Extended Data Figures

Extended Data Fig 1. Clones for detection of mtDNA variants.

a, A scatter plot showing mean sequencing coverage (x-axis) and peak variant allele frequency (VAF) of somatic mutations in the nuclear genome (y-axis) for each clone. **b**, Histogram depicting the average coverage of mtDNA within a comprehensive dataset of 2,096 normal clones. **c**, Median depth per mtDNA locus across the entire set of 2,096 normal clones. The outer circle represents mtDNA genes, while the inner circle represents the median depth, with vertical axes specifying values (2,000, 4,500, and 7,000). **d-e**, Detailed illustrations of the filtering process on a position-by-position basis. Heteroplasmy levels for each mutation in normal clones are aligned in descending order. The yellow bar denotes the heteroplasmy level of the called mutation in specific clones, while the surrounding gray bars denote heteroplasmy levels of the same mutation in other clones. **d**, A representative example demonstrating a true-positive call (m.1,197 G>A). **e**, A representative example illustrating a false-positive call (m.11,009 T>C). False-positive calls were subsequently excluded from further analyses. **f**, The experimental design for estimating the rate of mtDNA mutations during cell culture, as previously reported[6]. Briefly, a single crypt isolated from the surgical specimen (naturally clonal) was cultured for 27 days for whole-genome sequencing of clones. Subsequently, a second clonalization was conducted after an additional average culture of 46.5 days (ranging from 43 to 50 days). Among the 13 pairs of whole-genome sequences obtained from early and late clones, eight new mtDNA mutations were exclusively identified in the late clones, implying acquisition prior to the second clonalization (approximately ~73.5 days). This allowed us to calculate the culture-associated mtDNA variant rate: 8 mtDNA variants / 13 clones / 73.5 days = 0.008 per clone per day. Using this rate, we can estimate the upper boundary of the culture-associated mtDNA variant, which amounts to 0.008 per clone per day * 27 days = 0.216 variants, assuming the recent common ancestral cell (MRCA) expansion at day 27. Of note, the lower boundary remains at 0 if the expansion of the MRCA occurred at day 0. Consequently, it is inferred that the maximum proportion of the culture-associated mtDNA variants is 7%, calculated as 0.216/3.1*100 (where 3.1 denotes the average number of mtDNA variants per clone). **g**, Comparison between VAFs of early and late clones of the same variant. The gray line connects the same variant of two clones.

Extended Data Fig 2. Classification of Het_{FE} variants.

a, A schematic diagram showing the classification of shared mtDNA variants into four categories, with the respective variant counts in parentheses. **b**, A Venn diagram illustrating the number of Het_{FE} variants within each criterion, providing robust evidence for the classification of Het_{FE} variants. **c,** Bar plots showing the region preference of Het_{FE} variants compared with sporadic mutations. The x-axis indicates genomic regions, including D-loop, coding, rRNA, and tRNA regions. The y-axis indicates log2-transformed fold changes calculated by the ratio of Het_{FE} variant prevalence and sporadic mutation prevalence. **d**, A schematic diagram illustrating the calculation of caVAF using VAFs of each clone within an individual. Yellow circles represent individual normal clones, and the gray box contains the VAFs of the Het_{FE} variant in each clone. **e**, Comparison between the average number of Het_{FE} variants identified within clones and the average number of *de novo* mutations in bulk blood of family dataset. Blue bar plots represent the weighted average of Het_{FE} variants, while red diamonds represent the average of *de novo* mutations in bulk blood. Error bars show the 95% confidence interval, and the average variant counts are calculated for each specified range.

Extended Data Fig 3. Features of mtDNA hotspot mutations.

a, A scatter plot illustrating the association between the number of individuals with mtDNA hotspot mutations and their respective numbers of clones where the mutations were detected. The x-axis represents the number of individuals a mutation was detected, while the y-axis indicates the number of clones in which the same mutation was found. Mutations recurrently observed in 20 or more clones are specifically labeled. **b**, Illustrations of the mtDNA hotspot mutation, m.414 T>G, were recurrently detected in individuals' clones. Early clonal phylogenies of DB10 and DB9 are reconstructed using somatic mutations in the nuclear genome, with branch lengths proportional to the number of somatic mutations. VAFs of each clone are depicted in the bar plot at the bottom. **c**, Bar chart depicting the correlation between the proportion of clones carrying hotspot mutations and different age groups. Each color-coded bar corresponds to a specific frequently recurrent mutation.

Extended Data Fig 4. VAF distribution of Het_{FE} variants.

a, Histograms display the VAFs of Het_{FE} variants observed across different tissue types, representing each tissue type by color bars at the top. **b**, A heatmap representing the VAFs of Het_{FE} variants in each clone of the abortus. The columns correspond to individual clones, ordered according to their phylogenetic relationships. VAFs for each clone are indicated within

the cells of the heatmap.

Extended Data Fig 5. Mutational profiles of mtDNA sporadic mutations.

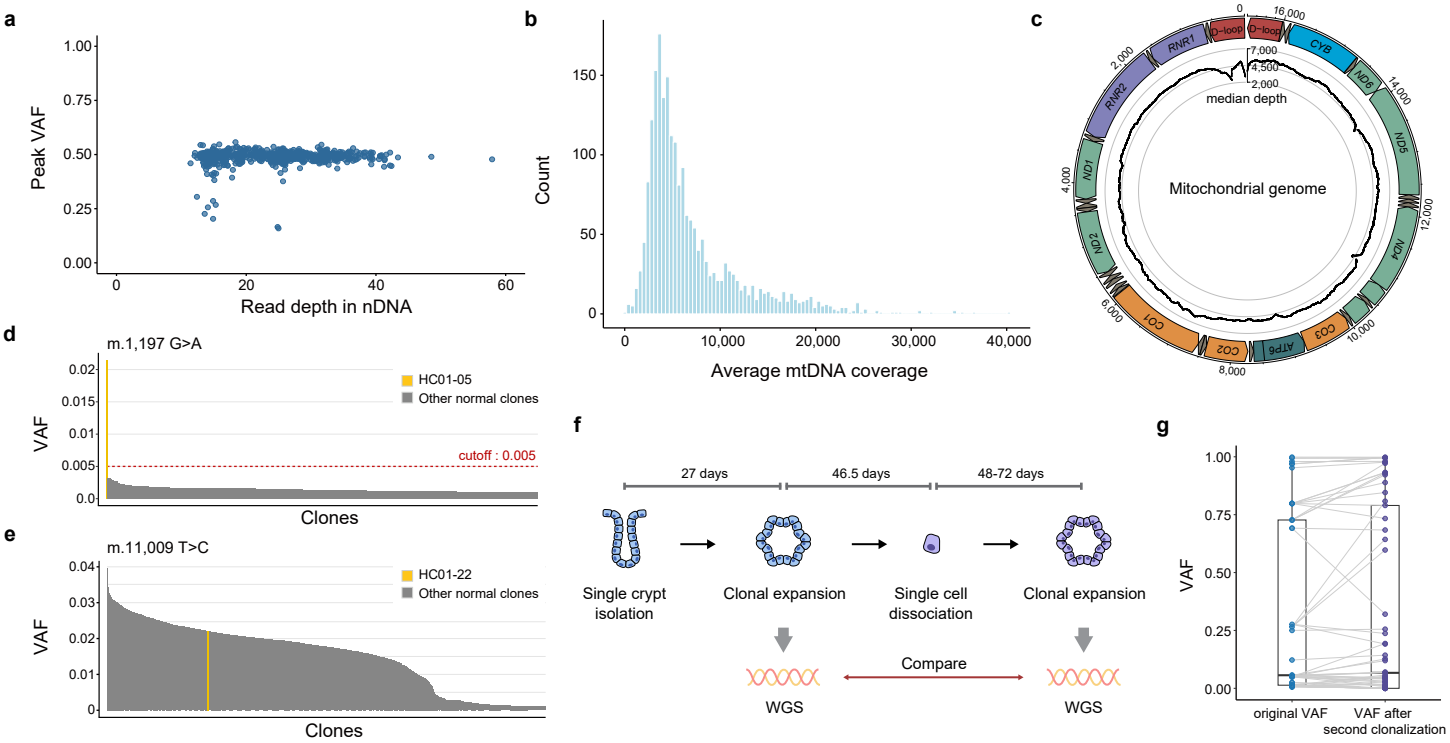
a, Linear correlation between the average mtDNA sporadic mutation count per individual and estimated turnover across 31 individuals. Estimated turnovers are calculated depending on tissue types for each individual. The Pearson's correlation coefficient and p-value from Pearson's correlation test are provided. The gray line represents the regression line, and the shaded area indicates its 95% confidence interval. Vertical lines crossing each dot indicate the range of the mtDNA mutation count in clones of the individual. Clones with UV-derived mutation burden below 1,500 are selectively included to mitigate the influence of UV radiation. **b**, Linear correlation between the average cumulative sum of VAFs per individual (S_{VAF}) and age across 31 individuals. The Pearson's correlation coefficient and p-value from Pearson's correlation test are provided. The gray line represents the regression line, and the shaded area indicates its 95% confidence interval. Vertical lines crossing each dot indicate the range of S_{VAF}. Clones with UV-derived mutation burden below 1,500 are selectively included to mitigate the influence of UV radiation. **c**, Histograms display the heteroplasmy levels of somatic mutations observed across different tissue types, representing each tissue type by color bars at the top. **d**, Spectrum of mtDNA sporadic mutations of clones with high and low UV exposure. The degree of UV exposure is determined using the SBS7 mutation counts in the nuclear genome. Six types of mutations are color-coded. Darker colors correspond to mutation types on the heavy strand, while lighter colors correspond to mutation types on the light strand.

Extended Data Fig 6. Distributions of truncating mutations and impact of mtDNA copy number on the transcriptome.

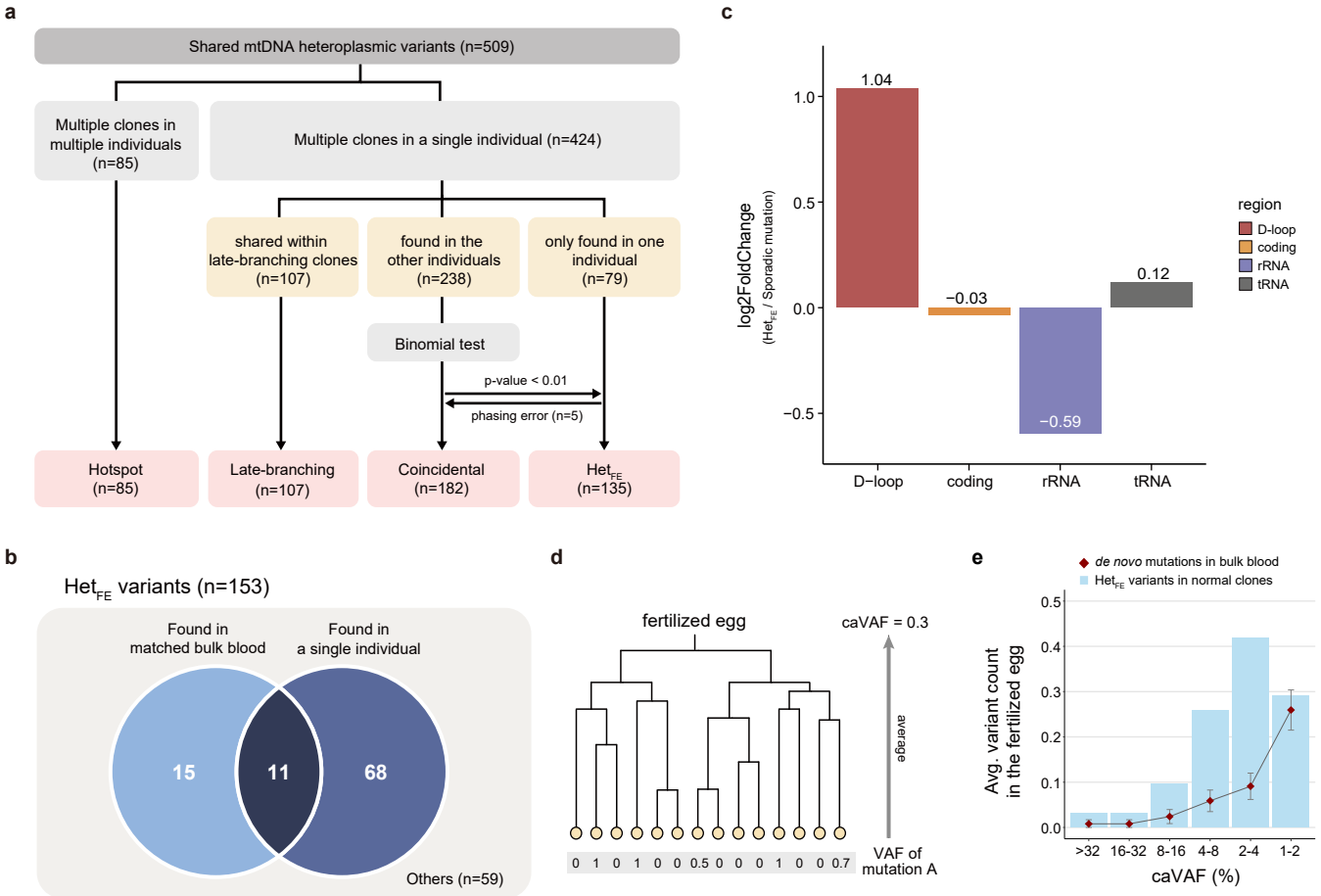
a, A total count of truncating mutations in mtDNA genes with VAFs exceeding 0.6. These values are normalized by dividing them by the length of the gene. Three different tissue types are color-coded. **b**, Violin plots showing the VAFs of truncating mutations per mtDNA genes. Each circle represents an individual truncating mutation, with colors indicating tissue types. **c,** A volcano plot showing differentially expressed genes identified between clones with high and low mtDNA copy number in HC17. The x-axis represents log2-transformed fold changes, and the y-axis represents -log10-transformed p-values. The gray dots indicate no significant difference, and the green dot indicates significantly downregulated genes in clones with high

mtDNA copy numbers. **d,** A dot plot representing mtDNA copy numbers of clones in HC17. Clones with high and low copy numbers are distinguished based on 1,200, indicated by a gray dashed line. **e-f**, IGV screenshots showing two mtDNA structural variations (HC06-14 and HC21-16). Notable changes in read depth and instances of discordant paired-end reads are visualized. **g**, Violin plots illustrating the VAFs of truncating mutations per mtDNA genes in colorectal cancers. Each circle represents an individual truncating mutation. **h**, Linear correlation plots demonstrating the relationships between diverse factors, including age, sporadic mutation count, S_{VAF}, and maximum VAF, and differences in mtDNA copy number (mtCN) between colorectal epithelium normal clones and matched colorectal cancers. Differences in mtCN are calculated by dividing the mtCN of colorectal cancer by the average mtCN of normal clones for each individual. Each data point corresponds to the mtCN differences of an individual. The Pearson's correlation coefficient and p-value from Pearson's correlation test are provided. S_{VAF}, the sum of the VAFs of all detected sporadic mtDNA mutations in a clone.

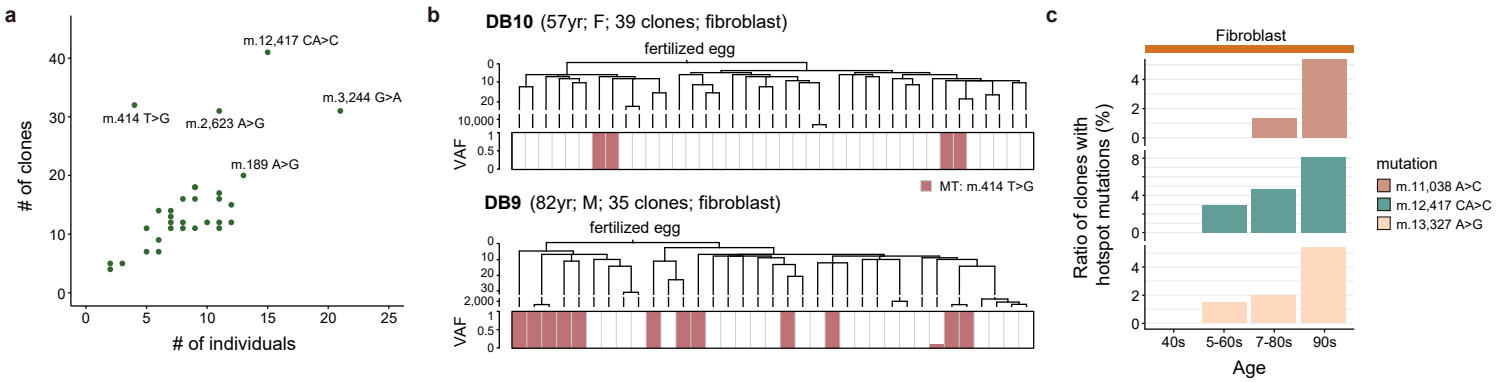
Extended Data Fig 1



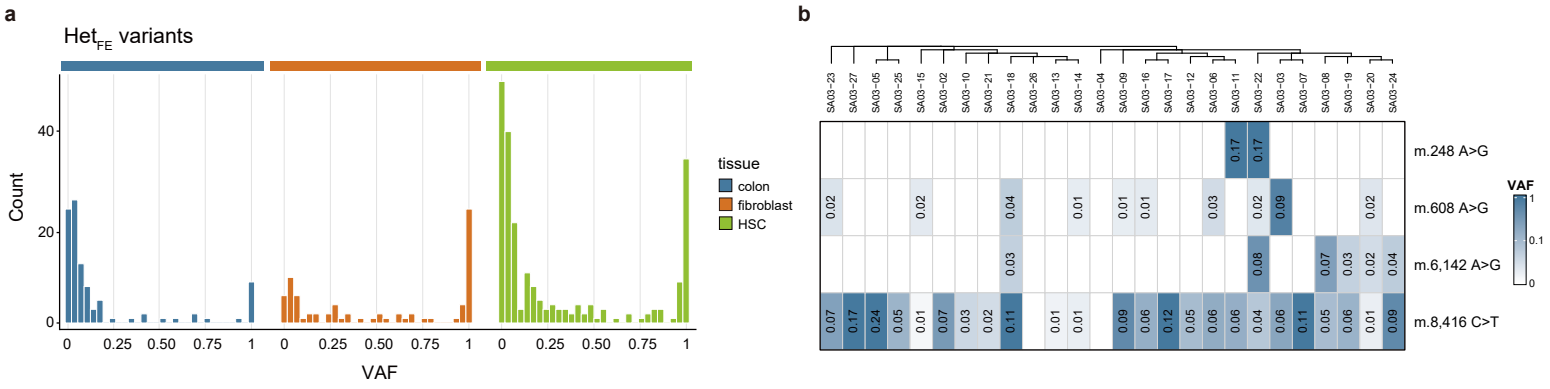
Extended Data Fig 2



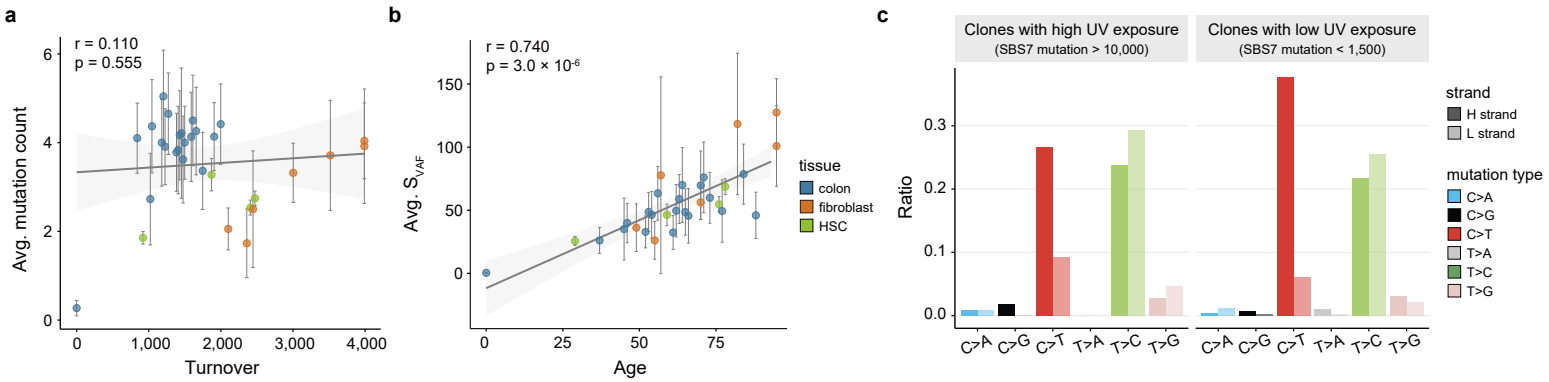
Extended Data Fig 3



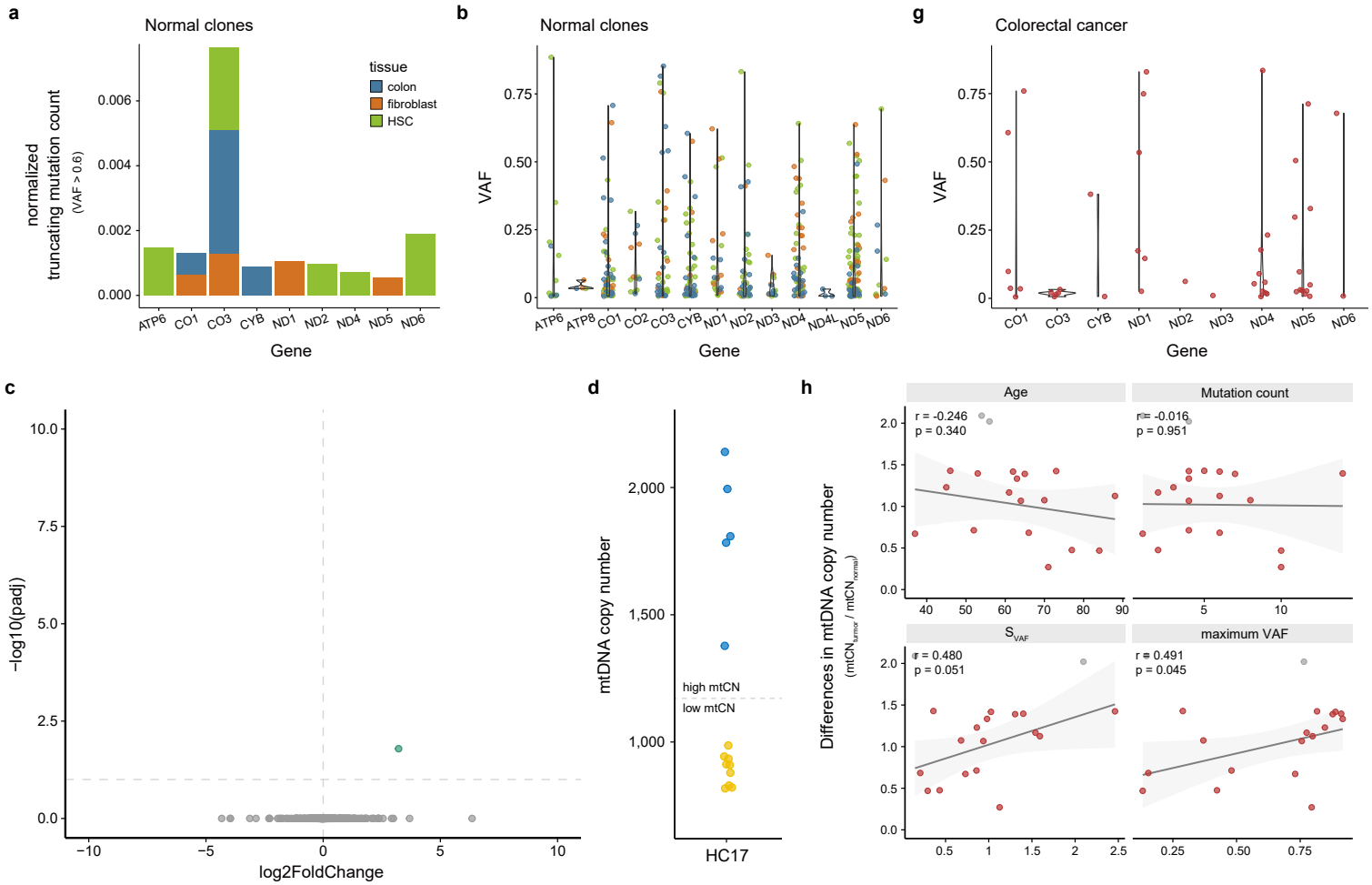
Extended Data Fig 4

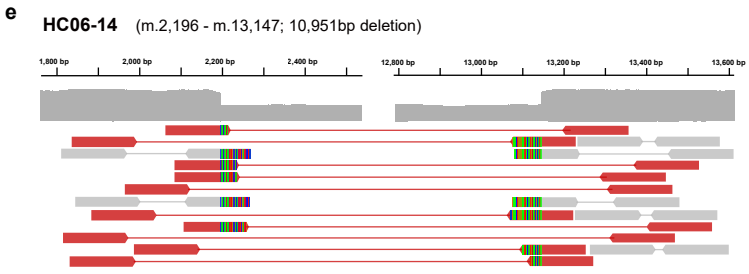


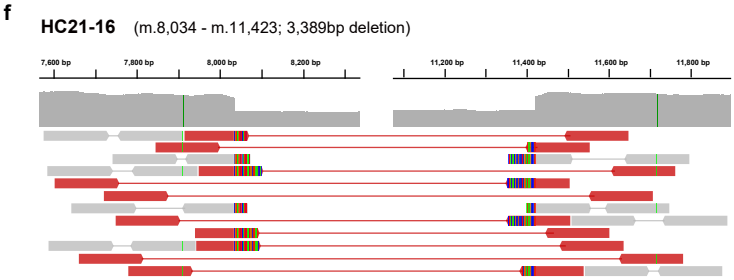
Extended Data Fig 5



Extended Data Fig 6







Supplementary Tables

Supplementary Table 1. Summarized WGS information of individuals.

Supplementary Table 2. Demographic and mutational characteristics of normal clones.

Supplementary Table 3. Background noise matrix of mtDNA variants.

Supplementary Table 4. Heteroplasmic mtDNA alterations identified in this study.

Supplementary Table 5. Summary of mtDNA alteration classification.

Supplementary Table 6. Fertilized egg originated variants classified in this study.

Supplementary Table 7. Hotspot mutations classified in this study.

Supplementary Table 8. Expansion of mtDNA population in simulations.

Supplementary Table 9. Expansion of mtDNA variant in simulations.

Supplementary Table 10. Demographic and mutational characteristics of adenomas and tumors.