Author's response to reviews

Title: Investigating the relationship between mitochondrial genetic variation and cardiovascular-related traits to develop a framework for mitochondrial phenome-wide association studies

Authors:

Sabrina L Mitchell (sabrina.l.mitchell@vanderbilt.edu)
Jacob B Hall (jacob.b.hall@vanderbilt.edu)
Robert J Goodloe (robert.j.goodloe@vanderbilt.edu)
Jonathan Boston (boston@chgr.mc.vanderbilt.edu)
Eric Farber-Eger (eric.h.farber-eger@vanderbilt.edu)
Sarah A Pendergrass (sap29@psu.edu)
William S Bush (william.s.bush@vanderbilt.edu)
Dana C Crawford (crawford@chgr.mc.vanderbilt.edu)

Version: 2
Date: 4 March 2014

Author's response to reviews: see over
March 4, 2014

RE: Manuscript ID: 6146102111020714

Dear Editor,

Thank you for your consideration of our manuscript titled “Investigating the relationship between mitochondrial variation and cardiovascular-related traits to develop a framework for mitochondrial phenome-wide association studies” for publication in the journal *Biodata Mining*. We thank the reviewer for useful comments that have significantly improved the quality of this manuscript. We have responded to specific reviewer comments in the following ways:

**Background**

Paragraph 2: instead of “nearly 16.6 kb” revise to exact number of base pairs. It now reads: “The human mitochondrial genome consists of a double-stranded, circular chromosome spanning 16,569 base pairs.”

**Methods**

Study population and phenotypes:

1) Clarify that only African Americans from EAGLE study are analyzed in the current study. This was clarified in the first sentence of paragraph 2 of the methods section: “In this pilot PheWAS, we accessed the African American samples in EAGLE BioVU and selected eight cardiovascular-related traits for analysis...”

2) For each individual was there a minimum number of measurements required for inclusion for any of the continuous trait data? There was no requirement for a minimum number of measurements for continuous traits per individual for inclusion in study. This has been indicated in paragraph 2 of the methods section: “For all continuous traits, only a single observation within the synthetic derivative was required for an individual to be included in the analysis.”

3) Why was the mean used for hemoglobin instead of the median value as used for all other continuous traits? To be consistent with the methods used for the other continuous traits, we determined the median of the median value for hemoglobin and repeated the analyses using the median value.

4) Clarify that the 150,000 de-identified DNA samples represent 150,000 unique individuals (assuming that is the case). This has been clarified in the first paragraph of the Methods section, as has been updated to reflect the current number of samples in BioVU, which exceeds 170,000 individual de-identified DNA samples.
Polygenic modeling:

1) The first 10 PCs were included as covariates in the fixed effects model. How were these calculated? Clarify number of SNPs used in the calculation of PCs and the nuclear GRM. How many mitochondrial SNPs were used to calculate the mitochondrial GRM? **Added text to paragraph 2 under Polygenic modeling (methods):** “Eigenstrat was used to generate nuclear PCs from the 192,139 autosomal SNPs that passed QC. The random effect is empirical genetic relatedness, estimated using GRMs, which were created via GCTA, for the nuclear and mitochondrial genomes, using 192,139 and 130 SNPs, respectively.”

2) How many SNPs passed QC? Are QC procedures the same as that in reference 14? **This was addressed under Methods – Study Population and Phenotypes** “…quality control was performed as previously described [Buyske et al reference], resulting in 192,139 autosomal SNPs and 130 mtSNPs for use in this analysis.” **Note:** To clarify, this is not the same dataset that is described in the Buyske et al manuscript, but follows the same QC procedures for Metabochip genotype data.

Single SNP Analysis:

1) Justify the use of only first 2 PCs here versus the use of 10 PCs in polygenic modeling. **To address this discrepancy, we re-ran GCTA including only 2 PCs, which are sufficient given that our study population consists only of African Americans.**

Results

Single SNP test of association:

1) Figure legends for fig. 3 and 4 are not referenced in the text. No legend for figure 4. **These comments have been addressed.** There was no figure 4, only figure 3a and 3b, which were initially uploaded separately so 3b was mislabeled as figure 4.

2) Discrepancy between the text and figures regarding the number of variants associated (at p<0.05) with total cholesterol and T2D. **This was a mistake that has been corrected.** The number of significant SNPs in the figures was correct while the numbers in the text were incorrect (as the text did not reflect the filtering of SNPs with MAF less than 0.01).

3) Single mtDNA SNPs tests of association are not corrected for multiple testing and this needs to be indicated in the text. **This has been addressed in both the Results –Single SNP test of association (“Assuming an uncorrected significance threshold of p <0.05 and considering the number of mtSNPs tested (86), 4.3 SNPs would be expected to be associated by chance alone.”) and Discussion (“Because the mt-PheWAS analyses presented herein were exploratory in nature we did not correct for multiple testing.”).**

Polygenic modeling:

1) Which mtSNPs were used to calculate mt PCs? All? **In the original submission, we briefly mention that inclusion of mt PCs did not change the results. We originally ran this analysis to examine**
the sensitivity of the model to cryptic mitochondrial structure, but because the results showed no significant differences we removed these analyses from the revision.

2) Edit final sentence to better reflect the point that GCTA does not require trait distributions to be normal. **We removed the statements regarding additional testing removing potential outliers; therefore we removed this statement as it is no longer necessary.**

**Discussion/Conclusions**

1) Expand discussion on single mtSNP associations with T2D; why do we highlight mt16189? Add discussion on single mtSNP associations with total cholesterol. **We have expanded the discussion on the single mtSNP association results for both T2D and total cholesterol. Variant mt16189 is highlighted in the manuscript because of its previous association with T2D in Asian and European-descent populations. It has also previously been associated with known risk factors for T2D including fasting insulin concentration, fasting glucose, and BMI. To our knowledge this is the first time an association between T2D and mt16189 has been identified in African Americans.**

2) Are the associated single variant results for either total cholesterol or T2D reflective of a common haplogroup background? **Yes, some of the mtSNPs associated with these phenotypes can be found on common mitochondrial haplogroups including L1c and L3e1. Text was added to the discussion describing this and previously reported associations with these haplogroups.**

**Figures**

1) Be consistent with how figures are reference in text (Fig.1 or Figure 1) – **addressed this; all are in the format Figure x.**

2) Figure 2 legend is incorrect; only notes blue bars not the red bars. **This figure has been revised (it is now Figure 1) and is correctly labeled in the figure legend.**

3) Additionally, we decided to remove Figure 1 of the original manuscript because it is sufficiently described in the text and inclusion of the figure does not add significantly to the manuscript. It is not necessary to understand the manuscript.

We hope that we have addressed the reviewer comments satisfactorily. We look forward to hearing from you soon. Please contact me should you have any questions or concerns regarding this revision.

Sincerely,

Sabrina Mitchell, PhD

sabrina.l.mitchell@vanderbilt.edu