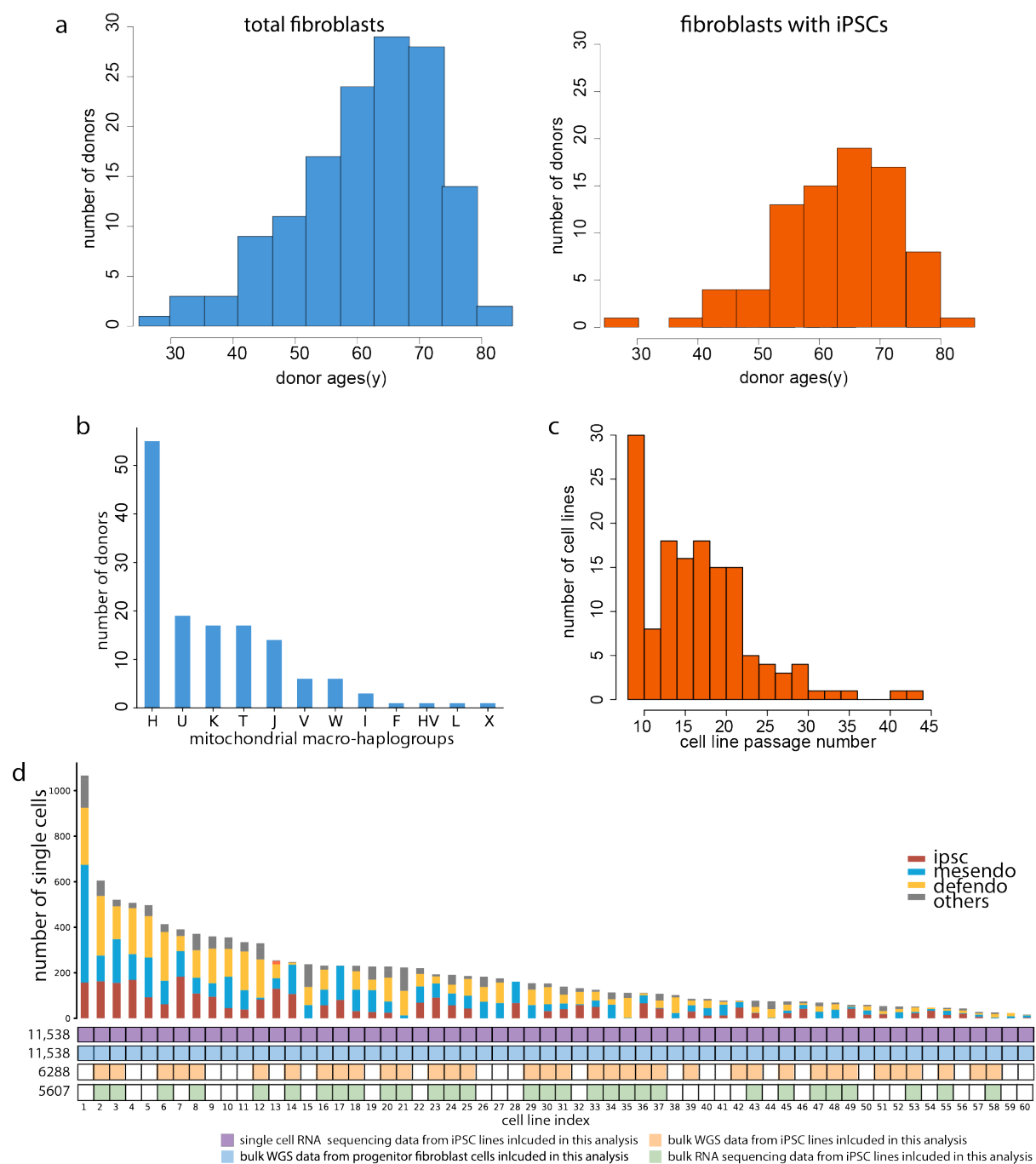


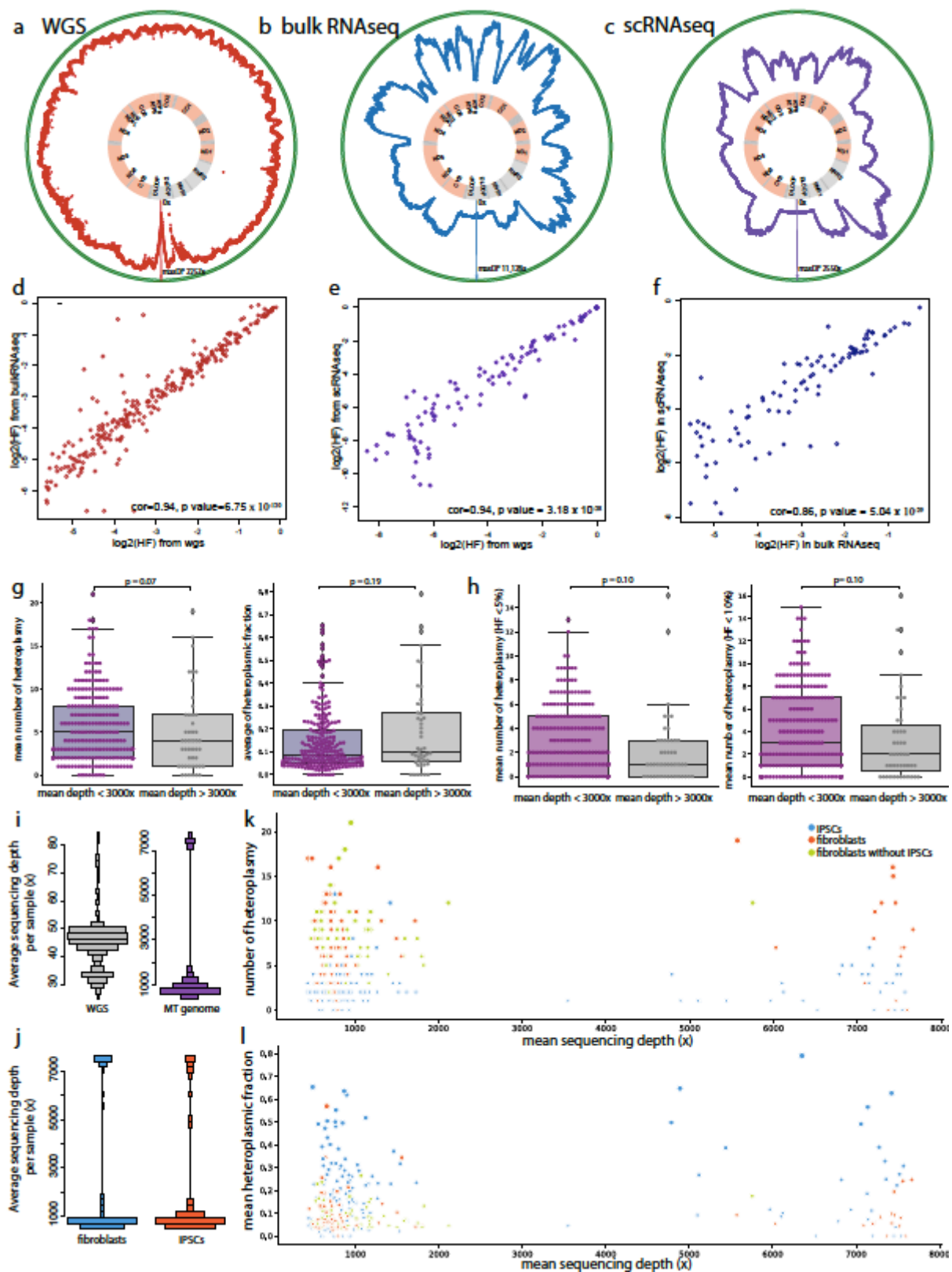
Supplementary Information

Cell reprogramming shapes the mitochondrial DNA landscape

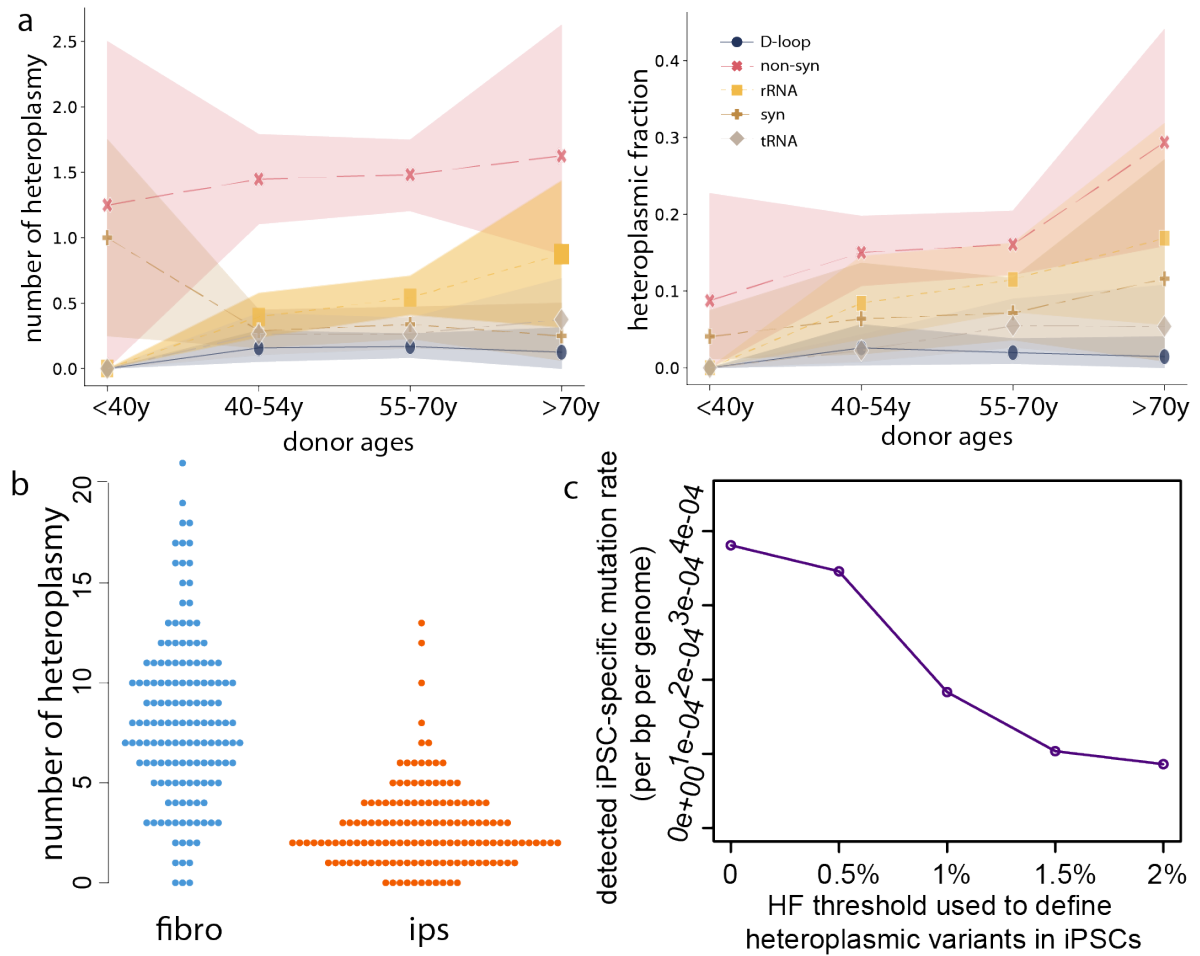
Wei Wei, Daniel J. Gaffney, Patrick F. Chinnery



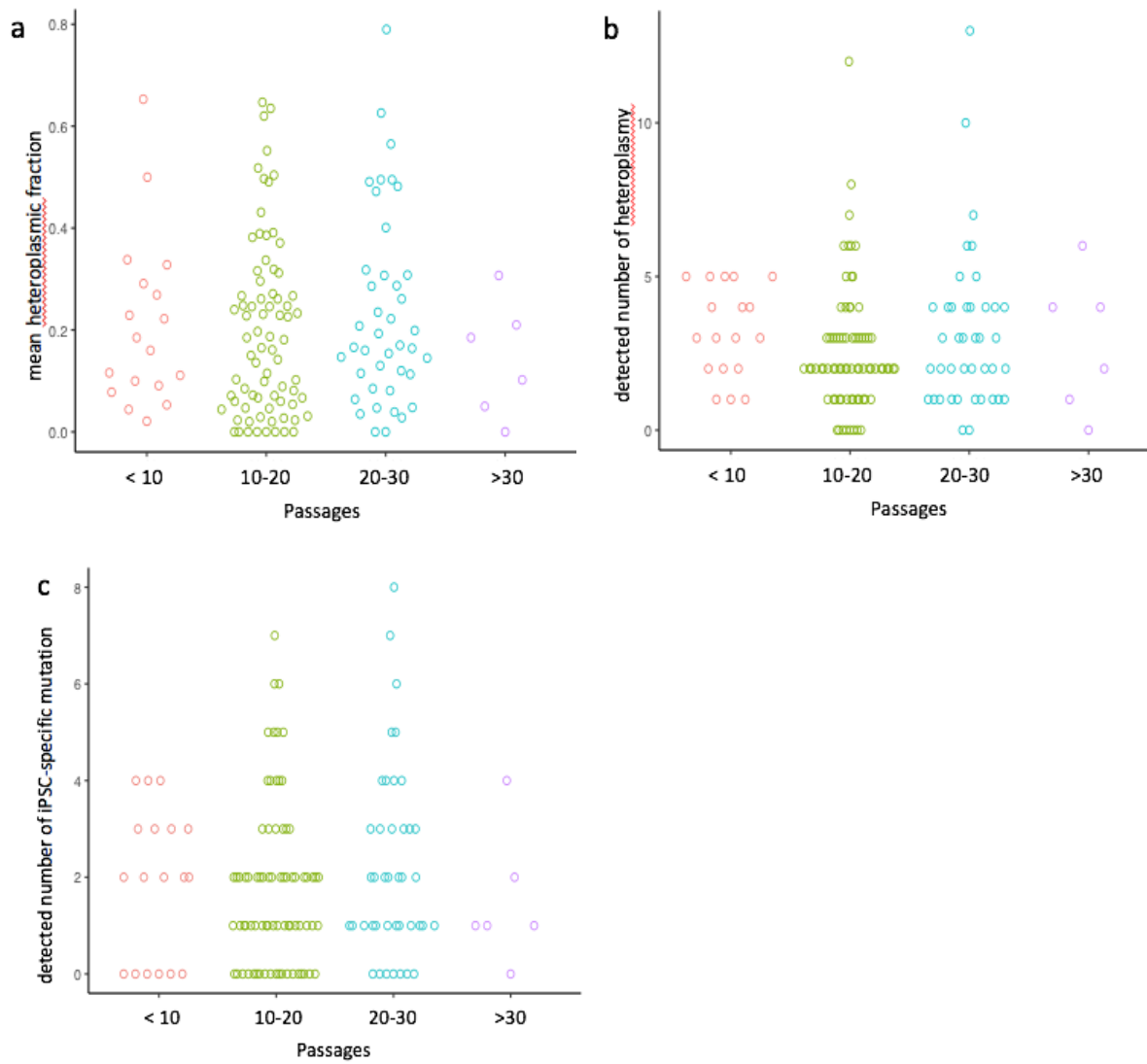
Supplementary Fig. 1 Summary of the cell lines included in this study. **a** Age distribution of the 146 donors contributing fibroblasts (left) and age distribution of the 83 donors contributing iPSCs (right). **b** mtDNA macro-haplogroups defined in the 146 donors. **c** Distribution of iPSC line passage number. **d** Top: the number of single cells from 60 donors analysed in this study. Each bar represents each donor, and cells from three cell stages iPSC, mesendo, defendo and undefined shown in different colours. Bottom: the types of sequencing data available in this analysis. Each square represents each cell line. The numbers of cells sequenced are shown on the left side.



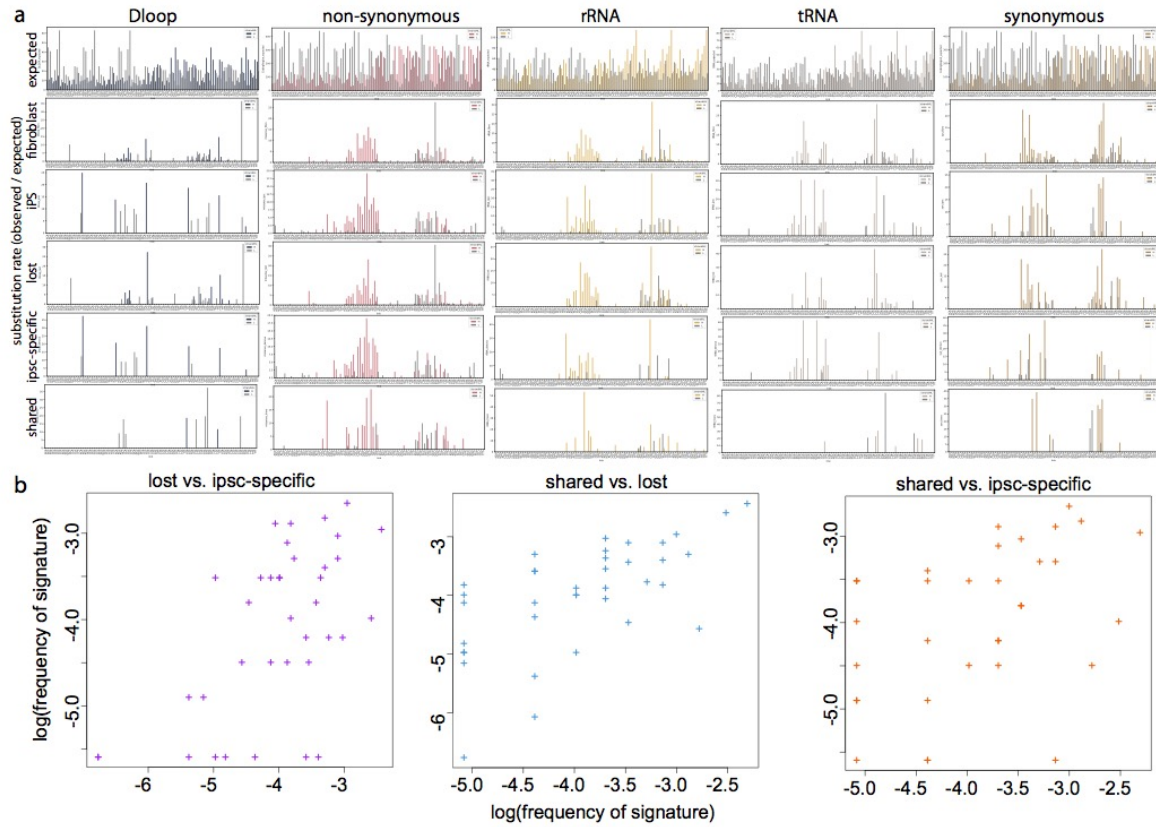
Supplementary Fig. 2 Coverage and depth of sequencing. **a-c** Average sequencing depth of the mitochondrial genome using different sequencing techniques. Inner circle, mitochondrial genome and regions; middle outline, average sequencing depth. From inner to outer, the arrows show the sequencing depth (DP) increasing. The max depth is labelled at the end of each arrow. **d-f** Scatter plot of $\log_2(\text{HF})$ in heteroplasmic variants from the same iPSC detected between WGS and bulk RNAseq (d), WGS and scRNAseq (e) and bulk RNAseq and scRNAseq (f). P values were calculated using Pearson's correlation test. Source data are provided as a Source Data file. **g** Combined box and swarm plots showing the mean number of heteroplasmic variants and average heteroplasmy fraction for cell lines sequenced at a low or high-depth. The middle “box” represents the median, lower and upper quartile of the data. The upper and lower whiskers represent the data outside the middle 50%. Each dot represents each mitochondrial genome. $n = 287$ mtDNA sequences. P values were calculated using two-sided Wilcoxon rank sum test. **h** Combined box and swarm plots showing the mean number of low level heteroplasmic variants (left plot $\text{HF} < 5\%$ and right plot $\text{HF} < 10\%$) for cell lines sequenced at a low or high-depth. The middle “box” represents the median, lower and upper quartile of the data. The upper and lower whiskers represent the data outside the middle 50%. Each dot represents each mitochondrial genome. $n = 287$ mtDNA sequences. P values were calculated using two-sided Wilcoxon rank sum test. **i** Left: Distribution of the mean whole-genome sequencing depth per sample. Right: Distribution of the mean mtDNA sequencing depth per sample (extracted from the whole genome). **j** Distribution of the mean mtDNA sequencing depth per fibroblast cell line (left) and iPSC line (right). **k** Correlation between the mean sequencing depth of the mtDNA genome with the mean number of heteroplasmic variants. **l** Correlation between the mean sequencing depth of the mtDNA genome and the average heteroplasmy fractions.



Supplementary Fig. 3 mtDNA variants detected by WGS. **a** Left: correlation between the mean number of heteroplasmic variants per iPSC line in each mtDNA region with the donors' age. Shaded regions show mean $\pm$ standard deviation. Right: correlation between the average of HF per iPSC line in each mtDNA region with the donors' age. Shaded regions show mean $\pm$ standard deviation. **b** Swarm plot shows the number of heteroplasmic variants detected in each fibroblast and iPSC cell line. $n = 287$ mtDNA sequences. Source data are provided as a Source Data file. **c** Detected iPSC-specific mutation rate (per base per genome) using different HF detection thresholds in the iPSC lines.



Supplementary Fig. 4 iPSC passages. **a** Mean heteroplasmic fraction in each iPSC passage group. $n = 141$ mtDNA sequences. **b** Detected mean number of heteroplasmic variant per iPSC line in each iPSC passage group. $n = 141$ mtDNA sequences. **c** Detected number of iPSC-specific mutation per iPSC line in each iPSC passage group. $n = 141$ mtDNA sequences.



Supplementary Fig. 5 Trinucleotide mutational signature of heteroplasmic variants. **a** Trinucleotide mutational signature of heteroplasmic variants in different mtDNA regions. The substitution rate of expected mutations, variants observed in fibroblasts / iPSCs, lost variants in fibroblasts, iPSC-specific variants in iPSCs, and shared variants between fibroblasts and iPSCs are shown separately. Mutations from the H- or L-strand are shown in different colours. **b** Scatter plot of the log(frequency of mutational signature) between lost and iPSC-specific heteroplasmic variants (left), shared and lost heteroplasmic variants (middle), and shared and iPSC-specific heteroplasmic variants (right).

Supplementary Fig. 6 a-c UMAP plots and hierarchical clustering of mtDNA mutation profiles from independent cell lines from single cell RNA sequencing. **a** UMAP plots of mtDNA mutation profiles with cells coloured by the heteroplasmy fraction of a specific mutation observed in a cell line. The mutations are labelled at the top of the UMAP plots. Hierarchical clustering of mtDNA genotyping (rows) for the single cells – coloured by their cell stages (columns). Colour bar, heteroplasmy fraction. **b** UMAP plots of mtDNA mutation profiles with cells from the same cell line coloured by cell stages. Three cell stages show the similar cluster profile. **c** 13327G and 1392G variants occurring in the same lineage were seen in the same 20% of the cells. **d** Left: Violin and box plots show the percentage of the variance for gene expression explained by cell stage, experiment, cell line, donor age and donor sex. Right: Bar plots show the number of gene with >10% >20% >30% and >40% variance explained by cell stage, experiment and cell line.