Published in final edited form as:

J Genet Genomics. 2013 December 20; 40(12): 607–615. doi:10.1016/j.jgg.2013.10.003.

Very Low-Level Heteroplasmy mtDNA Variations Are Inherited in Humans

Yan Guo^{1,*}, Chung-I Li², Quanhu Sheng¹, Jeanette F. Winther³, Qiuyin Cai⁴, John D. Boice Jr.⁵, and Yu Shyr^{1,*}

¹Vanderbilt Ingram Cancer Center, Center for Quantitative Sciences, Nashville, TN, 37232, USA

²Department of Applied Mathematics, National Chiayi University, Chiayi, 60004, Taiwan

³Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, DK-2100, Denmark

⁴Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, 37232, USA

⁵National Council on Radiation Protection & Measurements, Bethesda, MD, 20814, USA

Abstract

Little is known about the inheritance of very low heteroplasmy mitochondria DNA variations. Even with the development of new next-generation sequencing methods, the practical lower limit of measured heteroplasmy is still about 1% due to the inherent noise level of the sequencing. In this study, we sequenced the mitochondrial genome of 44 individuals using Illumina high-throughput sequencing technology and obtained high-coverage mitochondria sequencing data. Our study population contains many mother-offspring pairs. This unique study design allows us to bypass the usual heteroplasmy limitation by analyzing the correlation of mutation levels at each position in the mtDNA sequence between maternally related pairs and non-related pairs. The study showed that very low heteroplasmy variants, down to almost 0.1%, are inherited maternally and that this inheritance begins to decrease at about 0.5%, corresponding to a bottleneck of about 200 mtDNA.

Keywords

Maternal inheritance; Next-generation sequencing; High-depth sequencing; Heteroplasmy; mtDNA mutations; Bottleneck

INTRODUCTION

Typically, there are over 1,000 mitochondria in each mammalian cell, and each mitochondrion harbors 2–10 copies of mitochondria DNA (Robin and Wong, 1988). Thus, mtDNA mutations are often heteroplasmic, with a mixture of mutant and wild-type mtDNA copies within a cell (Durbin et al., 2010; Ng et al., 2010). Researchers have found that

heteroplasmy could promote tumor growth (Lewis et al., 2000; Park et al., 2009) and is closely associated with aging (Kann et al., 1998; Smigrodzki and Khan, 2005; Sondheimer et al., 2011) although both points are somewhat controversial. Due to limitations in the available sequencing technology, earlier studies of mitochondrial mutations (Holt et al., 1990; Lombes et al., 1992; Simon et al., 2003; Wong et al., 2002) had limited ability to detect low-level heteroplasmy. Low-level mtDNA heteroplasmies (less than 5% mutant) are not readily detected by the Human Mitochondrial Resequencing Array 2.0. Real-time polymerase chain reaction (RT-PCR) techniques may detect heteroplasmies down to near the 1% mutation level, but each measurement must be designed for just one specific heteroplasmic site, making this method unsuitable for the discovery of heteroplasmic mutations across the entire mitochondrial genome. In recent years, high-throughput sequencing technologies have matured in step with reduced costs. The high-coverage generated by the new high-throughput sequencing methods provides a powerful tool for the study of mtDNA heteroplasmy (Guo et al., 2012a; Tang and Huang, 2010).

Previous studies on heteroplasmy using high-depth sequencing usually chose a heteroplasmy threshold from 1% to 2% (Goto et al., 2011; Guo et al., 2012a; Ng et al., 2010) or even as high as 10% (Durbin et al., 2010). At very low heteroplasmy levels, it becomes impossible to distinguish between sequencing errors and the true mtDNA sequence due to the baseline error rates of high-throughput sequencing. For this reason, low-level heteroplasmy has not yet been studied widely. However, low-level mtDNA heteroplasmy is of fundamental importance since every *de novo* mtDNA mutation must start out at very low levels (Elson et al., 2001). The question of whether low-level mtDNA mutations survive the mitochondrial bottleneck between mother and child is an important key to understanding how mtDNA variations, including pathogenic variations, become established within families. Typically, studies of families with known pathogenic mtDNA mutations showed that the mutation levels can have large shifts between mother and offspring (Chinnery et al., 2000). However, these studies have been limited to mutations at levels > 1%, and the theory of random drift inheritance (Wonnapijit et al., 2008) predicts that the variance in the offspring mutation levels will decrease rapidly as the maternal mutation level decreases. This decrease in variance for very low-level heteroplasmic mutations may make them more stable across generations. For all of these reasons, we tested for the inheritance of very low-level mtDNA mutations. In the current study, using high-throughput sequencing technology, we showed that heteroplasmic variations at mutation levels as low as 0.15% are inherited in humans.

RESULTS

Heteroplasmy in Maternal Pair

High-quality sequencing data with a median depth for all samples of $3981 \times$ were generated. Homoplasmic SNPs were inherited without exception within each family, and all maternally related individuals within the same family shared the same detailed sub-haplogroup (Table 1), validating both the sequencing and the maternal relatedness of the subjects. In addition to the homoplasmic SNPs, we also found 12 different mutations that were heteroplasmic at mutation level of >1% in both mother and offspring (Table 2). In two families with multiple offspring, the heteroplasmic mutation was found in all offspring (Table 2). As has been

observed in families transmitting pathogenic mtDNA mutations (Chinnery et al., 2000), the heteroplasmic levels in the mother and offspring were different, with a minimum change of <1% and a maximum change of over 31%. Of the 15 heteroplasmic mutation transmissions observed, eight had lower heteroplasmy levels in the offspring while seven had higher heteroplasmy levels, consistent with neutral drift.

Heteroplasmy Inheritance Analysis between Related and Unrelated Pairs

We now turn to an analysis of the inheritance of low heteroplasmy-level mtDNA variants. The observation of inheritance of specific mutations is limited by the noise level of the sequencing method, which can be taken as approximately 1% for base-calling Phred scores of 20 or better. For each pair of samples, we calculated a correlation of the mutation levels at each site. Instead of analyzing the inheritance of heteroplasmy at specific sites, which may be compromised by sequencing errors, we analyzed the distribution of these correlation coefficients. We placed a series of upperlimit filters on the heteroplasmy levels used in the correlation (see Materials and Methods for details) in order to specifically test for the inheritance of low-level heteroplasmic sites. The overall depth of coverage and heteroplasmy information for all 44 samples can be found in Table S1. Table S1 contains depth and heteroplasmy information for all samples and all positions in rCRS. After filtering for quality, we observed, on average, 6000 instances of low-level heteroplasmy (<1% and >0.1%) per sample. At such low levels, it is difficult to distinguish between true signal and heteroplasmy.

Our set of 35 maternally related pairs consisted of 26 mother-child pairs and 9 sibling pairs. We first tested to see if there was a difference in correlation between mother-offspring pairs and sibling pairs. The correlations for mother-child pairs were not significantly different from those for sibling pairs (Fig. 1) at all heteroplasmy levels except arguably at the 100% level which only had a marginally significant *P*-value of 0.09 (Table S2). Since all lower thresholds on the mtDNA heteroplasmy level consistently showed no significant difference between mother-offspring and sibling pairs, we combined the mother-child pairs and sibling pairs as maternally-related pairs for the remainder of the analysis.

The heteroplasmy levels at each site in the mitochondrial genome were compared for each pair of individuals in the study (see Materials and Methods for details). We calculated correlation coefficients for the heteroplasmy levels at all sites that passed quality control criteria in both individuals of each pair. As expected, the Pearson's correlation coefficients of the related pairs were significantly higher than those of the unrelated pairs (Fig. 2). To our surprise, even the non-maternally related pairs had a median correlation coefficient greater than zero. This may be due to non-random sequencing errors that escaped our quality control criteria, or it could be due to low levels of true mtDNA mutations at specific mutation hotspots (Stoneking, 2000) that are common across even unrelated individuals, though the latter possibility is speculative. Since even the unrelated pairs had a positive median correlation coefficient, we used comparisons between maternally related pairs and unrelated pairs as a test for inheritance. When examining all positions without restricting the heteroplasmy level, the median Pearson's correlation coefficient in all related pairs was

0.716, while the median Pearson's correlation coefficient of all unrelated pairs was significantly lower at 0.129 (P -value = 8×10^{-18} , Fig. 2, left side).

In order to determine whether we could detect the inheritance of low-level heteroplasmy, we applied a series of decreasing upper thresholds on the heteroplasmy level (see Materials and Methods for details). As we lowered the threshold of heteroplasmy level from $< 100\%$ to $< 0.15\%$, the median Pearson's correlation in unrelated pairs stayed at approximately the same level near zero, while the median correlation for related pairs decreased gradually, eventually reaching zero at a heteroplasmy level of $< 0.1\%$. Even at the heteroplasmy level $< 0.15\%$, we were still able to observe a positive mean correlation in heteroplasmy in the related pairs (median $r = 0.154$), and the median correlation in the related pairs was still stronger at a highly significant level ($P = 7 \times 10^{-5}$) than in unrelated pairs ($r = 0.053$). We also counted the number of heteroplasmic sites between related and unrelated pairs (Fig. S1). Similar numbers of shared heteroplasmic sites between related and unrelated pairs of all heteroplasmy level cutoffs were observed. This suggests that many heteroplasmic sites could result from low-level noise and sequencing error. It is very difficult to distinguish between true heteroplasmy signal and error. Only by conducting correlation analysis were we able to draw the conclusion that the shared heteroplasmic sites between related pairs are closer than the sites of unrelated pairs, which can be explained as inheritance.

Bootstrap tests, performed 10,000 times, showed that related pairs were significantly ($P < 0.05$) more correlated than unrelated pairs at all heteroplasmy level thresholds (Table 3). For detail of bootstrap tests please see Materials and Methods section. Wilcoxon rank sum tests were concordant with the results of the bootstrap tests except at heteroplasmy level $< 0.1\%$ where the Pearson's correlation coefficient of related pairs was only marginally significant ($P = 0.09$ by Wilcoxon test).

Heteroplasmy Level in D-loop

Finally, we split the analysis into two regions: sites within the D-loop and sites outside the D-Loop. We observed that maternally related pairs have significantly stronger Pearson's correlation coefficients than unrelated pairs in both the D-loop and non-D-loop regions (Fig. 3). Within the maternally related pairs, the D-loop sites had a higher median correlation coefficient than non-D-loop sites (P -values = 0.037 to 0.0001) (Table S3). For unrelated pairs, the D-loop sites had an even higher median correlation coefficient than non-D-loop sites (P -values < 0.0001) (Table S4). The inheritance of low-level heteroplasmy for both related and un-related pairs, down to 0.1% heteroplasmy, was observed in both the D-loop and non-D-loop sections of the genome. The slightly higher correlation in the D-loop may be explained by the higher mutation frequency in the D-Loop compared to non-D-Loop regions (Pereira et al., 2009).

DISCUSSION

It is well established that mtDNA heteroplasmic variations at moderate to high mutation levels in humans are inherited from mother to children. However, previously there has been limited evidence for low-level heteroplasmy inheritance in humans. For example, Payne et al. (Payne et al., 2013) have recently shown that some low-level heteroplasmy can be

inherited using 454 pyrosequencing technology on two short amplicons of a few hundred base pairs. Yao et al. (Yao et al., 2013) have also demonstrated maternally inherited heteroplasmy at the level of 10% or greater.

In principle, the bottleneck in the inheritance of mtDNA from mother to offspring could act to filter out low-level mutations efficiently. Furthermore, sequencing errors confound the direct measurement of low mutation level variants, even with the new sequencing technology. In this study, it has been shown that significant correlations in mutation level can be observed even at heteroplasmy level $< 0.15\%$ and that these correlations are significantly stronger in maternally related pairs than those in arbitrary non-related pairs. The results suggest that not all heteroplasmy signals below 1% are sequencing errors. On the contrary, a significant portion of these heteroplasmy signals are inherited from the mother. However, the heteroplasmy signals become increasingly difficult to distinguish from sequencing errors as the heteroplasmy level decreases to approximately 0.1%. In addition to Pearson's correlation, we also conducted analysis using Spearman's correlation which resulted in weaker inheritance signals overall. However, related pairs still have stronger correlation than non-related pairs consistently across all heteroplasmy thresholds.

During the formation of the female germ line in early embryogenesis, the mtDNA copy number per cell drops to a relatively low value, forming a bottleneck in the inheritance of mtDNA. This bottleneck should act as a barrier limiting the inheritance of mtDNA mutations with very low mutation levels, as the probability of these low-level mutations passing through the bottleneck to an offspring should be low. Independent measurements (Cree et al., 2008; Wai et al., 2008) of this minimum mtDNA copy number in mice show a value of about 200 mtDNA molecules per cell at the initial formation of the primordial germ cells, shortly after implantation of the embryo. The very limited data in humans show that the variance in the mutation levels in both oocytes and offspring is approximately three times larger in humans than in mice (Wonnapijit et al., 2010), indicating an even deeper bottleneck in humans than in mice. A possible confounding factor here is that almost all of our current knowledge of mtDNA inheritance in humans comes from the study of pathogenic mutations in patient families. The new deep-sequencing methods and our new analysis method presented here allow us to investigate the inheritance of all heteroplasmic mtDNA mutations in the general population.

Our results show that for both related and unrelated pairs, the D-loop sites had a higher median correlation coefficient than non-D-loop sites. This could be explained by the high mutation rate and the small region (1124 base pairs) of the D-loop, which might result in highly recurrent mutation at the same position across multiple samples. Thus, pairs in D-loop regions produce higher pairwise correlation coefficients than do pairs in non-D-loop regions.

Given the small number of mtDNA molecules that pass from mother to offspring, the capacity for inheriting very low-level mtDNA mutations does appear to be quite limited. The evidence presented here shows that maternal inheritance of very low-level mtDNA mutations, down to the 0.15% mutation level (corresponding to only one in 670 molecules), clearly does occur in the general population. It is possible that low-level mtDNA mutations

are so numerous that even with a deep bottleneck, some of these rare variants are still inherited by the offspring, forming the correlations shown here. Indeed, our results in Fig. 2 show that the correlations in low-level mutations are reasonably steady down to a level of 0.5% before beginning to decrease. This is consistent with a bottleneck size of approximately 200 mtDNA (the inverse of 0.5%, where the correlations in Fig. 2 begin to decrease).

MATERIALS & METHODS

Data Collection

Permissions for the study were granted by the Danish Data Protection Agency (2001-41-1113) and the Danish Scientific Ethical Committee ([KF] 01-150/01 and [KF] 11-129/02). Informed consent of the Finnish family was obtained, as was local institutional review board approval. Approval was also obtained from the Ethical Committee of the Hospital District of Varsinais-Suomi in Finland. Blood samples were taken from 44 subjects from 17 Danish families and one Finnish family. The same samples have been used in a variety of previous studies (Curwen et al., 2010; Curwen et al., 2005; Tawn et al., 2011; Tawn et al., 2005) that describe in detail the sample collection and DNA extraction protocols. Because mitochondria are solely maternally inherited, and paternal samples were not available, only mothers and their offspring were included in this study. In the 44 samples, there are a total of 35 maternally related pairs (26 mother-child pairs and 9 sibling pairs), and there are 911 unrelated pairs. Details for each family and their haplogroup information are described in Table 1. These families were ascertained as a study of the potential effect of cancer radiation therapy on future offspring, and thus all of the mothers in this study were survivors of childhood cancer who underwent radiation therapy including radiation of the uterus. In that study, no significant associations were found between the radiation dosage and the amount of heteroplasmic mtDNA mutations in either the mothers or the offspring (Guo et al., 2012a).

Mitochondria Enrichment and Sequencing

mtDNA enrichment was done using the amplification kit from Affymetrix's Genechip Human Mitochondria Resequencing Array 2.0 (Affymetrix, USA). Genomic DNA was amplified using PCR with two primer sets, mito3 and mito1-2. The forward primer of mito3 is 5'-TCATTTTATTGCCACAACCTCCTCGGACTC-3', and the reverse primer is 5'-CGTGATGTCTTATTTAAGGGGAACGTGTGGGCTAT-3'; the forward primer of mito1-2 is 5'-ACATAGCACATTACAGTCAAATCCCTTCTCGTCCC-3', and the reverse primer is 5'-ATTGCTAGGGTGGCGCTTCCAATTAGGTGC-3'. The two primer sets generate fragments of 7814bp and 9307bp, respectively. This protocol specifically amplifies the entire mitochondrial genome from genomic DNA using overlapping primers to eliminate the bias that may be introduced from the PCR method. The enriched mtDNA were barcoded and sequenced using the Illumina GAII high-throughput sequencing platform (Illumina, USA).

Sequencing Data Processing

To deal with mtDNA sequence homologs (NUMT), we first aligned all reads against the human reference genome hg19 without mitochondria reference BWA (Li and Durbin, 2009). Then, we aligned the unmapped reads against the revised Cambridge mitochondrial reference sequence (GenBank: NC_012920) (Andrews et al., 1999). There is an artificially inserted “N” in the rCRS at position 3106 (Andrews et al., 1999). When performing alignment using rCRS, false heteroplasmies, SNPs, and small indels can be introduced around the position 3106 (Guo et al., 2012a). To avoid this, we deleted the “N” from rCRS before conducting alignment. Later, all positions after 3106 were shifted by one base to keep consistency with rCRS.

The aligned BAM (Li et al., 2009) files were locally realigned using the Genome Analysis Toolkit (GATK) (McKenna et al., 2010) developed by the Broad Institute. To further increase the local realignment accuracy, we built SNP and indel references in Variant Call Format (VCF) (Danecek et al., 2011) using the existing SNP and indel lists reported on mitomap.org and incorporated the information into the local realignment process. After local realignment, we performed base quality score recalibration on the realigned BAM files using GATK's recalibration tool. The recalibrated BAM files were filtered by removing all reads with mapping quality (MQ) Phred score (Cock et al., 2010) less than 20 and all bases with base quality (BQ) Phred score less than 20 (corresponding to a probability of base calling error of less than 1%). After applying these filters, a significant amount of heteroplasmy rate was adjusted. For example, for sample #22, 12,288 sites of the mitochondrial genome had heteroplasmy level changes. After applying MQ and BQ filters, a portion of the low quality bases were removed, which resulted in a decrease in the heteroplasmy rate of 9181 sites and an increase in the heteroplasmy rate of 3107 sites. Allele call information of each allele on both the forward and reverse strands separately was extracted for all 16,569 positions of the mtDNA reference sequence. Multiple thresholds for the filter were tested, and the conclusion of the study remained the same. The overall workflow can be seen in Fig. S2.

SNP Identification

SNP variant calling on the filtered BAM files was done using GATK's Unified Genotyper and GLFMMultiples, developed by the Abecasis Lab at the University of Michigan (<http://genome.sph.umich.edu/wiki/GLFMMultiples>). Both variance callers have been widely utilized in various studies, including the 1000 Genomes Project (Durbin et al., 2010). Final SNP lists were generated by forming a consensus from the results of the two base callers. Indels were also called using GATK's Indel Genotyper 2.0 and Samtools' Indel caller. The results of SNP and indel calling for this data have been previously reported (Guo et al., 2012a). For quality control purposes, SNPs and indels were individually examined through allele counting. Haplogroup information for each sample was accomplished by checking the SNP results against phylotree.org's mtDNA phylogeny tree (van Oven and Kayser, 2009). For quality control, we only analyzed sites with sequencing depth ≥ 400 and strand bias score ≤ 1 . Strand bias is a novel way of filtering out suspicious positions from sequencing data. The definition of strand bias score can be found in (Guo et al., 2012b), and it has been implemented in many existing tools such as SAMtools (Li et al., 2009), GATK (McKenna et al., 2010), VarScan (Koboldt et al., 2009), and MitoSeek (Guo et al., 2013). The requirement

that the strand bias score should be ≥ 1 removes the obvious sequencing errors due to short polynucleotide tracts.

Heteroplasmy Inheritance Analysis

For each site that passed quality control, we identified the major allele in each subject. The heteroplasmy level for a position of a subject was defined as 1 minus the frequency of the major allele. Subjects from the study were paired, and we compared maternally related pairs to unrelated pairs. We examined the inheritance of mtDNA variants grouped by different heteroplasmy levels (<100%, <5%, <1%, <0.5%, <0.25%, <0.20%, <0.15%, <0.1%). For each heteroplasmy-level threshold, a site was included in the analysis if one member of the pair had a measured heteroplasmy level below the threshold. For example, a site with a heteroplasmy level of 7% in one person and 0.23% in the other person of the pair would be included in all analyses from 100% down to the 0.25% threshold. To meet the criteria for analysis, a site must have passed quality control in both paired subjects, and the heteroplasmy level of the site must not have exceeded the heteroplasmy-level threshold in at least one member of the pair. Pearson's correlation coefficient was calculated over all sites meeting these criteria across the entire mitochondrial genome. As a secondary analysis, the correlation coefficient calculation was split into the D-loop (positions 16,024–16,569 and 1–576) and the coding region (577–16,023). Furthermore, heteroplasmy sites were verified using alternative analysis methods (Li and Stoneking, 2012) and (Li et al., 2010).

There were 35 maternally related pairs and 911 unrelated pairs in this study. Because the total number of possible unrelated pairs was significantly more than that of related pairs, we performed both a bootstrap test and a Wilcoxon rank sum test to assess the significance of the comparisons between related and unrelated pairs. The Wilcoxon rank sum test was performed between all related pairs and unrelated pairs. For the bootstrap test, we randomly drew 35 unrelated pairs from the unrelated pair pool with replacement and compared the median Pearson correlation coefficient of the unrelated pairs with that of the 35 related pairs. We repeated this process 10,000 times for each heteroplasmy-level threshold. The *P*-value was computed as the number of times that the median Pearson correlation coefficient of the unrelated pairs was higher than that of the related pairs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the families for participating in this study, and Brian Møllgren, Rigshospitalet, for collection of blood samples. The study was supported by the grant from the National Cancer Institute (RO1 CA104666). DNA sample preparation was conducted at Survey and Biospecimen Shared Resource, which is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485). We would also like to thank Peggy Schuyler and Margot Björing for their editing support.

REFERENCES

- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics*. 1999; 23:147. [PubMed: 10508508]
- Chinnery PF, Thorburn DR, Samuels DC, White SL, Dahl HHM, Turnbull DM, Lightowlers RN, Howell N. The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends Genet*. 2000; 16:500–505. [PubMed: 11074292]
- Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2010; 38:1767–1771. [PubMed: 20015970]
- Cree LM, Samuels DC, Lopes S, Rajasimha HK, Wonnapijit P, Mann JR, Dahl HHM, Chinnery PF. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nat Genet*. 2008; 40:249–254. [PubMed: 18223651]
- Curwen GB, Cadwell KK, Winther JF, Tawn EJ, Rees GS, Olsen JH, Rechnitzer C, Schroeder H, Guldberg P, Cordell HJ, Boice JD Jr. The heritability of G2 chromosomal radiosensitivity and its association with cancer in Danish cancer survivors and their offspring. *Int J Radiat Biol*. 2010; 86:986–995. [PubMed: 20807177]
- Curwen GB, Winther JF, Tawn EJ, Smart V, Whitehouse CA, Rees GS, Olsen JH, Guldberg P, Rechnitzer C, Schroder H, Bryant PE, Sheng X, Lee HS, Chakraborty R, Boice JD. G(2) chromosomal radiosensitivity in Danish survivors of childhood and adolescent cancer and their offspring. *Br J Cancer*. 2005; 93:1038–1045. [PubMed: 16234827]
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. The variant call format and VCFtools. *Bioinformatics*. 2011; 27:2156–2158. [PubMed: 21653522]
- Durbin RM, Altshuler DL, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Collins FS, De la Vega FM, Donnelly P, Egholm M, Flicek P, Gabriel SB, Gibbs RA, Knoppers BM, Lander ES, Lehrach H, Mardis ER, McVean GA, Nickerson D, Peltonen L, Schafer AJ, Sherry ST, Wang J, Wilson RK, Gibbs RA, Deiros D, Metzker M, Muzny D, Reid J, Wheeler D, Wang J, Li JX, Jian M, Li G, Li RQ, Liang HQ, Tian G, Wang B, Wang J, Wang W, Yang HM, Zhang XQ, Zheng HS, Lander ES, Altshuler DL, Ambrogio L, Bloom T, Cibulskis K, Fennell TJ, Gabriel SB, Jaffe DB, Shefler E, Sougnez CL, Bentley DR, Gormley N, Humphray S, Kingsbury Z, Koko-Gonzales P, Stone J, McKernan KJ, Costa GL, Ichikawa JK, Lee CC, Sudbrak R, Lehrach H, Borodina TA, Dahl A, Davydov AN, Marquardt P, Mertes F, Nietfeld W, Rosenstiel P, Schreiber S, Soldatov AV, Timmermann B, Tolzmann M, Egholm M, Affourtit J, Ashworth D, Attiya S, Bachorski M, Buglione E, Burke A, Caprio A, Celone C, Clark S, Connors D, Desany B, Gu L, Guccione L, Kao K, Keibel A, Knowlton J, Labrecque M, McDade L, Mealmaker C, Minderman M, Nawrocki A, Niazi F, Pareja K, Ramenani R, Riches D, Song W, Turcotte C, Wang S, Mardis ER, Dooling D, Fulton L, Fulton R, Weinstock G, Durbin RM, Burton J, Carter DM, Churcher C, Coffey A, Cox A, Palotie A, Quail M, Skelly T, Stalker J, Swerdlow HP, Turner D, De Witte A, Giles S, Gibbs RA, Wheeler D, Bainbridge M, Challis D, Sabo A, Yu F, Yu J, Wang J, Fang XD, Guo XS, Li RQ, Li YR, Luo RB, Tai S, Wu HL, Zheng HC, Zheng XL, Zhou Y, Yang HM, Marth GT, Garrison EP, Huang W, Indap A, Kural D, Lee WP, Leong WF, Huang WC, Indap A, Kural D, Lee WP, Leong WF, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu JT, Lee C, Mills RE, Shi XH, Daly MJ, DePristo MA, Altshuler DL, Ball AD, Banks E, Bloom T, Browning BL, Cibulskis K, Fennell TJ, Garimella KV, Grossman SR, Handsaker RE, Hanna M, Hartl C, Jaffe DB, Kernysky AM, Korn JM, Li H, Maguire JR, McCarroll SA, McKenna A, Nemesh JC, Philippakis AA, Poplin RE, Price A, Rivas MA, Sabeti PC, Schaffner SF, Shefler E, Shlyakhter IA, Cooper DN, Ball EV, Mort M, Phillips AD, Stenson PD, Sebat J, Makarov V, Ye K, Yoon SC, Bustamante CD, Clark AG, Boyko A, Degenhardt J, Gravel S, Gutenkunst RN, Kaganovich M, Keinan A, Lacroute P, Ma X, Reynolds A, Clarke L, Flicek P, Cunningham F, Herrero J, Keenen S, Kulesha E, Leinonen R, McLaren W, Radhakrishnan R, Smith RE, Zalunin V, Zheng-Bradley XQ, Korbel JO, Stutz AM, Humphray S, Bauer M, Cheetham RK, Cox T, Eberle M, James T, Kahn S, Murray L, Ye K, De La Vega FM, Fu YT, Hyland FCL, Manning JM, McLaughlin SF, Peckham HE, Sakarya O, Sun YA, Tsung EF, Batzer MA, Konkel MK, Walker JA, Sudbrak R, Albrecht MW, Amstislavskiy VS, Herwig R, Parkhomchuk DV, Sherry ST, Agarwala R, Khouri H, Morgulis AO, Paschall JE, Phan LD,

Rotmistrovsky KE, Sanders RD, Shumway MF, Xiao CL, McVean GA, Auton A, Iqbal Z, Lunter G, Marchini JL, Moutsianas L, Myers S, Tumian A, Desany B, Knight J, Winer R, Craig DW, Beckstrom-Sternberg SM, Christoforides A, Kurdoglu AA, Pearson J, Sinari SA, Tembe WD, Haussler D, Hinrichs AS, Katzman SJ, Kern A, Kuhn RM, Przeworski M, Hernandez RD, Howie B, Kelley JL, Melton SC, Abecasis GR, Li Y, Anderson P, Blackwell T, Chen W, Cookson WO, Ding J, Kang HM, Lathrop M, Liang LM, Moffatt MF, Scheet P, Sidore C, Snyder M, Zhan XW, Zollner S, Awadalla P, Casals F, Idaghdour Y, Keebler J, Stone EA, Zilversmit M, Jorde L, Xing JC, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Sahinalp SC, Sudmant PH, Mardis ER, Chen K, Chinwalla A, Ding L, Koboldt DC, McLellan MD, Dooling D, Weinstock G, Wallis JW, Wendl MC, Zhang QY, Durbin RM, Albers CA, Ayub Q, Balasubramaniam S, Barrett JC, Carter DM, Chen YA, Conrad DF, Danecek P, Dermitzakis ET, Hu M, Huang N, Hurles ME, Jin HJ, Jostins L, Keane TM, Keane TM, Le SQ, Lindsay S, Long QA, MacArthur DG, Montgomery SB, Parts L, Stalker J, Tyler-Smith C, Walter K, Zhang YJ, Gerstein MB, Snyder M, Abyzov A, Abyzov A, Balasubramanian S, Bjornson R, Du JA, Grubert F, Habegger L, Haraksingh R, Jee J, Khurana E, Lam HYK, Leng J, Mu XJ, Urban AE, Zhang ZD, Li YR, Luo RB, Marth GT, Garrison EP, Kural D, Quinlan AR, Stewart C, Stromberg MP, Ward AN, Wu JT, Lee C, Mills RE, Shi XH, McCarroll SA, Banks E, DePristo MA, Handsaker RE, Hartl C, Korn JM, Li H, Nemesh JC, Sebat J, Makarov V, Ye K, Yoon SC, Degenhardt J, Kaganovich M, Clarke L, Smith RE, Zheng-Bradley XQ, Korbel JO, Humphray S, Cheetham RK, Eberle M, Kahn S, Murray L, Ye K, De la Vega FM, Fu YT, Peckham HE, Sun YA, Batzer MA, Konkel MK, Xiao CL, Iqbal Z, Desany B, Blackwell T, Snyder M, Xing JC, Eichler EE, Aksay G, Alkan C, Hajirasouliha I, Hormozdiari F, Kidd JM, Chen K, Chinwalla A, Ding L, McLellan MD, Wallis JW, Hurles ME, Conrad DF, Walter K, Zhang YJ, Gerstein MB, Snyder M, Abyzov A, Du JA, Grubert F, Haraksingh R, Jee J, Khurana E, Lam HYK, Leng J, Mu XJ, Urban AE, Zhang ZD, Gibbs RA, Bainbridge M, Challis D, Coafra C, Dinh H, Kovar C, Lee S, Muzny D, Nazareth L, Reid J, Sabo A, Yu FL, Yu J, Marth GT, Garrison EP, Indap A, Leong WF, Quinlan AR, Stewart C, Ward AN, Wu JT, Cibulskis K, Fennell TJ, Gabriel SB, Garimella KV, Hartl C, Shefler E, Sougnez CL, Wilkinson J, Clark AG, Gravel S, Grubert F, Clarke L, Flicek P, Smith RE, Zheng-Bradley XQ, Sherry ST, Khouri HM, Paschall JE, Shumway MF, Xiao CL, McVean GA, Katzman SJ, Abecasis GR, Blackwell T, Mardis ER, Dooling D, Fulton L, Fulton R, Koboldt DC, Durbin RM, Balasubramaniam S, Coffey A, Keane TM, MacArthur DG, Palotie A, Scott C, Stalker J, Tyler-Smith C, Gerstein MB, Balasubramanian S, Chakravarti A, Knoppers BM, Peltonen L, Abecasis GR, Bustamante CD, Gharani N, Gibbs RA, Jorde L, Kaye JS, Kent A, Li T, McGuire AL, McVean GA, Ossorio PN, Rotimi CN, Su YY, Toji LH, Tyler-Smith C, Brooks LD, Felsenfeld AL, McEwen JE, Abdallah A, Christopher R, Clemm NC, Collins FS, Duncanson A, Green ED, Guyer MS, Peterson JL, Schafer AJ, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA, Consortium GP. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]

Elson JL, Samuels DC, Turnbull DM, Chinnery PF. Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age. *Am J Hum Genet*. 2001; 68:802–806. [PubMed: 11179029]

Goto H, Dickins B, Afgan E, Paul IM, Taylor J, Makova KD, Nekrutenko A. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. *Genome Biol*. 2011; 12:R59. [PubMed: 21699709]

Guo Y, Cai Q, Samuels DC, Ye F, Long J, Li CI, Winther JF, Tawn EJ, Stovall M, Lahteenmaki P, Malila N, Levy S, Shaffer C, Shyr Y, Shu XO, Boice JD Jr. The use of next generation sequencing technology to study the effect of radiation therapy on mitochondrial DNA mutation. *Mutat Res*. 2012a; 744:154–160. [PubMed: 22387842]

Guo Y, Li J, Li CI, Long J, Samuels DC, Shyr Y. The effect of strand bias in Illumina short-read sequencing data. *BMC genomics*. 2012b; 13:666. [PubMed: 23176052]

Guo Y, Li J, Li CI, Shyr Y, Samuels DC. MitoSeek: Extracting Mitochondria Information and Performing High Throughput Mitochondria Sequencing Analysis. *Bioinformatics*. 2013

Holt IJ, Harding AE, Petty RK, Morgan-Hughes JA. A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am J Hum Genet*. 1990; 46:428–433. [PubMed: 2137962]

- Kann LM, Rosenblum EB, Rand DM. Aging, mating, and the evolution of mtDNA heteroplasmy in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95:2372–2377. [PubMed: 9482892]
- Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009; 25:2283–2285. [PubMed: 19542151]
- Lewis PD, Baxter P, Paul Griffiths A, Parry JM, Skibinski DO. Detection of damage to the mitochondrial genome in the oncocytic cells of Warthin's tumour. *J Pathol*. 2000; 191:274–281. [PubMed: 10878549]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
- Li M, Schonberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet*. 2010; 87:237–249. [PubMed: 20696290]
- Li M, Stoneking M. A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biol*. 2012; 13:R34. [PubMed: 22621726]
- Lombes A, Diaz C, Romero NB, Ziegler F, Fardeau M. Analysis of the tissue distribution and inheritance of heteroplasmic mitochondrial DNA point mutation by denaturing gradient gel electrophoresis in MERRF syndrome. *Neuromuscul Disord*. 1992; 2:323–330. [PubMed: 1300181]
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010; 20:1297–1303. [PubMed: 20644199]
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*. 2010; 42:30–35. [PubMed: 19915526]
- Park JS, Sharma LK, Li H, Xiang R, Holstein D, Wu J, Lechleiter J, Naylor SL, Deng JJ, Lu J, Bai Y. A heteroplasmic, not homoplasmic, mitochondrial DNA mutation promotes tumorigenesis via alteration in reactive oxygen species generation and apoptosis. *Hum Mol Genet*. 2009; 18:1578–1589. [PubMed: 19208652]
- Payne BA, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels DC, Santibanez-Koref M, Chinnery PF. Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet*. 2013; 22:384–390. [PubMed: 23077218]
- Pereira L, Freitas F, Fernandes V, Pereira JB, Costa MD, Costa S, Maximo V, Macaulay V, Rocha R, Samuels DC. The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet*. 2009; 84:628–640. [PubMed: 19426953]
- Robin ED, Wong R. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol*. 1988; 136:507–513. [PubMed: 3170646]
- Simon DK, Friedman J, Breakefield XO, Jankovic J, Brin MF, Provias J, Bressman SB, Charness ME, Tarsy D, Johns DR, Tarnopolsky MA. A heteroplasmic mitochondrial complex I gene mutation in adult-onset dystonia. *Neurogenetics*. 2003; 4:199–205. [PubMed: 12756609]
- Smigrodzki RM, Khan SM. Mitochondrial microheteroplasmy and a theory of aging and age-related disease. *Rejuvenation Res*. 2005; 8:172–198. [PubMed: 16144471]
- Sondheimer N, Glatz CE, Tirone JE, Deardorff MA, Krieger AM, Hakonarson H. Neutral mitochondrial heteroplasmy and the influence of aging. *Human Molecular Genetics*. 2011; 20:1653–1659. [PubMed: 21296868]
- Stoneking M. Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet*. 2000; 67:1029–1032. [PubMed: 10968778]
- Tang S, Huang T. Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. *Biotechniques*. 2010; 48:287–296. [PubMed: 20569205]

- Tawn EJ, Rees GS, Leith C, Winther JF, Curwen GB, Stovall M, Olsen JH, Rechnitzer C, Schroeder H, Guldberg P, Boice JD. Germline minisatellite mutations in survivors of childhood and young adult cancer treated with radiation. *Int J Radiat Biol.* 2011; 87:330–340. [PubMed: 21087171]
- Tawn EJ, Whitehouse CA, Winther JF, Curwen GB, Rees GS, Stovall M, Olsen JH, Guldberg P, Rechnitzer C, Schroder H, Boice JD Jr. Chromosome analysis in childhood cancer survivors and their offspring--no evidence for radiotherapy-induced persistent genomic instability. *Mutat Res.* 2005; 583:198–206. [PubMed: 15914077]
- van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009; 30:E386–E394. [PubMed: 18853457]
- Wai T, Teoli D, Shoubridge EA. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nat Genet.* 2008; 40:1484–1488. [PubMed: 19029901]
- Wong LJ, Wong H, Liu A. Intergenerational transmission of pathogenic heteroplasmic mitochondrial DNA. *Genet Med.* 2002; 4:78–83. [PubMed: 11882784]
- Wonnapijit P, Chinnery PF, Samuels DC. The Distribution of Mitochondrial DNA Heteroplasmy Due to Random Genetic Drift. *Am J Hum Genet.* 2008; 83:582–593. [PubMed: 18976726]
- Wonnapijit P, Chinnery PF, Samuels DC. Previous Estimates of Mitochondrial DNA Mutation Level Variance Did Not Account for Sampling Error: Comparing the mtDNA Genetic Bottleneck in Mice and Humans. *Am J Hum Genet.* 2010; 86:540–550. [PubMed: 20362273]
- Yao YG, Kajigaya S, Feng X, Samsel L, McCoy JP Jr, Torelli G, Young NS. Accumulation of mtDNA variations in human single CD34+ cells from maternally related individuals: effects of aging and family genetic background. *Stem Cell Res.* 2013; 10:361–370. [PubMed: 23455392]

Fig. 1. Different heteroplasmy levels between sibling pairs and mother-child pairs

Box plots represent the probability distributions of the correlation coefficient for the mutation levels between sibling pairs and mother-child pairs. Progressively decreasing (from left to right) thresholds on the mtDNA mutation level indicate how the correlations decrease as the mutation level drops. See Material and Methods for details on the correlation calculation.

Fig. 2. Different heteroplasmy levels between related pairs and unrelated pairs

Distributions of the correlation coefficients for the mtDNA mutation level are plotted as in Fig. 1.

Fig. 3. Variants in the D-loop and outside of the D-loop

Distributions of the correlation coefficients for the mtDNA mutation level are plotted as in Fig. 1. **A:** sites outside of the D-loop. **B:** sites within the D-loop.

Table 1

Haplogroup information of all families

Family ID	Haplogroup	Number of children
Fin	I4	1
T07	K1a1b1	1
T10	H1c1	1
T13	H11a2	1
T16	H1a	1
T20	H36a	2
T21	U5a1a1	1
T25	H	1
T26	J1c5a	2
T27	U5a2a	1
T28	H1c3	1
T37	V	2
T38	T1a1	2
T39	J1c2	1
T41	HV9a	2
T45	H2a2a	1
T46	U5a1b	2
T54	K1a1b1	3

Table 2

Heteroplasmic sites with MAF > 1%

mtDNA variant	Family ID	Mother's mutation frequency	Child's mutation frequency	Reference allele	Alternative allele	Adjacent sequence
T146C	T45	19.20%	3.10%	T	C	GCCT/CCAT
T195C	T07	37.20%	35.60%	T	C	TACT/CTAC
T310C	T37	35.00%	35.90%	T	C	CCCT/CCCC
G316C	T10	2.80%	2.20%	G	C	CCCG/CCTT
T6292C	T45	30.30%	10.80%	T	C	CCCT/CCCC
C12091T	T39	12.50%	28.80%	C	T	ATTG/TTCC
C12972T	T26	48.40%	40.10%	C	T	CCAC/TTAC
G13927A	T13	3.00%	1.80%	G	A	CTAG/ACAT
G15354A	T07	10.10%	28.00%	G	A	AACG/AGGA
G15733A	T16	8.10%	6.30%	G	A	GCCG/ACAG
A16398G	T46	11.50%	27.30%	A	G	TTGA/GCCA
A16398G	T46	11.50%	43.00%	A	G	TTGA/GCCA
A4226G	T54	17.30%	29.10%	A	G	CATA/GCCC
A4226G	T54	17.30%	15.80%	A	G	CATA/GCCC
A4226G	T54	17.30%	19.40%	A	G	CATA/GCCC

Table 3

Test results of unrelated pairs vs. related pairs at different heteroplasmy levels

Heteroplasmy level threshold	<i>P</i> -value (Resampling Test *)	<i>P</i> -value (Wilcoxon Rank Sum Test)
100%	<0.0001	7.8×10^{-18}
5%	<0.0001	1.0×10^{-12}
1%	<0.0001	5.8×10^{-12}
0.5%	<0.0001	2.1×10^{-11}
0.25%	<0.0001	6.9×10^{-9}
0.2%	0.0001	1.7×10^{-7}
0.15%	0.0012	7.1×10^{-5}
0.1%	0.0119	0.09

* The resampling test was performed for 10,000 iterations