haematologica
Journal of the Ferrata Storti Foundation

Variations in mitochondrial genome as potential prognostic markers in sickle cell disease

by Rudra Ray, Haiou Li, Shijinqiu Gao, Nancy Asomaning, Kang Le, Maliha M Ahmad, Xunde Wang, Yuesheng Li, Sayuri Kamimura, Clifton L. Dalgard, Chengyu Liu, Zenaide M.N. Quezado, Justin Lack, Laxminath Tumburu and Swee Lay Thein

Received: November 18, 2025.
Accepted: March 11, 2026.

Citation: Rudra Ray, Haiou Li, Shijinqiu Gao, Nancy Asomaning, Kang Le, Maliha M Ahmad, Xunde Wang, Yuesheng Li, Sayuri Kamimura, Clifton L. Dalgard, Chengyu Liu, Zenaide M.N. Quezado, Justin Lack, Laxminath Tumburu and Swee Lay Thein. Variations in mitochondrial genome as potential prognostic markers in sickle cell disease.
Haematologica. 2026 Mar 19. doi: 10.3324/haematol.2025.300253 [Epub ahead of print]

Variations in mitochondrial genome as potential prognostic markers in sickle cell disease

Rudra Ray[1], Haiou Li[1], Shijinqiu Gao[1], Nancy Asomaning[1], Kang Le[1], Maliha M. Ahmad[1], Xunde Wang[1], Yuesheng Li[2], Sayuri Kamimura[3], Clifton L. Dalgard[4,5], Chengyu Liu[6], Zenaide M.N. Quezado[1,3], Justin Lack[7], Laxminath Tumburu[1], Swee Lay Thein[1]

[1]Sickle Cell Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD,USA.
[2]DNA Sequencing & Genomics Core, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA.

[3]Department of Perioperative Medicine, National Institutes of Health Clinical Center, Bethesda, MD, USA.

[4]The American Genome Center, Uniformed Services University of the Health Sciences, Bethesda, MD, USA.

[5]Department of Anatomy, Physiology & Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD, USA.

[6]Transgenic Core, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA.
[7]Bioinformatics (NCBR)/Integrated Data Sciences Section (IDSS)/ Research Technology Branch/DIR, NIAID, NIH, Bethesda, MD, USA.

Running heads: Mitochondrial DNA variations in SCD pathology

***Corresponding Author:**
Swee Lay Thein
National Heart, Lung and Blood Institute / National Institutes of Health
Building 10-CRC, Room 6S241
10 Center Drive, Bethesda, MD 20892
Office Line: +1-301-435-2345
Direct Line: +1-301-402-6699
Fax: +1-301-451-7091
Email: sl.thein@nih.gov

Data-sharing statement: Deidentified individual participant data in the reported results will be made available 3 months after publication for a period of 5 years after the publication date. Original data will be available from the corresponding author, Swee Lay Thein (sweelay.thein@nih.gov). Original data are available at dbgap.ncbi.nlm.nih.gov/home

Trial registration:

Human participants were enrolled under IRB approved protocols ClinicalTrials.gov identifiers: NCT00011648, NCT00081523, and NCT0368572. Mouse study protocols were approved by the NIH Clinical Center Animal Care and Use Committee (DPM 23-01).

Author contributions

SLT and LT conceived and designed the study; RR performed the mtDNA studies and data analyses in human and mouse samples; HL and SG provided statistical and bioinformatics support; NA provided updated survival data; KL provided support with mouse sample collection, and reviewing manuscript; MA and LT performed initial human mtDNA studies; XW provided support with mouse sample collection, lab experiments and nuclear variant analysis; YL provided support with NGS sequencing; CL, SK and ZMNQ provided SCD mouse support and mouse sample collection; CD performed whole genome sequencing studies; JL provided support with bioinformatics; SLT supervised the study. RR wrote the first draft of the manuscript; SLT edited and wrote the manuscript; all authors reviewed the final manuscript.

Disclosure of Conflict of interest:

All authors declare no competing interests that may be relevant to the submitted work

Acknowledgments:

Artificial Intelligence (AI)- ChatGPT (OpenAI, San Francisco, CA, USA) assisted in constructing and debugging the R-based pipeline used for mitochondrial DNA deletion detection from whole genome sequence data. All AI-assisted code outputs were critically reviewed, verified, and validated by the authors for accuracy and scientific integrity.

Funding:

The study was supported by the Division of Intramural Research of the National Heart, Lung, and Blood Institute (SLT) and the NIH Clinical Center (ZMNQ) at NIH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or DHSS.

Abstract

Alterations in the mitochondrial genome integrity, including changes in mitochondrial DNA copy number (mtDNA-CN) and accumulation of mtDNA mutations, are associated with aging and diverse disorders, often linked to underlying systemic inflammation and metabolic stress. In sickle cell disease (SCD), inflammation drives the pathology, resulting in organ damage and early mortality. The prognostic role of mitochondrial genomic variation in SCD is largely unexplored. This study investigated whole blood derived mtDNA alterations, including mtDNA-CN and mtDNA mutations, in adults with SCD, sickle trait, and healthy controls, and examined their associations with age and mortality in SCD. We also assessed mtDNA heteroplasmy distribution across tissues in a humanized mouse model of SCD. Elevated mtDNA-CN and mtDNA heteroplasmy burden were observed with increasing genotype severity across all cohorts (HbAA, HbAS < HbSB+ < HbSC < SCA (HbSS & HbSβ0)). In sickle cell anemia (SCA) patients, mtDNA mutation burden- including mtDNA heteroplasmy and mtDNA deletions increased with age, whereas mtDNA-CN level declined, indicating progressive deterioration of mtDNA integrity with age. In SCD patients, specific mtDNA variants showed strong positive correlations with mortality risk, lower mtDNA-CN correlated with higher NIH risk scores, and nuclear variants *CYB5R3* T117S and *PIEZO1* E756del influenced mtDNA mutation burden without affecting NIH risk score. Consistent patterns of mutational load were observed across specific regions in mitochondrial genome in both humans and mice, suggesting potential mtDNA mutational hotspots. We conclude that variations in the mitochondrial genome are potential prognostic markers for SCD.

Introduction

While the sickle pathology is initiated by polymerization of hemoglobin S (HbS), the multiple end-organ damage is inflicted by years of ongoing inflammation and vasculopathy,[1, 2] evidenced by the presence of multiple proinflammatory markers.[3] Degenerative changes in sickle cell disease (SCD) lack pathognomonic features and resemble those in non-SCD individuals, but occur earlier, indicating "accelerated" aging,[4] likely driven by the underlying ongoing inflammation. This convergence of chronic systemic inflammation and aging has been referred to "inflamm-aging" that contributes to the pathogenesis of age-related diseases.[5] One common feature of human aging and degenerative disease is an accumulation of cells with mitochondrial dysfunction,[6] linked to alterations in mitochondrial DNA copy number (mtDNA-CN), and accumulation of mtDNA heteroplasmy.[7, 8] mtDNA heteroplasmy refers to the coexistence of mutated and wild type mitochondrial DNA within a cell or individual, where the accumulation of mutated mtDNA beyond a critical threshold can disrupt mitochondrial function and lead to significant physiological dysfunction.[9] Mitochondria have a crucial role in oxidative metabolism and the synthesis of adenosine triphosphate (ATP) via oxidative phosphorylation (OXPHOS), but they are also a potent source of inflammatory triggers, particularly circulating cell-free mtDNA (cf-mtDNA).[10] Of relevance, we previously showed that paired plasma samples of SCD patients contained significantly elevated and hypomethylated cf-mtDNA during acute pain compared with steady state, which were not only fragmented and proinflammatory but also were linked to abnormally retained mitochondria in enucleated mature sickle red blood cells (RBCs).[11] Mitochondria that are retained in the sickle RBCs are depolarized, in a state that is associated with increased mitochondrial reactive oxygen species (mROS) generating more oxidative stress and mitochondria dysfunction.[12]

It is unclear whether changes in mtDNA integrity, such as heteroplasmy and mtDNA-CN, directly drive mitochondrial dysfunction in SCD or result from chronic oxidative and metabolic stress, potentially creating a vicious cycle that worsens the disease. Nonetheless, these mitochondrial genomic changes may serve as biomarkers of the 'inflamm-aging' process and disease severity in SCD, as demonstrated in many other diseases.[9] mtDNA-CN, an indicator of mitochondrial biogenesis and function, is linked to aging-related diseases and serves as a biomarker in conditions like cognitive impairment, cardiovascular disease prognosis, and cancer-related metabolic

changes.[13-15] Conversely, mtDNA heteroplasmies and deleterious variants predict mortality risk and disease progression in cancers, myeloid neoplasms, neurodegenerative diseases, and renal dysfunctions.[16-19] Collectively, these mtDNA signatures provide valuable insights into disease progression and potential as prognostic biomarkers.

This study analyzed mtDNA from blood-derived whole-genome sequences (WGS) of 1043 individuals, including SCD patients with various genotypes, carriers of the sickle allele (HbAS), and healthy ethnic-matched controls (HbAA) from three cohorts. We investigated associations between mtDNA variation and demographic factors, hemoglobin level, hydroxyurea (HU) treatment, alpha thalassemia status, relevant nuclear variants, disease severity (NIH phenotypic risk score), and mortality. Tissue-specific mtDNA heteroplasmy was also assessed in humanized Townes SCD mice.

mtDNA variation correlated with SCD genotype severity, with higher mtDNA-CN and mutation burden seen in more severe genotypes across all 3 human cohorts. mtDNA variants were associated with increased mortality, while mtDNA-CN was inversely associated with NIH risk score. Certain nuclear variants influenced mtDNA heteroplasmy burden. In mice, mtDNA heteroplasmy varied in tissue and genotype. Analysis of mtDNA mutation patterns across mitochondrial genome suggested the presence of mutational hotspots. Overall, these findings indicate that mtDNA-CN and heteroplasmy burden may serve as biomarkers and prognostic indicators in SCD.

Methods

Human subjects

Adult human participants enrolled under IRB approved protocols NCT00011648, NCT00081523, and NCT03685721 in three cohorts, were included in this study. All participants were of African descent. Individuals with HbSS and HbS-Beta-thalassemia^0 (HbSB0) genotypes were combined as sickle cell anemia (SCA), the most severe form of SCD. Cohort1 included a total of 673 SCD patients (554 SCA, 91 HbSC, 25 HbS-Beta-thalassemia$^+$ (SB+), and 3 other SCD genotypes); Cohort 2 included 171 individuals (113 SCA, 30 HbAS, 16 HbAA, and 12 other SCD genotypes); Cohort 3 included 199 individuals (96 SCA, 47 HbAS, 44 HbAA, and 12 other SCD genotypes). DNA from each human subject was extracted using peripheral blood buffy coat and was subjected to whole genome sequencing (WGS). Survival analysis was based on Cohort 1 where samples were collected over 23 years (February 2001-July 2024), with a mean follow-up time of 6.26 years

and a maximum follow-up duration of 20.16 years. Details on human cohorts, sample distributions, and their demographics are described in *Online Supplementary Methods, Supplementary Figure S1A, Supplementary Table S1*.

Mouse samples

Eighteen Townes humanized mouse model of SCD- littermates, equally distributed by sex and across three genotypes (AA, AS, SS), were included in the study. All experimental protocols were approved by the NIH Clinical Center Animal Care and Use Committee (DPM 23-01). Genomic DNA was extracted from 9 tissue samples per mouse (total 162 samples) from which mtDNA was enriched and subjected to NGS. Details on mouse sample distribution and generation of humanized mouse model of SCD are described in *Online Supplementary Methods, Supplementary Figure S1B.* mtDNA was enriched from mouse genomic DNA samples by long range PCR (LR-PCR) and confirmed by agarose gel electrophoresis, as detailed in *Online Supplementary Methods, Supplementary Figure S2.*

Assessment of Mitochondrial Copy Number (mtDNA-CN), mtDNA Heteroplasmy and large mtDNA deletions

We applied the equation given below to quantify the mtDNA-CN for the human samples.[20, 21]

$$\text{mtDNA copy number per cell} = 2X \left(\frac{\text{mitochondrial coverage}}{\text{autosomal coverage}}\right)$$

Mouse samples were excluded from the mtDNA-CN analysis as mtDNA was enriched prior to NGS.

Variant calling was done using MitoHPC workflow (https://github.com/dpuiu/MitoHPC) with LoFreq variant caller tool (version 2.4) and VEP/106 annotation. NuMT (nuclear sequences of mitochondrial origin), coverage depth threshold, and variant allele frequency VAF (%) filters were applied to eliminate false positives, homoplasmies, and low-confidence heteroplasmies. Details on NGS and analysis, variant calling, and heteroplasmy detection pipeline, and identification of large mtDNA deletions (≥100 bp) from WGS BAM files using an R-based pipeline, are described in detail in *Online Supplementary Methods.*

Statistical Analysis

We used the Cox proportional hazards model to analyze the relationship between mtDNA heteroplasmy and all-cause mortality. Overall phenotype risk was calculated using the NIH risk score for SCD. Comparison between two groups was conducted using Mann Whitney test, and

three or more groups using Kruskal-Wallis test. Data were analyzed using GraphPad Prism (version 10.2.0), R (version 4.2.1), and Microsoft Excel.

All details are provided in the *Online Supplementary Methods*.

Results

mtDNA-CN level altered with genotypic severity and age

mtDNA-CN was estimated from WGS data as the ratio of mitochondrial to nuclear genomic reads. Across all three cohorts, mtDNA-CN increased with genotypic severity, with SCA showing significantly higher levels than other genotypes. In Cohort 1, SCA group had significantly higher mtDNA-CN (mean=483.9) than HbSC (mean= 379.2; $P = 0.0017$), and HbSB+ (mean= 299.4; $P = 0.001$) (Figure 1A). In Cohort 2, SCA showed significantly higher mtDNA-CN (mean=804.7) than HbAS (mean=246.3; $P < 0.0001$) and HbAA (mean=244.9; $P = 0.0019$), and a similar pattern was observed in Cohort 3, where SCA had higher mtDNA-CN (mean=796.7) than HbAS (mean=217.1, $P < 0.0001$) and HbAA (mean=239; $P < 0.0001$) (Figure 1B, C). When all 3 cohorts were combined, mtDNA-CN increased with genotypic severity: HbAA, HbAS< HbSB+ < HbSC< SCA; and SCA (n=763) had significantly higher mtDNA-CN ($P < 0.001$) compared to all other genotypes (Figure 1D).

We next examined whether mtDNA-CN level varies with age. Among all subjects combined from the three cohorts (n=1043), mtDNA-CN decreased with increasing age (*Online Supplementary Figure S3*). Further, when all SCA patients combined from three cohorts (n=763) were dichotomized into two age groups, the older group (≥ 30-year) had lower mean mtDNA-CN (559.8) compared to the younger (< 30-year) age group (582.3), the difference was not significant (Figure 1E). Stratifying the SCA genotypic group into four age groups further shows an overall decrease in mtDNA-CN with age, with mean mtDNA-CN values of 594.0, 609.3, 512.3, and 512.1 in the <25, 25–34, 35–44, and ≥ 45-year age groups, respectively (Figure 1F).

mtDNA heteroplasmy burden increased with genotypic severity and age

mtDNA heteroplasmies were detected using LoFreq variant caller. Potential homoplasmies (VAF > 90%) were removed, and to ensure consistent detection of heteroplasmic variants across different cohorts with varying mtDNA sequencing depths, lower VAF (%) thresholds were defined based on the mean mitochondrial coverage for each cohort (*Online Supplementary Table S2*, and *Supplementary Figure S4A-C*).

The total mtDNA heteroplasmy level per individual tended to increase with genotypic severity across all three cohorts, although the trend was not statistically significant (*Online Supplementary Figure S5A-C*). The increase in the burden of heteroplasmies was particularly noted in non-synonymous (NS) mutations across the three cohorts (*Online Supplementary Figure S5D-F*, and *Supplementary Table S3*). When all 3 cohorts were combined (n=1043), mean NS heteroplasmy burden increased with genotypic severity: AA (0.01) < AS (0.03) <SB+ (0.03) < SC (0.08) < SCA (0.12) (*Online Supplementary Figure S5G*). Although the differences across individual genotypes were not statistically significant (*Online Supplementary Figure S5D-F)*, the mean NS heteroplasmy burden was found to be significantly higher in SCA when compared to other two major genotypes combined in each cohort (Figure 2A-D). In Cohort-1, mean NS heteroplasmy was significantly higher in SCA at 0.11 compared to less severe genotypes (HbSB+ & HbSC) combined at 0.05 (*P = 0.02*) (Figure 2A). In Cohort 2, SCA showed significantly higher NS heteroplasmy burden (mean=0.14) than HbAS & HbAA combined (0.02) (*P = 0.04*) (Figure 2B). Similarly, in Cohort 3, SCA had higher NS heteroplasmy burden (mean=0.10) than HbAS & HbAA combined (mean= 0.03), with the difference approaching statistical significance (*P = 0.08*) (Figure 2C). When subjects from all 3 cohorts were combined (n=1043), total SCA patients (n=763) showed a significantly higher mean NS heteroplasmy burden at 0.12 compared to all other non-SCA individuals (n=280) at 0.05 (*P = 0.001*) (Figure 2D).

We further analyzed whether NS heteroplasmy burden varied with age. Among the total SCA patients (n=763), the mean NS heteroplasmy burden was significantly higher in the older age group (≥ 30 years) compared to the < 30 years age group (0.14 vs 0.09, *P = 0.03*) (Figure 2E). Further stratification into four age groups showed a gradual increase in NS heteroplasmy burden with age, with mean NS heteroplasmy burden of 0.10, 0.11, 0.11, and 0.17 in the <25, 25–34, 35–44, and ≥45-year age groups, respectively (Figure 2F).

mtDNA-CN and mtDNA heteroplasmy variations were independent of hydroxyurea (HU) treatment and cell counts

We evaluated if HU treatment affects mtDNA heteroplasmy burden and mtDNA-CN level in Cohort 1 which includes only SCD patients (n=673). HU treatment had no impact on heteroplasmy burden, but mtDNA-CN was found to be slightly increased upon HU treatment (*Online Supplementary Table S4, and Supplementary Figure S6A,B*). When mtDNA-CN level was compared by genotypes within the HU-treated and non-HU treated groups separately, mtDNA-CN

increased with genotypic severity in both the groups, mean mtDNA-CN was significantly higher in SCA than SC and SB+ in the non-HU treated group (*Online Supplementary Figure S6C*).

Whole blood derived buffy coat primarily consists of leucocytes containing both nuclear and mitochondrial DNA, but it can also include traces of platelets and reticulocytes that contain mitochondrial DNA and lack nuclear DNA and thereby may influence the mtDNA-CN estimations. To assess whether genotype-wise variation in mtDNA-CN was confounded by platelet or reticulocyte count, multivariate regression analysis was performed in Cohort 1. Platelet and reticulocyte counts exhibited negligible effects on mtDNA-CN variation with regression coefficients of 0.12 and 0.07, respectively, whereas genotype (regression coefficient =97.78; $P = 0.03$) contributed to mtDNA-CN variation to a much greater extent (*Online Supplementary Table S5, Supplementary Figure S6D*).

Sex differences and hemoglobin levels did not affect mtDNA-CN and mtDNA heteroplasmy burden, whereas α- thalassemia specifically impacted mtDNA-CN

We evaluated whether sex differences influenced mtDNA-CN and mtDNA heteroplasmy. All participants (n=1043) from 3 human cohorts were stratified by sex. No significant differences were observed between males and females in mtDNA-CN levels, total mtDNA heteroplasmy burden, or non-synonymous (NS) heteroplasmy burden. Similarly, focusing on the total SCA patients (n=763) from the 3 human cohorts, no significant sex based differences were detected (*Online Supplementary Figure S7*).

Next, we assessed whether hemoglobin (Hb) levels were associated with mtDNA readouts. This analysis was restricted to the SCA genotypes in Cohort 1 to ensure a clinically homogeneous phenotype. Spearman correlation analysis revealed no significant association between Hb levels and mtDNA-CN, total mtDNA heteroplasmy, NS mtDNA heteroplasmy (*Online Supplementary Figure S8*).

We further examined whether alpha thalassemia status was associated with mtDNA-CN levels and mtDNA heteroplasmy burden in SCD (*Online Supplementary Table S6, Supplementary Figure S9*). Among the SCD patients in Cohort 1 (n=673), α globin genotyping data were available for 632 patients. Spearman correlation analysis demonstrated a significant positive association between mtDNA-CN and α globin gene count (Spearman r=0.089, $P=0.02$, n=632). In contrast, neither total mtDNA heteroplasmy nor NS heteroplasmy showed a significant association with α globin gene count. The patients were further stratified into four groups based on alpha (α) globin

genotypes: -α/-α (2 normal α globin genes), -α/$\alpha\alpha$ (3 normal α globin genes), $\alpha\alpha$/$\alpha\alpha$ (4 normal α globin genes), and $\alpha\alpha$/$\alpha\alpha\alpha$ or more (normal α globin gene ≥ 5). mtDNA-CN increased with increasing α globin gene count in all SCD patients and SCA patients. Patients with -α/-α (two globin genes) had significantly lower mtDNA-CN compared to $\alpha\alpha$/$\alpha\alpha$ (wild type) among both SCD patients (*P=0.029*) and SCA patients (*P=0.038*) (*Online Supplementary Figure S9B, C*). No differences were found in total mtDNA heteroplasmy burden or NS heteroplasmy burden according to number of α globin genes (*Online Supplementary Figure S9D, E*).

mtDNA deletion burden increased with genotypic severity, mtDNA-CN, and age

A total of 116 mtDNA deletions (≥100 bp) were identified in Cohort 1 which included 673 SCD patients. In Cohort 2 and Cohort 3, no deletions ≥100 bp were detected; all observed deletions were <100 bp and were excluded from analysis, as such small deletions are likely mapping artifacts. In cohort 1, the mean mtDNA deletion (≥100 bp) burden increased with genotypic severity, with SCA having the highest burden compared to SC (0.19 vs. 0.06, *P = 0.04*) and SB+ (0.19 vs.0.04, *P= ns*) (Figure 3A). mtDNA deletion burden was significantly higher in SCA than all non-SCA genotypes combined (mean of 0.05, *P = 0.008*) (Figure 3B). The mtDNA deletion burden increased with mtDNA-CN, showing a positive correlation ($R^2 = 0.42$; $P < 0.0001$) (Figure 3C). Among the SCA patients in Cohort 1, the deletion burden also rose with age, being higher in those aged ≥ 30-years (0.21) compared to those < 30-years (0.17), although the difference was not significant (Figure 3D). Protein coding mtDNA genes including *ND3, ND4, ND5, ND6, and CYTB*, showed a higher deletion burden, with SCA exhibiting the greatest burden across genotypes (Figure 3E).

mtDNA heteroplasmy burden was predominantly composed of transition mutations enriched in the D-loop and Complex I regions

In Cohort 1 (SCD patients, n=673), we identified 268 heteroplasmic occurrences, with the displacement loop (D-loop) region contributing the most (36.94%) followed by Complex I mutations (23.51%) (*Online Supplementary Table S7*). Among SCA patients (n=554), Complex IV (10.1%), tRNA (13.2%), and Complex V (3.5%) regions displayed higher burden compared to the other genotypes (*Online Supplementary Table S7, Supplementary Figure S10*). Of the 268 occurrences, 165 were insertions deletions (Indels) and 103 were single nucleotide polymorphisms (SNPs), predominantly transition-type substitutions (98/103, 95.2%) compared to transversions (5/03, 4.8%) (*Online Supplementary Table S8*).

Specific mtDNA variants were associated with increased all-cause mortality in SCD

In Cohort 1 (SCD patients, n=673), we explored the impact of mtDNA heteroplasmy on survival using Cox proportional hazards model. Samples were collected over 23 years (February 2001-July 2024), with a mean follow-up time of 6.26 years and a maximum follow-up duration of 20.16 years (*Online Supplementary Table S9).* While not statistically significant, there was a trend indicating increased risk of all-cause mortality with higher mtDNA heteroplasmy (hazard ratio, HR = 1.01). Categorization by cumulative burden of heteroplasmy further increased the risk (HR =1.17 for 1 heteroplasmy vs. none; HR = 1.3 for 2 heteroplasmies vs none) (*Online Supplementary Table S10A*). This trend persisted after adjusting for age, suggesting a potential association despite limited sample size. Additionally, 7 mito-variants were linked to higher mortality risk, with HRs from 22.09 to 153.98 ($P < 0.05$), mainly in the genes involved in oxidative phosphorylation (Complex I, Complex IV, and Complex V) (Table 1). The highest-risk variant, MT: 8483 (HR = 153.98, $P = 0.0008$) was in the ATP8 gene (Complex V). Of note, six of these seven mito-variants were found in HbSS patients and one in HbSC patient (*Online Supplementary Table S10B*).

In keeping with the inverse association between mtDNA-CN and age, there was a negative trend between mtDNA-CN and mortality (HR = 0.964), consistent after corrected for age (*Online Supplementary Table S10C*).

Correlation of mtDNA genomic variants with NIH Risk Score.

We evaluated the relationship between mtDNA metrics and phenotypic severity in SCD using the NIH risk score, a previously published phenotypic risk prediction model integrating 9 clinical and laboratory variables.[22] A total of 452 SCD patients in Cohort 1 with complete data for all component variables required to calculate NIH risk score were included in this analysis.

Spearman correlation analysis revealed no significant association between NIH risk score and mtDNA-CN, total mtDNA heteroplasmy, or NS heteroplasmy (*Online Supplementary Table S11, Supplementary Figure S11A, C, E).* Stratification of the patients into two risk score groups: high risk (NIH risk score >3) and low risk (NIH risk score ≤ 3), showed that patients with high risk (n=71) had significantly lower mean mtDNA-CN than those with low risk (n=381) (mean mtDNA-CN 348.40 vs. 459.48, P<0.001) (*Online Supplementary Figure S11B),* consistent with the inverse trend between mtDNA-CN and mortality observed in survival analysis. In contrast, neither total mtDNA heteroplasmy nor NS heteroplasmy differed significantly between the two groups (*Online Supplementary Figure S11D, F).*

Specific nuclear variants impacted mtDNA heteroplasmy burden

We examined a panel of nuclear genetic variants with known relevance to SCD red cell biology, oxidative stress, mitochondrial function, and hemolytic process- including *SOD2* V16A, *PIEZO1* E756del, *CYB5R3* T117S, *G6PD* A376G, and *G6PD* G202A (*Online Supplementary Table S12*), to assess their associations with mtDNA variants in SCD patients (Cohort 1, n=673). Minor allele frequency (MAF) of these genetic variants ranged from 0.11 to 0.42 among the patients (*Online Supplementary Table S13*). For association analyses, variant status was encoded as 0 (wild type), 1 (heterozygous), 2 (homozygous). Spearman correlation identified two variants that were significantly (*P<0.05*) associated with mtDNA heteroplasmy. *PIEZO1* E756del showed a weak but statistically significant positive correlation with total mtDNA heteroplasmy (Spearman r=0.077, *P=0.045*). Similarly, *CYB5R3* T117S was positively correlated with total mtDNA heteroplasmy (Spearman r=0.11, *P=0.002*) and NS heteroplasmy (Spearman r=0.09, *P=0.009*) (*Online Supplementary Figure S12**,** Supplementary Table S14*). The remaining variants showed no significant associations, and mtDNA-CN was not significantly associated with any of these variants (*Online Supplementary Figure S12**,** Supplementary Table S14*).

To further investigate these associations, patients were stratified into wild type (WT) or mutant (heterozygous and homozygous) groups. Both mutant groups carrying *PIEZO1* E756del and *CYB5R3* T117S exhibited higher total mtDNA and NS heteroplasmy burden than WT individuals (Figure 4). Specifically, in the *PIEZO1* E756del group, mean mtDNA heteroplasmy was 0.49 in mutant vs. 0.36 in WT (*P=0.045*); in *CYB5R3* T117S, mean total mtDNA heteroplasmy was 0.48 in mutant vs. 0.33 in WT (*P=0.002*). NS heteroplasmy was also higher in *PIEZO1* E756del mutant group, 0.14 vs. 0.09 in WT (*P=ns*) and *CYB5R3* T117S mutant (0.15 vs. 0.07 in WT; *P=0.010*). The remaining variants showed no significant differences in mtDNA heteroplasmy between WT and mutant groups. mtDNA-CN levels did not differ between the mutant and WT groups for any of the variants (*Online Supplementary Table S15*).

None of the nuclear variants showed any significant association with NIH phenotypic risk score among the SCD patients in Cohort 1 (*Online Supplementary Table S16).*

SCD mice exhibited tissue-specific and genotype-dependent mtDNA heteroplasmy burden

Of the 162 mouse tissue samples, 1 sample (AA, male, lung) was removed from analysis due to low coverage (*Online Supplementary Figure S4D*). mtDNA was enriched prior to NGS to achieve a uniform average mito-coverage across the tissues (*Online Supplementary Figure S4E*). A total

of 11 heteroplasmic mito-variants with 106 total heteroplasmic occurrences were detected across 161 mouse tissue samples (*Online Supplementary Table S17*).

Similar to observations in humans, mtDNA heteroplasmy burden increased with increasing severity of the genotypes, with HbSS mice showing the highest mean total mtDNA heteroplasmy burden (0.76) followed by HbAS (0.70) and HbAA (0.51) (Figure 5A). mtDNA heteroplasmy burden differed between tissues within and across genotypes, with spleen having the highest overall mtDNA heteroplasmy burden (17/106 occurrences) (Figure 5B). Among the 11 mito-variants, the variant at mito-position 9820 (MT: 9820, *mt-tr*) was the most prevalent (53/106 occurrences) shared between all 3 genotypes (*Online Supplementary Table S17*, and Figure 5C). Four variants- MT:6753 (Phe476-Leu in COX1), MT:6974 (intergenic), MT:1220 (intergenic), and MT:425 (D-Loop) were found only among the SS mice (Figure 5C), which occurred 22 times across the tissues with spleen having the highest burden (Figure 5D, *Online Supplementary Table S18*). In all 3 genotypes, female mice tended to have higher mtDNA heteroplasmy burden than males (*Online Supplementary Figure S13)*; however, the differences were not statistically significant, the small sample size was insufficient for a reliable and robust comparison.

Mitochondrial genome displayed potential mutation hotspots

We compared mtDNA heteroplasmy distribution across the 37 mitochondrial genes in the 3 human cohorts and mouse samples using NCBI reference sequences (human: NC_012920.1 and mouse: NC_005089.1). The analysis showed a similar distribution of mtDNA heteroplasmy load across cohorts (Figure 6) and across genotypes (*Online Supplementary Figure S14A*). The D-loop, *RNR1, RNR2, ND1, CO1, ND5, ATP6, and CYT-B* genes displayed higher heteroplasmy burdens compared to other mitochondrial genes, suggesting shared hotspots in the mitochondrial genome for both human and mouse samples (*Online Supplementary Figure S14B*).

Discussion

Mitochondrial dysfunction and chronic inflammation are two determinants of the complex aging process and have been implicated in the pathogenesis of a variety of diseases.[4, 5, 23, 24] Inflammatory processes, both acute and chronic, play a key role in SCD pathology leading to high background levels of ROS and oxidative stress that contribute to mitochondrial dysfunction and mtDNA damage, generating more oxidative stress driving the pathological inflammatory cycle in SCD.

Here we provide a comprehensive analysis of alterations in mitochondrial genome in SCD, integrating human cohorts and mouse models. This study demonstrates the impact of two key defects in mitochondrial genomic alterations– changes in mtDNA content (mtDNA-CN) and accumulation of mtDNA mutations in SCD.

We showed for the first time that the mtDNA-CN level and mtDNA heteroplasmy burden per individual increased progressively from HbAA, HbAS < HbSβ+ thalassemia < HbSC < SCA (HbSS & HbSβ0), with genotypic groups that are associated with increasing phenotypic severity. In general, cell types with high ATP demand exhibit higher mtDNA copy numbers.[25, 26] Given that ATP depletion is a characteristic feature of SCD,[27, 28] progressive increase in mtDNA-CN with genotype severity could reflect an adaptive response to meet heightened energy requirements and mitigate oxidative stress, particularly in the more severe HbSS and HbSβ^0 (SCA) genotypes. Consistent with this, elevated mtDNA content is a recognized adaptive mechanism against oxidative stress,[29, 30] as also evidenced by an increase in mtDNA-CN in whole blood from patients with transfusion-dependent β-thalassemia, a condition associated with iron overload and oxidative injury.[31] Non-synonymous (NS) heteroplasmy burden was significantly higher among SCA patients across three cohorts in this study, indicating an accumulation of potentially deleterious mutations in coding regions of mtDNA, which are likely to impact mitochondrial function, including ATP production, and contribute to SCD pathophysiology. In keeping with other studies, we also found a higher heteroplasmic burden in the D-loop.[32, 33] D-loop, being the control region of mtDNA replication initiations is more likely to be exposed to oxidative stress and prone to accumulate mutations, contributing to its higher mutational load.[33, 34] Among the coding mutations, we found a higher ratio of transition to transversion mutations consistent with other reports.[35] The low abundance of transversion mutations, which are mainly caused by oxidative damage and are deleterious, is known to result from repair mechanisms by mitochondrial 8-oxoguanine DNA glycosylase (OGG) which corrects the oxidative lesions, and strong purifying selection in mitochondria that hinders their clonal expansion.[36, 37] Large mtDNA deletions have long been associated with oxidative stress, ageing, and many disease conditions including Kearns-Sayre syndrome (KSS), Pearson syndrome, metabolic diseases, and cancers.[38] In the current study, SCA patients exhibited an increased burden of mtDNA deletions, predominantly affecting genes involved in oxidative phosphorylation- *ND3, ND4, ND5, ND6, and CYTB*. Consistent with our

findings, increased mtDNA deletion burden has also been found in transfusion-dependent β-thalassemia associated with iron overload and oxidative stress.[31]

Hydroxyurea (HU) affects mtDNA-CN and mitochondrial genome stability in model organism such as yeast.[39] HU, standard care therapy for SCD, reduces acute pain and improves overall clinical outcomes but also acts as a myelosuppressive agent.[40, 41] In this study of SCD patients, many of whom are on long-term HU therapy, we explored HU's impact on the mitochondrial genome. HU treatment did not influence the increase in mtDNA heteroplasmy burden associated with genotypic severity. Although mtDNA-CN slightly increased with HU treatment, a genotype-specific increase of mtDNA-CN was observed in both HU treated and untreated ($P < 0.001$) groups. The genotype-dependent increase was independent of contributions from non-nucleated, mtDNA-containing cell types like platelets and reticulocytes, and was not influenced by sex or Hb levels. Together, these findings suggest that intrinsic genotypic severity primarily drives the increase in mtDNA-CN, with a positive correlation with mtDNA deletion burden indicating a compensatory response to mtDNA damage. The number of α globin genes did not have an impact on mtDNA heteroplasmy, but was associated with lower mtDNA-CN in both SCD and SCA patients, potentially reflecting the established beneficial effect of alpha thalassemia coinheritance in ameliorating outcomes of SCD.[42, 43] The increased burden of non-synonymous mtDNA heteroplasmy in SCD patients with *PIEZO1* E756del[44-46] and *CYB5R3* T117S variants underscores their impact on mitochondrial dysfunction in SCD. *PIEZO1* E756del promotes RBC dehydration and calcium overload, disrupting mitochondrial dynamics, while *CYB5R3* T117S[47] impairs NADH-dependent redox balance and hydroxyurea-mediated HbF induction; together exacerbating SCD severity.

In patients with sickle cell anemia, mtDNA heteroplasmy, particularly deleterious non-synonymous variants, and mtDNA deletion burden tended to increase with age, indicating cumulative mitochondrial damage. Concurrently, mtDNA-CN declined with age, and reduced mtDNA-CN was associated with higher NIH risk score, further strengthening the compensatory role of mtDNA-CN in SCD. This study also found that higher mtDNA heteroplasmy burden correlates with increased all-cause mortality, independent of age, and specific mito-variants have greater impact on mortality than cumulative heteroplasmies. These findings align with a study using data from the UK Biobank showing a dose-response relationship between mtDNA

heteroplasmy burden and mortality risk.[16] Other studies have also linked specific mito-variants to increased mortality risks in breast cancer,[48] dementia and stroke among elderly individuals.[49]

In SCD mice, despite a limited sample size, we observed varying mtDNA heteroplasmy across tissues, a phenomenon also reported in other studies.[36, 50] Notably, splenic mtDNA heteroplasmy was highest in HbAA controls and lowest in HbAS mice, highlighting the complexity of mtDNA dynamics across tissues. Given the exploratory nature of the mouse study and limited sample size, tissue specific differences across genotypes would require validation in larger cohorts. The elevated splenic heteroplasmy observed in HbAA may be contributed by mouse-specific splenic environment, and splenic erythrophagocytosis process that may influence mtDNA dysfunction independent of disease genotype. Such fluctuations could be better attenuated in larger sample cohorts.

We also found that certain mitogenome regions consistently showed high mutation loads across genotypes in both humans and mice, indicating recurrent mutational hotspots. Several studies have demonstrated that these hotspots are more prone to accumulating mtDNA mutations, and are linked to disease prognosis, oxidative stress, and inflammation.[51-54]

This study, though limited by sample size, includes SCD patients, sickle carrier, and ethnically matched healthy controls from a single center. We are mindful that our results represent a global assessment of mtDNA defects, without specifying cell type contributions. In conclusion, our findings suggest that variations in mitochondrial genome integrity, such as alterations in mtDNA-CN and mtDNA heteroplasmy burden, could serve as potential biomarkers for SCD. A recent study in SCD mice[55] demonstrated increased clonal hematopoiesis and mutational burden with aging in SS mice, especially in the hematological malignancy related genes, underscoring their prognostic relevance in SCD. Likewise, mtDNA genome variations- including changes in mtDNA-CN, mtDNA heteroplasmy burden, and specific mtDNA variants could serve as biomarker of disease severity and accelerated aging in SCD, with a potential utility for assessing individual risk prognosis and inform clinical management. mtDNA alterations may serve as prognostic markers and reflect mitochondrial dysfunction in SCD progression. Elevated oxidative stress and ROS in SCD cause cumulative mtDNA damage, impair mtDNA and mitochondrial function, and further increase ROS production, creating a vicious cycle that worsens disease severity. Future longitudinal studies are needed to clarify the mechanisms and rate of dysfunctional mtDNA

accumulation, identify critical heteroplasmy thresholds, and determine how mtDNA hotspots are linked to SCD pathophysiology.

References

1. Conran N, Belcher JD. Inflammation in sickle cell disease. Clin Hemorheol Microcirc. 2018;68(2-3):263-299.
2. Sundd P, Gladwin MT, Novelli EM. Pathophysiology of Sickle Cell Disease. Ann Rev Pathol. 2019;14:263-292.
3. Barbu EA, Mendelsohn L, Samsel L, Thein SL. Pro-inflammatory cytokines associate with NETosis during sickle cell vaso-occlusive crises. Cytokine. 2020;127:154933.
4. Idris IM, Botchwey EA, Hyacinth HI. Sickle cell disease as an accelerated aging syndrome. Exp Biol Med (Maywood). 2022;247(4):368-374.
5. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. Nat Rev Endocrinol. 2018;14(10):576-590.
6. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. Cell. 2013;153(6):1194-1217.
7. Jang JY, Blum A, Liu J, Finkel T. The role of mitochondria in aging. J Clin Invest. 2018;128(9):3662-3670.
8. Filograna R, Mennuni M, Alsina D, Larsson NG. Mitochondrial DNA copy number in human disease: the more the better? FEBS Lett. 2021;595(8):976-1002.
9. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat Rev Genet. 2015;16(9):530-542.
10. Pinti M, Cevenini E, Nasi M, et al. Circulating mitochondrial DNA increases with age and is a familiar trait: Implications for "inflamm-aging". Eur J Immunol. 2014;44(5):1552-1562.
11. Tumburu L, Ghosh-Choudhary S, Seifuddin FT, et al. Circulating mitochondrial DNA is a proinflammatory DAMP in sickle cell disease. Blood. 2021;137(22):3116-3126.
12. Gallivan A, Alejandro M, Kanu A, et al. Reticulocyte mitochondrial retention increases reactive oxygen species and oxygen consumption in mouse models of sickle cell disease and phlebotomy-induced anemia. Exp Hematol. 2023;122:55-62.
13. Choi J, Beroncal EL, Chernega T, et al. Exploring mitochondrial blood-based and genetic markers in older adults with mild cognitive impairment and remitted major depressive disorder. Transl Psychiatry. 2024;14(1):457.
14. Wei R, Ni Y, Bazeley P, et al. Mitochondrial DNA Content Is Linked to Cardiovascular Disease Patient Phenotypes. J Am Heart Assoc. 2021;10(4):e018776.
15. Reznik E, Miller ML, Senbabaoglu Y, et al. Mitochondrial DNA copy number variation across human cancers. Elife. 2016;5:e10769.
16. Hong YS, Battle SL, Shi W, et al. Deleterious heteroplasmic mitochondrial mutations are associated with an increased risk of overall and cancer-specific mortality. Nat Commun. 2023;14(1):6113.
17. Hong YS, Pasca S, Shi W, et al. Mitochondrial heteroplasmy improves risk prediction for myeloid neoplasms. Nat Commun. 2024;15(1):10133.

18. Risi B, Imarisio A, Cuconato G, Padovani A, Valente EM, Filosto M. Mitochondrial DNA (mtDNA) as fluid biomarker in neurodegenerative disorders: A systematic review. Eur J Neurol. 2025;32(1):e70014.
19. Feng J, Chen Z, Liang W, Wei Z, Ding G. Roles of Mitochondrial DNA Damage in Kidney Diseases: A New Biomarker. Int J Mol Sci. 2022;23(23):15166.
20. Battle SL, Puiu D, TOPMed mtDNA Working Group, et al. A bioinformatics pipeline for estimating mitochondrial DNA copy number and heteroplasmy levels from whole genome sequencing data. NAR Genom Bioinform. 2022;4(2):lqac034.
21. Ding J, Sidore C, Butler TJ, et al. Assessing Mitochondrial DNA Variation and Copy Number in Lymphocytes of ~2,000 Sardinians Using Tailored Sequencing Analysis Tools. PLoS Genet. 2015;11(7):e1005306.
22. Sachdev V, Tian X, Gu Y, et al. A phenotypic risk score for predicting mortality in sickle cell disease. Br J Haematol. 2021;192(5):932-941.
23. Sanada F, Taniyama Y, Muratsu J, et al. Source of Chronic Inflammation in Aging. Front Cardiovasc Med. 2018;5:12.
24. Bartman S, Coppotelli G, Ross JM. Mitochondrial Dysfunction: A Key Player in Brain Aging and Diseases. Curr Issues Mol Biol. 2024;46(3):1987-2026.
25. Lee W, Johnson J, Gough DJ, et al. Mitochondrial DNA copy number is regulated by DNA methylation and demethylation of POLGA in stem and cancer cells and their differentiated progeny. Cell Death Dis. 2015;6(2):e1664.
26. Castellani CA, Longchamps RJ, Sumpter JA, et al. Mitochondrial DNA copy number can influence mortality and cardiovascular disease via methylation of nuclear DNA CpGs. Genome Med. 2020;12(1):84.
27. McMahon TJ, Darrow CC, Hoehn BA, Zhu H. Generation and Export of Red Blood Cell ATP in Health and Disease. Front Physiol. 2021;12:754638.
28. Banerjee T, Kuypers FA. Reactive oxygen species and phosphatidylserine externalization in murine sickle red cells. Br J Haematol. 2004;124(3):391-402.
29. Lee HC, Yin PH, Lu CY, Chi CW, Wei YH. Increase of mitochondria and mitochondrial DNA in response to oxidative stress in human cells. Biochem J. 2000;348 Pt 2(Pt 2):425-432.
30. Liu CS, Tsai CS, Kuo CL, et al. Oxidative stress-related alteration of the copy number of mitochondrial DNA in human leukocytes. Free Radic Res. 2003;37(12):1307-1317.
31. Lal A, Gomez E, Calloway C. Increased mitochondrial DNA deletions and copy number in transfusion-dependent thalassemia. JCI Insight. 2016;1(12):e88150.
32. Khrapko K, Coller HA, Andre PC, Li XC, Hanekamp JS, Thilly WG. Mitochondrial mutational spectra in human cells and tissues. Proc Natl Acad Sci U S A. 1997;94(25):13798-13803.
33. Nicholls TJ, Minczuk M. In D-loop: 40 years of mitochondrial 7S DNA. Exp Gerontol. 2014;56:175-181.

34. Rothfuss O, Gasser T, Patenge N. Analysis of differential DNA damage in the mitochondrial genome employing a semi-long run real-time PCR approach. Nucleic Acids Res. 2010;38(4):e24.
35. Gupta R, Kanai M, Durham TJ, et al. Nuclear genetic control of mtDNA copy number and heteroplasmy in humans. Nature. 2023;620(7975):839-848.
36. Sanchez-Contreras M, Sweetwyne MT, Tsantilas KA, et al. The multi-tissue landscape of somatic mtDNA mutations indicates tissue-specific accumulation and removal in aging. Elife. 2023;12:e83395.
37. Cote-L'Heureux A, Maithania YNK, Franco M, Khrapko K. Are some mutations more equal than others? Elife. 2023;12:e87194.
38. Fontana GA, Gahlon HL. Mechanisms of replication and repair in mitochondrial DNA deletion formation. Nucleic Acids Res. 2020;48(20):11244-11258.
39. Li J, Stenberg S, Yue JX, et al. Genome instability footprint under rapamycin and hydroxyurea treatments. PLoS Genet. 2023;19(11):e1011012.
40. Yawn BP, Buchanan GR, Afenyi-Annan AN, et al. Management of sickle cell disease: summary of the 2014 evidence-based report by expert panel members. JAMA. 2014;312(10):1033-1048.
41. Wong TE, Brandow AM, Lim W, Lottenberg R. Update on the use of hydroxyurea therapy in sickle cell disease. Blood. 2014;124(26):3850-3857
42. Raffield LM, Ulirsch JC, Naik RP, et al. Common alpha-globin variants modify hematologic and other clinical phenotypes in sickle cell trait and disease. PLoS Genet. 2018;14(3):e1007293.
43. Darbari DS, Nouraie M, Taylor JG, Brugnara C, Castro O, Ballas SK. Alpha-thalassaemia and response to hydroxyurea in sickle cell anaemia. Eur J Haematol. 2014;92(4):341-345.
44. Rooks H, Brewin J, Gardner K, et al. A gain of function variant in PIEZO1 (E756del) and sickle cell disease. Haematologica. 2019;104(3):e91-e93.
45. Ilboudo Y, Bartolucci P, Garrett ME, et al. A common functional PIEZO1 deletion allele associates with red blood cell density in sickle cell disease patients. Am J Hematol. 2018;93(11):E362-E365.
46. Romero LO, Bade M, Elsherif L, et al. Enhanced PIEZO1 function contributes to the pathogenesis of sickle cell disease. Proc Natl Acad Sci U S A. 2025;122(40):e2514863122.
47. Chowdhury FA, Sharma M, Schafer D, et al. CYB5R3 T117S tempers fetal hemoglobin induction by hydroxyurea in patients with sickle cell disease. Blood Adv. 2024;8(23):6098-6103.
48. Li Y, Sundquist K, Vats S, et al. Mitochondrial heteroplasmic shifts reveal a positive selection of breast cancer. J Transl Med. 2023;21(1):696.
49. Tranah GJ, Katzman SM, Lauterjung K, et al. Mitochondrial DNA m.3243A > G heteroplasmy affects multiple aging phenotypes and risk of mortality. Sci Rep. 2018;8(1):11887.
50. Li M, Schroder R, Ni S, Madea B, Stoneking M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. Proc Natl Acad Sci U S A. 2015;112(8):2491-2496.

51. Galtier N, Enard D, Radondy Y, Bazin E, Belkhir K. Mutation hot spots in mammalian mitochondrial DNA. Genome Res. 2006;16(2):215-222.
52. Hertweck KL, Dasgupta S. The Landscape of mtDNA Modifications in Cancer: A Tale of Two Cities. Front Oncol. 2017;7:262.
53. Stoneking M. Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet. 2000;67(4):1029-1032.
54. Chinnery PF, Brown DT, Andrews RM, et al. The mitochondrial ND6 gene is a hot spot for mutations that cause Leber's hereditary optic neuropathy. Brain. 2001;124(Pt 1):209-218.
55. Mengna C, Lucas F, Husami A, et al. Replicative stress-induced aging of hematopoietic stem progenitor cells increases oncogenic mutation burden and incidence of myelodysplasia in sickle cell disease mice. Blood. 2025;146(Supplement 1):11.

Tables

Table 1. Specific mito-variants associated with higher hazard ratio for all-cause mortality.

Mito-Variants	Hazard Ratio			Significance (P value)		Mutation Type	Base Change	Impact	Gene Affected	Functional Category
	HR	HR.L	HR.R	*P*	*P*.Adj					
MT: 8483	153.98	17.20	1378.18	6.66E-06	0.000839	SNP	A/G	Coding	ATP8	Complex V
MT:16015	40.65	5.36	308.04	0.0003363	0.021190	INS	Indel	Non-Coding	MT-TT	MT- tRNA
MT: 8156	32.54	4.34	243.78	0.0007002	0.029408	SNP	A/G	Coding	ND2	Complex I
MT: 4916	26.37	3.55	195.62	0.0013740	0.042967	SNP	A/G	Coding	ND2	Complex I
MT: 4658	24.05	3.25	177.81	0.0018338	0.042967	SNP	T/C	Coding	ND2	Complex I
MT: 4695	23.04	3.12	170.01	0.0020980	0.042967	SNP	G/T	Coding	COX2	Complex IV
MT: 6158	22.09	3.00	162.82	0.0023870	0.042967	SNP	A/G	Coding	COX1	Complex IV

Figure legends

Figure 1: mtDNA copy number (mtDNA-CN) increased with genotypic severity and declined with age. (A) mtDNA-CN across genotypes in Cohort 1. Mean mtDNA-CN in SCA (n=554) compared with HbSC (n=91) were 483.9 vs 379.2 (P = 0.0017, **), and with HbSB+ (n=26), were 483.9 vs 299.4 (P = 0.0010, **). **(B)** mtDNA-CN across genotypes in Cohort 2. SCA (n=113) had significantly higher mean mtDNA-CN (804.7) compared to HbAS (n=30, mean mtDNA-CN 246.3) (P < 0.0001, ****), and HbAA (n=16, mean mtDNA-CN 244.9) (P = 0.0019, **). **(C)** mtDNA-CN across genotypes in Cohort 3. mtDNA-CN was significantly higher among SCA (n=96) (mean =796.7) compared to HbAS (n=47) (mean = 217.1) (P < 0.0001, ****) and HbAA (n=43) (mean =239) (P < 0.0001, ****). **(D)** Comparison of mtDNA-CN levels across combined genotypes in cohorts 1, 2, and 3. Among the total samples (n=1043) from three human cohorts combined, SCA (n=763) had significantly higher mtDNA-CN compared to other genotypes: SC (mean = 367.8, n=109) (*P<0.0001*,****), SB+ (mean=313.30, n=30) (*P=0.0017*, **), AS (mean=228.5, n=77) (*P<0.0001*, ****), and AA (mean=240.6, n=60) (*P<0.0001*, ****). (**E**) mtDNA-CN number stratified by age among the SCA patients (n=763) combined from cohorts 1, 2 and 3. Mean mtDNA-CN was higher in the younger age group (< 30 years; n=373) (mean= 582.3) compared to the older group (≥ 30 years, n=390) (mean = 559.8); (P = ns). (**F**) mtDNA-CN levels among the SCA patients (n=763) across four age groups among the SCA patients (n=763) combined from cohorts 1, 2, and 3. mtDNA-CN level decreased gradually with age: < 25 years (n = 228, mean = 594.0), 25–34 years (n = 268, mean = 609.3), 35–44 years (n = 148, mean = 512.3), and ≥45 years (n = 119, mean = 512.1). Statistical significance (P value) was assessed by Kruskal-Wallis test (between three or more groups) or Mann Whitney test (between two groups). If no other indication, results were not significant.

Figure 2: Burden of non-synonymous mtDNA heteroplasmy (NS Heteroplasmy) increased with severity of genotypes and age. (**A**) Burden of NS heteroplasmy across genotypes in Cohort-1. SCA (n=554) group had higher burden of NS heteroplasmy (mean =0.11) compared to HbSB+ and HbSC genotypes together (n= 116, mean =0.05) (*P= 0.0263*, *). **(B)** Burden of NS heteroplasmy across genotypes in Cohort 2. The burden of NS heteroplasmy was higher in SCA (n=113, mean= 0.14) compared to HbAS & HbAA together (n= 46, mean= 0.02) (*P=0.0434*, *).

(C) Burden of NS heteroplasmy across genotypes in Cohort 3. SCA (n=96) had higher burden of NS heteroplasmy (mean=0.10) compared to HbAS & HbAA together (n=91, mean 0.03) (*P =0.05826*) (**D**) Burden of NS heteroplasmy across genotypes in combined human cohorts (Cohort 1, Cohort 2, and Cohort 3 combined). SCA subjects (n=763) combined from all cohorts had significantly higher burden of NS heteroplasmy (mean=0.12) compared to rest of the genotypes together (n=280, mean=0.05) (*P=0.0017*,**). (**E**) NS heteroplasmy burden with age among the SCA patients (n=763) combined from three human cohorts. NS heteroplasmy burden increased in the older age group (≥ 30 years; n=390, mean = 0.14) compared to the younger age group (< 30 years; n=373, mean=0.09) (*P= 0.0397*, *). (**F**) NS heteroplasmy burden across four age groups among the SCA patients (n=763) combined from three human cohorts. NS heteroplasmy increased with age: < 25 years (n = 228, mean = 0.10), 25–34 years (n = 268, mean = 0.11), 35–44 years (n = 148, mean = 0.11), and ≥45 years (n = 119, mean = 0.17). NS Heteroplasmy: non-synonymous heteroplasmy. Statistical significance (*P* value) was assessed by Kruskal-Wallis test (between three or more groups) or Mann Whitney test (between two groups). If no other indication, results were not significant (*P≥0.05*).

Figure 3: Burden of mtDNA Deletions among the SCD patients in Cohort 1. (A) Burden of mtDNA deletions (>100bp) increased with genotypic severity in Cohort-1. SCA subjects (n=554) had higher burden of mtDNA deletions (mean=0.19) compared to other genotypes: SC (n=91, mean= 0.06) (*P=0.0467*, *) , SB+ (n=25, mean= 0.04) individually (**B**) Comparison among SCA (n=554) and rest of the samples combined (Rest, n= 119) in Cohort 1 (n=673) showed that SCA subjects (n=554) had a significantly higher burden of mtDNA deletions (mean=0.19) compared to the Rest (mean=0.05) (*P=0.008*, **). (**C**) Pearson correlation between mtDNA deletions and mtDNA-CN in Cohort 1 (n=673). mtDNA deletions displayed a positive correlation ($R^2 = 0.42$; *P* < 0.0001, **) with mtDNA-CN. (**D**) Burden of mtDNA deletions with age among the SCA patients (n=554) in Cohort 1. mtDNA deletions burden was higher in the older age group (≥ 30 years; n=296, mean = 0.21) compared to the < 30 years age group (n=258, mean =0.17). (**E**) Distribution of mtDNA deletions burden across mito-genome. Circular representation (Circos plot) of the human mitochondrial genome where the outer most ring represents different mito-genes indicated in different color. Inner bars indicate genomic locations with mtDNA deletion burden, and the bar height reflects burden frequency. Bar colors represent genotypes (SCA,

orange; SC, blue; SB^+, green). A higher burden of mtDNA deletions was observed in protein-coding genes: *ND3, ND4, ND5, ND6, CYTB*. The overall pattern of mtDNA deletion burden was similar between genotypes; with SCA having higher burden. Statistical significance (*P* value) was assessed by Kruskal-Wallis test (between three or more groups) or Mann Whitney test (between two groups). If no other indication, results were not significant ($P \geq 0.05$).

Figure 4: Association of nuclear variants *PIEZO1* E756del (rs572934641) and *CYB5R3* T117S (rs1800457) with mtDNA heteroplasmy burden. SCD patients in Cohort 1(n=673) were stratified into two groups: WT (wild type) and Mutant (heterozygous and homozygous). Mean of mtDNA-CN, total mtDNA heteroplasmy, and NS heteroplasmy was compared between the two groups. (**A**) Mean total mtDNA heteroplasmy burden was significantly higher in *PIEZO1* E756del mutant (n=207, mean=0.39) compared to WT (n=466, 0.36); *P=0.045*. (**B**) Mean NS heteroplasmy burden was higher in *PIEZO1* E756del mutant (n=207, mean=0.14) compared to WT (n=466, 0.09); *P=ns*. (**C**) Mean total mtDNA heteroplasmy burden was significantly higher in *CYB5R3* T117S mutant (n=356, mean=0.48) compared to WT (n=317, 0.33); *P=0.002*. (**D**) Mean NS heteroplasmy burden was significantly higher in *CYB5R3* T117S mutant (n=356, mean=0.15) compared to WT (n=317, 0.07); *P=0.010*. Statistical significance (*P* value) was assessed by Mann Whitney test. If no other indication, results were not significant ($P \geq 0.05$).

Figure 5: mtDNA heteroplasmy burden among the mouse samples. In total 106 heteroplasmic occurrences originating from 11 unique mito-variants were obtained among all mice samples (n=161, 1 sample removed due to dropped coverage). (**A**) mtDNA heteroplasmy burden among three genotypes. HbSS mice (n=54) showed highest mtDNA heteroplasmy (mean=0.76), followed by HbAS (n=54; mean=0.70), and HbAA (n=53, 1 sample removed due to dropped coverage; mean=0.51) (*P = ns*). (**B**) Tissue wise comparison of mtDNA heteroplasmy burden across three genotypes. The number above each bar in the graph indicates the total heteroplasmic burden for the respective tissue, while the number within each bar represents the genotype-specific heteroplasmic burden within each tissue. The burden of mtDNA heteroplasmy varied across different tissues both within and across genotypes. Spleen had the overall highest burden (n=17) of mtDNA heteroplasmy. (**C**) The distribution of 11 unique mito- variants across three genotypes. The number above each bar in the graph indicates the occurrences of the respective mito-variants

in each genotype. (**D**) The distribution of the HbSS unique variants (MT: 6753, MT: 6794, MT: 1220, and MT: 425) across all nine tissues. The number above each bar in the graph indicates the occurrences of HbSS unique variants across tissues. Statistical significance (*P* value) was assessed by Kruskal-Wallis test (between three or more groups) or Mann Whitney test (between two groups). If no other indication, results were not significant ($P \geq 0.05$).

Figure 6: Pattern of mtDNA mutations load in mito-genome across cohorts and genotypes. The burden of mtDNA heteroplasmy across the mito-genome was assessed across 4 cohorts (3 human cohorts and mouse cohort) and each genotype within respective cohorts. All 4 cohorts were subjected to same variant calling pipeline (LoFreq variant caller, VAF: 1% - 90%, Coverage depth > 1000x) to obtain mtDNA heteroplasmic variants. mtDNA heteroplasmic load across the mitochondrial genes was quantified and visualized for each genotype within their respective cohorts. The top bar with different colors represents the mitochondrial genome, consisting of 39 regions- 37 mitochondrial genes, the D loop, and intergenic regions. All the heteroplasmies originating from the unspecified stretch of sequences between two genes are grouped as "Intergenic" and represented at the right-side end of mitogenome bar in the figure. D-loop, RNR1, RNR2, ND1, CO1, ND5, ATP6, and CYT-B exhibited higher mtDNA heteroplasmy burden compared to the rest of the other genes in the mito-genome, consistent across the cohorts and genotypes.

Figure 1

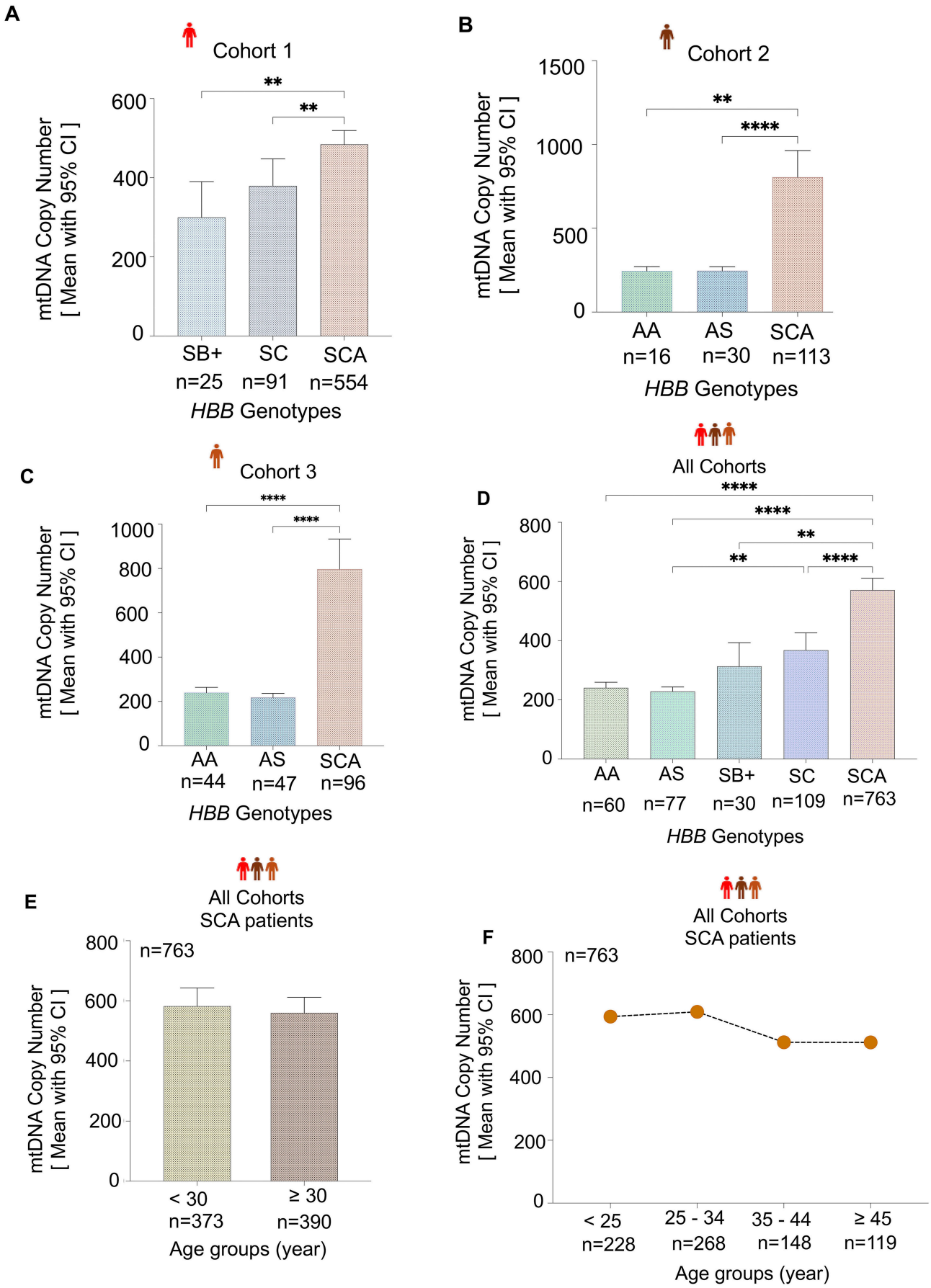

A
Cohort 1
mtDNA Copy Number
[Mean with 95% CI]
600
400
200
0
**
**
SB+
n=25
SC
n=91
SCA
n=554
HBB Genotypes
B
Cohort 2
mtDNA Copy Number
[Mean with 95% CI]
1500
1000
500
0
**
****
AA
n=16
AS
n=30
SCA
n=113
HBB Genotypes
C
Cohort 3
mtDNA Copy Number
[Mean with 95% CI]
1000
800
600
400
200
0
****
****
AA
n=44
AS
n=47
SCA
n=96
HBB Genotypes
D
All Cohorts
mtDNA Copy Number
[Mean with 95% CI]
800
600
400
200
0
****
****
**
**
****
AA
n=60
AS
n=77
SB+
n=30
SC
n=109
SCA
n=763
HBB Genotypes
E
All Cohorts
SCA patients
n=763
mtDNA Copy Number
[Mean with 95% CI]
800
600
400
200
0
< 30
n=373
≥ 30
n=390
Age groups (year)
F
All Cohorts
SCA patients
n=763
mtDNA Copy Number
[Mean with 95% CI]
800
600
400
200
0
< 25
n=228
25 - 34
n=268
35 - 44
n=148
≥ 45
n=119
Age groups (year)

Figure 2

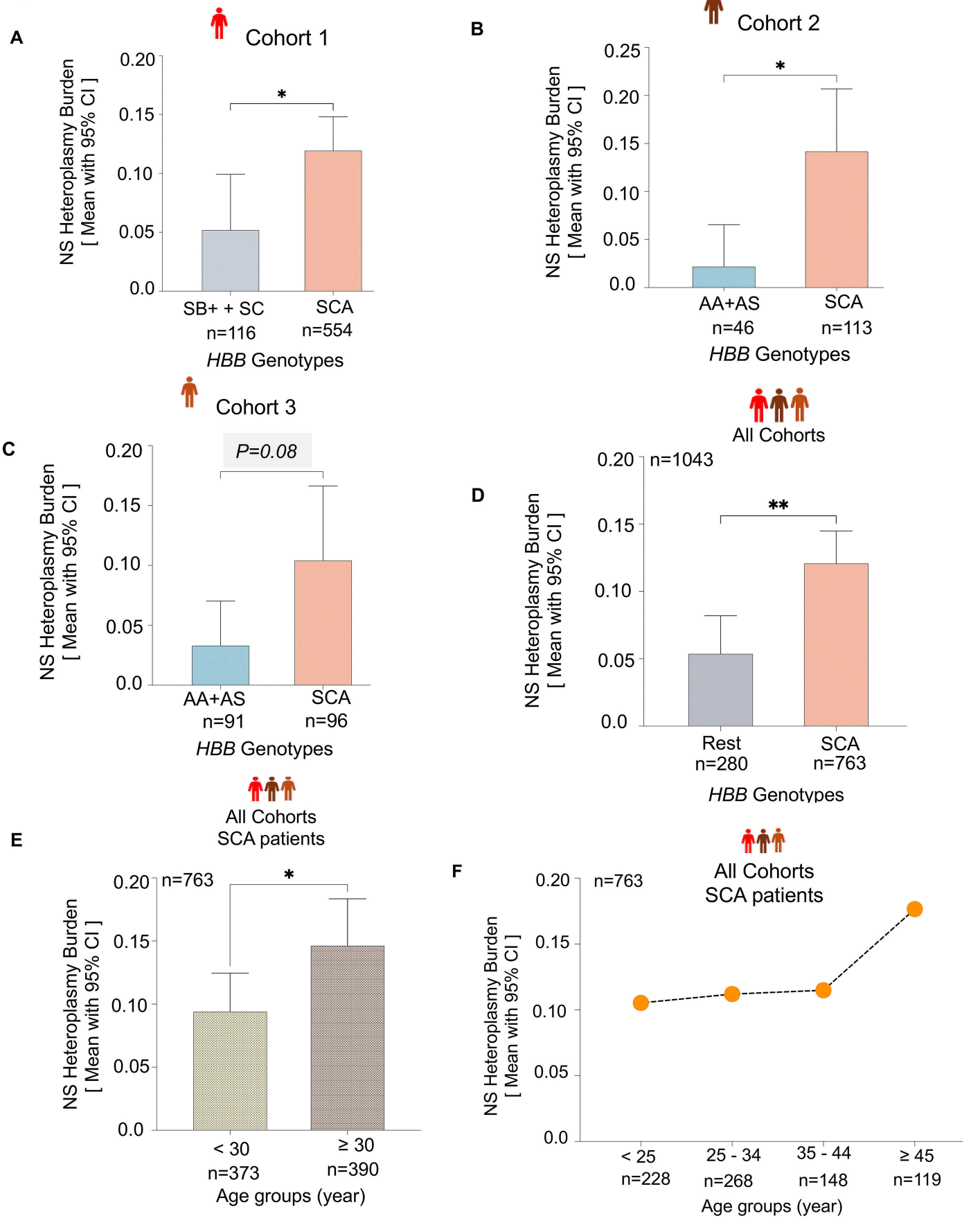

A
Cohort 1
NS Heteroplasmy Burden
[Mean with 95% CI]
0.20
0.15
0.10
0.05
0.0
*
SB+ + SC
n=116
SCA
n=554
HBB Genotypes
B
Cohort 2
NS Heteroplasmy Burden
[Mean with 95% CI]
0.25
0.20
0.15
0.10
0.05
0.0
*
AA+AS
n=46
SCA
n=113
HBB Genotypes
C
Cohort 3
P=0.08
NS Heteroplasmy Burden
[Mean with 95% CI]
0.20
0.15
0.10
0.05
0.0
AA+AS
n=91
SCA
n=96
HBB Genotypes
D
All Cohorts
n=1043
**
NS Heteroplasmy Burden
[Mean with 95% CI]
0.20
0.15
0.10
0.05
0.0
Rest
n=280
SCA
n=763
HBB Genotypes
E
All Cohorts
SCA patients
n=763
*
NS Heteroplasmy Burden
[Mean with 95% CI]
0.20
0.15
0.10
0.05
0.0
< 30
n=373
≥ 30
n=390
Age groups (year)
F
All Cohorts
SCA patients
n=763
NS Heteroplasmy Burden
[Mean with 95% CI]
0.20
0.15
0.10
0.05
0.0
< 25
n=228
25 - 34
n=268
35 - 44
n=148
≥ 45
n=119
Age groups (year)

Figure 3

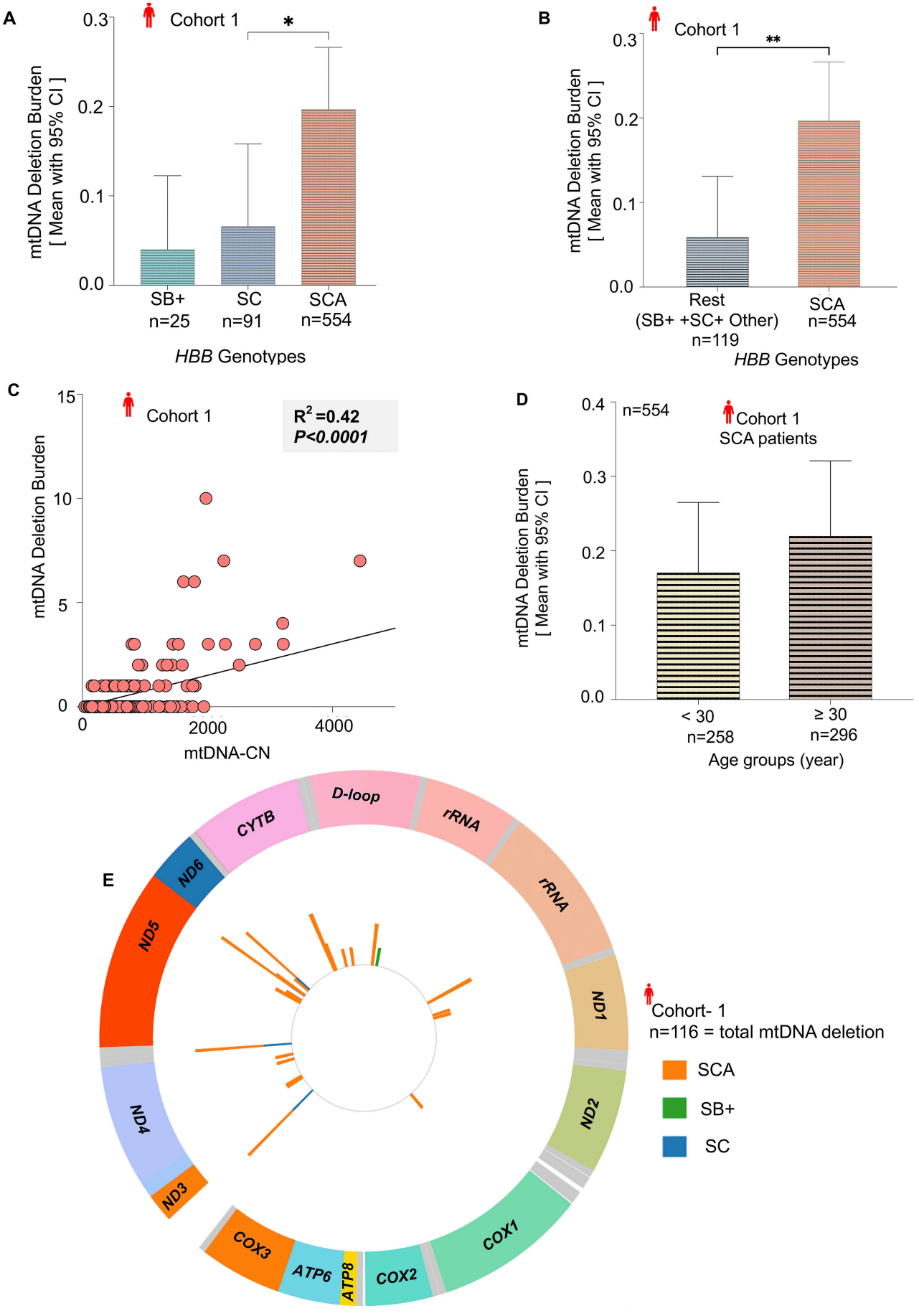

A
Cohort 1
*
mtDNA Deletion Burden
[Mean with 95% CI]
0.3
0.2
0.1
0.0
SB+
n=25
SC
n=91
SCA
n=554
HBB Genotypes
B
Cohort 1
**
mtDNA Deletion Burden
[Mean with 95% CI]
0.3
0.2
0.1
0.0
Rest
(SB+ +SC+ Other)
n=119
SCA
n=554
HBB Genotypes
C
Cohort 1
R2 =0.42
P<0.0001
mtDNA Deletion Burden
15
10
5
0
0
2000
4000
mtDNA-CN
D
n=554
Cohort 1
SCA patients
mtDNA Deletion Burden
[Mean with 95% CI]
0.4
0.3
0.2
0.1
0.0
< 30
n=258
≥ 30
n=296
Age groups (year)
E
D-loop
rRNA
rRNA
ND1
ND2
COX1
COX2
ATP8
ATP6
COX3
ND3
ND4
ND5
ND6
CYTB
Cohort- 1
n=116 = total mtDNA deletion
SCA
SB+
SC

Figure 4

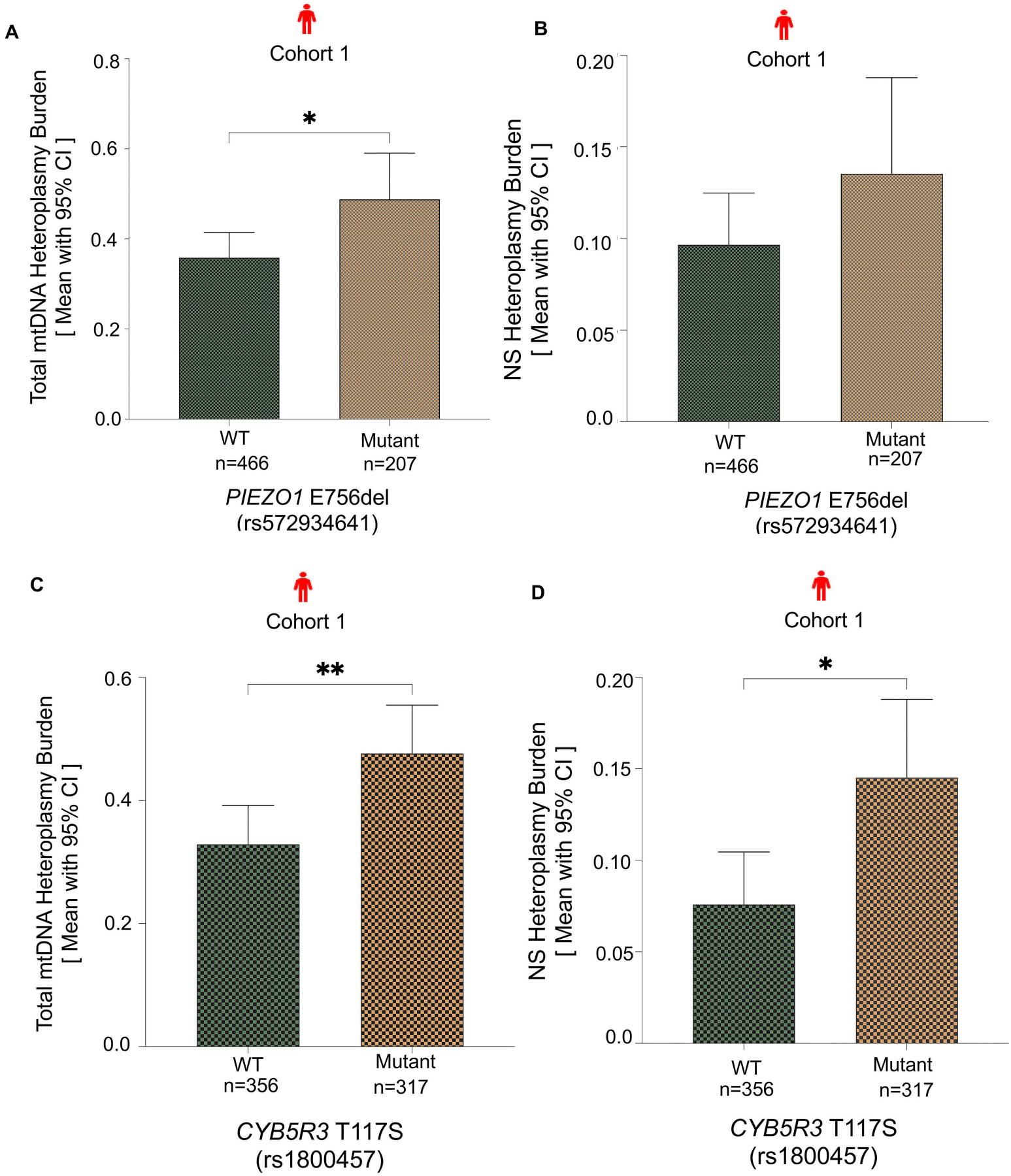

A
Cohort 1
Total mtDNA Heteroplasmy Burden
[Mean with 95% CI]
0.8
0.6
0.4
0.2
0.0
*
WT
n=466
Mutant
n=207
PIEZO1 E756del
(rs572934641)
B
Cohort 1
NS Heteroplasmy Burden
[Mean with 95% CI]
0.20
0.15
0.10
0.05
0.0
WT
n=466
Mutant
n=207
PIEZO1 E756del
(rs572934641)
C
Cohort 1
Total mtDNA Heteroplasmy Burden
[Mean with 95% CI]
0.6
0.4
0.2
0.0
**
WT
n=356
Mutant
n=317
CYB5R3 T117S
(rs1800457)
D
Cohort 1
NS Heteroplasmy Burden
[Mean with 95% CI]
0.20
0.15
0.10
0.05
0.0
*
WT
n=356
Mutant
n=317
CYB5R3 T117S
(rs1800457)

Figure 5

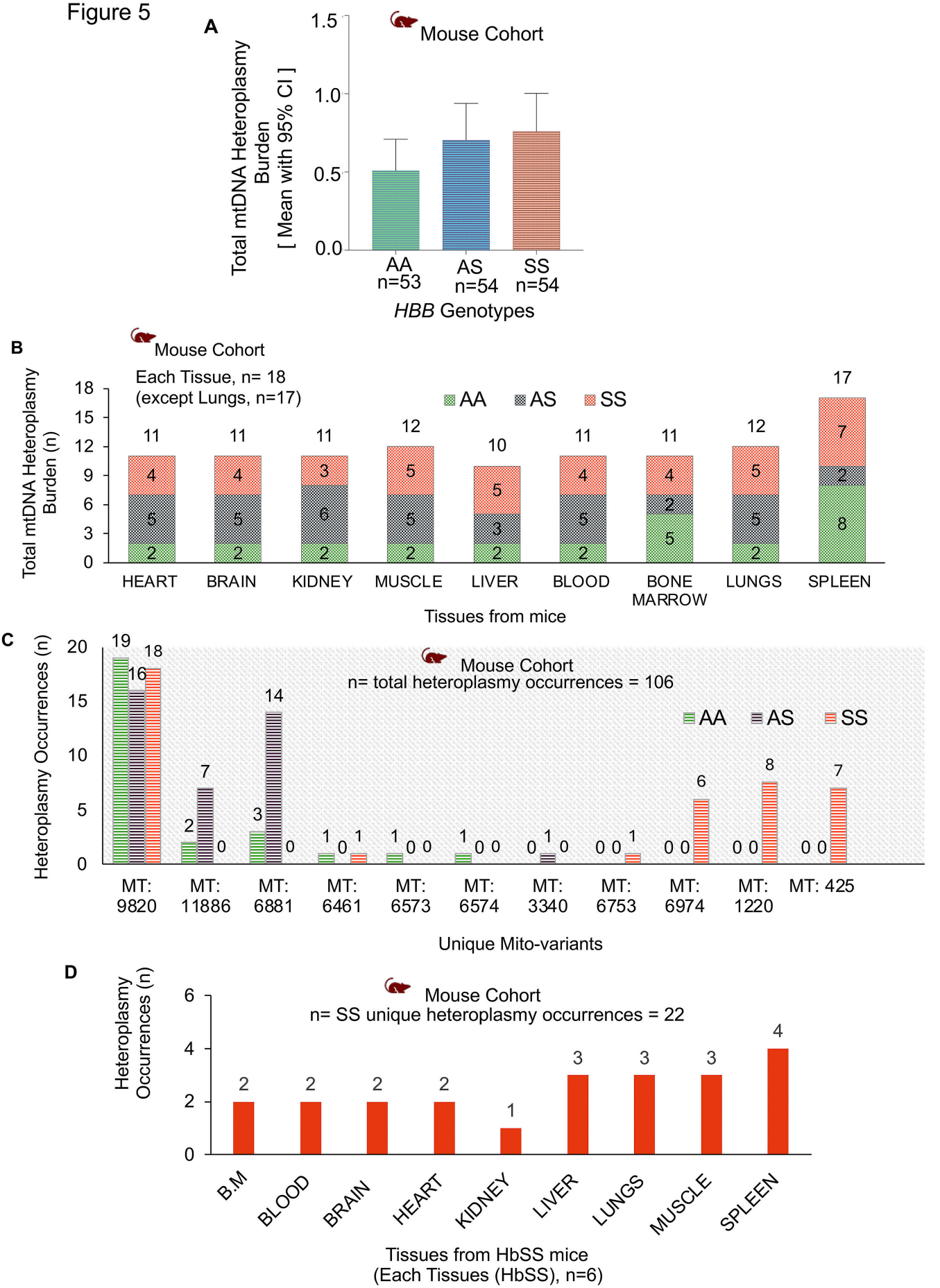
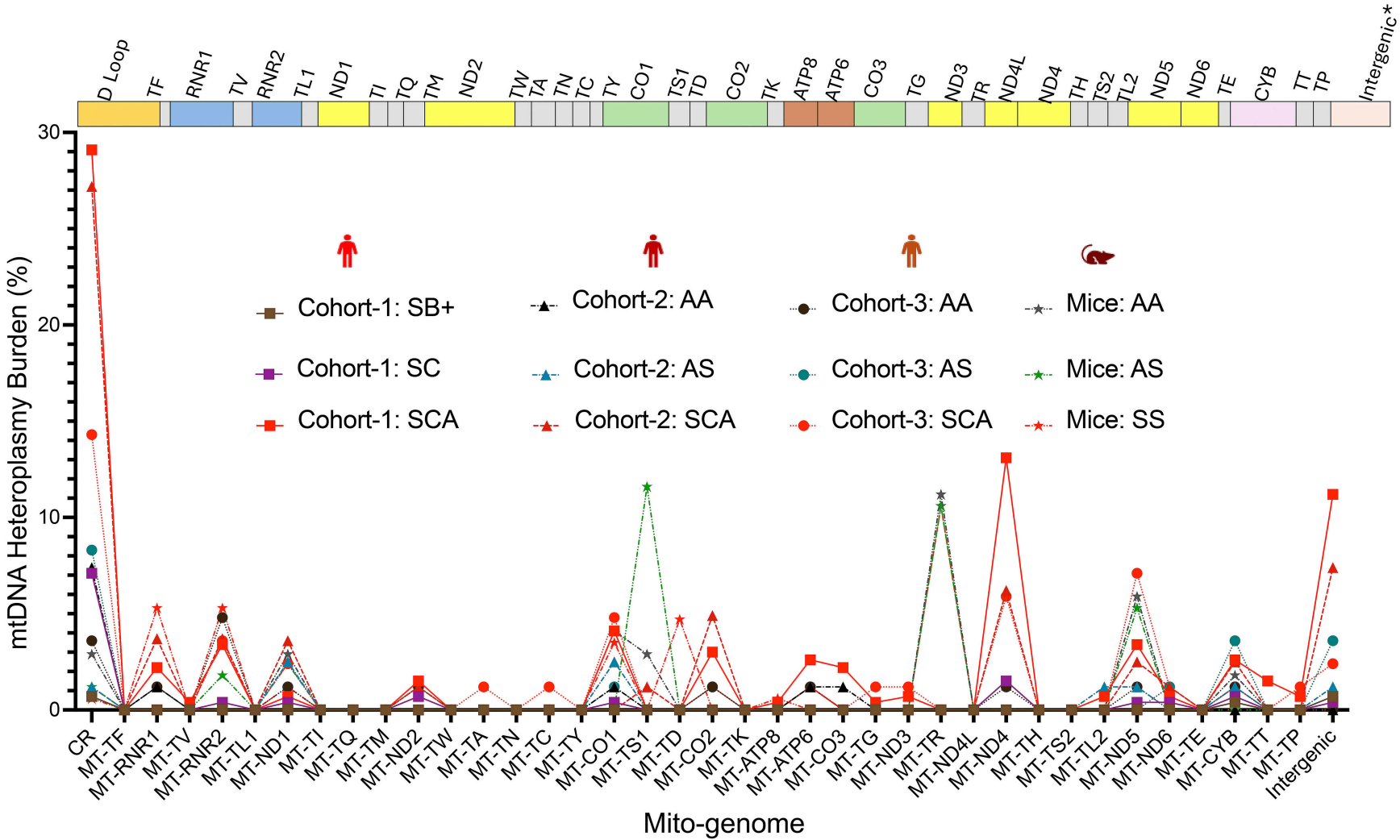

A
Mouse Cohort
Total mtDNA Heteroplasmy Burden [Mean with 95% CI]
1.5
1.0
0.5
0.0
AA n=53
AS n=54
SS n=54
HBB Genotypes
B
Mouse Cohort
Each Tissue, n= 18 (except Lungs, n=17)
AA
AS
SS
Total mtDNA Heteroplasmy Burden (n)
HEART
BRAIN
KIDNEY
MUSCLE
LIVER
BLOOD
BONE MARROW
LUNGS
SPLEEN
Tissues from mice
C
Mouse Cohort
n= total heteroplasmy occurrences = 106
Heteroplasmy Occurrences (n)
MT: 9820
MT: 11886
MT: 6881
MT: 6461
MT: 6573
MT: 6574
MT: 3340
MT: 6753
MT: 6974
MT: 1220
MT: 425
Unique Mito-variants
D
Mouse Cohort
n= SS unique heteroplasmy occurrences = 22
Heteroplasmy Occurrences (n)
B.M
BLOOD
BRAIN
HEART
KIDNEY
LIVER
LUNGS
MUSCLE
SPLEEN
Tissues from HbSS mice
(Each Tissues (HbSS), n=6)

Figure 6


D Loop
TF
RNR1
TV
RNR2
TL1
ND1
TI
TQ
TM
ND2
TW
TA
TN
TC
TY
CO1
TS1
TD
CO2
TK
ATP8
ATP6
CO3
TG
ND3
TR
ND4L
ND4
TH
TS2
TL2
ND5
ND6
TE
CYB
TT
TP
Intergenic*
Cohort-1: SB+
Cohort-1: SC
Cohort-1: SCA
Cohort-2: AA
Cohort-2: AS
Cohort-2: SCA
Cohort-3: AA
Cohort-3: AS
Cohort-3: SCA
Mice: AA
Mice: AS
Mice: SS
mtDNA Heteroplasmy Burden (%)
30
20
10
0
CR
MT-TF
MT-RNR1
MT-TV
MT-RNR2
MT-TL1
MT-ND1
MT-TI
MT-TQ
MT-TM
MT-ND2
MT-TW
MT-TA
MT-TN
MT-TC
MT-TY
MT-CO1
MT-TS1
MT-TD
MT-CO2
MT-TK
MT-ATP8
MT-ATP6
MT-CO3
MT-TG
MT-ND3
MT-TR
MT-ND4L
MT-ND4
MT-TH
MT-TS2
MT-TL2
MT-ND5
MT-ND6
MT-TE
MT-CYB
MT-TT
MT-TP
Intergenic
Mito-genome

Variations in mitochondrial genome as potential prognostic markers in sickle cell disease

Rudra Ray[1], Haiou Li[1], Shijinqiu Gao[1], Nancy Asomaning[1], Kang Le[1], Maliha Ahmad[1], Xunde Wang[1], Yuesheng Li[2], Sayuri Kamimura[3], Clifton L. Dalgard[4,5], Chengyu Liu[6], Zenaide M.N. Quezado[1,3], Justin Lack[7], Laxminath Tumburu[1], Swee Lay Thein[1]

[1]Sickle Cell Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.
[2]DNA Sequencing & Genomics Core, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.
[3]Department of Perioperative Medicine, National Institutes of Health Clinical Center, Bethesda, MD.
[4]The American Genome Center, Uniformed Services University of the Health Sciences, Bethesda, MD.
[5]Department of Anatomy, Physiology & Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD.
[6]Transgenic Core, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.
[7]Bioinformatics (NCBR)/Integrated Data Sciences Section (IDSS)/ Research Technology Branch/DIR, NIAID, NIH, Bethesda, MD.

SUPPLEMENTARY MATERIALS.

Contents:

- **Supplementary Table S8:** Distribution of Indels and SNPs, Transition and Transversion within SNPs in human cohort 1.
- **Supplementary Table S9:** Sample collection and follow-up for the survival study in Cohort 1 (n=673).
- **Supplementary Table S10A:** Hazard ratio (95% confidence interval) for all-cause mortality by mtDNA heteroplasmy in Cohort 1.
- **Supplementary Table S10B:** Demographics of the patients with high hazard ratio mito-variants in Cohort 1.
- **Supplementary Table S10C:** Hazard ratio (95% confidence interval) for all-cause mortality by mtDNA-CN in Cohort 1 (n=673).
- **Supplementary Table S11:** Correlation between NIH phenotypic risk scores and mtDNA metrics among the SCD patients in Cohort 1.
- **Supplementary Table S12:** Description of the nuclear variants relevant to mitochondrial function and SCD.
- **Supplementary Table S13:** Distribution and prevalence of the nuclear variants among SCD patients in Cohort 1 (n=673).
- **Supplementary Table S14:** Spearman correlation of the nuclear variants with mtDNA metrics among the SCD patients in Cohort 1 (n=673) .
- **Supplementary Table S15:** Association of the nuclear variant status (wild type vs mutant) with mtDNA metrics among the SCD patients in Cohort 1 (n=673).
- **Supplementary Table S16:** Correlation between NIH phenotypic risk score and nuclear variants among the SCD patients in Cohort 1.
- **Supplementary Table S17:** Distribution mito-variants across tissues and genotypes in mouse cohort.
- **Supplementary Table S18:** Description and distribution of SS-unique variants.

Supplementary Figures

- **Supplementary Figure S1:** Description of human and mouse cohorts.
- **Supplementary Figure S2:** mtDNA enrichment by long range PCR (LR-PCR).
- **Supplementary Figure S3:** Estimation of mtDNA-CN with age in combined samples from 3 cohorts.
- **Supplementary Figure S4:** Details of obtained mito-coverage among different cohorts.

- **Supplementary Figure S5:** mtDNA heteroplasmy burden across genotypes and cohorts.
- **Supplementary Figure S6:** mtDNA heteroplasmy burden and mtDNA-CN levels are independent of Hydroxyurea (HU) treatment and cell counts.
- **Supplementary Figure S7:** mtDNA-CN and mtDNA heteroplasmy are not influenced by sex.
- **Supplementary Figure S8:** mtDNA copy number and mtDNA heteroplasmy burden are not influenced by hemoglobin levels.
- **Supplementary Figure S9:** Alpha globin genotype impacted mtDNA copy number but not mtDNA heteroplasmy burden in SCD.
- **Supplementary Figure S10:** mtDNA heteroplasmy burden in the functional regions of the mito-genome.
- **Supplementary Figure S11**: Correlation of NIH phenotypic risk score with mtDNA copy number and mtDNA heteroplasmy burden in SCD.
- **Supplementary Figure S12**: Heatmap showing correlation between the nuclear variants and mtDNA metrics.
- **Supplementary Figure S13:** mtDNA heteroplasmy burden in HbAA, HbAS and HbSS mice stratified by sex.
- **Supplementary Figure S14:** Pattern of mtDNA mutations burden in mito-genome across genotypes and cohorts.

Supplementary Methods

Description of sample distributions and demographics of human cohorts

A total of 1043 adult human subjects enrolled in 3 distinct cohorts were included in this study (*Online Supplementary Figure S1A*). The samples were collected over 23 years (February 2001-July 2024), with a mean follow-up time of 6.26 years and a maximum follow-up duration of 20.16 years. Cohort 1 (n= 673) includes patients with HbSS (SS, n=538), HbS-Beta-thalassemia^0 (SB0, n=16), HbSC (SC, n=91), HbS-Beta-thalassemia$^+$ (SB+, n=25), and other SCD genotypes (SCD Other, n=3: SD, n=2; Hb SO Arab, n=1) ; Cohort 2 (n=171) includes individuals with HbSS (n=111), HbSB0 (n=2), HbAS (AS, n=30), HbAA (AA, n=16), and other SCD genotypes (SCD Other n=12: HbSC, n=7; HbSB+, n=4; Hb SO Arab, n=1). Cohort 3 (n=199) includes individuals with HbSS (n=95), HbSB0 (n=1), HbAS (n=47), HbAA (n=44), and other SCD genotypes (SCD

Other n=12: HbSC, n=11, HbSB+, n=1). We combined the HbSS and HbSB0 samples in a single group as sickle cell anemia (SCA) owing to their similarity in phenotypic severity (*Online Supplementary Figure S1A*). The average ages of the subjects were 33.87, 32.81, and 37.38 years in Cohort 1, Cohort 2, and Cohort 3, respectively (*Online Supplementary Table S1*). Among the HbSS subjects, the distribution of age was similar, with an average of 32.87 years in Cohort 1, 28.90 years in Cohort 2, and 32.82 years in Cohort 3, respectively. Total study population (n=1043) comprised 45.83% male and 54.17% female subjects (*Online Supplementary Table S1*).

Description and distribution of mouse samples

Humanized Townes SS mice recapitulate many phenotypes of human SCD including a state of chronic inflammation and multi-organ damage.[1] We generated AA, AS, and SS age-matched male and female mice using in vitro fertilization. Donor AS females were superovulated and AS oocytes were incubated with AS sperm. Embryos were then transferred to pseudopregnant recipient mice. Eighteen littermate mice, 3 male and 3 female mice from 3 genotypes- AA, AS, and SS- were included in the study. Mice were euthanized at 16 weeks when full SCD phenotypes should have manifested and 9 tissue samples (heart, lung, brain, kidney, liver, muscle, blood, bone marrow [B.M.], spleen) were collected from each mouse. Thus, a total of 162 tissue samples equally distributed between the 3 genotypes, both sexes, and 9 tissue types were collected (*Online Supplementary Figure S1B*).

mtDNA enrichment

mtDNA was enriched from mouse genomic DNA samples by long range PCR (LR-PCR) following manufacturer's protocol using two sets of mtDNA specific primers to obtain two overlapping PCR products (8.6 kb and 7.7 kb) of the complete 16,295-bp mitochondrial genome. PCR enrichment conditions and primer sequences were adapted following Hirose, M. *et al.* (2018)[2] and PCR integrity was confirmed with agarose gel imaging (*Online Supplementary Figure S2*). Equimolar concentrations of the two enriched mtDNA fragments generated from the two primer sets for each sample, were combined to obtain the entire enriched mtDNA genome that was utilized for library preparation using Nextera XT DNA Library Preparation Kit (Illumina Inc., CA, USA) following manufacturer's guidelines. The quality control (QC) of the library and enrichment efficacy of mtDNA were assessed using the Illumina iSeq platform, prior to full depth sequencing using Illumina NovaSeq platform. mtDNA enrichment efficacy of each sample was evaluated by

calculating the ratio between the total reads mapped to entire GRCm39 (mm39) reference genome and the number of reads mapped to chrM only.

NGS, variant calling, and mtDNA heteroplasmy detection

Next generation sequencing (NGS) was performed for all the human and mice samples. For human samples, WGS was performed in different batches for the 3 cohorts at different facilities. The average sequencing depth (genomic coverages) were- 33.2x for Cohort 1, 43.1x for Cohort 2, and 71.4x for cohort 3. Enriched mtDNA samples were subjected to NGS for the mouse samples.

The raw sequencing files were demultiplexed, trimmed, and converted to FASTQ format. The FASTQ reads were evaluated for QC and analyzed using MultiQC (version 1.14) (https://multiqc.info/). The reads were mapped to the GRCh37/GRCh38 human and GRCm39 (mm39) house mouse (*Mus musculus*) reference genomes, for human and mouse samples, respectively. Mapped reads were further filtered using mitochondrial genome (chrM) reference sequence from their respective reference genomes. The estimation of coverage depth against mito-genome (chrM) was conducted using Mosdepth (https://github.com/brentp/mosdepth). Variant calling was done using MitoHPC workflow (https://github.com/dpuiu/MitoHPC) with LoFreq variant caller tool (version 2.4) using the resulting BAM files. Annotation was then carried out using VEP/106. MitoHPC workflow used for variant calling (https://github.com/dpuiu/MitoHPC) has inbuilt function to remove NuMTs. Additionally, for the human samples, nuclear sequences of mitochondrial origin (NuMTs), homopolymers, and repeat sequences were excluded from the analysis as described in previous reports.[3, 4] In the mouse samples, the use of two overlapping sets of mito-genome specific PCR primers avoids amplification of NuMTs[5] and PCR specificity was confirmed by agarose gel analyses (*Online Supplementary Figure S2*).

For both mtDNA sequences extracted from blood-derived WGS in humans and PCR-enriched mtDNA sequences from mouse tissue samples, analysis included only variants with coverage depth of 1000X or above. Variants with allele frequency (VAF) above 90% were excluded in order to remove potential homoplasmies. The lower VAF cutoffs for the respective cohorts were estimated using a software (http://app.olgen.cz/clc/) which takes into consideration sequencing error rate, probability of false positive result, and probability of true positive result to calculate the VAF (%) for a particular genomic coverage.

Sequencing reads obtained from both human and mouse samples were subjected to the same pipeline, sequencing reads for the mouse samples were not subjected to additional NuMts removal as they were obtained from enriched mtDNA.

Human nuclear genotype variants were extracted from WGS generated VCF files using VariantAnnotation package in R. Alpha globin genotypes were determined using a combination of machine-learning derived algorithm[6] and droplet digital PCR (ddPCR) (Bio-Rad QX200 ddPCR system).[7]

Detection of large mtDNA deletions from WGS data

Large mtDNA deletions (≥100 bp) were identified from whole-genome sequencing (WGS) BAM files using a R-based pipeline (detailed code attached in a separate *Online Supplementary File-2*). The pipeline was constructed and iteratively debugged with assistance from ChatGPT (OpenAI), and all code was reviewed and validated by the authors. Additionally, the identified mtDNA deletions were visually inspected and confirmed by IGV Integrative Genomics Viewer (IGV).

Briefly, BAM reads were aligned and mapped to mitochondrial chromosome (chrM) using Rsamtools and GenomicAlignments. To prevent false positives arising from NuMts, reads overlapping from known NuMT regions were removed. Deletions were identified from the remaining reads. CIGAR strings were parsed to identify deletion ("D") operations ≥100 bp in length. The identified deletions were summarized by genomic coordinates, length, read count, and mean mapping quality (MapQ). Only the deletions supported by at least 10 reads and with a mean MapQ ≥ 45 were included for subsequent analysis.

Survival analysis

Survival analysis was conducted in Cohort 1, which included SCD patients exclusively. Survival data was ascertained between the consent date and the date of the last follow-up, proxy interview, medical records, National Death Index, and Social Security Death Index search. The Cox proportional hazards model was applied to analyze the relationship between mtDNA heteroplasmy and all-cause mortality. Hazard ratio (HR) with 95% confidence interval (CI) was computed. False discovery rate (FDR) was applied for multiplicity adjustment when studying the relationship between specific heteroplasmy variant with mortality.

NIH risk score prediction

NIH risk score was calculated using a previously published phenotypic risk model for predicting disease severity and mortality in sickle cell disease.[8] Nine significant clinical and laboratory variables were selected using a regularized Cox proportional hazards model: age, body mass index (BMI), heart rate, alkaline phosphatase, blood urea nitrogen, septal thickness, mitral E velocity, tricuspid regurgitation velocity, and right atrial (RA) pressure. Each variable contributed one point when the corresponding risk threshold was met, and the total NIH risk score was obtained by summing all component points, with higher scores indicating greater disease severity and mortality risk. Body mass index was calculated as weight in kilograms divided by height in meters squared (kg/m^2). The same variable definitions and scoring criteria described in the original publication were applied in the current study.

Supplementary Table S1: Details of the human sample cohorts.

Cohorts	Genotypes		Number of subjects (n)			Average Age (years)
			Male (n)	Female (n)	Total (n)	
Cohort 1	SB+		6	19	25	39.04
	SC		42	49	91	38.19
	SCA	SS	261	277	538	32.87
		SB0	10	6	16	34.56
	Other	SD	2	0	2	36
		S O-Arab	0	1	1	36
	Total		321 (47.70%)	352 (52.30%)	673 (100%)	33.87
Cohort 2	AA		5	11	16	37.69
	AS		9	21	30	42.83
	SCA	SS	62	49	111	28.9
		SB0	2	0	2	37.5
	Other	S O-Arab	1	0	1	23
		SB+	1	3	4	39
		SC	2	5	7	37.29
	Total		82 (47.95%)	89 (52.05%)	171 (100%)	32.81
Cohort 3	AA		12	32	44	39.84
	AS		16	31	47	44.06
	SCA	SS	41	53	94	32.82
		SB0	1	1	2	30.5
	Other	SB+	1	0	1	33
		SC	4	7	11	39.64
	Total		75 (37.69%)	124 (62.31%)	199 (100%)	37.38
All Cohorts	Total		478 (45.83%)	565 (54.17%)	1043 (100%)	34.36

Supplementary Table S2: Details of mito-coverage and VAF cutoff for each cohort.

	Number of samples	Mean mtDNA coverage (X)	Lower VAF cutoff	Upper VAF cutoff
Cohort-1	673	7663	1%	90%
Cohort-2	171	1985	5%	90%
Cohort-3	199	17830	1%	90%
Mouse Cohort	162	4948	4%	90%

Footnote: VAF, variant allele frequency.

Supplementary Table S3: Distribution and stratification of mtDNA heteroplasmies in three human cohorts.

Cohort	Genotypes	Total Heteroplasmic Occurrences (n)	Non- Coding (n)	Coding	
				Synonymous (n)	Non- Synonymous (NS) (n)
Cohort 1	SB+ (n=25)	5	4	1	0
	SC (n=91)	33	21	6	6
	SCA (n=554)	227	136	25	66
	Other (n=3)	3	2	0	1
	Total (n=673)	268	163	32	73
Cohort 2	AA (n=16)	4	2	2	0
	AS (n=30)	8	4	3	1
	SCA (n=113)	42	22	4	16
	Other (n=12)	6	5	1	0
	Total (n=171)	60	33	10	17
Cohort 3	AA (n=44)	15	12	2	1
	AS (n=47)	18	13	3	2
	SCA (n=96)	44	21	13	10
	Other (n=12)	7	1	2	4
	Total (n=199)	84	47	20	17

Supplementary Table 4: Distribution of mDNA heteroplasmy burden and mtDNA-CN stratified by hydroxyurea (HU) treatment in Cohort 1.

Genotypes	Mean mtDNA Heteroplasmy		Mean mtDNA-CN	
	HU	Non-HU	HU	Non-HU
SCA (n=554)	0.40 (n=244)	0.40 (n=310)	514.20 (n=244)	460.20 (n=310)
SC (n=91)	0.38 (n=11)	0.27 (n=80)	403.54 (n=11)	375.84 (n=80)
SB+ (n=25)	0.25 (n=9)	0.11 (n=16)	384.86 (n=9)	251.26 (n=16)
SD (n=2)	1.00 (n=1)	2.00 (n=1)	334.23 (n=1)	194.36 (n=1)
SO Arab (n=1)	0.00 (n=1)	n.a (n=0)	240.58 (n=1)	n.a (n=0)
Total (n=673)	**0.40** (n=266)	**0.40** (n=407)	**503.55** (n=266)	**434.68** (n=407)

Supplementary Table S5: Multivariate regression co-efficient for platelet and reticulocyte counts and genotype in Cohort 1 (n=650).

	Regression coefficients (β)	p value
Platelet Count	0.126048185	0.235237815
Reticulocyte Count	0.075979532	0.523619889
Genotype	97.78539695	0.038777218

Footnote: 23 samples with missing platelet and reticulocyte counts were excluded; R^2 =0.016; Adjusted R^2 = 0.011.

Supplementary Table S6: Correlation between alpha (α) globin gene number and mtDNA metrics among SCD patients in Cohort 1.

Outcome Comparison	Sample (n)	Spearman Correlation (r)	95% CI	*P* value	Statistical Significance
α globin gene number vs. mtDNA-Copy Number	632	0.089	0.009–0.168	0.0256	Significant
α globin gene number vs. Total heteroplasmy	632	0.02	−0.061–0.100	0.619	NS
α globin gene number vs. NS heteroplasmy	632	0.06	−0.020–0.140	0.131	NS

Footnote: NS heteroplasmy- Non-synonymous heteroplasmy; n= 632 Samples included in analysis; n= 41 individuals with missing alpha globin genotyping data were excluded.

Supplementary Table S7: Mutational burden across functional regions of the mito-genome among different genotypes in human Cohort 1.

Type of Mutation	Mito-regions and functional classification	Affected mito-genes	Heteroplasmic occurrences found per genotypes (n)				Total Heteroplasmic occurrences (n) among all samples
			Other	SB+	SC	SCA (SS+SB0)	
Non-Coding (n=163)	Control Region	D Loop		2	19	78	99 (36.94%)
	Intergenic	Intergenic	1	2	1	30	34 (12.69%)
	tRNA	MT- TF				1	14 (5.22%)
		MT-TG				1	
		MT-TL2				2	
		MT-TP				2	
		MT-TS1				1	
		MT-TS2				2	
		MT-TT				4	
		MT-TV				1	
	rRNA	RNR1				5	16 (5.97%)
		RNR2	1		1	9	
Coding (n=105)	Complex I	ND1			1	2	63 (23.51%)
		ND2			2	4	
		ND3				2	
		ND4	1		4	34	
		ND5			1	9	
		ND6			1	2	
	Complex III	CYT-B		1	2	7	10 (3.73%)
	Complex IV	COX1			1	9	24 (8.96%)
		COX2				8	
		COX3				6	
	Complex V	ATP6				7	8 (2.99%)
		ATP8				1	
Total			3	5	33	227	268 (100%)

Footnote: Intergenic - regions are unspecified stretch of sequences between two genes

Supplementary Table S8: Distribution of Indels and SNPs, Transition and Transversion within SNPs in human Cohort 1.

Total Heteroplasmic Occurrences (n= 268)		
Indels (n=165)	SNPs (n=103)	
	Transition (n= 98)	Transversion (n=5)
	G/A (n=32)	A/T (n=1)
	A/G (n=27)	T/A (n=1)
	T/C (n=26)	C/A (n=1)
	C/T (n=13)	G/T (n=1)

Footnote: SNPs - Single nucleotide polymorphisms.

Supplementary Table S9: Sample collection and follow-up for the survival study in Cohort 1 (n=673).

Description	**Value**
Sample Collection Period	February 2001 – July 2024
Mean Follow-up time	6.26 years
Median Follow-up time	4.91 years
Longest Follow-up time	20.16 years

Supplementary Table S10A: Hazard ratio (95% confidence interval) for all-cause mortality by mtDNA heteroplasmy in Cohort 1.

	Univariable HR (95% CI) (n=673)	Age Control HR (95% CI) (n=673)
Heteroplasmy: continuous		
Heteroplasmy	1.01 (0.853 – 1.197)	1.018 (0.855 – 1.212)
Heteroplasmy: categorical		
1 Heteroplasmy Vs 0 Heteroplasmy	1.173 (0.873 – 1.576)	1.114 (0.828 – 1.499)
2 Heteroplasmy Vs 0 Heteroplasmy	1.3 (0.806 – 2.096)	1.28 (0.794 – 2.064)

Supplementary Table S10B: Demographics of the patients with high hazard ratio mito-variants in Cohort 1.

Variants with higher HR	Description of the subjects harboring the variant		
	Genotype	Age	Sex
MT: 8483	SS	31	Female
MT:16015	SS	31	Male
MT: 8156	SS	44	Female
MT: 4916	SC	51	Male
MT: 4658	SS	38	Female
MT: 4695	SS	52	Female
MT: 6158	SS	19	Female

Supplementary Table S10C: Hazard ratio (95% confidence interval) for all-cause mortality by mtDNA-CN in Cohort 1 (n=673).

	Univariable (HR 95% CI)	Age Control (HR 95% CI)
mtDNA Copy number (mtDNA-CN)		
mtDNA-CN	0.964 (0.809 – 1.15)	0.997 (0.833 – 1.193)

Supplementary Table S11: Correlation between NIH phenotypic risk scores and mtDNA metrics among the SCD patients in Cohort 1.

Outcome Comparison	Sample (n)	Spearman Correlation (r)	*P* value	Statistical Significance
NIH Risk Score vs. mtDNA-Copy Number	452	−0.069	0.144	NS
NIH Risk Score vs. Total heteroplasmy	452	−0.012	0.802	NS
NIH Risk Score vs. NS heteroplasmy	452	0.003	0.951	NS

Footnote: NS heteroplasmy- Non-Synonymous Heteroplasmy. Among the 673 SCD patients in Cohort 1, component variables required to calculate the NIH risk score were available in 452.

Supplementary Table S12: Description of the nuclear variants relevant to mitochondrial function and SCD.

Nuclear Variants	Variant ID	Gene Name	Location (GRCh37)	Substitution	Functional association	References
SOD2 V16A	rs4880	*SOD2*	chr6:160113872	A>C, G, T	Oxidative stress, mitochondrial activity, SCD	9, 10
PIEZO1 E756del	rs572934641	*Piezo1*	chr16:88800372-383	TCC Indels	Red cell dehydration, SCD, Malaria	11, 12
CYB5R3 T117S	rs1800457	*CYB5R3*	chr22:43024271	G>A, C	Methemoglobin Reduction, Oxidative Stress, Anemia	13
G6PD A376G	rs1050829	*G6PD*	X:153763492	T>A, C	Hemolytic Anemia	14
G6PD G202A	rs1050828	*G6PD*	X:153764217	C>T	Hemolytic Anemia	14

Supplementary Table S13: Distribution and prevalence of the nuclear variants among SCD patients in Cohort 1 (n=673).

Nuclear Variants	Variant ID	Total Samples n (%)	WT n (%)	Mutation n (%)	Heterozygous n (%)	Homozygous n (%)	MAF
SOD2 V16A	rs4880	673 (100%)	230 (34.18)	443 (65.82)	317 (47.10)	126 (18.72)	0.42
PIEZO1 E756del	rs572934641	673 (100%)	466 (69.24)	207 (30.76)	185 (27.49)	22 (3.27)	0.17
CYB5R3 T117S	rs1800457	673 (100%)	356 (52.90)	317 (47.10)	263 (39.08)	54 (8.02)	0.28
G6PD A376G	rs1050829	673 (100%)	387 (57.50)	286 (42.50)	149 (22.14)	137 (20.36)	0.31
G6PD G202A	rs1050828	673 (100%)	556 (82.62)	117 (17.38)	80 (11.89)	37 (5.50)	0.11

Supplementary Table S14: Spearman correlation of the nuclear variants with mtDNA metrics among the SCD patients in Cohort 1 (n=673).

Nuclear Variants	mtDNA-CN		Total Heteroplasmy		NS Heteroplasmy	
	Spearman r	*P* Value	Spearman r	*P* Value	Spearman r	*P* Value
SOD2 V16A	−0.040	0.295	0.011	0.766	−0.009	0.812
PIEZO1 E756del	−0.047	0.222	0.077	***0.045***	0.045	0.245
CYB5R3 T117S	−0.028	0.466	0.117	***0.0023***	0.099	***0.0099***
G6PD A376G	0.018	0.64	0.001	0.981	−0.009	0.816
G6PD G202A	−0.057	0.138	0.051	0.184	0.066	0.089

Footnote: In Cohort 1 (n=673), the status of nuclear gene variants *SOD2* V16A, *CYB5R3* T117S, *G6PD* A376G, *G6PD* G202A, and *PIEZO1* E756 del were evaluated and encoded as 0 (wild type), 1(heterozygous), 2 (homozygous) for Spearman correlation association analysis.

Supplementary Table S15: Association of the nuclear variant status (wild type vs mutant) with mtDNA metrics among the SCD patients in Cohort 1 (n=673).

Nuclear Variants	Group	mtDNA-CN (Mean)	Total Heteroplasmy (Mean)	NS Heteroplasmy (Mean)
SOD2 V16A	WT (n=443)	441	0.4	0.1
	Mutant (n=230)	502	0.4	0.12
***PIEZO1* E756del**	WT (n=466)	469	***0.36***	0.09
	Mutant (n=207)	446	***0.49***	0.14
***CYB5R3* T117S**	WT (n=356)	475	***0.33***	***0.07***
	Mutant (n=317)	447	***0.48***	***0.15***
G6PD-combined	WT (n=384)	450	0.41	0.11
	Mutant (n=289)	478	0.38	0.11

Footnote: Patients were stratified into 2 groups - WT (wild type) and Mutant (includes both heterozygous and homozygous) according to the status of nuclear gene variants- *SOD2* V16A, *CYB5R3* T117S, *G6PD* A376G, *G6PD* G202A, and *PIEZO1* E756del. As *G6PD* A376G and *G6PD* G202A variants are biologically associated, they were combined together as '*G6PD*-combined' for this analysis. Mean mtDNA-CN, total mtDNA heteroplasmy, and NS heteroplasmy were compared between WT and Mutant groups.

Supplementary Table S16: Correlation between NIH phenotypic risk score and nuclear variants among the SCD patients in Cohort 1.

Outcome Comparison	Spearman Correlation (r)	95% CI	*P* value	Statistical Significance
NIH Risk Score vs. *SOD2* V16A	-0.0175	-0.1123 to 0.0775	0.71	No
NIH Risk Score vs. *PIEZO1* E756del	-0.0002	-0.0952 to 0.0947	0.9959	No
NIH Risk Score vs. *CYB5R3* T117S	-0.0265	-0.1211 to 0.0686	0.5746	No
NIH Risk Score vs. *G6PD* A376G	-0.0783	-0.1719 to 0.0168	0.0965	No
NIH Risk Score vs. *G6PD* G202A	-0.012	-0.1068 to 0.0831	0.7997	No

Footnote: Among the 673 SCD patients in Cohort 1, 452 with no missing component variables required to calculate the NIH risk score, were included in the risk score analysis.

Supplementary Table S17: Distribution mito-variants across tissues and genotypes in mouse cohort.

	MT: 9820			MT: 11886			MT: 6881			MT: 6461			MT: 6573			MT: 6574			MT: 3340			MT: 6753			MT: 6974			MT: 1220			MT: 425		
	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS	AA	AS	SS
Heart	2	2	2		1			2																			1			1			
Brain	2	2	2		1			2																							1			1
Kidney	2	1	3		2			3																			1						
Muscle	2	2	1		1			2																			1			1			1
Liver	2	2	2					1																			1			1			1
Blood	2	2	2		1			2																						1			1
B.M	2	2	2				2																							1			1
Lungs	1	2	2	1	1			2																			1			1			1
Spleen	4	1	2	1			1			1		1	1			1				1				1			1			1			1
Total Per Genotype	19	16	18	2	7	0	3	14	0	1	0	1	1	0	0	1	0	0	0	1	0	1	0	0	0	0	6	0	0	8	0	0	7
Total Per Variants	53			9			17			2			1			1			1			1			6			8			7		
Total Occurrences	106																																

Footnote: B.M - bone marrow

Supplementary Table S18: Description and distribution of SS-unique variants.

	SS -Unique Variants	Affected Mito-genome regions	Amino acid change	Total occurrences	Tissues involved
1.	MT: 425	D-Loop	N/A	n=7	Brain, Muscle, Liver, Blood, B.M, Lungs, Spleen
2.	MT: 1220	Upstream to ND1 / Intergenic	N/A	n=8	Heart, Brain, Muscle, Liver, Blood, BM, Lungs, Spleen
3.	MT: 6753	Cox1	p.Phe476Leu	n=1	Spleen
4.	MT: 6974	Downstream to Cox2/ Intergenic	N/A	n=6	Heart, Muscle, Spleen, Kidney, Liver, Lungs

Supplementary Figure Legends:

Supplementary Figure S1: Description of human and mouse cohorts. (**A**)The distribution of samples with different genotypes among 3 human cohorts: Cohort 1, Cohort 2, and Cohort 3 is described. HbSS and HbSB0 are grouped together as SCA (sickle cell anemia) because of similarity in disease severity. Cohort 1 includes patients with SCA (HbSS & SB0), HbSC, HbSB+, and other SCD genotypes (HbSD, n=2 and HbSO Arab, n=1). Cohort 2 includes individuals with SCA (HbSS & HbSB0), HbAS, HbAA, and other SCD genotypes (HbSC, n=7; HbSB+, n=4; HbSO Arab, n=1). Cohort 3 includes individuals with SCA (SS & SB0), AS, AA, and other SCD genotypes (HbSC, n=11, HbSB+, n=1). The sample size for each cohort, and for each genotype within each cohort, is indicated in the figure.

(**B**) Distribution of 162 mice samples across tissues, genotypes, and sexes. Humanized Towne's SCD mice comprised 18 total littermates from 3 genotypes- HbAA, HbAS, HbSS. Nine tissue samples- heart, lung, brain, kidney, liver, muscle, blood, bone marrow (B.M), and spleen from each of the 18 age-matched mice were collected at week 16.

Supplementary Figure S2: mtDNA enrichment by long range PCR (LR-PCR). Mitochondrial DNA was enriched from genomic DNA using the LR-PCR method with two sets of overlapping mtDNA-specific primers generating two amplicons, fragment-1 (8.6kb) and fragment-2 (7.7kb), covering the entire 16.3KB mitochondrial genome. This representative figure includes six samples derived from six different mouse tissues, lanes 1 and 7- heart, 2 and 8- brain, lanes 3 and 9- kidney, lanes 4 and 10 - muscle, lanes 5 and 11 - liver, and lanes 6 and 12 – peripheral blood, respectively.

Supplementary Figure S3: Estimation of mtDNA-CN with age in combined samples from 3 cohorts. Total subjects (n=1043) from 3 human cohorts were combined and divided into 4 age groups: < 25 years (n = 279), 25–34 years (n = 339), 35–44 years (n = 196), and ≥45 years (n = 229). mtDNA-CN level decreased with age. Age group ≥45 years had the lowest mtDNA-CN level (mean 399.5) compared to 35–44 years (mean = 461.2), 25–34 years (mean = 546.3, *P=0.0101*, *), and < 25 years (mean = 540.9, *P=0.0129,* *). Statistical significance (*P* value) was assessed by Kruskal-Wallis test (between three or more groups) or Mann Whitney test (between two groups). If no other indication, results were not significant.

Supplementary Figure S4: Details of obtained mito-coverage among different cohorts. Mean mtDNA coverage was 7663X for Cohort 1 (**A**), 1985X for Cohort 2 (**B**), 17830X for Cohort 3 (**C**), and 4948X for the mouse cohort (**D**). One sample with dropped coverage (indicated by arrow) was excluded from heteroplasmy analysis in the mouse cohort. (**E**) Similar average mito-coverage was obtained across 9 different tissue samples from mice. The red line represents the mean value. The number of sample (n) for each cohort is indicated on the X-axis.

Supplementary Figure S5: mtDNA heteroplasmy burden across genotypes and cohorts. Evaluation of total mtDNA heteroplasmies (A, B and C) and non-synonymous (NS) mtDNA heteroplasmies (D, E and F) across genotypes in each cohort. (**A**) Total mtDNA heteroplasmy burden between genotypes in Cohort 1. SCA (n=554) had highest mtDNA heteroplasmy burden (mean =0.41) followed by HbSC (n=91) (mean =0.36), and HbSB+ (n=25) (mean =0.19); (*P* = ns). SD (n=2) and SO-Arab (n=1) genotypes were excluded from the comparison in Cohort 1 due to small sample sizes. **(B)** Total mtDNA heteroplasmy burden in AA, AS and SCA in Cohort 2. Mean heteroplasmies in SCA (n=113), HbAS (n=30), and HbAA (n=16), were 0.37, 0.27, and 0.25, respectively (*P*=ns). SO-Arab (n=1), SB+ (n=4), and SC (n= 7) genotypes were excluded from the comparison due to small sample sizes. **(C)** Total mtDNA heteroplasmy burden in AA, AS and SCA in Cohort 3. Mean heteroplasmies in SCA (n=96), HbAS (n=47), and HbAA (n=44), were 0.46, 0.38, and 0.35, respectively (*P*=ns). SB+ (n=1), and SC (n= 11) genotypes were excluded from the comparison due to small sample sizes. (**D**). Burden of NS heteroplasmy in Cohort 1. Mean heteroplasmy burden in SCA (n=554), HbSC (n=91), and HbSB+ (n=25), were 0.11, 0.06, and 0, respectively; (*P* = ns). SD (n=2) and SO-Arab (n=1) genotypes were excluded from the comparison due to small sample sizes. (**E**). Burden of NS heteroplasmy in Cohort 2. Mean heteroplasmy burden in SCA (n=113), HbAS (n=30), and HbAA (n=16), were 0.14, 0.03, and 0, respectively; (*P* = ns). SO-Arab (n=1), SB+ (n=4), and SC (n= 7) genotypes were excluded from the comparison due to small sample sizes. **(F)** Burden of NS heteroplasmy in Cohort 3. Mean heteroplasmy burden in SCA (n=96), HbAS (n=47), and HbAA (n=44), were 0.10, 0.04, and 0.02, respectively; (*P* = ns). SB+ (n=1), and SC (n= 11) genotypes were excluded from the comparison due to small sample sizes. (**G**) Burden of NS heteroplasmy between genotypes in total subjects (n=1043) from 3 human cohorts. Mean NS heteroplasmy burden increased with sickle burden: AA (n=60, mean=0.01), AS (n=77, mean = 0.03), SB+ (n=30, mean=0.03), SC (n=109,

mean= 0.08), SCA (n=763, mean= 0.12); (*P* = ns). Four samples (n=4) were excluded from the comparison due to small sample sizes: SD (n=2) and SO-Arab (n=1) in Cohort 1; and SO-Arab (n=1) in Cohort 2. Statistical significance (*P* value) was assessed by Kruskal-Wallis test (between three or more groups) or Mann Whitney test (between two groups). If no other indication, results were not significant (*P≥0.05*).

Supplementary Figure S6: mtDNA heteroplasmy burden and mtDNA-CN levels are independent of Hydroxyurea (HU) treatment and cell counts. (**A**) SCD patients in Cohort 1 (n=673) were divided in two groups based on their HU treatment status: HU-treated (n= 266), and non-HU-treated (n=407). Mean mtDNA heteroplasmy was same (0.40) for both HU-treated and non-HU-treated groups. (**B**) HU-treated group mtDNA-CN had a higher mean mtDNA-CN of 503.55 compared to non-HU-treated group with mean mtDNA-CN 434.68 (*P* = ns). (**C**) mtDNA-CN level increased with genotypic severity within both HU-treated, and non-HU treated groups. Within the HU-treated group, SCA individuals had the highest mean mtDNA-CN compared to SC and SB+, with values of 514.20, 403.54 and 384.86, respectively but the difference sdid not achieve significance. Among the non-HU-treated group, mean mtDNA-CN was significantly higher in SCA (n=310, mean mtDNA-CN 460.20) compared to SC (n=80, mean mtDNA-CN 375.84, *P =0.004*, **), and SB+ (n= 16, mean mtDNA-CN = 251.26, *P =0.0002*, ***). (**D**) Influence of platelet count and reticulocyte on mtDNA-CN levels. In buffy coat, platelets and reticulocytes are non-nucleated cells that contain mitochondria and, therefore, may have cofounding effect of mtDNA -CN estimation. Regression analysis was performed to evaluate the influence of platelet count, absolute reticulocyte count, and genotypes (SCA & non-SCA) on mtDNA-CN levels. Subjects with cell count data missing (n=23) in cohort 1 (n=673) were excluded from the analysis. Regression analysis showed that platelet count (regression coefficients (β)= 0.12, *P=ns*) and reticulocyte count (regression coefficients: β) = 0.07, *P=ns*) had negligible effect on mtDNA-CN variation, whereas genotypes contributed to mtDNA-CN variation to a much greater extent (regression coefficients: β) = 97.78, *P=0.038,* **).

Supplementary Figure S7: mtDNA-CN and mtDNA heteroplasmy are not influenced by sex. Total subjects from 3 human cohorts were stratified by sex in total SCD (n=1043, 476 male and 565 female; Figures A, C and E) and SCA (n=763, 377 male and 386 female; Figures B, D and F)

patients. (**A**) SCD: Mean mtDNA-CN was higher in males than females at 544 vs 456 (*P= ns*); (**B**) SCA: males had higher mean mtDNA-CN than females at 615 vs. 527 (*P= ns*) . (**C**) SCD: Mean total heteroplasmy was higher in males than females at 0.43 vs. 0.37 (*P= ns*) among all subjects (n=1043); and (**D**) SCA: both males and females had mean total heteroplasmy at 0.41. (**E**) SCD: Mean non-synonymous (NS) heteroplasmies were similar in males and females at 0.10. (**F**) SCA: Females had higher mean NS heteroplasmy than males at 0.12 vs 0.11 (*P= ns*). None of the differences were statistically significant. Statistical significance (*P* value) was assessed by Mann Whitney test. If no other indication, results were not significant (*P≥0.05*).

Supplementary Figure S8: mtDNA copy number and mtDNA heteroplasmy burden are not influenced by hemoglobin levels. Spearman correlation analysis was done to evaluate the association between mtDNA-CN and mtDNA heteroplasmy (total and non-synonymous) and hemoglobin (Hb, g/dL) levels in Cohort 1. The analysis was restricted to the SCA to ensure a clinically homogeneous phenotype. Among the 554 SCA patients, 4 individuals lacked hemoglobin measurements and were excluded, resulting in a final analytic sample of 550 patients. (**A**) Correlation between Hb levels and mtDNA-CN: Spearman r = −0.047, *P = 0.274*, n=550. (**B**) Correlation between Hb levels and total mtDNA heteroplasmy: Spearman r = −0.007, *P=0.874*, n=550. (**C**) Correlation between Hb levels and total NS heteroplasmy: Spearman r = −0.012, *P=0.779*, n=550.

Supplementary Figure S9: Alpha globin genotype impacted mtDNA copy number but not mtDNA heteroplasmy burden in SCD. Of the 673 patients in Cohort 1, 632 were included in the analysis, and 43 were excluded due to missing α globin genotyping data. (**A**) Correlation between alpha thalassemia (α globin gene count) and mtDNA-CN. Spearman correlation analysis showed a significant positive correlation between α globin gene count and mtDNA-CN (Spearman r=0.089, *P=0.02*, n=632). (**B-E**) Subjects in Cohort 1 were divided in four groups based on their alpha (α) globin genotype status: Homozygotes with α 3.7 deletions (α globin gene count=2), heterozygotes with α 3.7 deletion (α globin gene count=3), wild type (α globin gene count=4), and individuals with α 3.7 triplication or more (α globin gene count ≥ 5). (**B**) mtDNA-CN increased with number of α globin genes: $-\alpha/-\alpha$ (n=33; mean mtDNA-CN= 334.0), $\alpha\alpha/-\alpha$ (n=204; mean mtDNA-CN= 440.8), $\alpha\alpha/\alpha\alpha$ (n=375; mean mtDNA-CN= 446.2), and $\alpha\alpha/\alpha\alpha\alpha$ or more (n=20;

mean mtDNA-CN= 888.6). SCD patients with –α/–α had significantly lower mtDNA-CN compared to αα/αα (*P=0.029*). (**C**) Among the SCA patients in Cohort 1(n= 517 ; excluded= 37), a similar pattern was observed: –α/–α (n=25; mean mtDNA-CN= 313.2), –α/αα (n=171; mean mtDNA-CN= 461.5), αα/αα (n=304; mean mtDNA-CN= 463.7), and αα/ααα or more (n=17; mean mtDNA-CN= 969.8). SCA patients with –α/–α had significantly lower mtDNA-CN than αα/αα (*P=0.038*). (**D**) No statistically significant differences were observed in total mtDNA heteroplasmy burden with *α* globin genotype status in SCD patients (n=632): –α/–α (n=33; mean total mtDNA heteroplasmy=0.42), –α/αα (n=204; mean=0.35), wild type (n=375; mean=0.42), and αα/ααα or more (n=20; mean=0.40); (*P= ns*). (**E**) No statistically significant differences were observed in non-synonymous (NS) heteroplasmy burden with *α* globin genotype status in SCA patients (n=517): –α/–α (n=25; mean NS heteroplasmy=0.09), –α/αα (n=171; mean=0.07), αα/αα (n=304; mean=0.12), and αα/ααα or more (n=17; mean=0.15); (*P=ns*). Statistical significance (*P* value) was assessed by Kruskal-Wallis test (between three or more groups) or Mann Whitney test (between two groups).

Supplementary Figure S10: mtDNA heteroplasmy burden in the functional regions of the mito-genome. The prevalence of mtDNA heteroplasmy burden in SCA genotype (n=554) was compared with rest of the genotypes (n=119) (Rest: SC, n=91; SB+, n=25 ; SCD Other, n=3: SD, n=2 ; SO Arab, n=1) across different functional mitogenome regions: Control Region (D loop), intergenic regions, tRNA, rRNA, and OXPHOS genes (Complex I, Complex III, Complex IV, Complex V). SCA group displayed higher mtDNA heteroplasmy burden in Complex V (3.52%), Complex IV (10.13%), and tRNA (6.17%) compared to Rest.

Supplementary Figure S11: **Correlation of NIH phenotypic risk score with mtDNA copy number and mtDNA heteroplasmy burden in SCD.** Among the 673 SCD patients in Cohort 1, complete data for all component variables required for NIH risk score calculation were available for 452 patients, who constituted the final analytic sample NIH risk score analyses. Spearman correlation analysis (A,C,E) was conducted between NIH risk score and mtDNA-CN, total mtDNA heteroplasmy, and NS heteroplasmy. The number of subjects in each risk score category is

indicated below the x-axis. An additional exploratory analysis was conducted by dichotomizing the NIH risk score in two groups: Low risk score (NIH risk score ≤ 3) and High risk score (NIH risk score >3). Mean mtDNA-CN, mtDNA heteroplasmy, NS heteroplasmy were compared between the two risk score groups (B,D,F). (**A**) Spearman correlation analysis showed no correlation between NIH risk score and mtDNA-CN: Spearman r=-0.069, *P=0.14*. Although the association did not reach statistical significance, the correlation curve suggested a mild downward trend at high risk score. (**B**) Mean mtDNA-CN was significantly lower in the high risk score (n=71, mean=348.40) than low-risk score group (n=381, mean=459.48); Welch's t-test *P*<0.001. (**C**) Spearman correlation demonstrated no significant association between NIH risk score and total mtDNA heteroplasmy: Spearman r = −0.012, *P = 0.802*. (**D**) Mean total mtDNA heteroplasmy did not differ significantly between low-risk score (n=381, mean=0.38) and high risk score (n=71, mean=0.35); (*P=ns*). (**E**) Spearman correlation demonstrated no significant association between NIH risk score and total NS heteroplasmy: Spearman r = 0.003*, P=0.951*. (**F**) Mean NS heteroplasmy did not differ significantly between low-risk score (n=381, mean=0.10) and high risk score (n=71, mean=0.11); (*P=ns*). Welch's t-test was performed to compare the means between two risk score groups and assess *P* values.

Supplementary Figure S12: **Heatmap showing correlation between the nuclear variants and mtDNA metrics.** Spearman correlation analysis between the nuclear mutations and mtDNA-CN, total mtDNA heteroplasmy, and NS heteroplasmy. Among SCD patients in Cohort 1 (n=673), status of the nuclear gene variants- *SOD2* V16A, *CYB5R3* T117S, *G6PD* A376G, *G6PD* G202A, and *PIEZO1* E756del were evaluated and encoded as 0 (wild type), 1 (heterozygous), 2 (homozygous) for association analysis with mtDNA-CN, total mtDNA heteroplasmy, and NS heteroplasmy. The color intensity represents Spearman correlation coefficients (r). Asterisks (*), highlighted in red, denote statistical significance (*P*<0.05). The exact Spearman correlation coefficients (r) and *P* values are provided in *Online Supplementary Table S14*.

Supplementary Figure S13: mtDNA heteroplasmy burden in HbAA, HbAS and HbSS mice stratified by sex.

A total of 18 mice were included in this study which were equally distributed between sexes and genotypes, with 3 mice per sex per genotype. Mean mtDNA heteroplasmy was higher in females

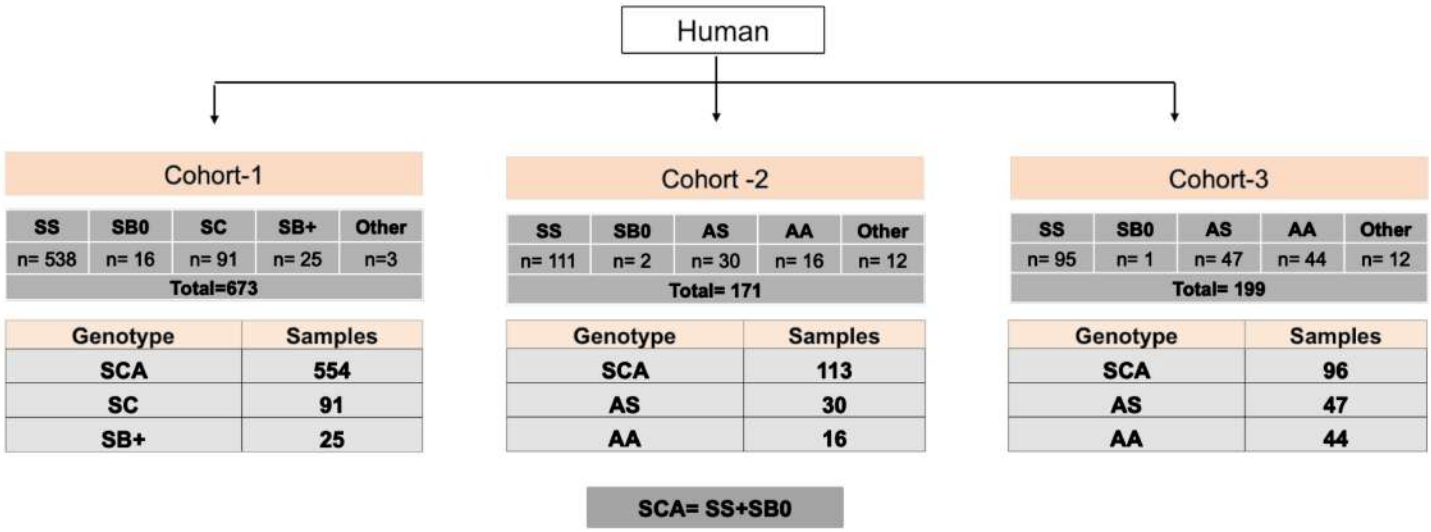
compared to males in all three genotypes- AA: Female, 8.3 vs Male, 0.7 (*P=ns*); AS: Female, 10.0 vs Male, 2.7 (*P=ns*); SS: Female, 9.0 vs Male, 4.7 (*P=ns*); the differences were not statistically significant. Statistical significance (*P* value) was assessed by Mann Whitney test.

Supplementary Figure S14: Pattern of mtDNA mutations burden in mito-genome across genotypes and cohorts. (A) Genotype wise pattern of mtDNA mutations across mito-genome. mtDNA heteroplasmic load across the mitochondrial genes was quantified and visualized for each genotype. The top bar with different colors represents the mitochondrial genome, consisting of 39 regions- 37 mitochondrial genes, the D loop, and intergenic regions. mtDNA heteroplasmic burden across the mitochondrial genes was compared between genotypes across four different cohorts. The pattern of mutational load in the mitogenome was consistent across the genotypes from each cohort. **(B)** Heat-map showing mtDNA heteroplasmy burden across the mito-genome among the different genotypes in the four cohort. The top bar with different colors represented the mitochondrial genome, consisting of 39 regions (indicated in different colors): 37 mitochondrial genes, the D loop, and intergenic regions. Each row indicated the mtDNA heteroplasmy burden in one genotype of one cohort. The highlighted box below the heat-map indicated the regions (D loop, RNR1, RNR2, CO1, ATP6, ND5, CYB) with higher mtDNA heteroplasmy burden in all four cohorts, indicating potential hotspots in the mito-genome.

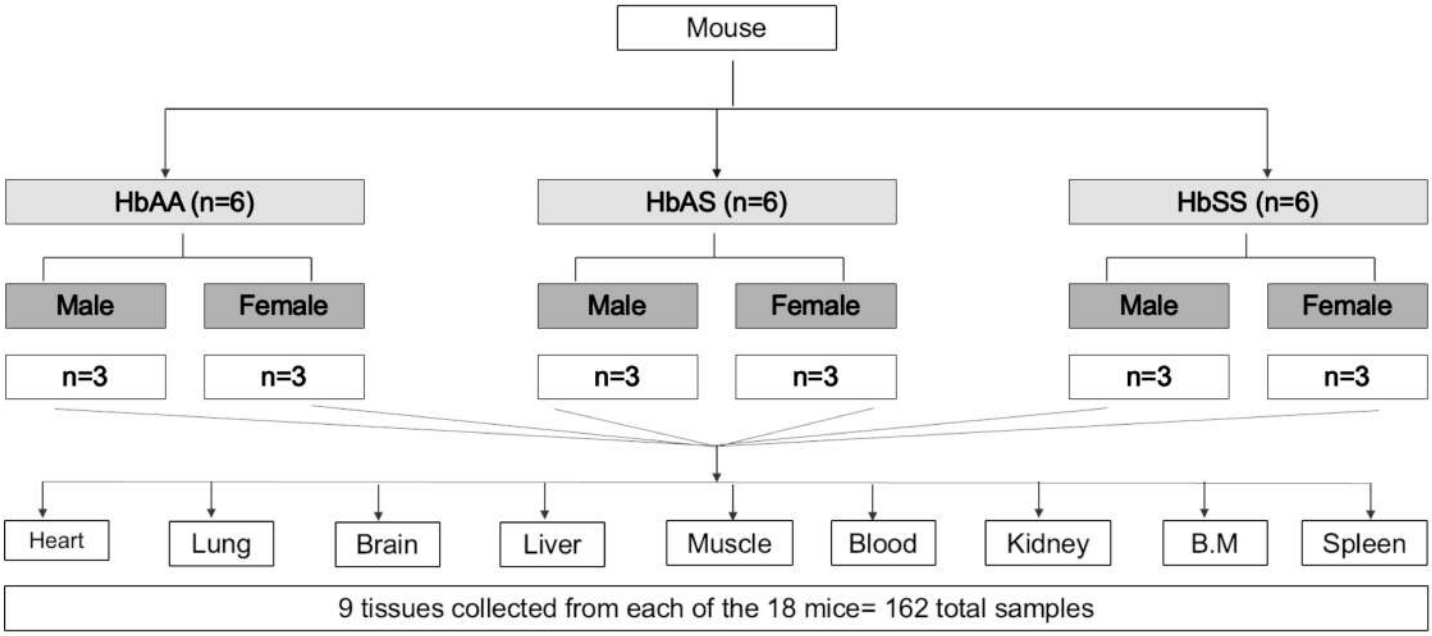
References

1. Kamimura S, Smith M, Vogel S, Almeida LEF, Thein SL, Quezado ZMN. Mouse models of sickle cell disease: Imperfect and yet very informative. Blood Cells, Molecules, and Diseases. 2024;104(102776.
2. Hirose M, Schilf P, Gupta Y, et al. Low-level mitochondrial heteroplasmy modulates DNA replication, glucose metabolism and lifespan in mice. Scientific reports. 2018;8(1):5872.
3. Battle SL, Puiu D, Group TOmW, et al. A bioinformatics pipeline for estimating mitochondrial DNA copy number and heteroplasmy levels from whole genome sequencing data. NAR Genom Bioinform. 2022;4(2):lqac034.
4. Guo S, Zhou K, Yuan Q, et al. An innovative data analysis strategy for accurate next-generation sequencing detection of tumor mitochondrial DNA mutations. Mol Ther Nucleic Acids. 2021;23(232-243.
5. Douglas AP, Vance DR, Kenny EM, Morris DW, Maxwell AP, McKnight AJ. Next-generation sequencing of the mitochondrial genome and association with IgA nephropathy in a renal transplant population. Scientific reports. 2014;4(1):7379.
6. Hansen NF, Wang X, Tegegn MB, et al. Random forest classifiers trained on simulated data enable accurate short read-based genotyping of structural variants in the alpha globin region at Chr16p13.3. bioRxiv. 2023;
7. Lee TY, Lai MI, Ramachandran V, et al. Rapid detection of alpha-thalassaemia variants using droplet digital PCR. Int J Lab Hematol. 2016;38(4):435-443.
8. Sachdev V, Tian X, Gu Y, et al. A phenotypic risk score for predicting mortality in sickle cell disease. Br J Haematol. 2021;192(5):932-941.
9. Dosunmu-Ogunbi A, Yuan S, Reynolds M, et al. SOD2 V16A amplifies vascular dysfunction in sickle cell patients by curtailing mitochondria complex IV activity. Blood. 2022;139(11):1760-1765.
10. Farias ICC, Mendonça-Belmont TF, da Silva AS, et al. Association of the SOD2 Polymorphism (Val16Ala) and SOD Activity with Vaso-occlusive Crisis and Acute Splenic Sequestration in Children with Sickle Cell Anemia. Mediterr J Hematol Infect Dis. 2018;10(1):e2018012.
11. Rooks H, Brewin J, Gardner K, et al. A gain of function variant in PIEZO1 (E756del) and sickle cell disease. Haematologica. 2019;104(3):e91-e93.
12. Romero LO, Bade M, Elsherif L, et al. Enhanced PIEZO1 function contributes to the pathogenesis of sickle cell disease. Proc Natl Acad Sci U S A. 2025;122(40):e2514863122.
13. Chowdhury FA, Sharma M, Schafer D, et al. CYB5R3 T117S tempers fetal hemoglobin induction by hydroxyurea in patients with sickle cell disease. Blood Adv. 2024;8(23):6098-6103.
14. Gordeuk VR, Shah BN, Zhang X, et al. The CYB5R3(c) (.350C>G) and G6PD A alleles modify severity of anemia in malaria and sickle cell disease. Am J Hematol. 2020;95(11):1269-1279.

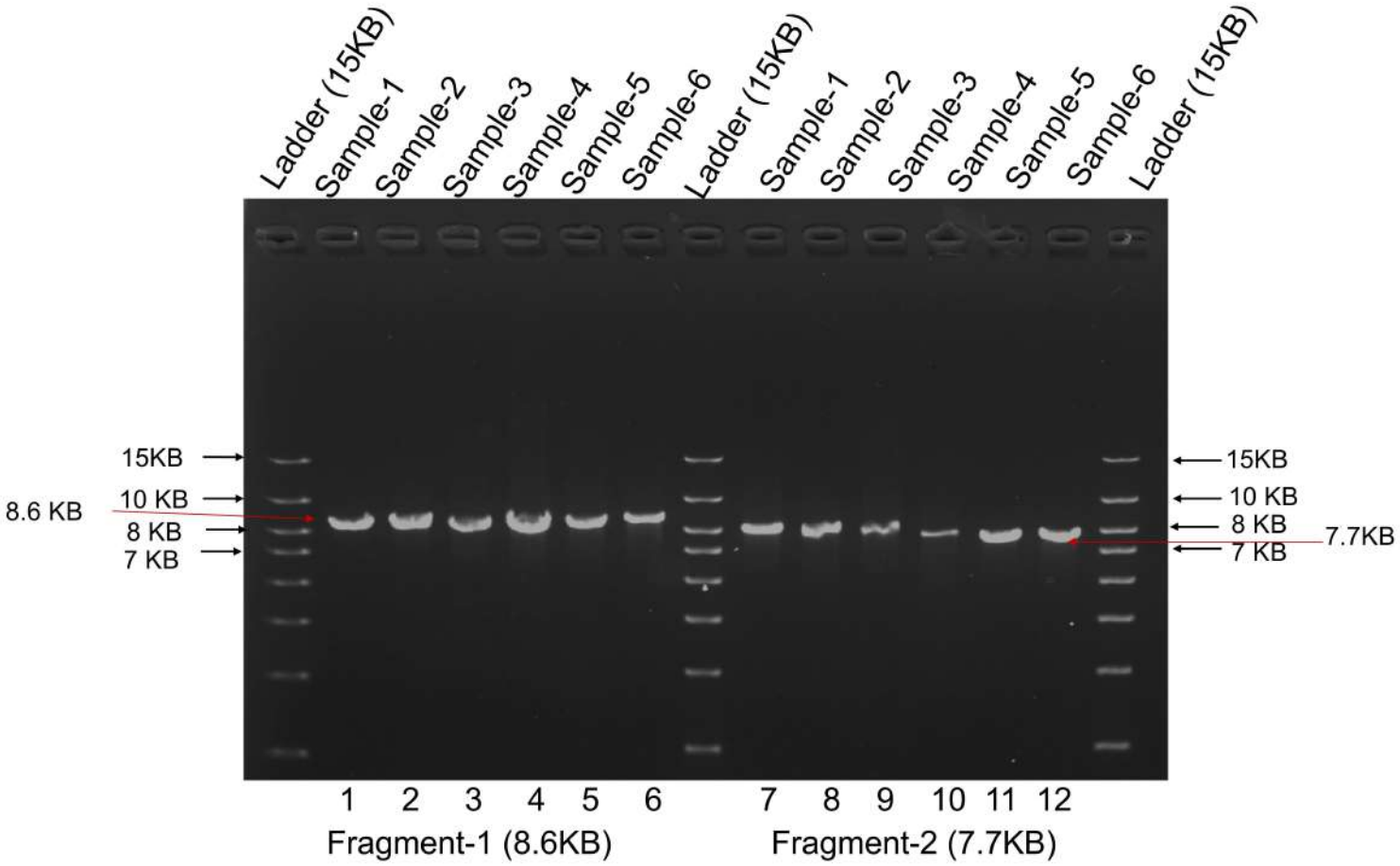
A



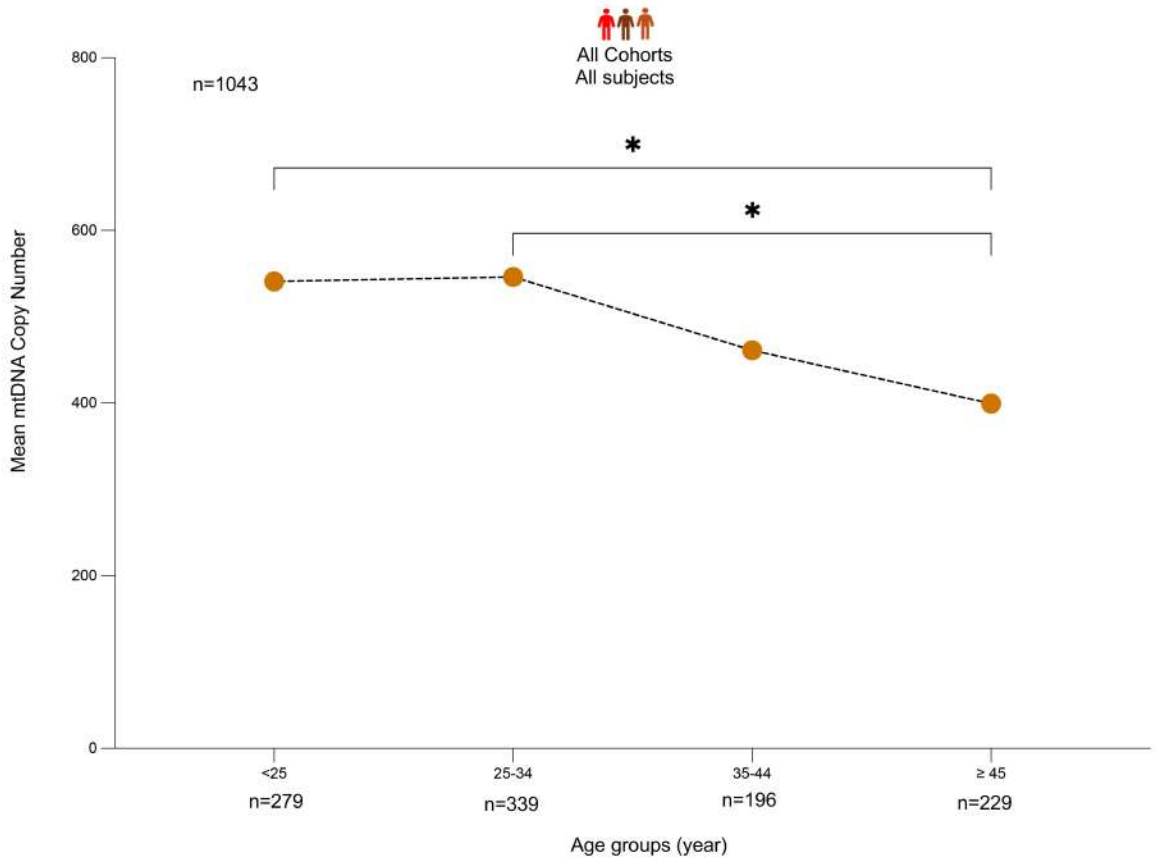
B



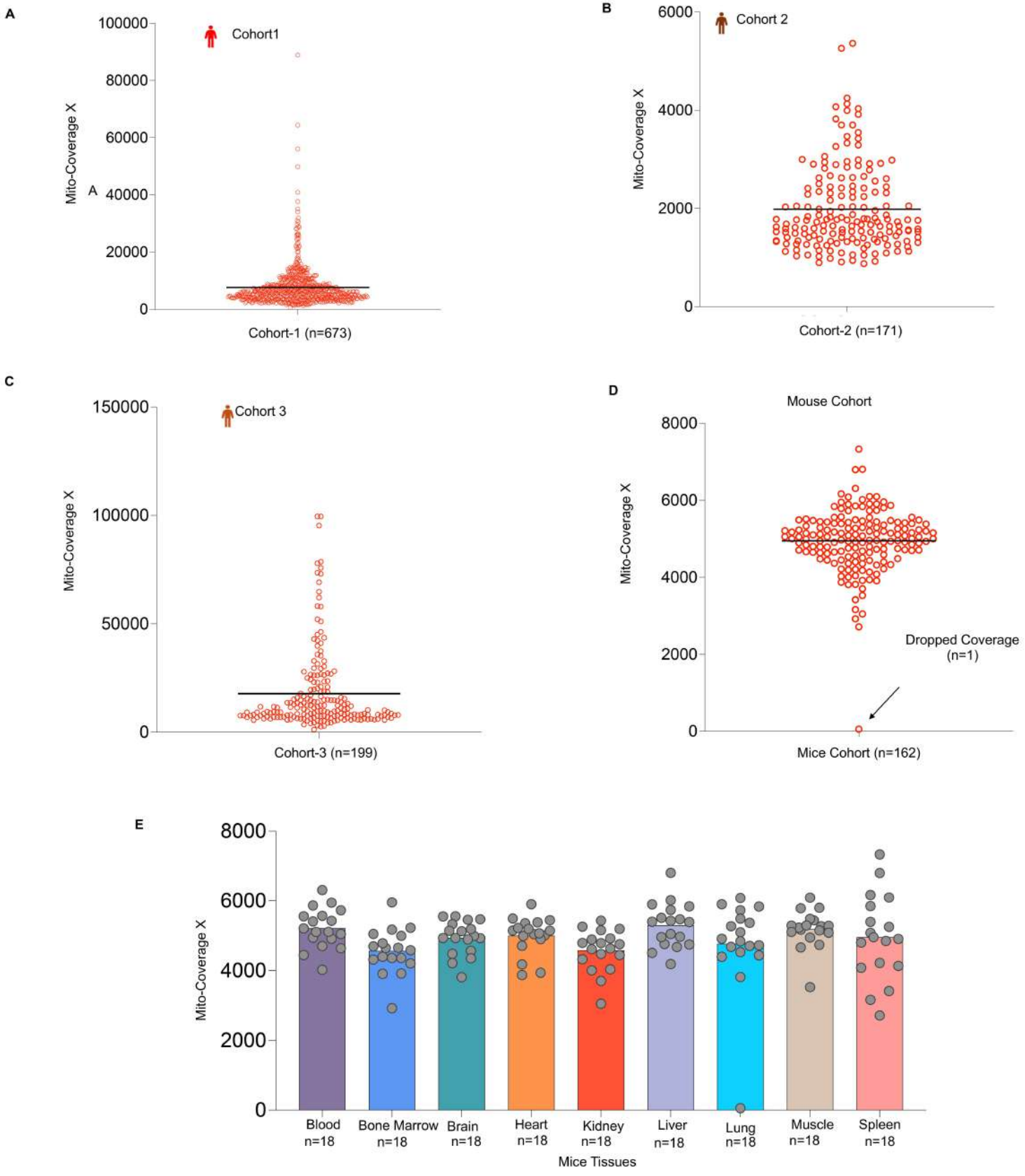
Ray, R. et al, Supplementary Figure S1



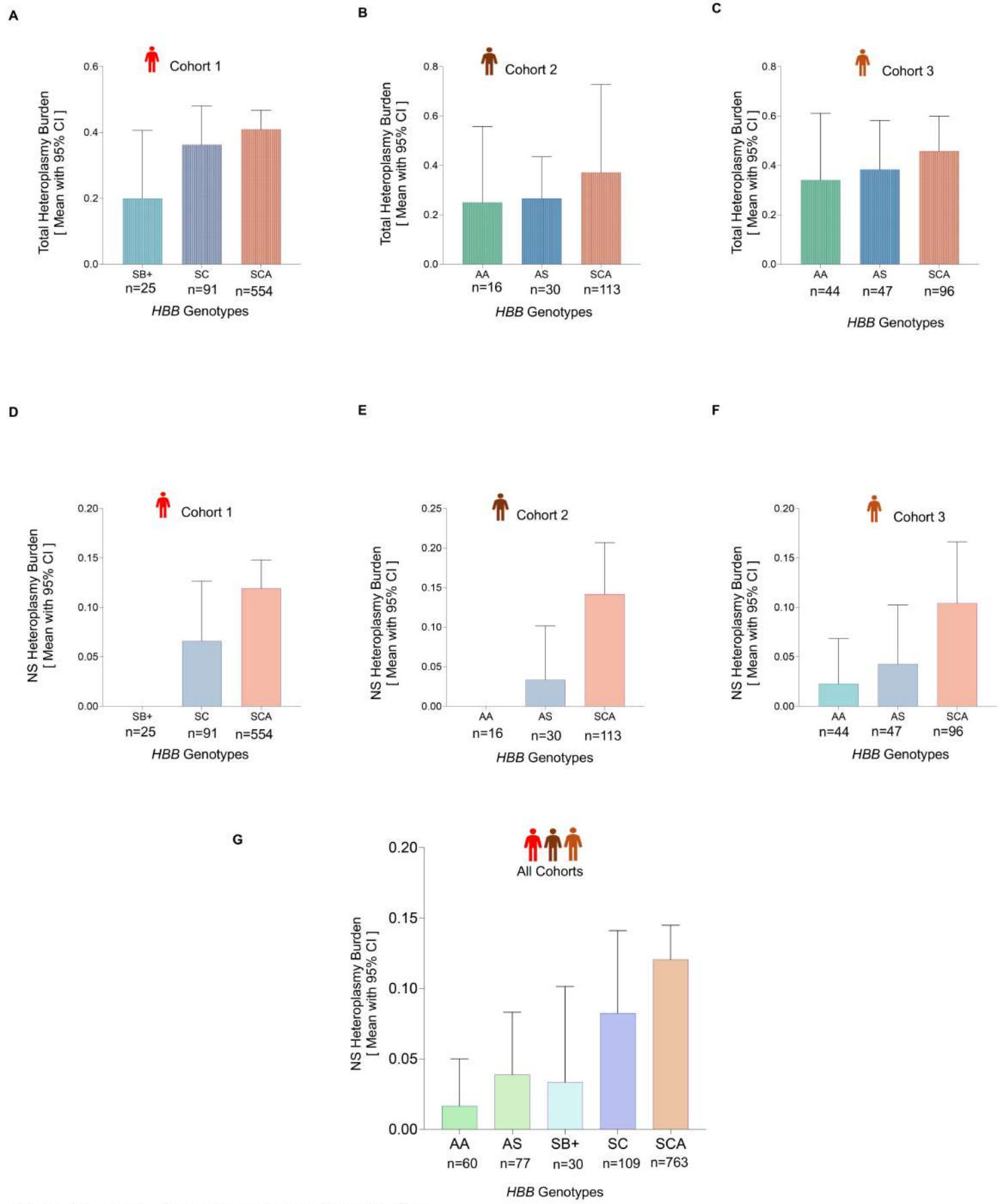
Ray, R. et al, Supplementary Figure S2



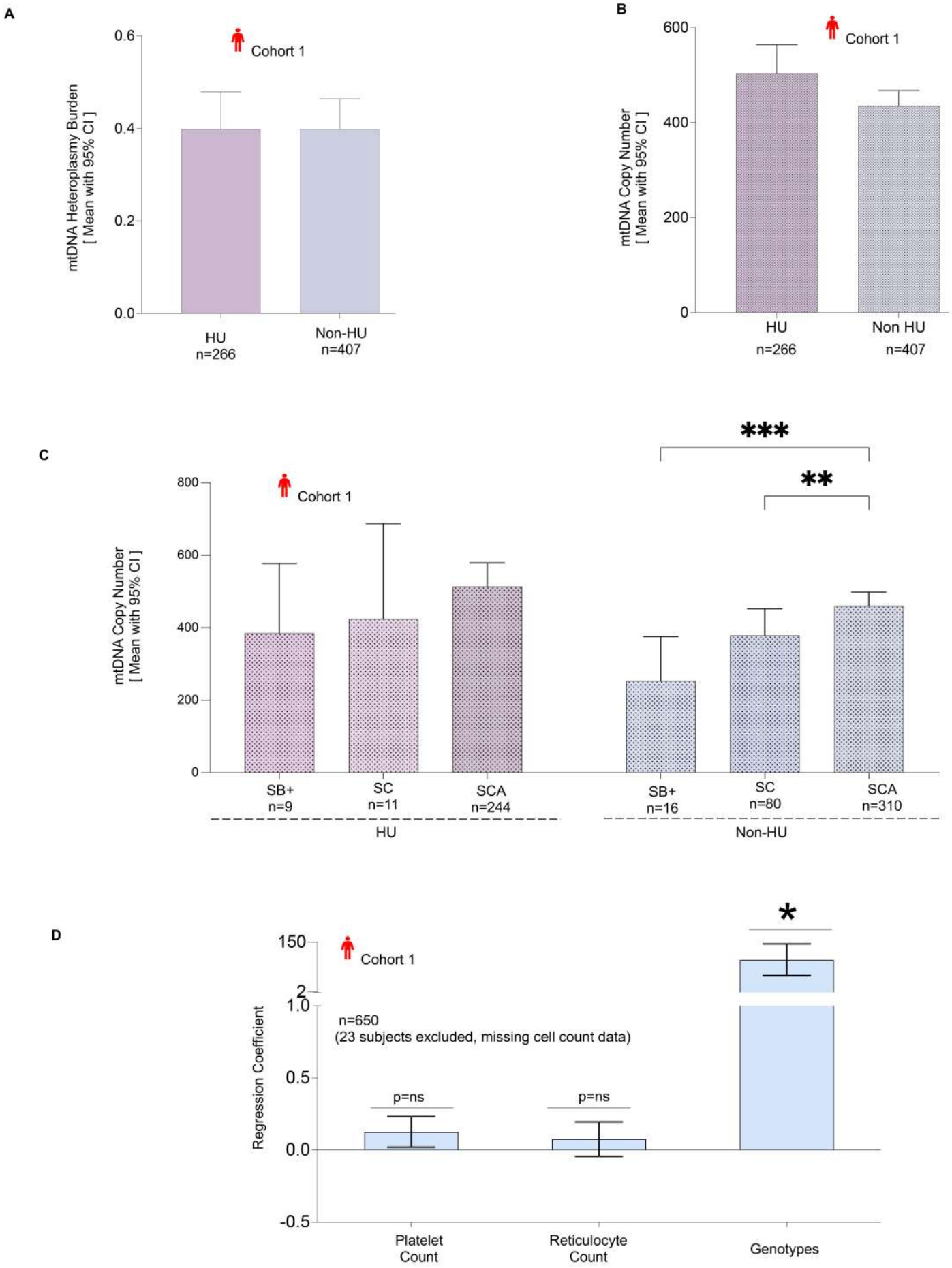
Ray, R. et al, Supplementary Figure S3



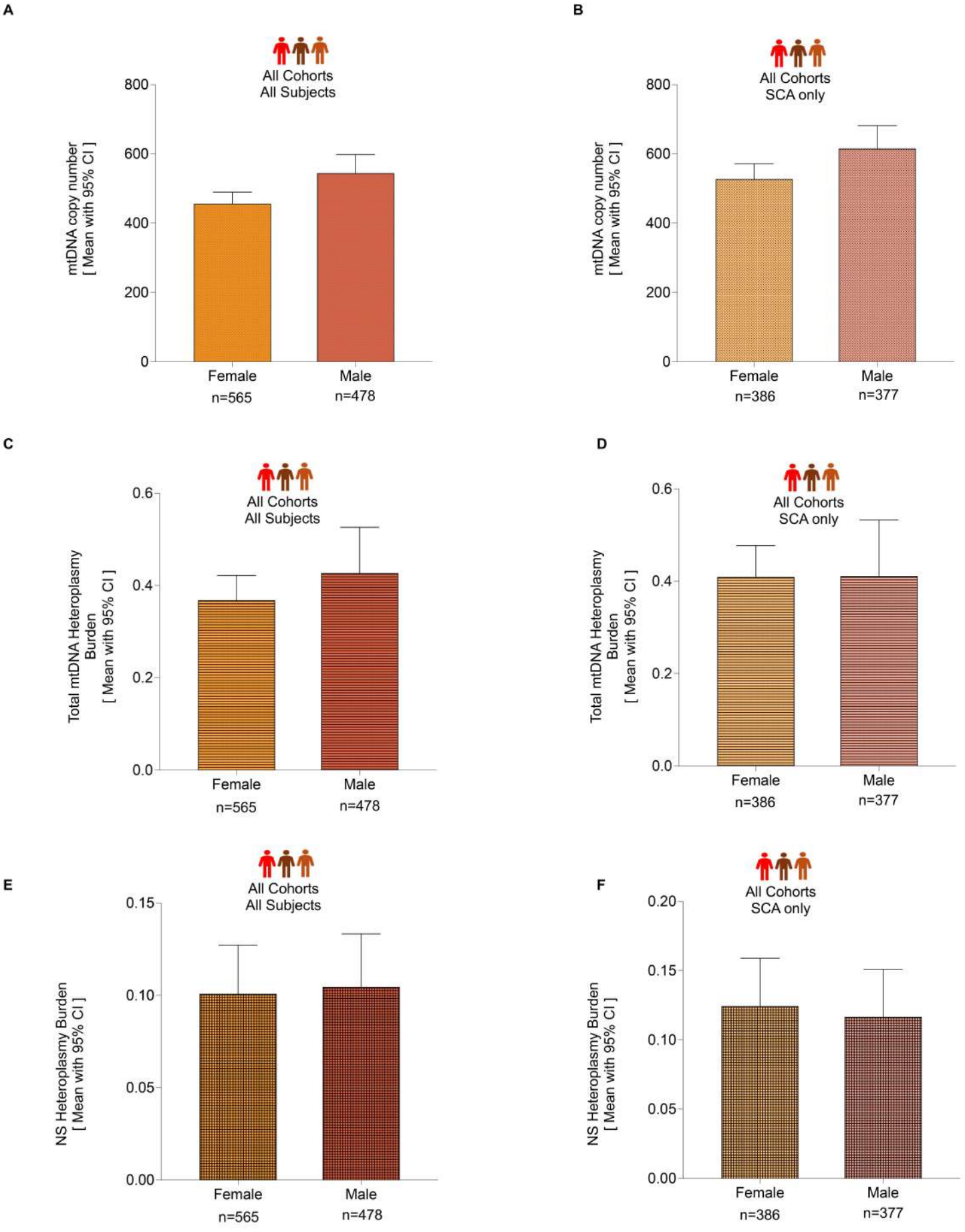
Ray, R. et al, Supplementary Figure S4



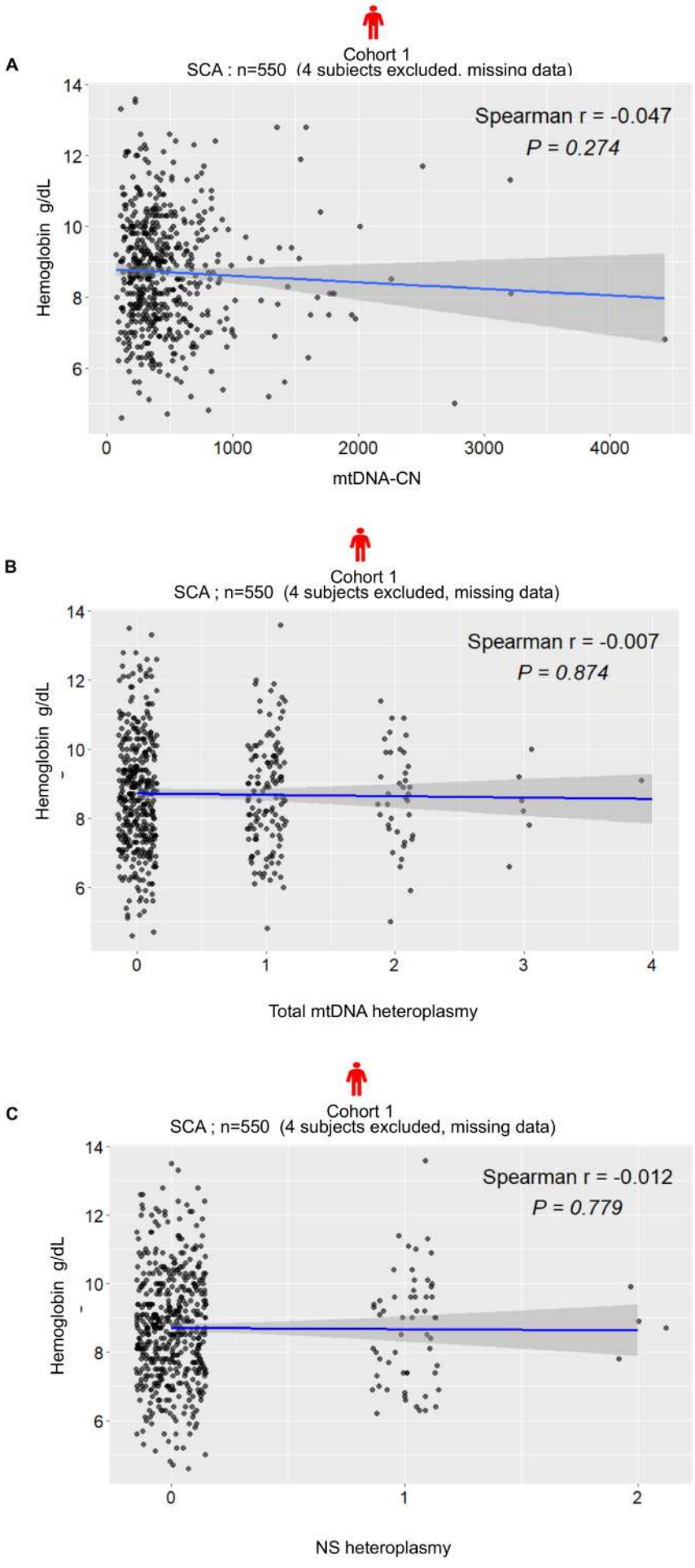
Ray, R. et al, Supplementary Figure S5



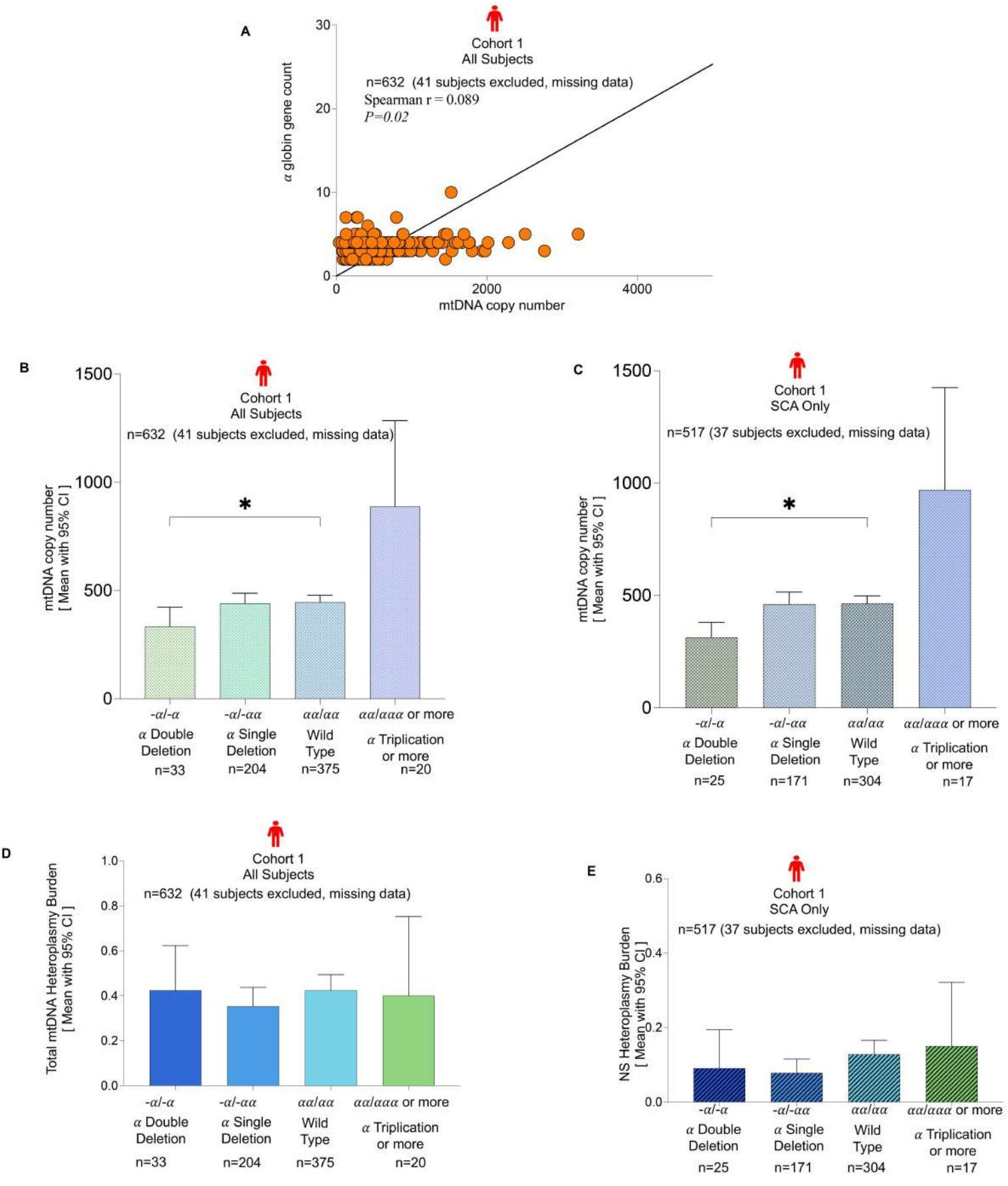
Ray, R. et al, Supplementary Figure S6



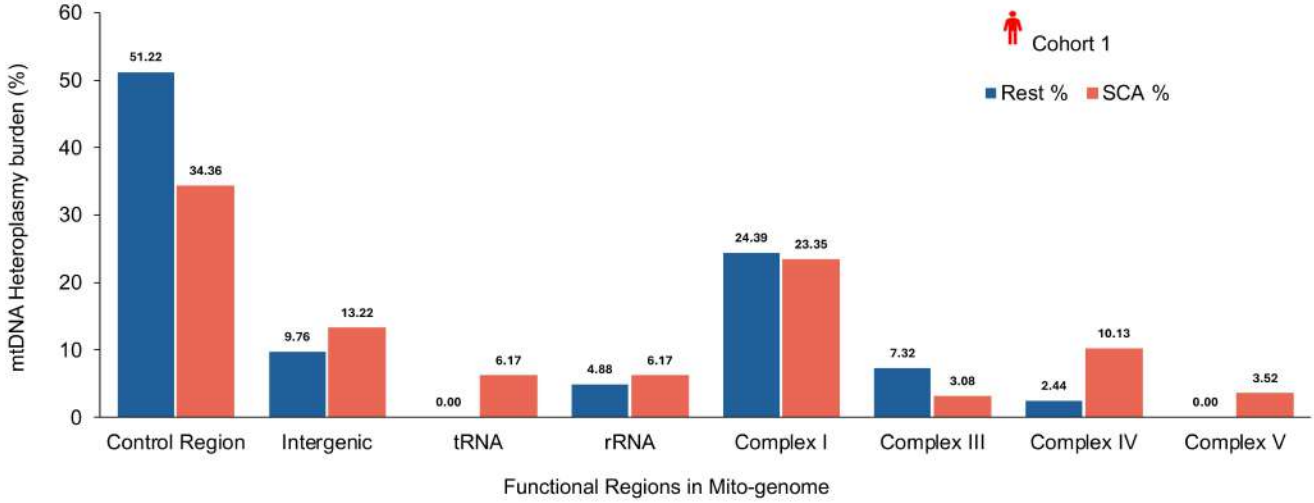
Ray, R. et al, Supplementary Figure S7



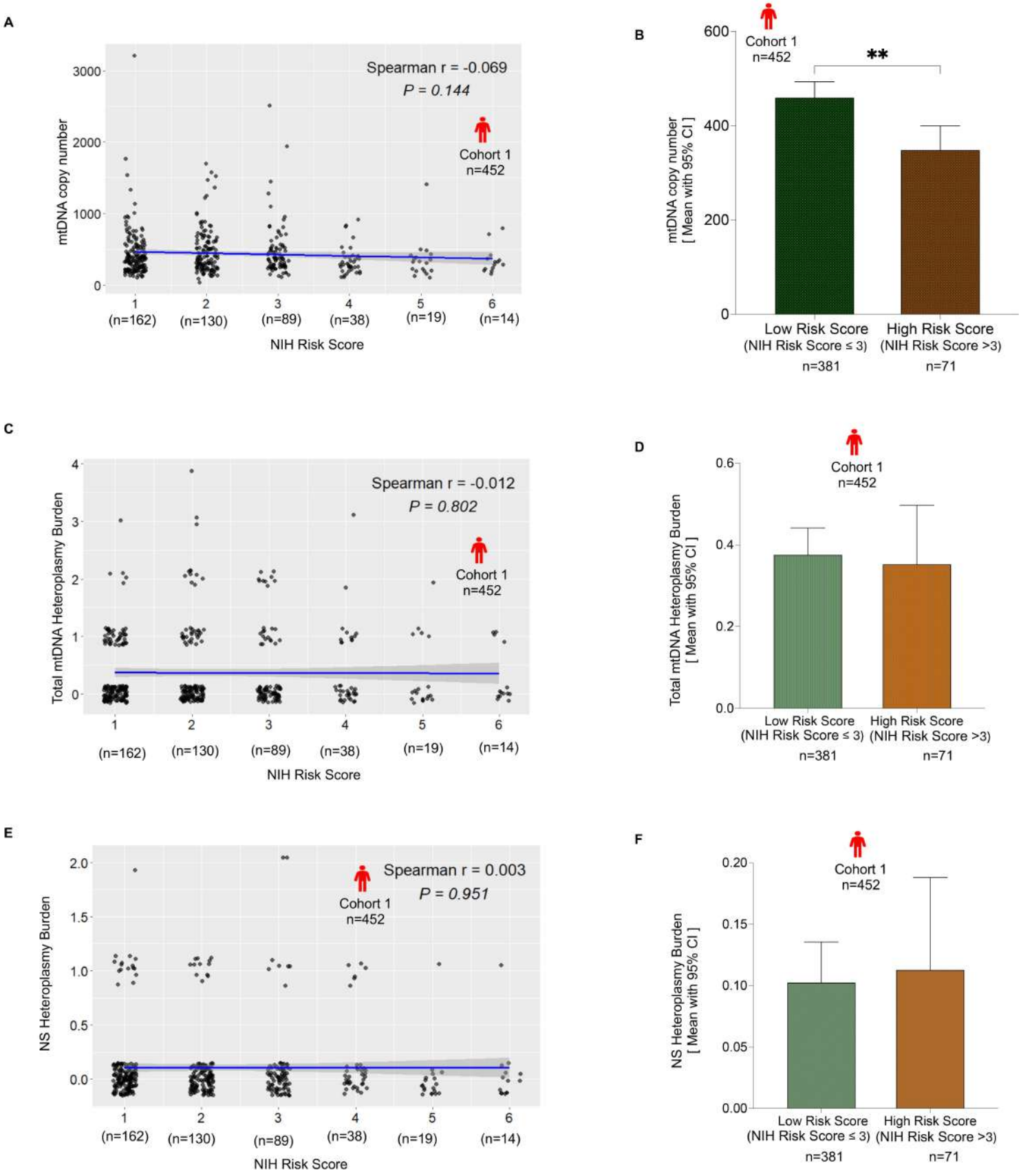
Ray, R. et al, Supplementary Figure S8



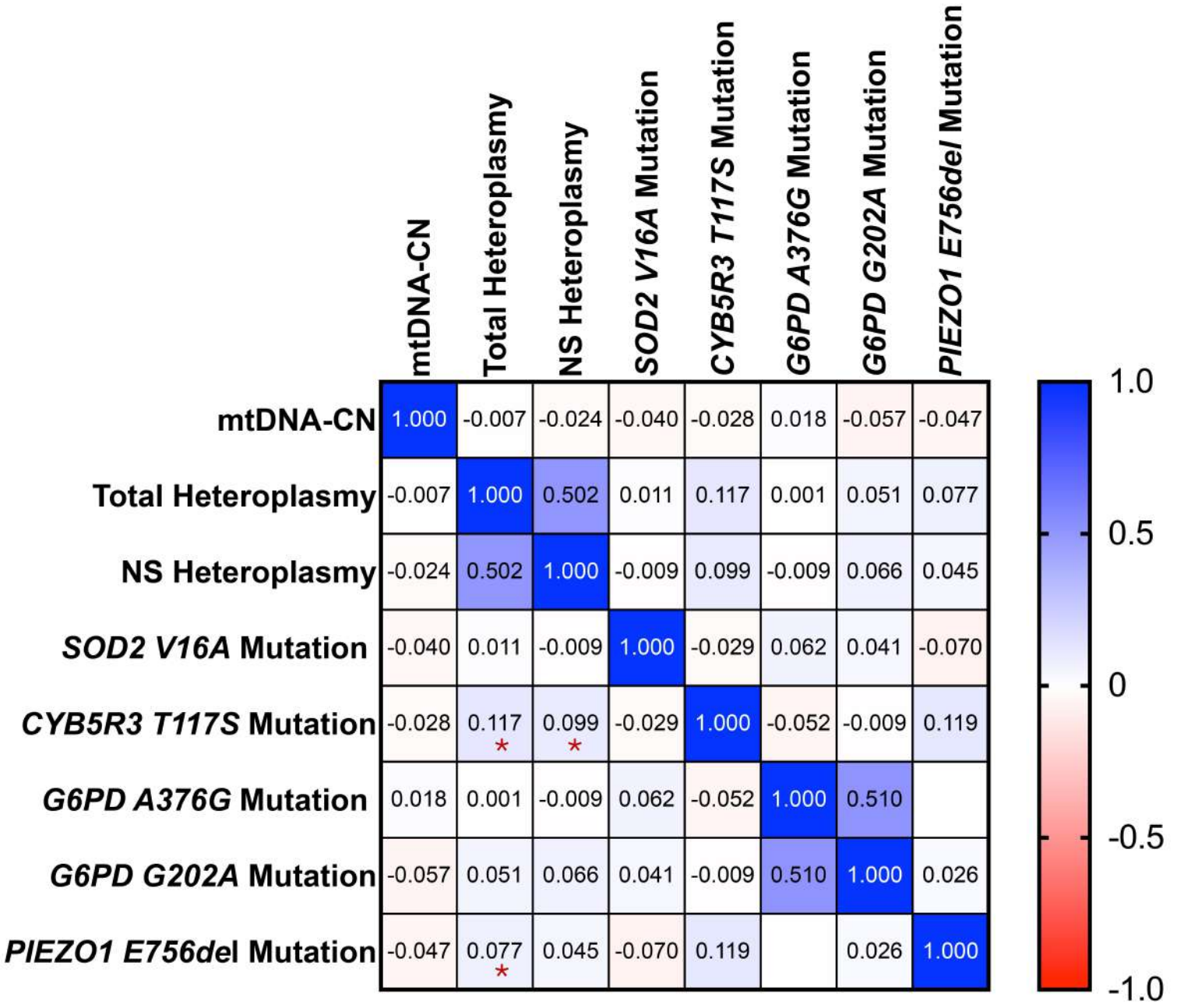
Ray, R. et al, Supplementary Figure S9



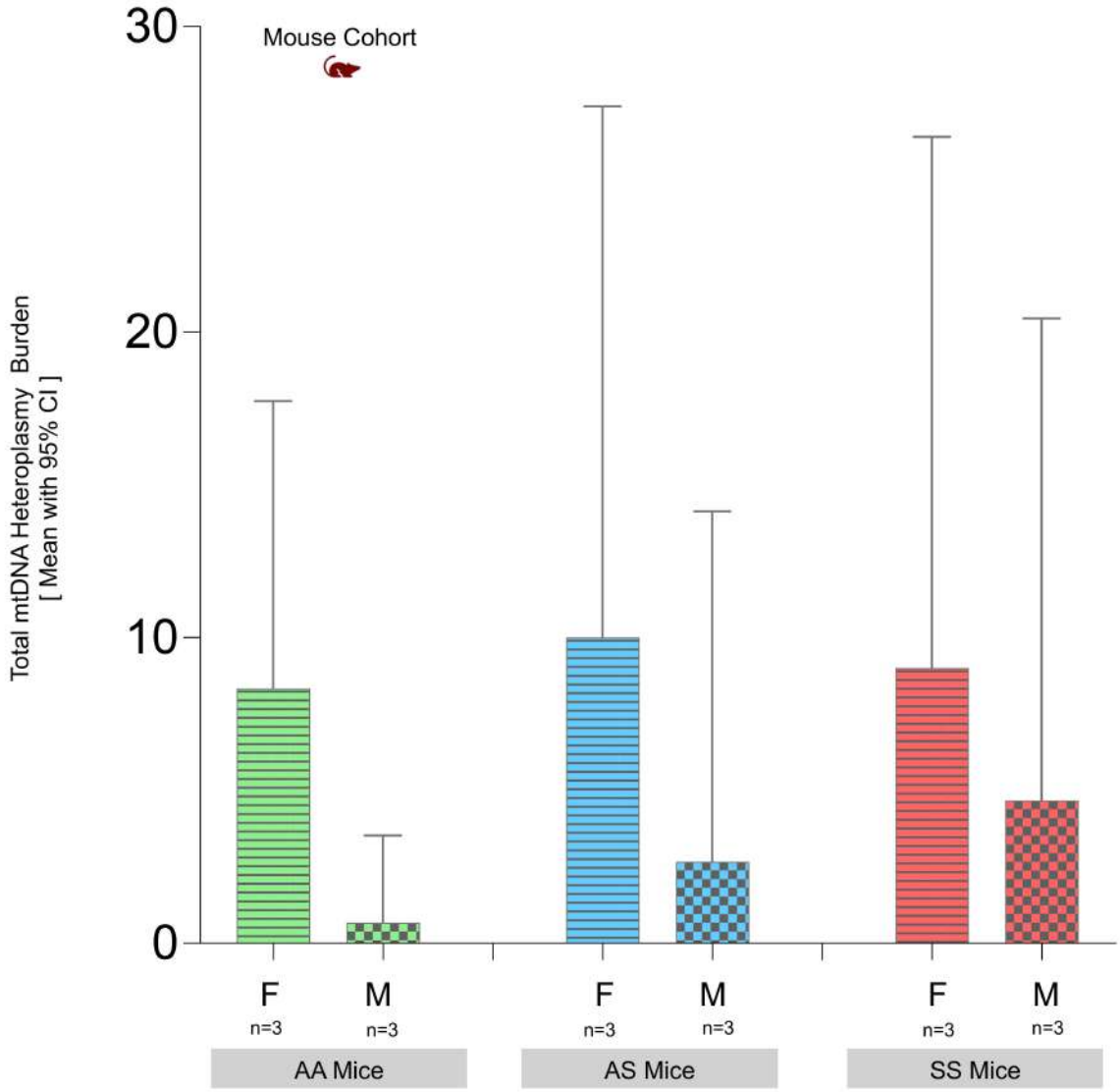
Ray, R. et al, Supplementary Figure S10



Ray, R. et al, Supplementary Figure S11

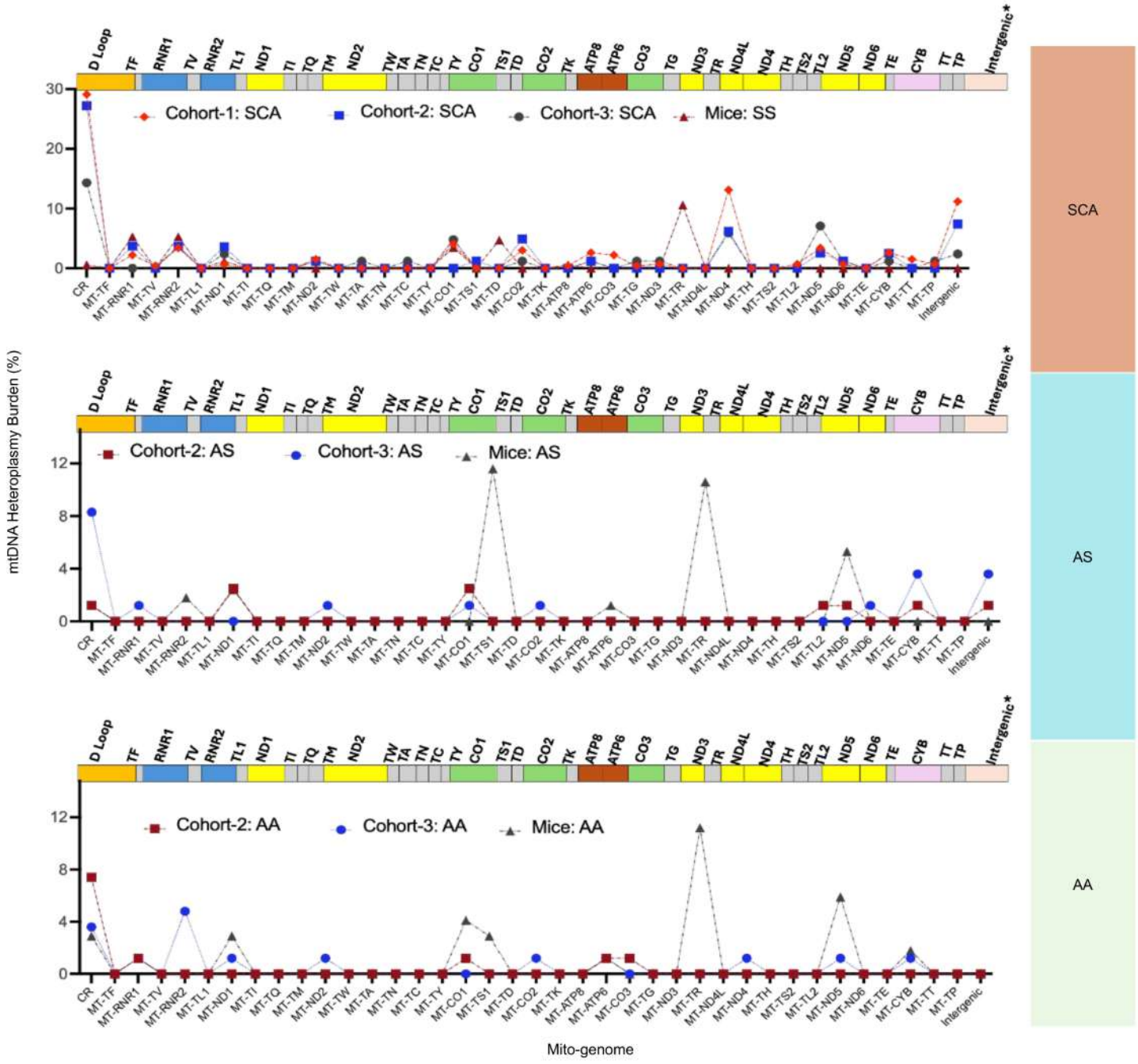


Ray, R. et al, Supplementary Figure S12



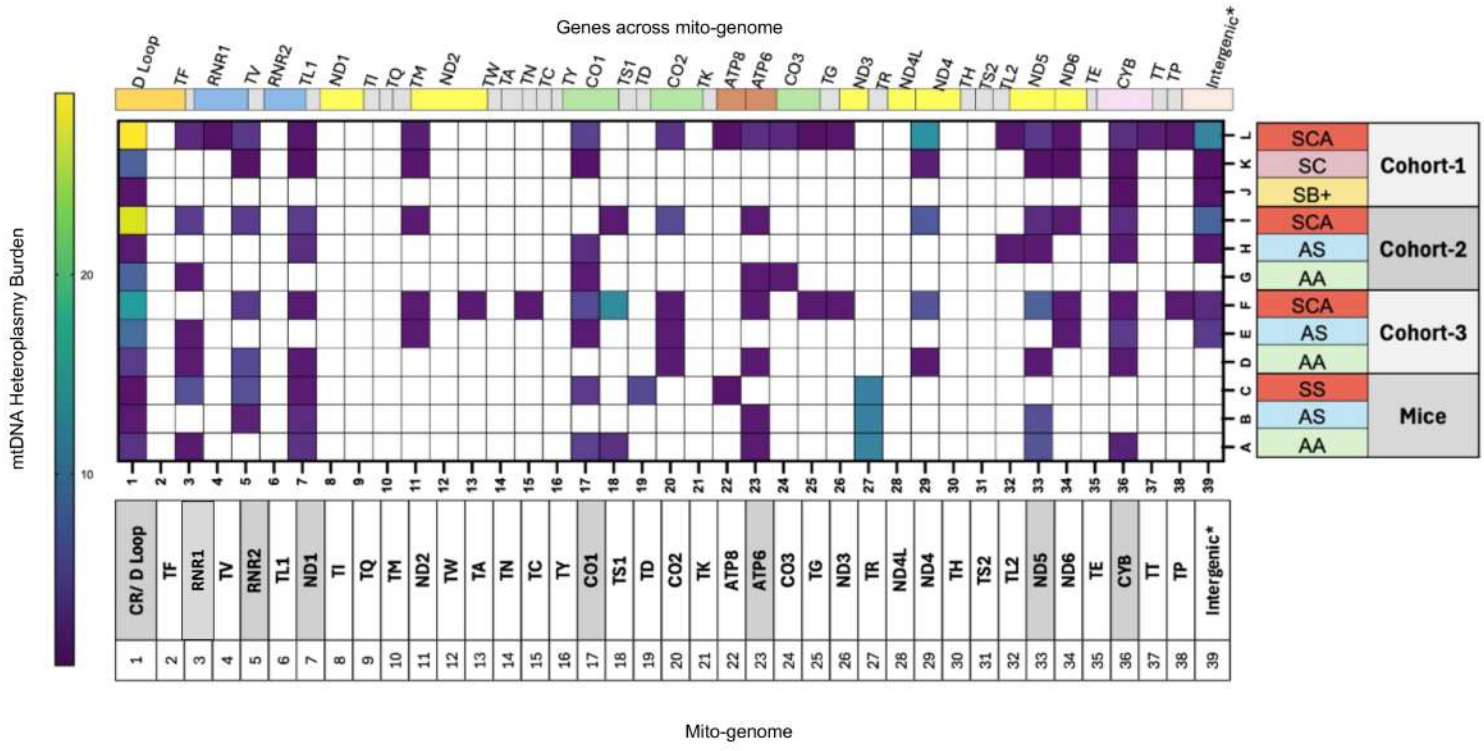
Ray, R. et al, Supplementary Figure S13

A



Ray, R. et al, Supplementary Figure S14A

B



Ray, R. et al, Supplementary Figure S14B

NIH Publishing Agreement & Manuscript *Cover Sheet*

By signing this Cover Sheet, the Author, on behalf of NIH, agrees to the provisions set out below, which **modify and supersede,** solely with respect to NIH, any conflicting or ambiguous provisions that are in the Publisher's standard copyright agreement (the "Publisher's Agreement"). NIH and Author accept only those provisions of the Publisher's Agreement that are consistent with the provisions herein, and only to the extent that they are directly related to the disposition of copyrights or duties of the parties that are directly related to the subject "Work."

1. **Indemnification.** No Indemnification or "hold harmless" obligation is provided by either party.

2. **Governing Law.** This agreement will be governed by the law of the court in which a claim is brought.

3. **Copyright.** Author's contribution to the Work was done as part of the Author's official duties as a NIH employee and is a Work of the United States Government. Therefore, copyright may not be established in the United States. 17 U.S.C. § 105. If Publisher intends to disseminate the Work outside of the U.S., Publisher may secure copyright to the extent authorized under the domestic laws of the relevant country, subject to a paid-up, nonexclusive, irrevocable worldwide license to the United States in such copyrighted work to reproduce, prepare derivative works, distribute copies to the public and perform publicly and display publicly the work, and to permit others to do so.

4. **No Compensation.** No royalty income or other compensation may be accepted for work done as part of official duties. The author may accept for the agency a limited number of reprints or copies of the publication.

5. **NIH Representations.** NIH represents to the Publisher that the Author is the sole author of the Author's contribution to the Work and that NIH is the owner of the rights that are the subject of this agreement; that the Work is an original work and has not previously been published in any form anywhere in the world; that to the best of NIH's knowledge the Work is not a violation of any existing copyright, moral right, database right, or of any right of privacy or other intellectual property, personal, proprietary or statutory right; that where the Author is responsible for obtaining permissions or assisting the Publishers in obtaining permissions for the use of third party material, all relevant permissions and information have been secured; and that the Work contains nothing misleading, obscene, libelous or defamatory or otherwise unlawful.

6. **Disclaimer.** NIH and the Author expressly disclaim any obligation that is not consistent with the Author's official duties or the NIH mission.

7. **PubMed Central.** The Author is a US government employee who must comply with the NIH Public Access Policy, and the Author or NIH will deposit, or have deposited, in NIH's PubMed Central archive, an electronic version of the final, peer-reviewed manuscript upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication. PubMed Central may tag or modify the work consistent with its customary practices.

Thein, Swee Lay (NIH/NHLBI) [E] Digitally signed by Thein, Swee Lay (NIH/NHLBI) [E] Date: 2026.03.09 12:28:45 -04'00'

_____ _____9 March 2026_____

Signature of Author Date

Dr Swee Lay Thein

_____

Name of Author