

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 Tumburu¹ and Swee Lay Thein¹

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

Correspondence: S.L. Thein
sl.thein@nih.gov

Received: November 18, 2025.
Accepted: March 11, 2026.
Early view: March 19, 2026.

<https://doi.org/10.3324/haematol.2025.300253>

©2026 NIH (National Institutes of Health)

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

Rudra Ray¹, Haiou Li¹, Shijinqiu Gao¹, Nancy Asomaning¹, Kang Le¹, Maliha Ahmad¹, Xunde Wang¹, Yuesheng Li², Sayuri Kamimura³, Clifton L. Dalgard^{4,5}, Chengyu Liu⁶, Zenaide M.N. Quezado^{1,3}, Justin Lack⁷, Laxminath Tumburu¹, Swee Lay Thein¹

¹Sickle Cell Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.

²DNA Sequencing & Genomics Core, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.

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

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

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

⁶Transgenic Core, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD.

⁷Bioinformatics (NCBR)/Integrated Data Sciences Section (IDSS)/ Research Technology Branch/DIR, NIAID, NIH, Bethesda, MD.

SUPPLEMENTARY MATERIALS.

Contents:

Supplementary Method

Supplementary Tables

- **Supplementary Table S1:** Details of the human sample cohorts.
- **Supplementary Table S2:** Details of mito-coverage and VAF cutoff for each cohort.
- **Supplementary Table S3:** Distribution and stratification of mtDNA heteroplasmies in three human cohorts.
- **Supplementary Table 4:** Distribution of mtDNA heteroplasmy burden and mtDNA-CN stratified by hydroxyurea (HU) treatment in Cohort 1.
- **Supplementary Table S5:** Multivariate regression co-efficient for platelet and reticulocyte counts and genotype in cohort 1 (n=650).
- **Supplementary Table S6:** Correlation between alpha (α) globin gene number and mtDNA metrics among SCD patients in Cohort 1.
- **Supplementary Table S7:** Mutational burden across functional regions of the mitochondrial genome among different genotypes in human Cohort 1.

- **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-thalassemia⁰ (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/dpuii/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/dpuii/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 NuMTs⁵ 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 algorithm⁶ 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

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)	<i>P</i> 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
<i>SOD2</i> V16A	rs4880	<i>SOD2</i>	chr6:160113872	A>C, G, T	Oxidative stress, mitochondrial activity, SCD	9, 10
<i>PIEZO1</i> E756del	rs572934641	<i>Piezo1</i>	chr16:88800372-383	TCC Indels	Red cell dehydration, SCD, Malaria	11, 12
<i>CYB5R3</i> T117S	rs1800457	<i>CYB5R3</i>	chr22:43024271	G>A, C	Methemoglobin Reduction, Oxidative Stress, Anemia	13
<i>G6PD</i> A376G	rs1050829	<i>G6PD</i>	X:153763492	T>A, C	Hemolytic Anemia	14
<i>G6PD</i> G202A	rs1050828	<i>G6PD</i>	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
<i>SOD2</i> V16A	rs4880	673 (100%)	230 (34.18)	443 (65.82)	317 (47.10)	126 (18.72)	0.42
<i>PIEZO1</i> E756del	rs572934641	673 (100%)	466 (69.24)	207 (30.76)	185 (27.49)	22 (3.27)	0.17
<i>CYB5R3</i> T117S	rs1800457	673 (100%)	356 (52.90)	317 (47.10)	263 (39.08)	54 (8.02)	0.28
<i>G6PD</i> A376G	rs1050829	673 (100%)	387 (57.50)	286 (42.50)	149 (22.14)	137 (20.36)	0.31
<i>G6PD</i> 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
<i>SOD2</i> V16A	−0.040	0.295	0.011	0.766	−0.009	0.812
<i>PIEZO1</i> E756del	−0.047	0.222	0.077	0.045	0.045	0.245
<i>CYB5R3</i> T117S	−0.028	0.466	0.117	0.0023	0.099	0.0099
<i>G6PD</i> A376G	0.018	0.64	0.001	0.981	−0.009	0.816
<i>G6PD</i> 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)
<i>SOD2</i> V16A	WT (n=443)	441	0.4	0.1
	Mutant (n=230)	502	0.4	0.12
<i>PIEZO1</i> E756del	WT (n=466)	469	0.36	0.09
	Mutant (n=207)	446	0.49	0.14
<i>CYB5R3</i> T117S	WT (n=356)	475	0.33	0.07
	Mutant (n=317)	447	0.48	0.15
<i>G6PD</i> -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. <i>SOD2</i> V16A	-0.0175	-0.1123 to 0.0775	0.71	No
NIH Risk Score vs. <i>PIEZO1</i> E756del	-0.0002	-0.0952 to 0.0947	0.9959	No
NIH Risk Score vs. <i>CYB5R3</i> T117S	-0.0265	-0.1211 to 0.0686	0.5746	No
NIH Risk Score vs. <i>G6PD</i> A376G	-0.0783	-0.1719 to 0.0168	0.0965	No
NIH Risk Score vs. <i>G6PD</i> 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			
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			
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 heteroplasmy (A, B and C) and non-synonymous (NS) mtDNA heteroplasmy (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 heteroplasmy 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 heteroplasmy 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 ($P \geq 0.05$).

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 did 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 confounding 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) heteroplasmy 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 ($P \geq 0.05$).

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 $-\alpha/-\alpha$ had significantly lower mtDNA-CN compared to $\alpha\alpha/\alpha\alpha$ ($P=0.029$). (C) Among the SCA patients in Cohort 1 ($n=517$; excluded= 37), a similar pattern was observed: $-\alpha/-\alpha$ ($n=25$; mean mtDNA-CN= 313.2), $-\alpha/\alpha\alpha$ ($n=171$; mean mtDNA-CN= 461.5), $\alpha\alpha/\alpha\alpha$ ($n=304$; mean mtDNA-CN= 463.7), and $\alpha\alpha/\alpha\alpha\alpha$ or more ($n=17$; mean mtDNA-CN= 969.8). SCA patients with $-\alpha/-\alpha$ had significantly lower mtDNA-CN than $\alpha\alpha/\alpha\alpha$ ($P=0.038$). (D) No statistically significant differences were observed in total mtDNA heteroplasmy burden with α globin genotype status in SCD patients ($n=632$): $-\alpha/-\alpha$ ($n=33$; mean total mtDNA heteroplasmy=0.42), $-\alpha/\alpha\alpha$ ($n=204$; mean=0.35), wild type ($n=375$; mean=0.42), and $\alpha\alpha/\alpha\alpha\alpha$ 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 ($n=517$): $-\alpha/-\alpha$ ($n=25$; mean NS heteroplasmy=0.09), $-\alpha/\alpha\alpha$ ($n=171$; mean=0.07), $\alpha\alpha/\alpha\alpha$ ($n=304$; mean=0.12), and $\alpha\alpha/\alpha\alpha\alpha$ 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

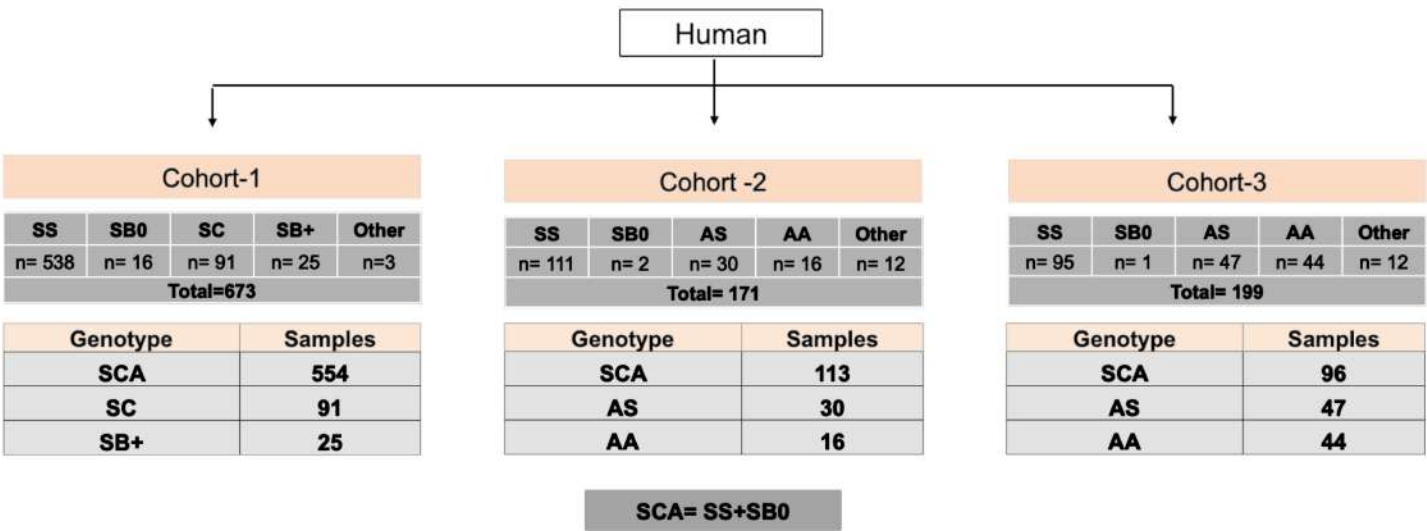
compared to males in all three genotypes- AA: Female, 8.3 vs Male, 0.7 ($P=ns$); AS: Female, 10.0 vs Male, 2.7 ($P=ns$); SS: Female, 9.0 vs Male, 4.7 ($P=ns$); the differences were not statistically significant. Statistical significance (P value) was assessed by Mann Whitney test.

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.

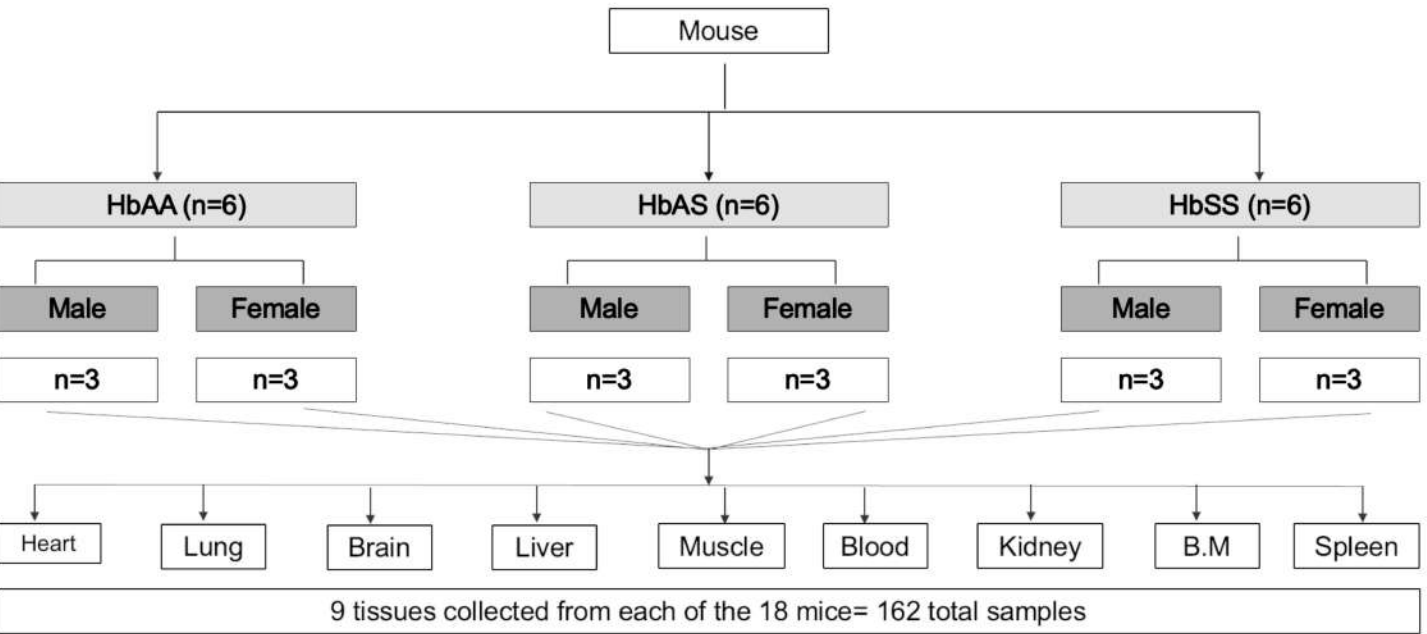
References

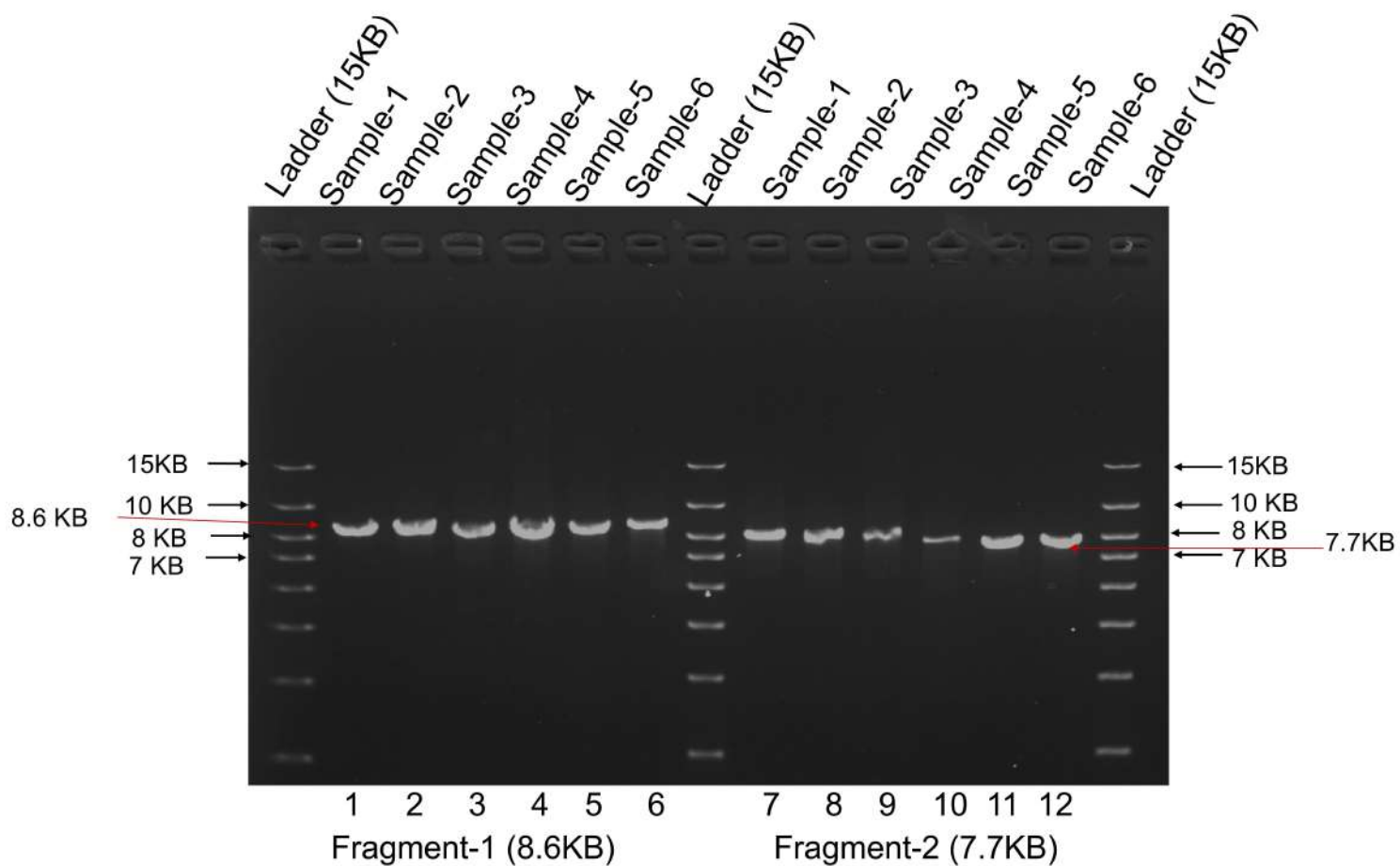
1. Kamimura S, Smith M, Vogel S, Almeida LEF, Thein SL, Quezado ZMN. Mouse models of sickle cell disease: Imperfect and yet very informative. *Blood Cells, Molecules, and Diseases*. 2024;104(102776).
2. Hirose M, Schilf P, Gupta Y, et al. Low-level mitochondrial heteroplasmy modulates DNA replication, glucose metabolism and lifespan in mice. *Scientific reports*. 2018;8(1):5872.
3. Battle SL, Puiu D, Group TOW, 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

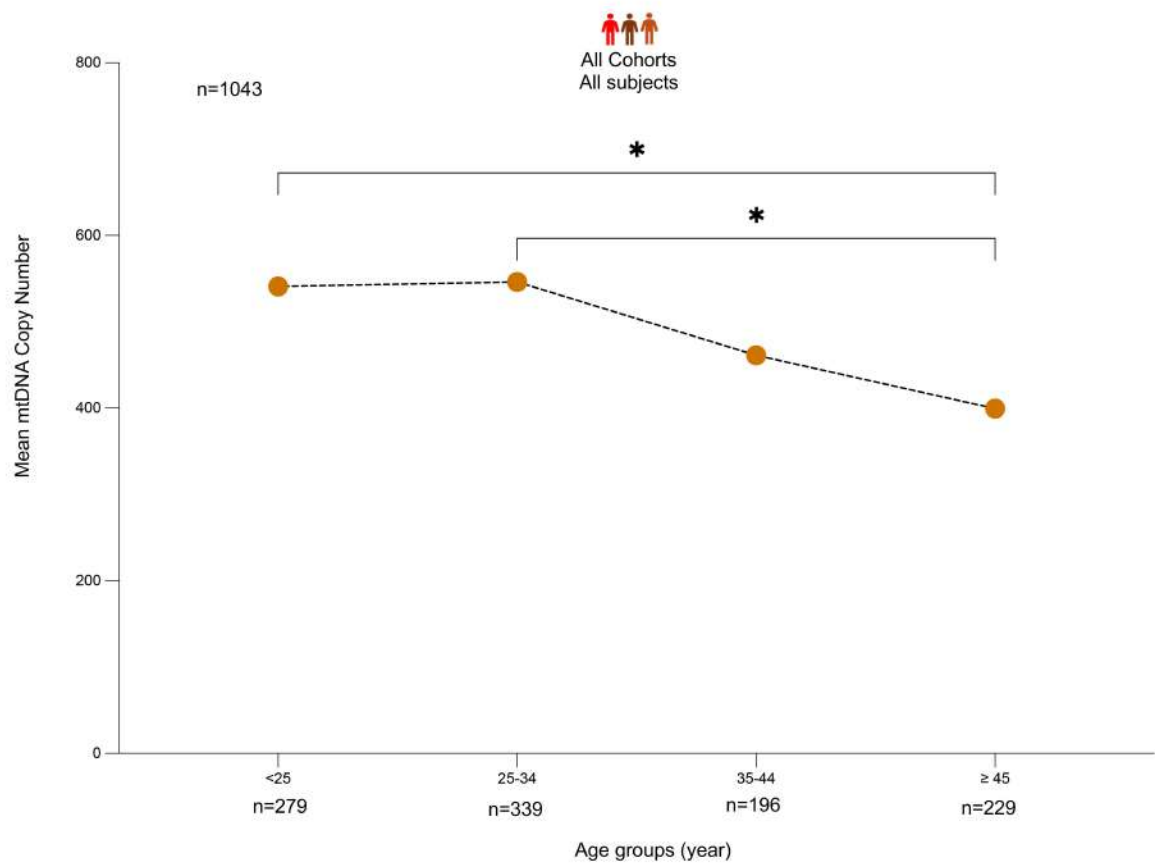


B

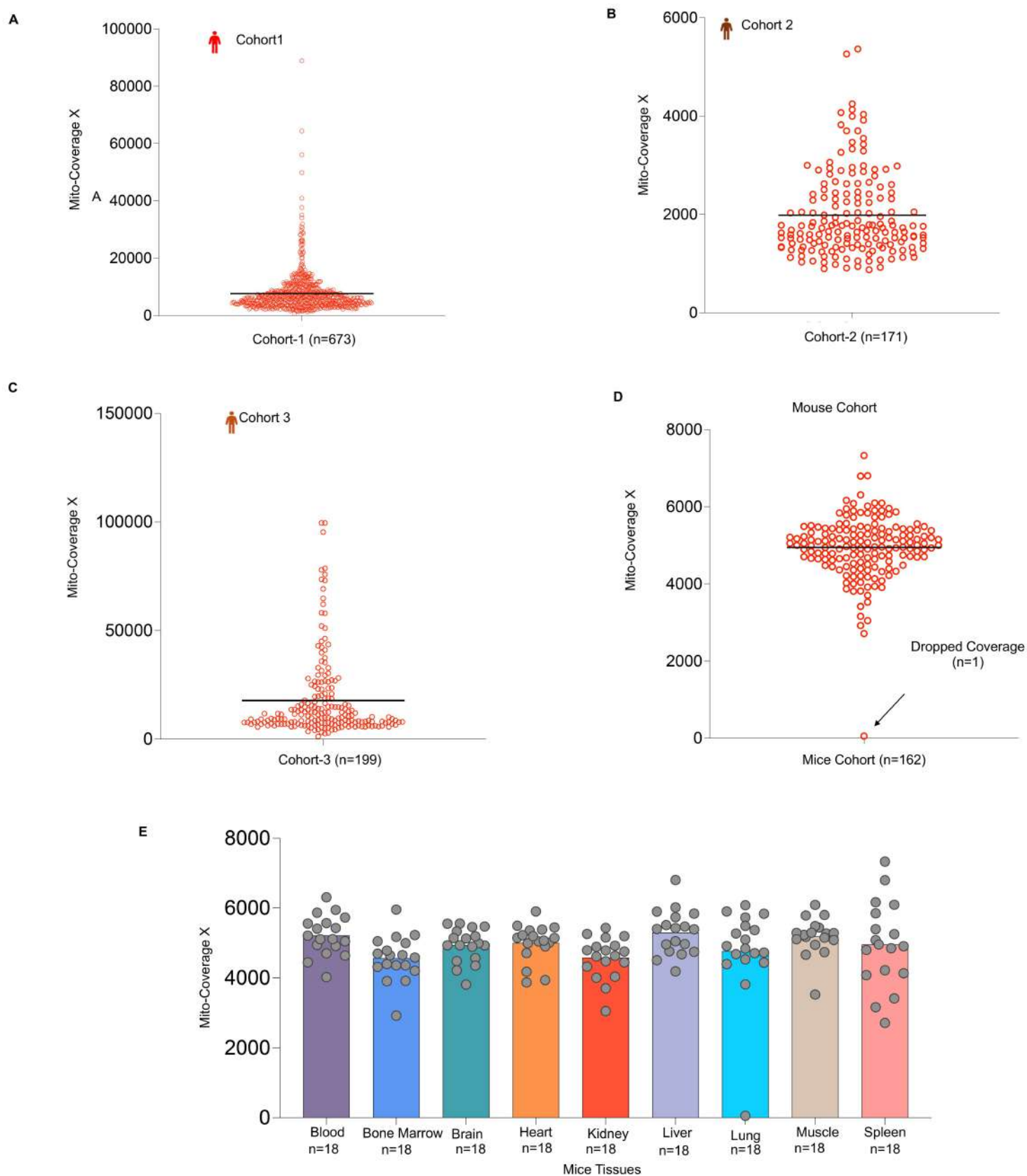




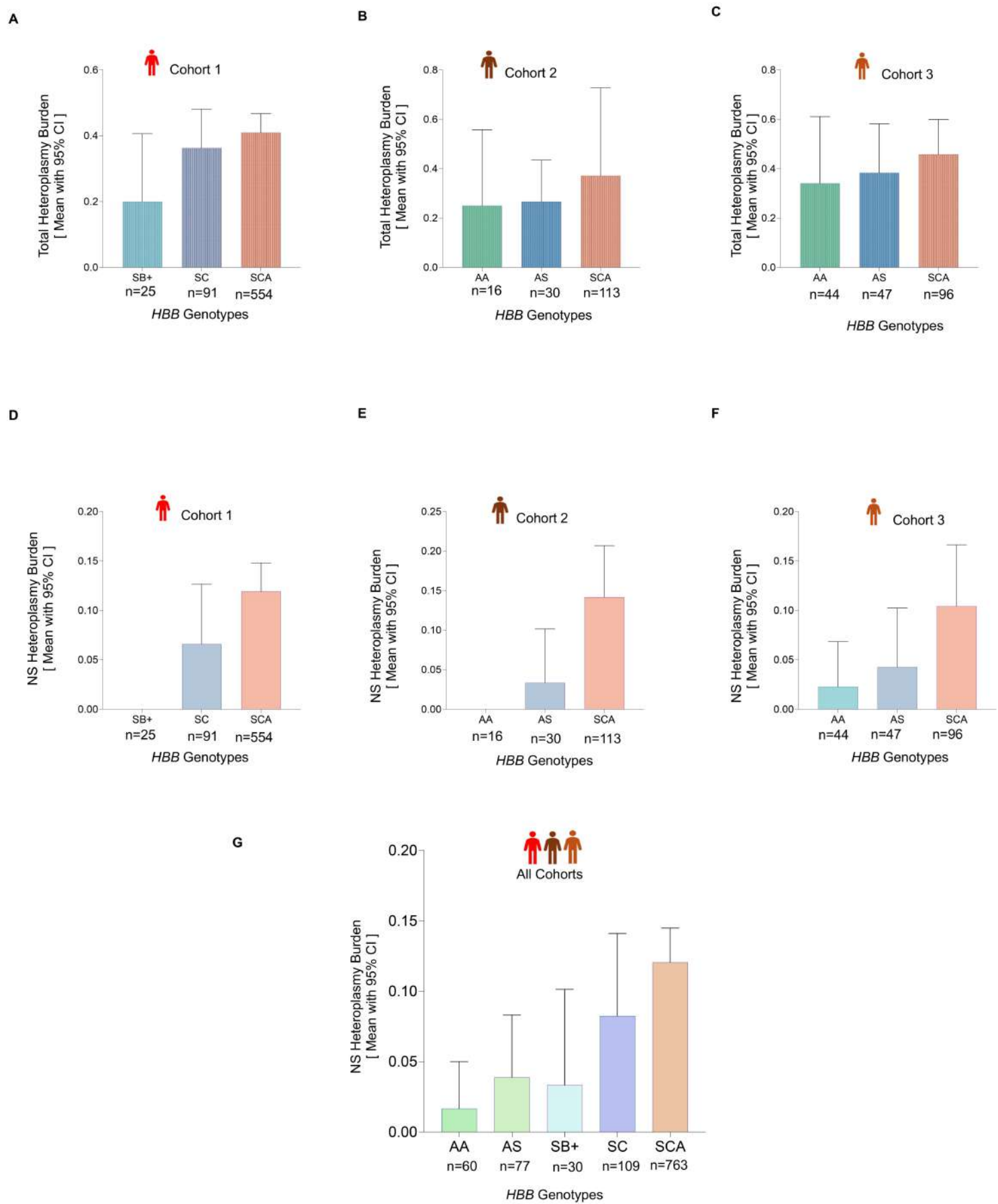
Ray, R. et al, Supplementary Figure S2

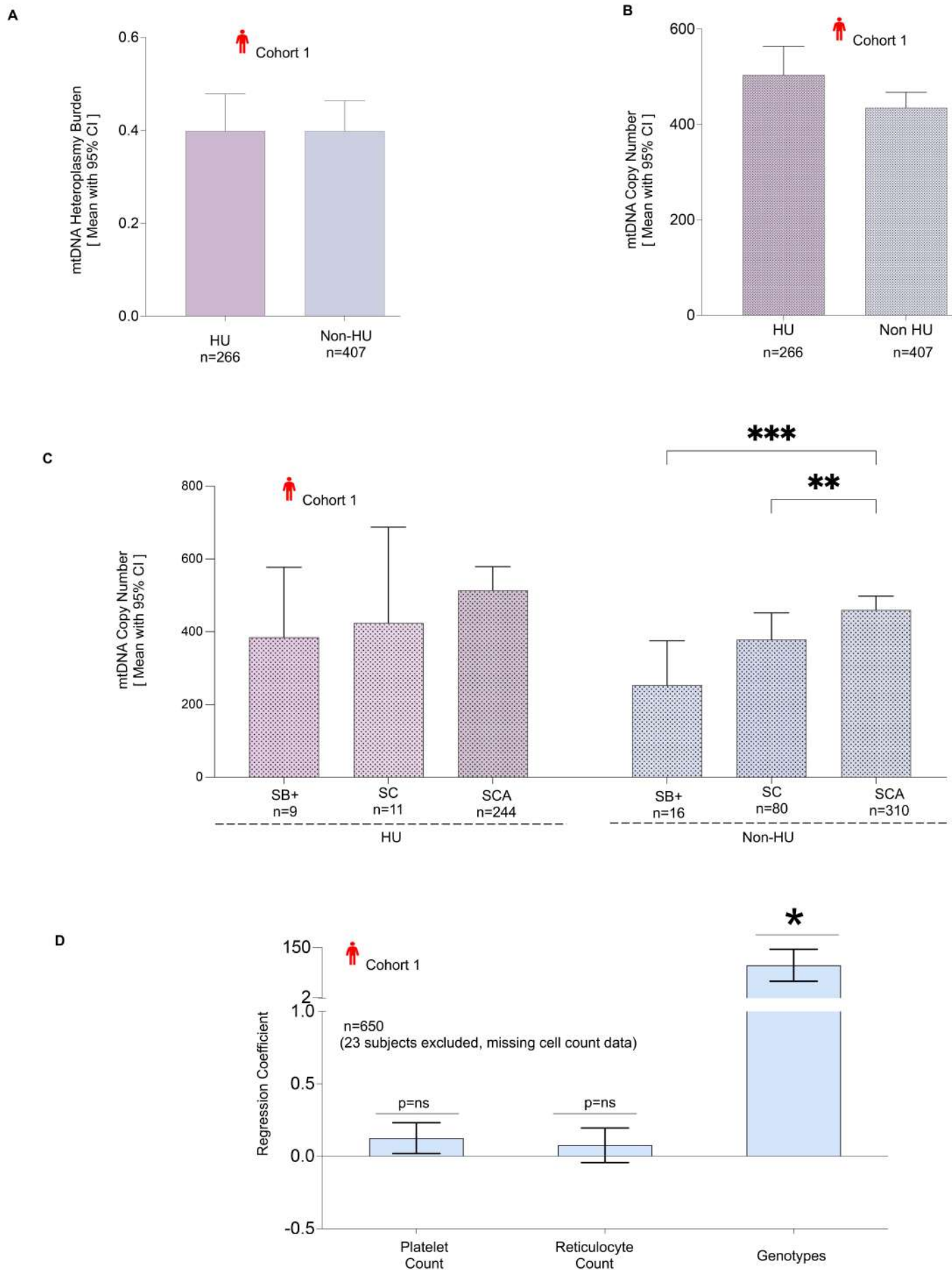


Ray, R. et al, Supplementary Figure S3

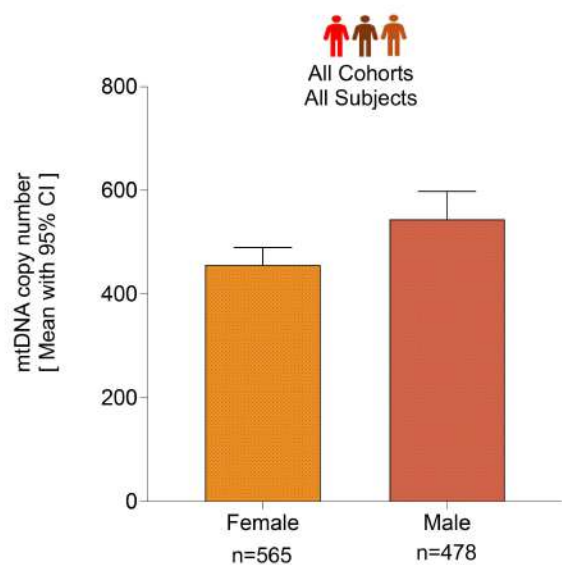


Ray, R. et al, Supplementary Figure S4

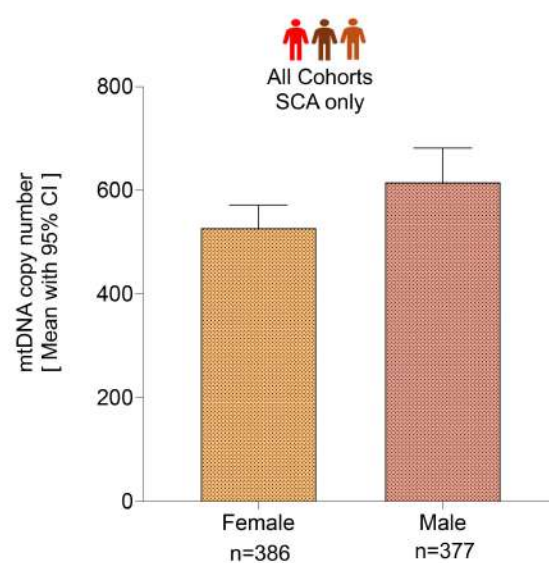




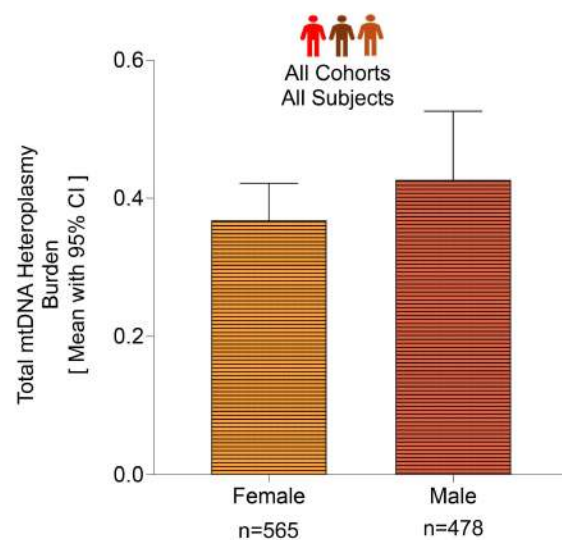
A



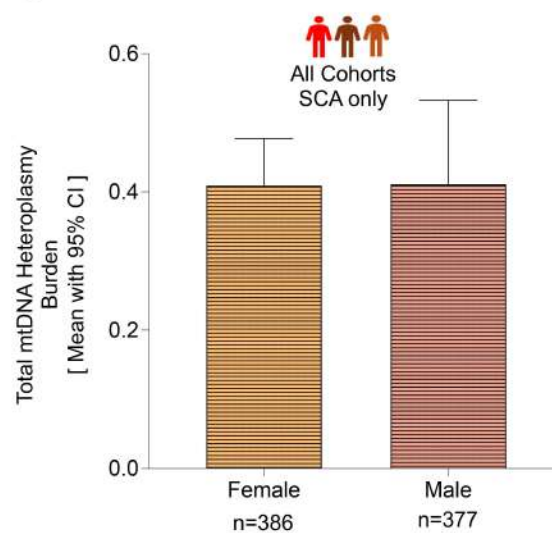
B



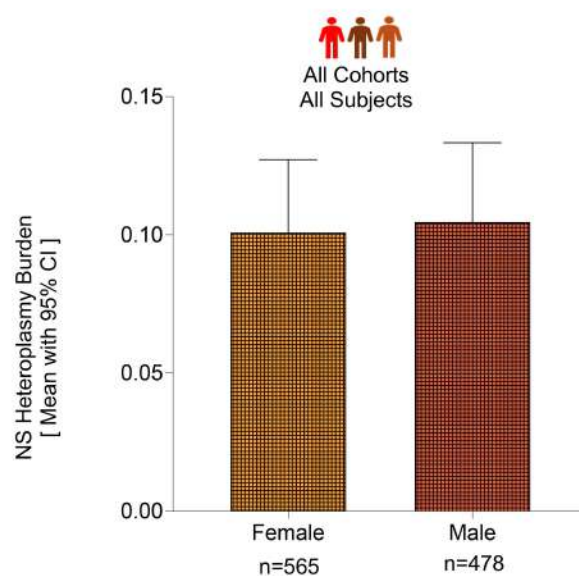
C



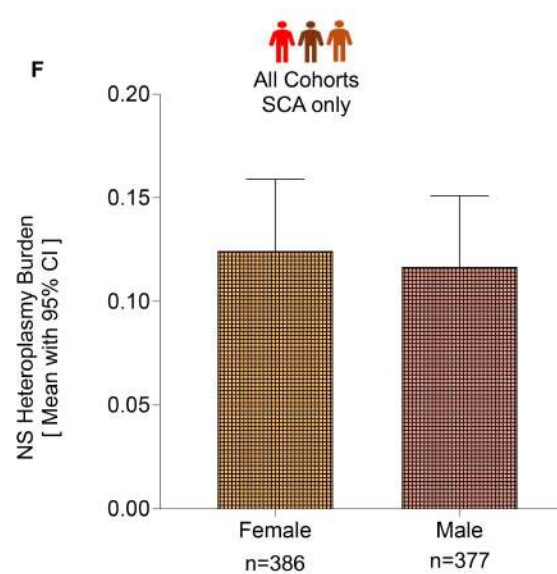
D

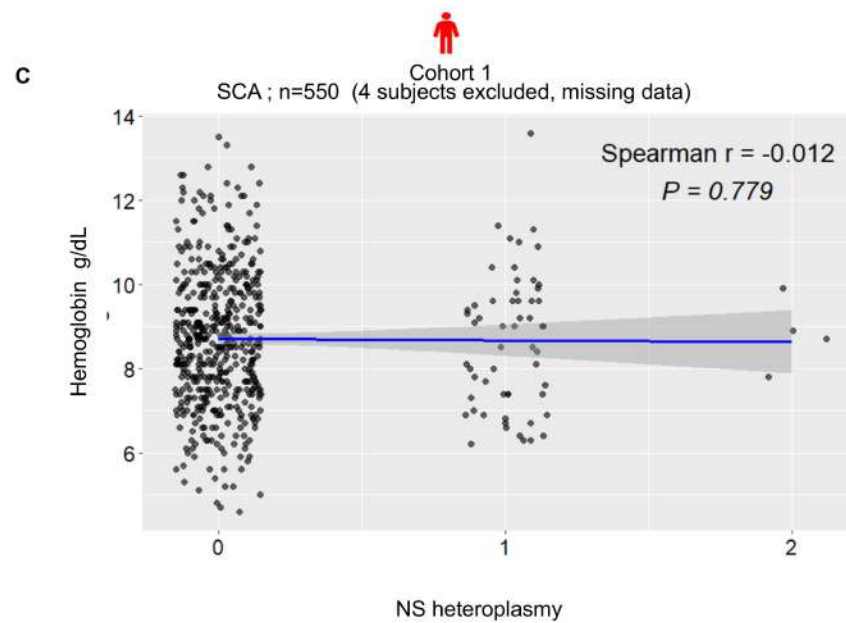
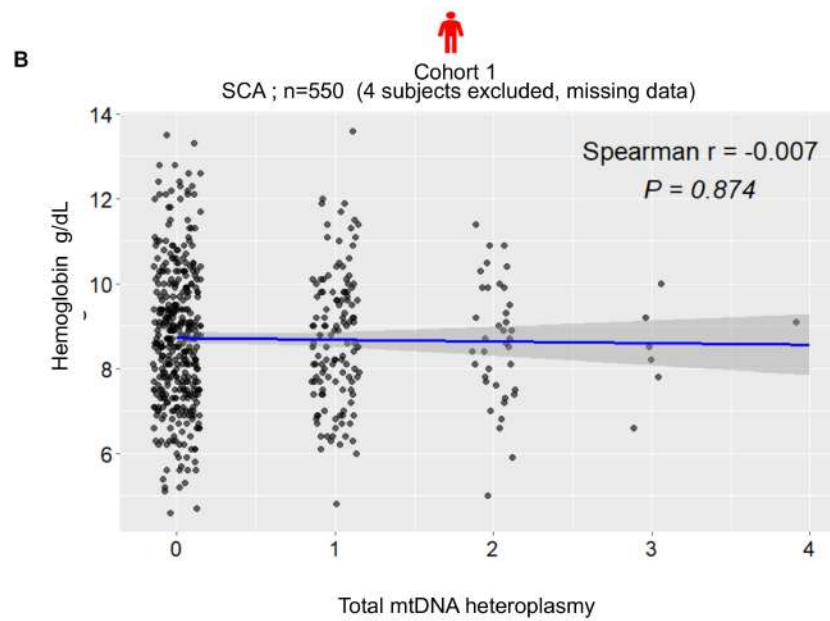
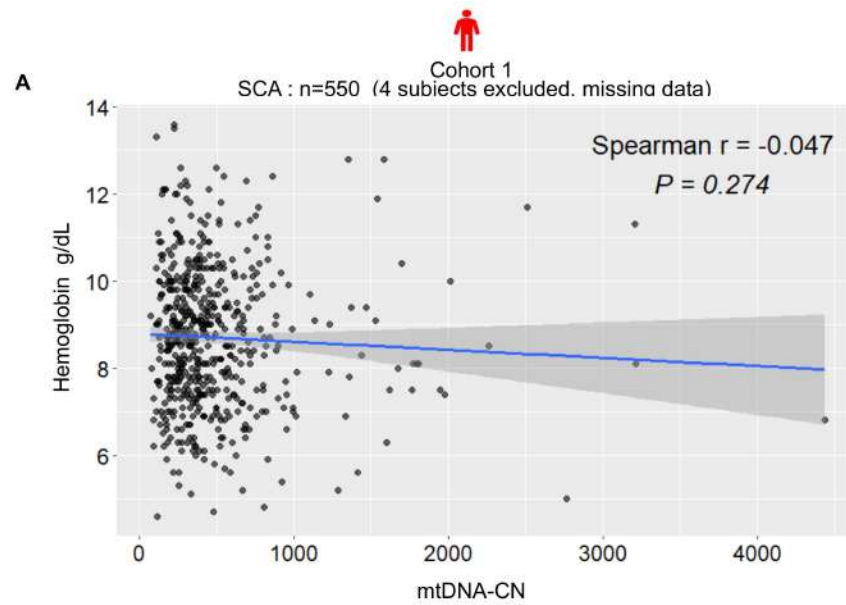


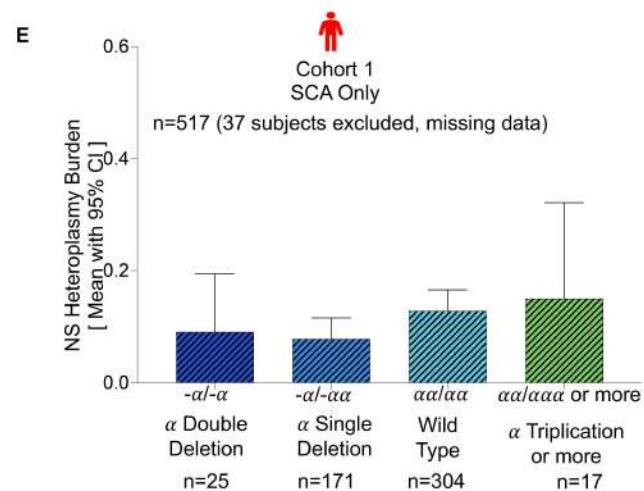
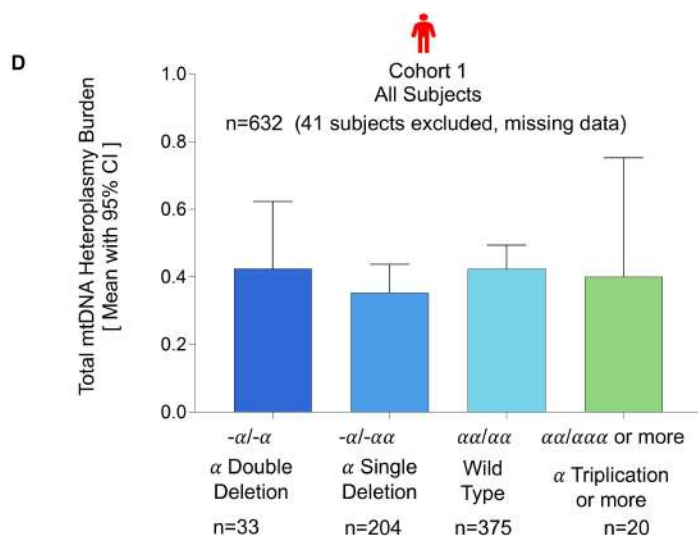
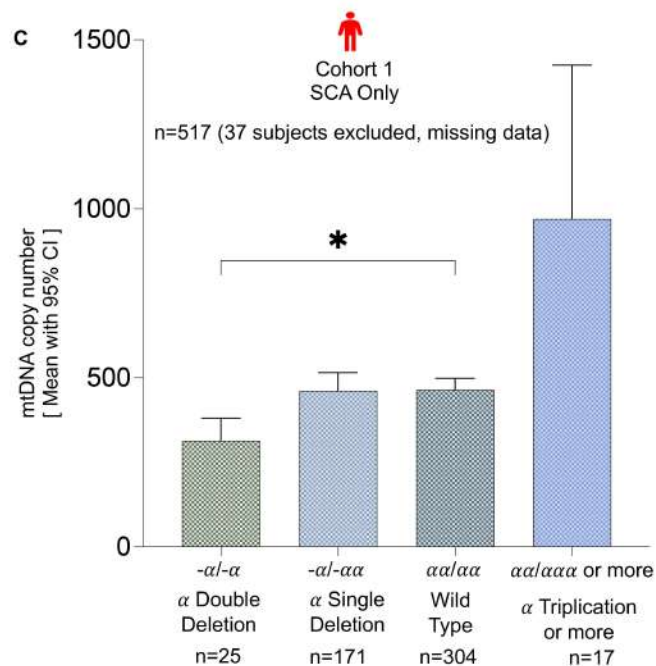
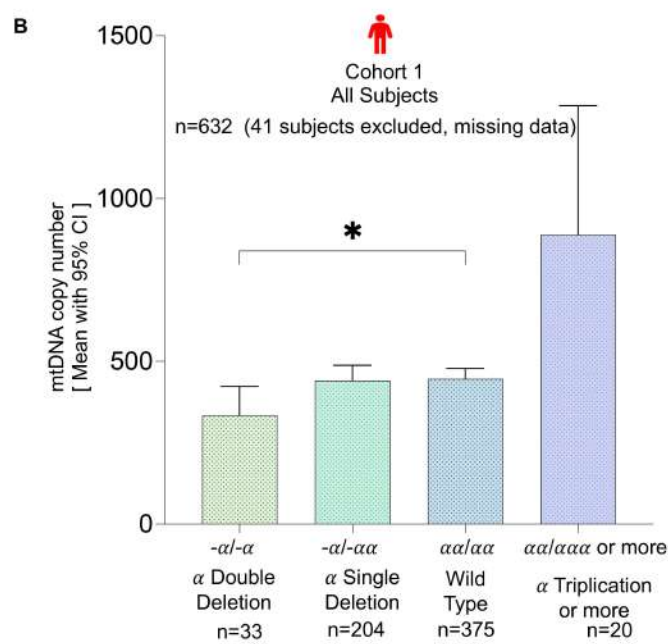
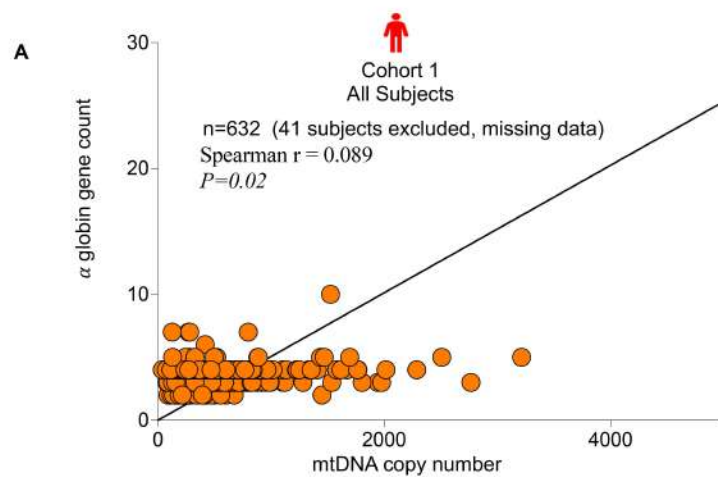
E

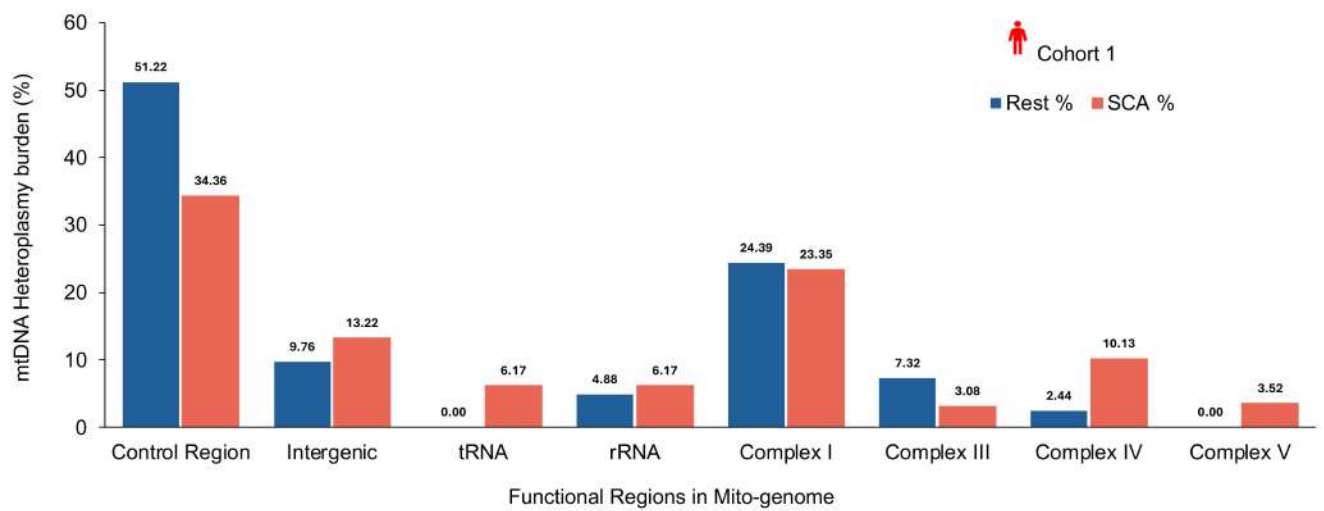


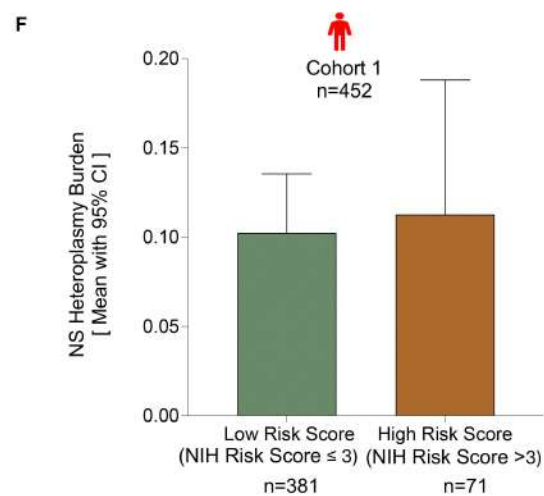
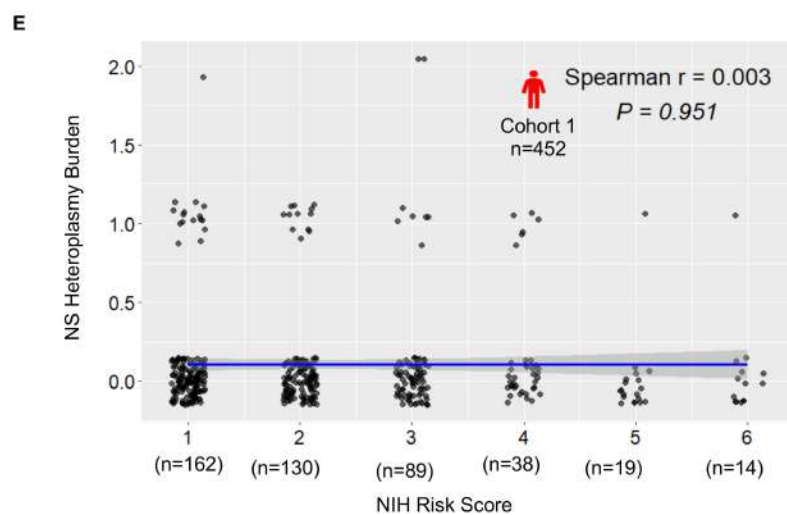
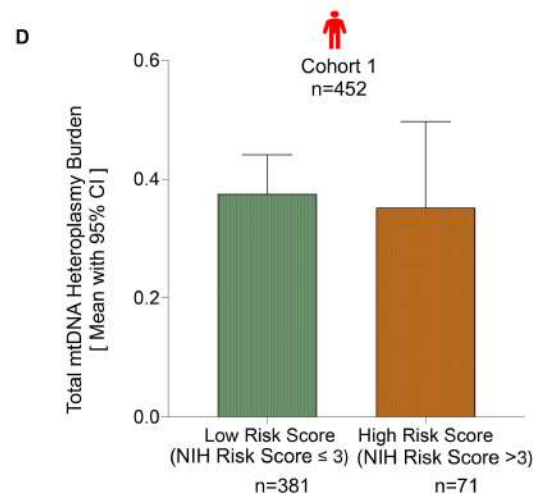
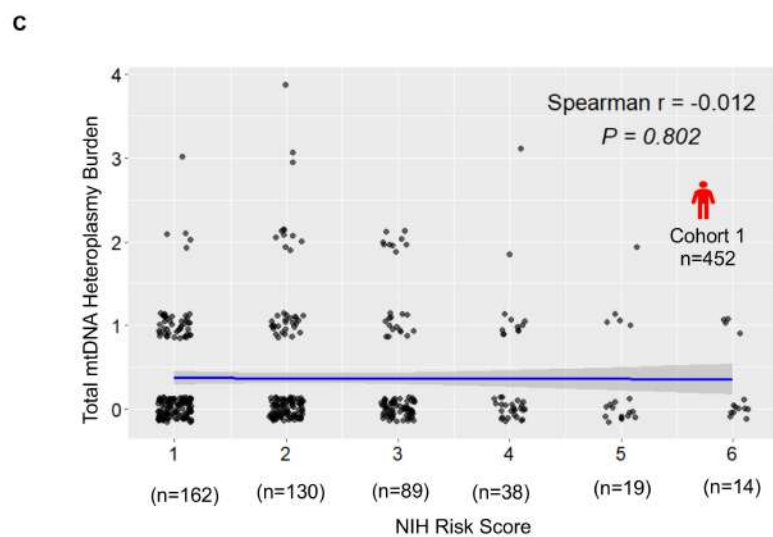
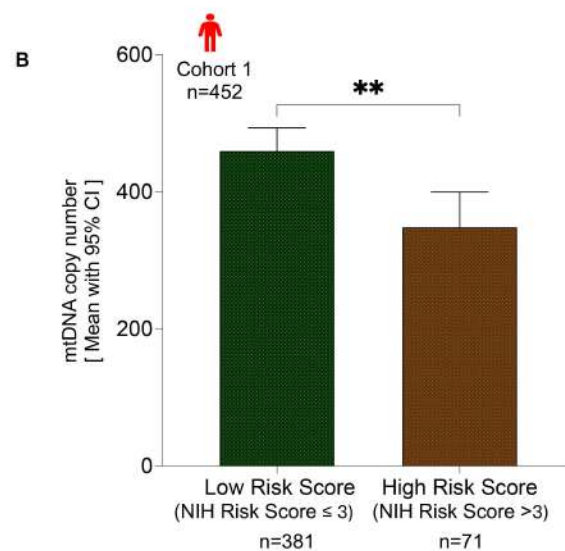
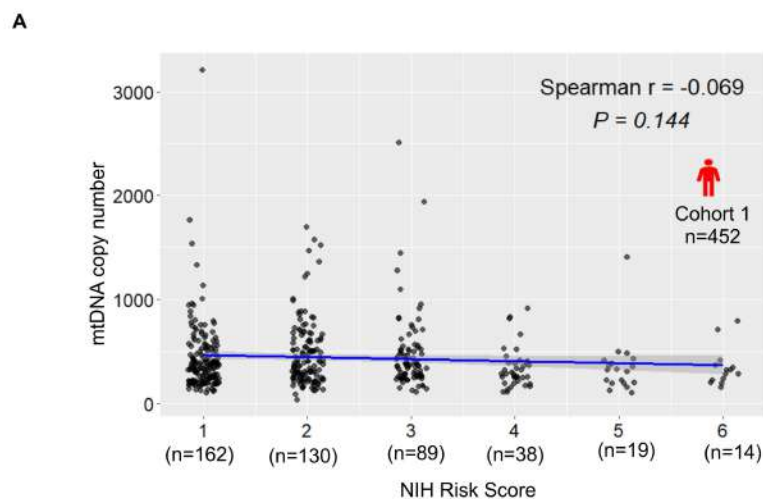
F

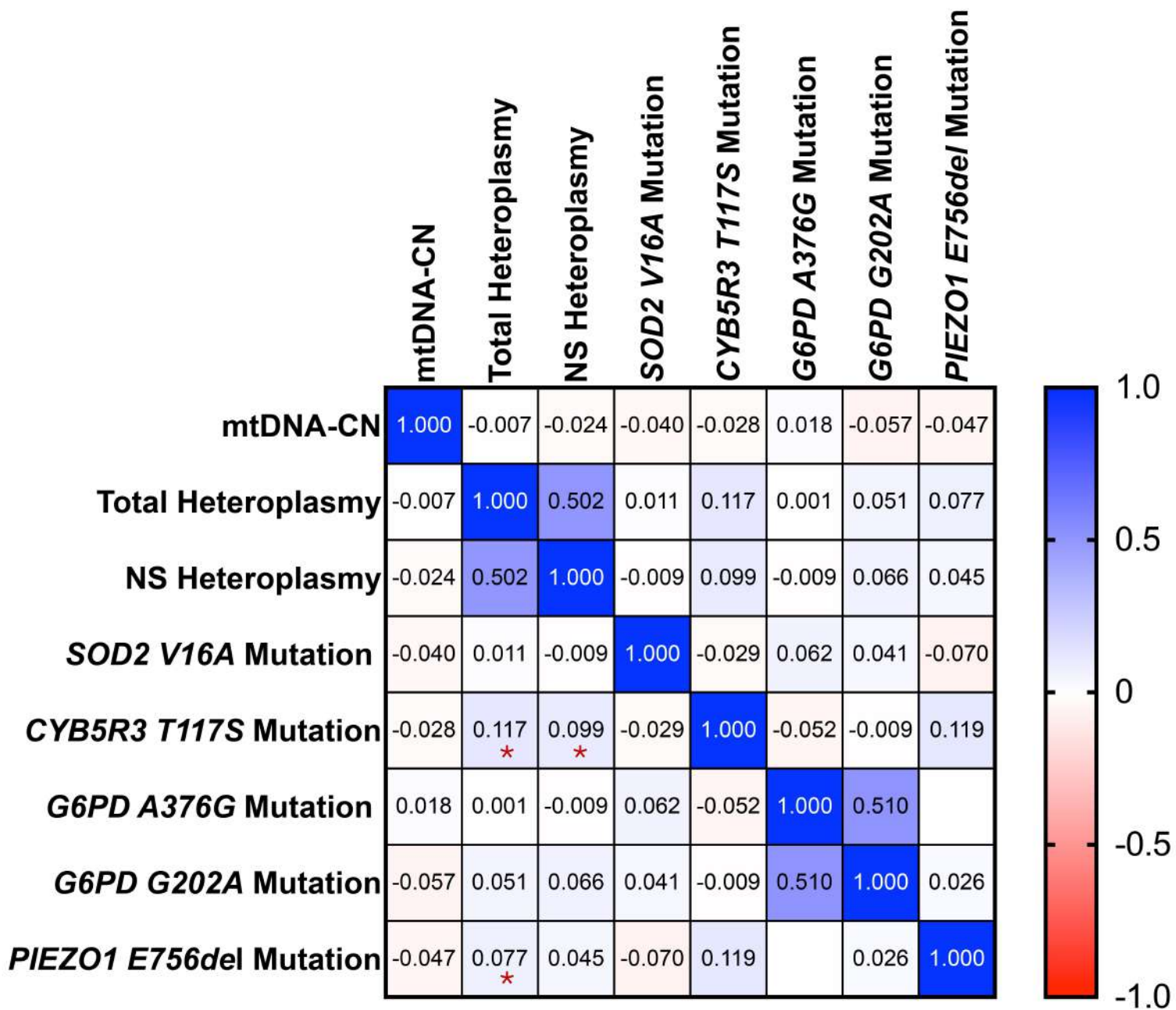


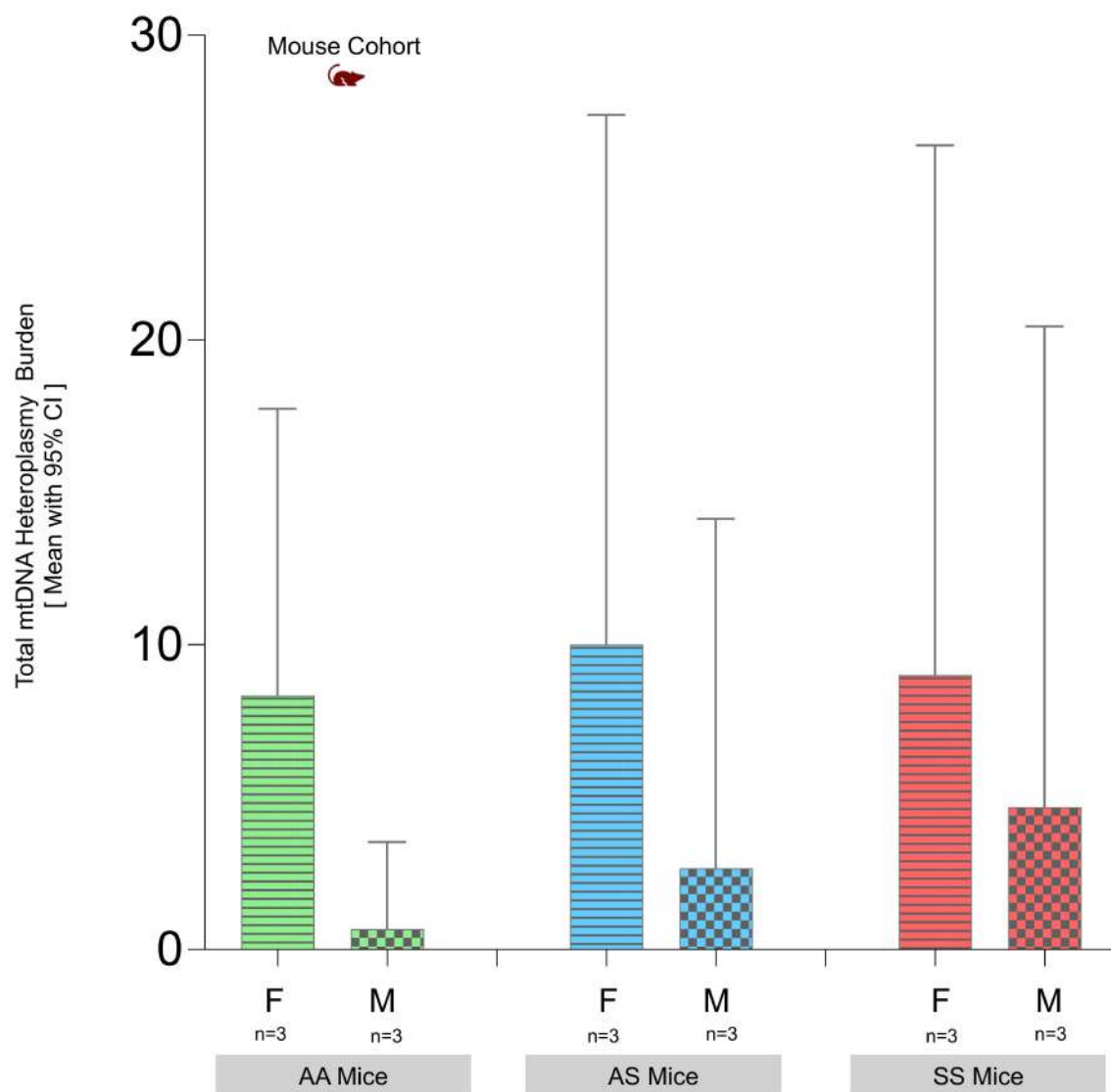






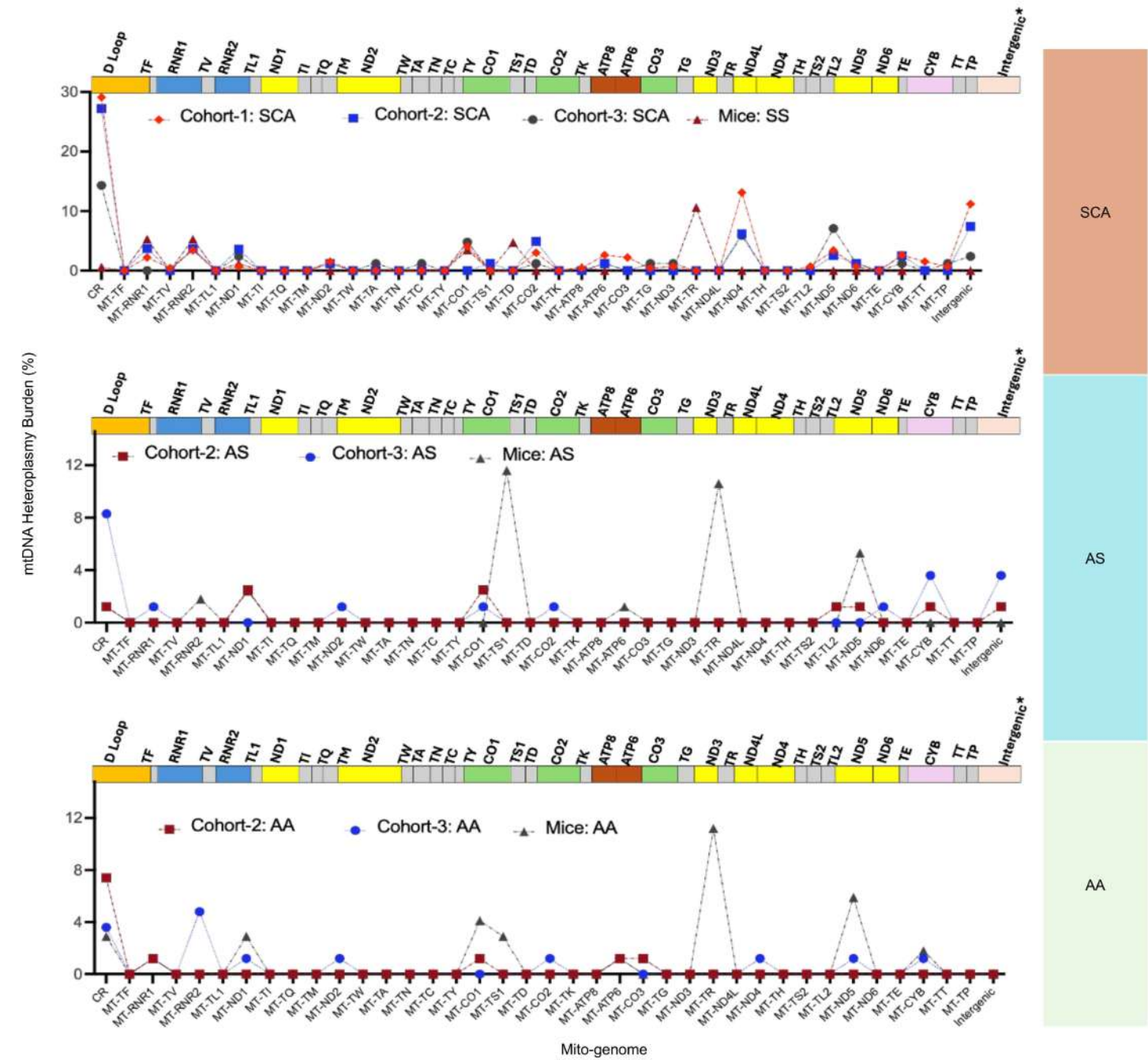






Ray, R. et al, Supplementary Figure S13

A



Ray, R. et al, Supplementary Figure S14A

B

