDNA copy number cures male infertility despite unaltered mtDNA mutation load. *Cell. Metab.* **26**, 429–436.e4 (2017).
51. Cai, N. et al. Molecular signatures of major depression. *Curr. Biol.* **25**, 1146–1156 (2015).
52. Mengel-From, J. et al. Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. *Hum. Genet.* **133**, 1149–1159 (2014).
53. Hosc{"o}gd, H. D. et al. Mitochondrial DNA copy number and lung cancer risk in a prospective cohort study. *Carcinogenesis** **31**, 847–849 (2010).
54. Yu, M. Generation, function and diagnostic value of mitochondrial DNA copy number alterations in human cancers. *Life Sci.** **89**, 65–71 (2011).
55. Shen, J., Platek, M., Mahasneh, A., Ambroseon, C. B. & Zhao, H. Mitochondrial copy number and risk of breast cancer: a pilot study. *Mitochondrion** **10**, 62–68 (2010).
56. Larsen, S. et al. Biomarkers of mitochondrial content in skeletal muscle of healthy young humans. *J. Physiol** **390**, 3170–3177 (2015).
57. Wachsmuth, M., Huebner, A., Li, M., Madaea, B. & Stoneking, M. Age-related and heteroplasmic-related variation in human mtDNA copy number. *PLoS Genet.* **12**, e1005939 (2016).
58. Cai, N. et al. Genetic control over mtDNA and its relationship to major depressive disorder. *Curr. Biol.* **25**, 3170–3177 (2015).
59. Mishmar, D. et al. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl Acad. Sci. USA** **100**, 171–176 (2003).
60. Balloux, F., Handley, L. J. L., Jombart, T., Liu, H. & Manica, A. Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proc. R. Soc. B** **276**, 3447–3455 (2009).
61. Sudlow, C. et al. UK Biobank: an open access resource for identifying the genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am. J. Hum. Genet.* **96**, 37–53 (2015).
62. Lind, J. M. et al. Elevated male European and female African contributions to the genomes of African American individuals. *Hum. Genet.** **120**, 713–722 (2007).
63. Jobling, M., Hellot, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. *Human Evolutionary Genetics* 2nd edn (Garland Science, New York and London, 2013).
64. Pfaal, C. L. et al. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am. J. Hum. Genet.* **68**, 198–207 (2001).
65. Tang, H. et al. Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am. J. Hum. Genet.* **81**, 626–633 (2007).
66. Goldberg, A. & Rosenberg, N. A. Beyond 2/3 and 1/3: the complex signatures of sex-biased admixture on the X chromosome. *Genetics** **201**, 263–279 (2013).
67. Schacht, K., Durant, E. Y., Macpherson, J., Reich, D. & Mountain, J. L. The genetic ancestry of African Americans, Latinos, and European Americans to the genomes of African American individuals. *Hum. Genet.** **120**, 713–722 (2007).
68. Jobling, M., Hellot, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. *Human Evolutionary Genetics: Origins, Peoples & Disease* (Garland Science, New York and London, 2013).
69. Brown, G. R., Laland, K. N. & Mulder, M. B. Rateman's principles and human sex roles. *Trends Ecol. Evol.* **24**, 297–304 (2009).
70. Bezit, L. Means, variances, and ranges in reproductive success: comparative evidence. *Evol. Hum. Behav.* **33**, 309–317 (2012).
71. Quinlan, A. R. *BEDTools: The swiss-army tool for genome feature analysis* (Garland Science, New York and London, 2013).
72. Nielsen, R. & Statkin, M. An Introduction to Population Genetics: Theory and Applications (Sinauer Associates, Sunderland, 2013).

### Acknowledgements
We thank R. Nielsen, M. Shriver, W. Chuse and S. Craig for their comments on the manuscript. We would also like to thank E. Torres Gonzalez for reviewing the code. This project was supported by a seed grant awarded to A.A.Z. and K.D.M. from the Center of Human Evolution and Diversity (CHED) at The Pennsylvania State University and by a grant from NIH (R01GM116044). Additional funding was provided by the Eberly College of Sciences, The Huck Institute of Life Sciences and the Institute for CyberScience at Penn State, as well as, in part, under grants from the Pennsylvania Department of Health using Tobacco Settlement and CURE Funds. The department specifically disclaims any responsibility for any analyses, responsibility or conclusions.

### Author contributions
A.A.Z. and K.D.M. conceived the study. A.A.Z. carried out analyses. A.A.Z. and K.D.M. wrote and edited the paper.

### Competing interests
The authors declare no competing interests.

### Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0766-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to K.D.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2018
Corresponding author(s): Kateryna Makova

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- **n/a**: Confirmed
- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] An indication of 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
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- [ ] 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 statistics 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. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- [ ] 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 d, Pearson's r), indicating how they were calculated
- [ ] Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about [availability of computer code]

**Data collection**: No software was used. Data are publicly available.

**Data analysis**

- R version 3.4.3
- PLINK 1.9
- samtools 1.3.1
- bedtools 2.27.1
- ADMIXTURE 1.3.0
- vcftools 0.1.14

Code is available on https://github.com/makovalab-psu/Mito_nuclear_incompatibility

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All analyses were conducted using publicly available data. The 1000 Genomes Project data are available on their ftp site. Intermediate files and code have also been made publicly available on github: (https://github.com/makovalab-psu/Mito_nuclear_incompatibility).

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size
No sample size calculations were carried out. We analyzed publicly available data. As a result, we were restricted to the sample size of this dataset. We carried out simulations to test whether the available sample sizes provide sufficient power for analyses. These are detailed in the paper.

Data exclusions
Some of the data points were excluded. Specifically, we removed samples which were sequenced from blood samples, as opposed to lymphoblastoid cell lines, to limit heterogeneity in tissue type. This is described in more detail in the manuscript.

Replication
This is an exploratory study and was limited by the sample size available. Thus, no explicit replication experiments were carried out. However, there are implicit replications in our experimental setup. For example, the association between ancestry and mtDNA copy number is independently reproducible across mtDNA haplotypes.

Randomization
As this is an observational study, experimental randomization is not applicable.

Blinding
As this is an observational study, blinding (w.r.t treatment) is not applicable.

Reporting for specific materials, systems and methods

Materials & experimental systems

Methods

n/a Involved in the study

n/a Involved in the study

ChIP-seq
Flow cytometry
MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics
Our analyses are based on samples of admixed ancestry, which are part of the 1,000 Genomes Project Dataset. The samples derive their ancestry from Native American, European, and African source populations.

Recruitment
These samples were collected by the 1,000 Genomes Project Group. The data are publicly available.