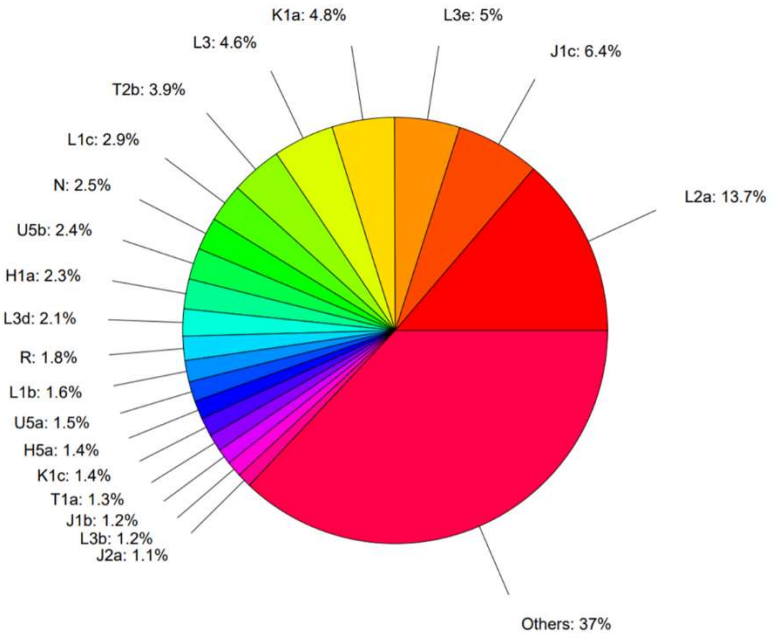
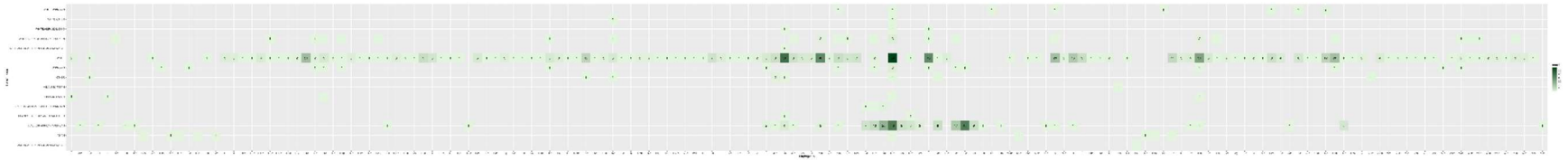


Haplogroup	Frequency	Percent
L2a	127	13.71
J1c	59	6.37
L3e	46	4.97
K1a	44	4.75
L3	43	4.64
T2b	36	3.89
L1c	27	2.92
N	23	2.48
U5b	22	2.38
H1a	21	2.27
L3d	19	2.05
R	17	1.84
L1b	15	1.62
U5a	14	1.51
H5a	13	1.40
K1c	13	1.40
T1a	12	1.30
J1b	11	1.19
L3b	11	1.19
J2a	10	1.08
Others	343	37.04
Total	926	100.00

Supp. Figure S1. Breakdown of the 929 preterm samples based on haplogroup. (A) 929 preterm samples were analyzed for their haplogroup by the automated pipeline, using the Mitomaster API (https://www.mitomap.org/mitomaster/websrvc.cgi). 3 of the samples failed the haplogroup analysis due to low quality and low mitochondrial genome coverage, and were excluded from further analysis. For the remaining 926 samples, the prevalence of each haplogroup was calculated and presented as shown. The top five haplogroups were as follows: L2a (13.7%), J1c (6.4%), L3e (5.0%), K1a (4.8%), and L3 (4.6%).



Supp. Figure S2. Distribution of all haplogroups in the preterm patient cohort by race. Heat map shows distribution of mtDNA haplogroups by self-identified race or ethnic group in the preterm patient cohort. The labels on the left axis represent the various self-identified ethnic or ancestry groupings, while the labels along the bottom axis represent all detected haplogroups in the preterm patient cohort. The numbers in black within each square indicate the number of samples where a particular mtDNA haplogroup was detected for a particular racial or ethnic category. Each square is also color-coded based on the number of samples with that particular combination of mtDNA haplogroup and self-identified racial category, based on the key on the right side of the heat map.

Sample ID	Variant	Heteroplasmy % Mutect2 (lcWGS)	Heteroplasmy % PCR-NGS	Difference in Heteroplasmy %
Sample 1	m.3243A>G	28	28	0
Sample 3	m.8344A>G	96	96	0
Sample 6	m.13513G>A	66	64	2
Sample 8	m.3243A>G	13	13	0
Sample 9	m.10197G>A	53	54	1
Sample 10	m.8344A>G	92	92	0
Sample 11	m.3243A>G	71	70	1
Sample 12	m.3243A>G	69	68	1
Sample 13	m.3243A>G	66	65	1

Supp. Table S2: **Validation of Mutect2 for the estimation of mtDNA heteroplasmy.** The automated pipeline for analyzing low-coverage whole genome sequencing (lcWGS) using Mutect2 was first tested against a set of nine patient samples previously tested using the gold-standard PCR-NGS method (see "Materials and Methods). After analyzing the PCR-NGS data using the Mutect2 pipeline, the results were found to be very well correlated, with a maximum difference of 2% for any given sample (average difference 0.67%, standard deviation 0.67%).

Sample Name	Mutation	Heteroplasmy % from LCWGS	Heteroplasmy % from PCR-NGS
Patient_116	m.1494C>T	85.1	99.2
Patient_152	m.7471dup	16.4	17.0
Patient_203	m.3243A>G	7.1	8.3
Patient_242	m.3243A>G	53	52.3
Patient_727	m.1555A>G	68.6	99.6
Patient_823	m.1555A>G	70.8	99.2
Patient_875	m.11778G>A	91.7	99.2
Patient_877	m.3243A>G	85.8	93.6

Supp. Table S3. Pathogenic variants identified by the automated lcWGS pipeline confirmed by PCR-NGS. DNA samples from the 8 patients with known pathogenic variants were resequenced using the PCR-NGS approach, which is the gold standard for mtDNA sequenching. The estimated heteroplasmy values obtained by PCR-NGS showed a high level of correlation with the results of the automated lcWGS pipeline, with an R-squared value of 0.9228 (**Figure 4**).