Associations Between Maternal Lifetime Stress and Placental Mitochondrial DNA Mutations in an Urban Multiethnic Cohort

SUPPLEMENT 1

Methods/Results:

The online data supplement provides more details on the methods/materials for placental sampling, sequencing, strand bias interrogation, and complete mutational data for PRISM and its association with previously reported mutations in Mitomap. Results are also provided for all models of mutational load and haplogroup-specific effects.

Placental Processing

Placenta samples (~1-2 cm3) were taken on the fetal side ~1 – 1.5 cm below the fetal membrane to avoid membrane contamination and approximately 4 cm from the cord insertion site, taking care to avoid large vessels. Samples were cut into smaller pieces (~0.1 cm3), placed into 1 ml of RNAlater™ RNA Stabilization Reagent (Qiagen) and stored at -4°C for <24 h; excess RNAlater was then removed and samples were stored at -80º C until DNA extraction. The origin of placental tissue from the fetal side of the organ was confirmed by the previously reported near-perfect agreement of placenta and cord blood samples in 64 genotyping probes used for identity verification.

Mitochondrial DNA Sequencing

First, whole mitochondrial DNA sequencing (complete genome sequence ID AY495156.2, 16569 bp) utilized two pairs of primers that cover the entire mtDNA sequence in two individual amplifications: MTL-F1 AAAGCACATACCAAGGCCAC (9397 - 9416)/MTL-R1 TTGGCTCTCCTTGCAAAGTT (1893 – 1874), amplicon size 9065 bp; MTL-F2 TATCCGCCATCCCATACATT (15195 - 15214)/MTL-R2 AATGTTGAGCCGTAGATGCC (9796 - 9777), amplicon size 11170 bp. To check the amplification specificity and yield, 8 µl PCR product together with negative control were added to an agarose gel to run electrophoresis. Expected specific amplification from gDNA samples were shown in the gel with higher DNA yield from primer MTL-F1/R1 in general. A subset of participants (~20%, n=92) were re-sequenced to interrogate the reproducibility of our variant calls.

Second, we interrogated strand bias. The reason for this is that the two strands of mitochondrial DNA have very different nucleotide composition. Because sequencing efficiency depends on the nucleotide composition, we suspected that the two strands of DNA may lead to very different depth of coverage. However, as shown in **Figure S1**, sequences collected from our experiments covered both L-strand and H-strand almost equally. We also show that the coverage of the mitochondrial genome is comprehensive as depicted by the **figure S2**-the reads correspond to two long-range PCR products, leaving no gap. Further, previous research suggests around 90% somatic mutations in mitochondria are caused by C>T and T>C substitutions and most of them occur on the H-strand. (35) We observed a similar pattern in our data with 96.3% of the mutations belong to the G>A, A>G, T>C, C>T categories (**Figure S3**). Note that a G>A mutation in the L-strand is equivalent to the C>T mutation in the H-strand. Similarly, an A>G mutation in the L-strand is equivalent to the T>C mutation in the H-strand. Given the similarities with previous published literature, this finding supports the validity of our variant calling.

Lastly, we have compiled all variants identified in the PRISM cohort and their corresponding variant allele frequencies in **Table S1**. **Table S2** provides a list of the mutations that are reported in Mitomap and **table S3** provides details on the 801 mutations in PRISM that have previously been associated with disease(s) (45).

Results

In the re-sequenced subset, we observed that over 95% of all mutations are common to both samples and that most discrepant mutations have low read counts (<5); mutations with low reads (< 5) make up ~ 25% of our total mutations. Concordance was lower (80%) for heteroplasmic mutations with low read counts (between 2 and 4) but improved as the number of reads increased (i.e., 88% concordance among heteroplasmic mutations with at least 3 reads rather than 2 or more). Although we still have a good amount of concordance among mutations with low read counts, there is a greater chance of discrepancy among mutations with low read counts which could be due to sequencing errors.

Tables S5-S8 provide all statistical model details for the main effect of maternal lifetime stress on total, nonsynonymous, heteroplasmic, and homoplasmic mutational load. Interestingly, 21% of the variants identified in MT-ND1 (i.e. 126 out of 597) have been linked to adverse metabolic, cardiac, and neurodevelopmental outcomes across the life course (i.e. diabetes, insulin resistance, developmental delay, myalgia, hypertension, atherosclerosis, Alzheimer's, and Parkinson's disease) and the placental mutational load of MT-ND1 was positively and significantly associated with increased maternal cumulative stress in the PRISM cohort (**Figure 4**). We observed race-specific effects of maternal lifetime stress on total (Figure 5A), nonsynonymous (Figure 5B), and heteroplasmic (**Figure S4**) mutations; all relationships were similar with Black women exhibiting more stress-related mutations. We also observe similar results in our haplogroup-specific analyses (**Figure S5**). Compared to the Native American/Asian haplogroup, women in the African (L) haplogroup were more likely to exhibit a greater degree of total mutations associated with lifetime psychosocial stress (Figure S5A, $\chi 2= 7.80$, $p = 0.005$) and borderline significant among European haplogroups ($\chi 2= 3.69$, $p =0.06$). It should be noted that 71% of women self-reporting as Black belong to the African (L) haplogroup (**Table S9**).

See Supplement 2 (Excel) for all supplemental tables.

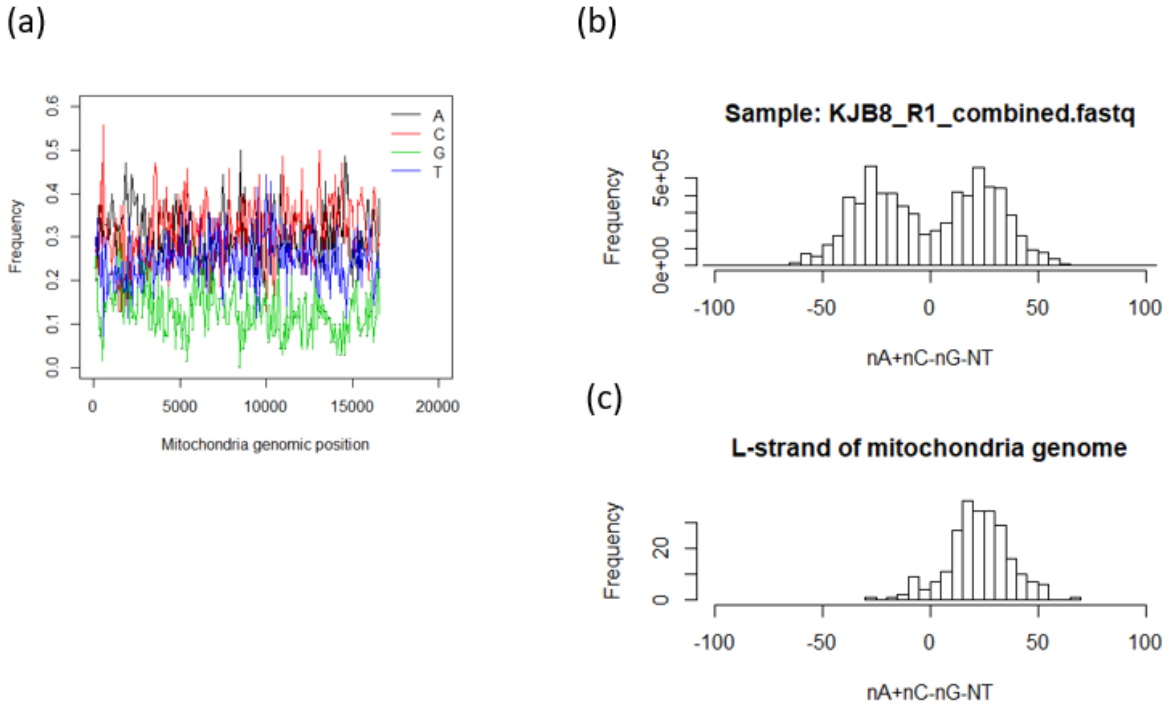

Figure S1. Strand bias in the mitochondria genome and in the collected short sequence reads. (a) Nucleotide composition of human mitochondria genome computed from the standard reference sequence. (b) Histogram of strand bias, which is calculated as nA+nC-nT-nG, where nA, nC, nG, nT were computed as the number of A, C, G, Ts in a 100-nt short-read sequence respectively. (c) Histogram of strand bias calculated from the light strand of the standard mitochondria genomic sequence.

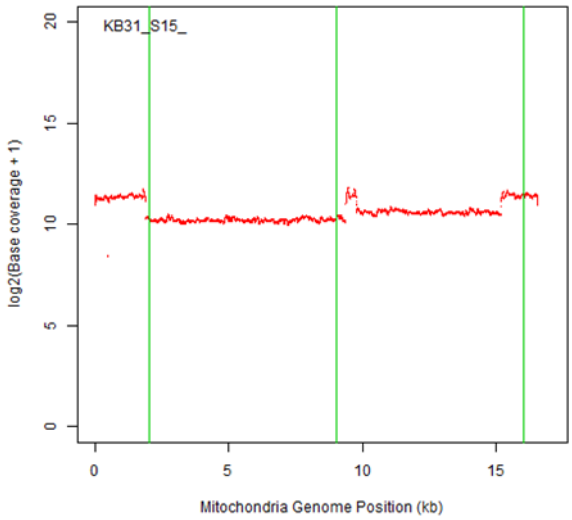

Figure S2. Coverage of the mitochondrial genome. The reads correspond to two long-range PCR products and similar coverage was observed across individuals.

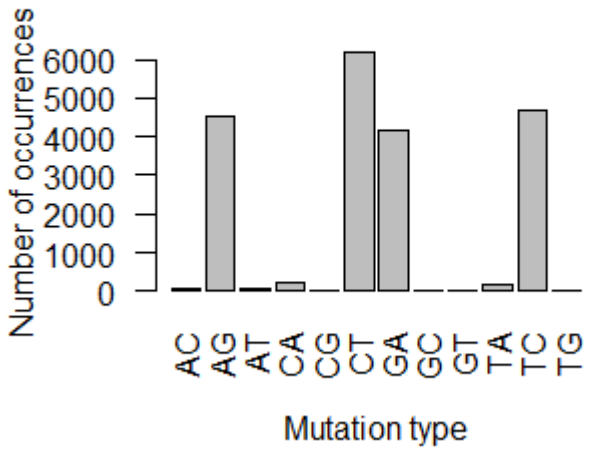

Figure S3. Number of mutations stratified by the mutation type. Only single nucleotide substitutions are included in the figure.

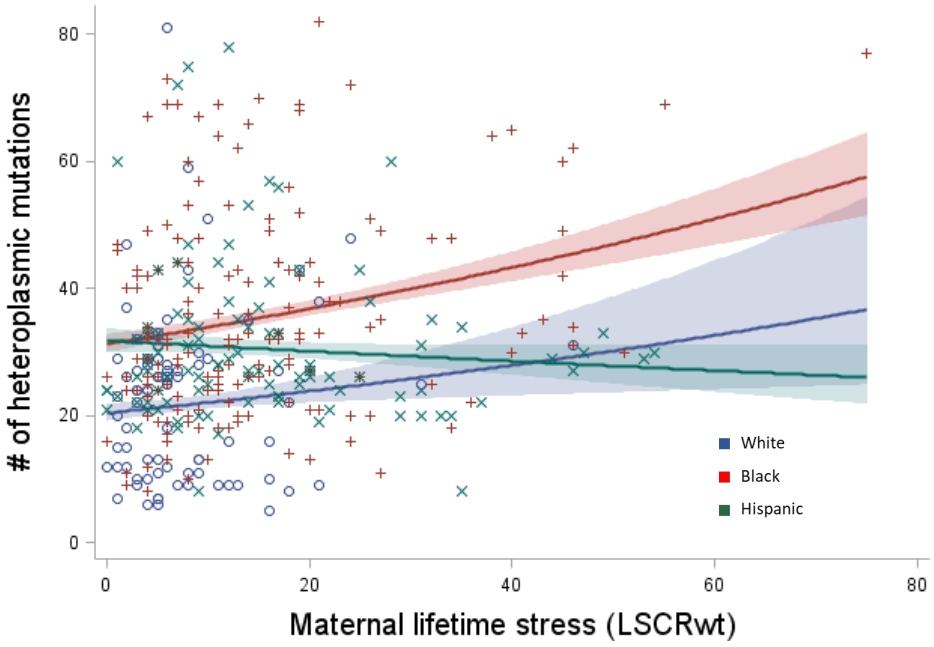

Figure S4. Race/ethnic differences in the association between maternal lifetime stress and placental mitochondrial heteroplasmy load. The likelihood ratio type 3 analysis for the three-way interaction (i.e., LSCRwt x race/ethnicity) was significant (p<0.0001). Contrast statements were used to test for differences in slopes across levels of race/ethnicity. Model adjusted for maternal age at birth (p < 0.0001) and maternal education (p=0.23).

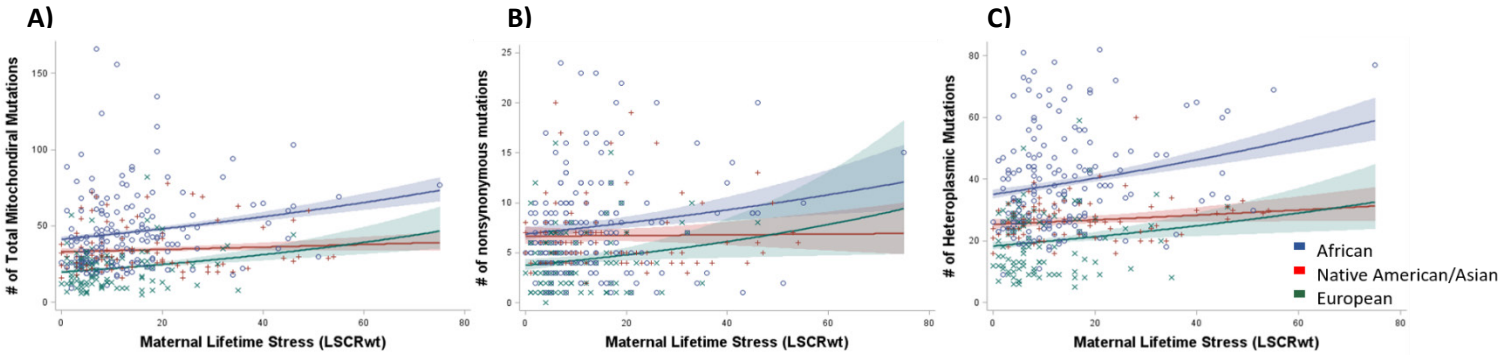

Figure S5. Haplogroup differences in the association between maternal lifetime stress and placental mitochondrial mutational load. A) Total (heteroplasmic + homoplasmic) mutational load, B) nonsynonymous mutations, and C) heteroplasmic mutations only. The likelihood ratio type 3 analysis for the three-way interaction (i.e., LSCRwt x haplogroup) was significant for total (p=0.0002) and borderline for heteroplasmic (p=0.05) and nonsynonymous (p=0.08) mutations. Models adjusted for maternal age at birth and maternal education.