

Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA

Anitha D. Jayaprakash¹, Erica K. Benson¹, Swapna Gone³, Raymond Liang¹, Jaehee Shim¹, Luca Lambertini², Masoud M. Toloue³, Mike Wigler⁴, Stuart A. Aaronson¹ and Ravi Sachidanandam^{1,*}

¹Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, NY 10029, USA, ²Department of Preventive Medicine and Department of Obstetrics, Gynecology and Reproductive Science, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, NY 10029, USA, ³Bioo Scientific Corporation, 7050 Burleson Road, Austin, TX 78744, USA and ⁴Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, NY 11724, USA

Received November 11, 2014; Revised January 13, 2015; Accepted January 14, 2015

ABSTRACT

Eukaryotic cells carry two genomes, nuclear (nDNA) and mitochondrial (mtDNA), which are ostensibly decoupled in their replication, segregation and inheritance. It is increasingly appreciated that heteroplasmy, the occurrence of multiple mtDNA haplotypes in a cell, plays an important biological role, but its features are not well understood. Accurately determining the diversity of mtDNA has been difficult, due to the relatively small amount of mtDNA in each cell (<1% of the total DNA), the intercellular variability of mtDNA content and mtDNA pseudogenes (Numts) in nDNA. To understand the nature of heteroplasmy, we developed Mseek, a novel technique to purify and sequence mtDNA. Mseek yields high purity (>90%) mtDNA and its ability to detect rare variants is limited only by sequencing depth, providing unprecedented sensitivity and specificity. Using Mseek, we confirmed the ubiquity of heteroplasmy by analyzing mtDNA from a diverse set of cell lines and human samples. Applying Mseek to colonies derived from single cells, we find heteroplasmy is stably maintained in individual daughter cells over multiple cell divisions. We hypothesized that the stability of heteroplasmy could be facilitated by intercellular exchange of mtDNA. We explicitly demonstrate this exchange by co-culturing cell lines with distinct mtDNA haplotypes. Our results shed new light on the maintenance of heteroplasmy and provide a novel platform to investigate features of heteroplasmy in normal and diseased states.

INTRODUCTION

Mitochondria are organelles present in almost every eukaryotic cell (1). They enable aerobic respiration (2) to efficiently generate ATP and play an important role in oxygen sensing, inflammation, autophagy and apoptosis (3,4). Mitochondrial activity relies on over a thousand proteins, mostly coded by the nuclear DNA in humans (5), but proteins from the mitochondrial genome, a small circular DNA (mtDNA), play a critical role in their function. In humans, the reference mtDNA is 16 569 bp long and codes 13 proteins critical for the electron transport chain, along with 22 tRNAs, two rRNAs and a control region, called the displacement loop (D-loop) (Supplementary Figure S1) (6). Each mitochondrion carries multiple mitochondrial genomes (5–10) (7) and each cell contains hundreds to thousands of mitochondria, depending on the tissue (8). The mtDNA replicate without recombination. mtDNA is inherited solely from the mother; inherited mutations in mtDNA have been linked to several genetic disorders including *diabetes mellitus and deafness* and *Leber's hereditary optic neuropathy* (9). De novo mutations in mtDNA have also been linked to diseases (10–13).

Heteroplasmy, which is the occurrence of multiple mtDNA haplotypes, has been documented in various studies, cancer cells (14,15), blood samples from families (16) blood and muscle biopsies from identical twins (17) and cells from the 1000 genomes project (18,19). Though extensive, these studies have not established the nature of heteroplasmy at the *level of the single cell*. Accurate determination of heteroplasmy, especially the low-frequency haplotypes, is needed for disease-association studies with mtDNA, as well as studies of metabolic activity of cancer cells (14,20). Deep sequencing provides the only means to identify novel mtDNA haplotypes as well as comprehensively screen for somatic mutations in tissues. This is important for associa-

*To whom correspondence should be addressed. Tel: +1 212 824 9336; Fax: +1 212 987 2240; Email: ravi.mssm@gmail.com

tion studies that can link haplotypes to disease states. However, measurements of heteroplasmy are compromised by copies of large segments of mtDNA, called nuclear mitochondrial DNA (Numts), present in the mammalian nDNA (21) (Supplementary Figures S2 and S3).

Without purification of mtDNA, Numts introduce unpredictable inaccuracies in the estimates of heteroplasmy, especially because they exhibit variations in sequence and copy numbers. Numts have been annotated in the reference human genome (22,23), and there are tools to analyze high-throughput sequencing data in light of these annotations (24), but a comparison of two recent versions of the reference human genome (hg19 and hg38 on the UCSC genome browser, Supplementary Figures S2 and S3) suggests that this annotation is not complete and changes significantly with the reference genome. Additionally, the distribution of Numts might be specific to each individual's nuclear genome. A recent study of mtDNA from twins has highlighted the need for further investigation of Numts and their potential to confound analyses of heteroplasmy (25).

Isolating mtDNA has long been a challenge. In forensics and genealogy, allele-specific primer extensions (SNaPshot) are used for genotyping mtDNA (26). Hyper variable regions in the D-loop have been amplified using polymerase chain reaction (PCR) (27). Entire mtDNA has been accessed using primers specific to mtDNA to either perform long-range PCR (28), or amplify overlapping fragments (14). Isolation of organelles by ultra-high-speed centrifugation (29) has also been used, but the yields are low and contaminated with fragmented nuclear DNA (30). Computational methods have also been used to infer heteroplasmy from whole-exome (19,31) and whole-genome data (15), but such data contains Numts and furthermore, generating such data for new samples is expensive. A new approach uses methyl-specific endonucleases MspJI and AbaSI to deplete nDNA that is likely to be methylated (32), but Numts as well as large parts of nDNA are likely to be unmethylated, leading to substantial contamination of the mtDNA. Heteroplasmy derived from PCR-based methods are error-prone, due to variability in amplification. Clonal amplification of errors introduced by polymerases is also a problem in PCR-amplicon sequencing. Additionally, sequence and copy number variations of Numts confound results from computational and PCR-based methods in unpredictable ways. Thus, it is difficult to ascertain if the methods outlined above are accurate in their measurement of the frequencies of common mtDNA variants and accurate in their identification of rare mtDNA variants that occur at frequencies below 5%.

We present here Mseek, a novel method to enzymatically purify mtDNA by depleting linear nDNA and inexpensively sequencing it (Supplementary Figure S4). Mseek uses exonuclease V to digest away linear nDNA, leaving behind circular mtDNA. A major benefit of this method is the ability to call extremely rare variants, with the sensitivity of calls only limited by the sequencing depth. By applying Mseek to several cell lines and human peripheral blood mononuclear cells (PBMC), we identified mixtures of different mtDNA haplotypes in the samples. Additionally, through clonal expansion of single cells from a variety of cell lines, we establish, for the first time, that heteroplasmy

is stably maintained at a single-cell level over multiple divisions. We infer the stability is due to intercellular exchange of mtDNA and experimentally demonstrate this exchange by co-culturing pairs of cell lines with distinct mtDNA haplotypes to demonstrate the transfer of mtDNA from one cell type to another.

MATERIALS AND METHODS

Mseek

Our method of isolating and sequencing mtDNA, dubbed Mseek (Supplementary Figure S4), consists of the following steps. (i) Total DNA is isolated from the sample. (ii) The nDNA is digested using Exonuclease V. (iii) The products are purified using Ampure beads to remove short fragments. (iv) To test the results of the digestion, PCR primers specific to mtDNA and nDNA are used on 1 μ l of the digested sample (Figure 1B). (v) The rest of the sample is fragmented using Covaris and end-repaired. (vi) Barcoded adapters compatible with the sequencing platform are ligated to the fragments. (vii) The universal adapters are used to amplify the library for loading on deep-sequencing instruments.

Sample processing

Total DNA was isolated from 500 μ l of whole blood or cell lines using the Epicenter protocol (MC85200). DNA was eluted in 100 μ l of TE buffer and checked for quality and quantity using a 1% agarose and nanodrop respectively. The eluted DNA was further heated at 70°C for 30 min to inactivate any left over proteinase K that was introduced in the DNA isolation protocol. A first digestion was carried out by adding to the total DNA Sample (4–8 μ g in 35 μ l), the following, NEB4 10 \times Buffer (6 μ l), 10 mM ATP (12 μ l), ExoV from NEB-M0345S (4 μ l) and H₂O (3 μ l). The digests were left at 37°C for 48 h, heat inactivated at 70°C for 30 min and purified using AMPure beads (Beckman Coulter). A second digestion was carried out by adding to the ExoV treated DNA (in 35 μ l) the following, NEB4 10 \times Buffer (6 μ l), 10 mM ATP (12 μ l), ExoV from NEB-M0345S (4 μ l) and H₂O (3 μ l). The digests were left at 37°C for 16 h, heat inactivated at 70°C for 30 min and purified using AMPure beads (Beckman Coulter).

The primers listed in Supplementary Tables ST1 and ST2 were used to detect the presence of nuclear and mtDNA. The samples were processed for deep sequencing only if the digested DNA did not show significant PCR product for nuclear DNA. In cases with nuclear DNA contamination, the digestions were repeated to improve mtDNA purification. We used Covaris for shearing and the Rapid DNA kit from Bioo Scientific for DNA library prep (5144-01).

Cell culture

293T (embryonic kidney cells containing the SV40 T-antigen), 501T (normal adult fibroblasts), U2OS (osteosarcoma), Saos-2 (osteosarcoma), HCC1806 (breast cancer), MDA-MB-157 (breast cancer) and A382 (glioblastoma) cells were grown in Dulbecco's Modified Eagle's Medium (Sigma-Aldrich, St Louis, MO, USA) supplemented with 10% fetal bovine serum (Sigma-Aldrich), 50 units/ml of

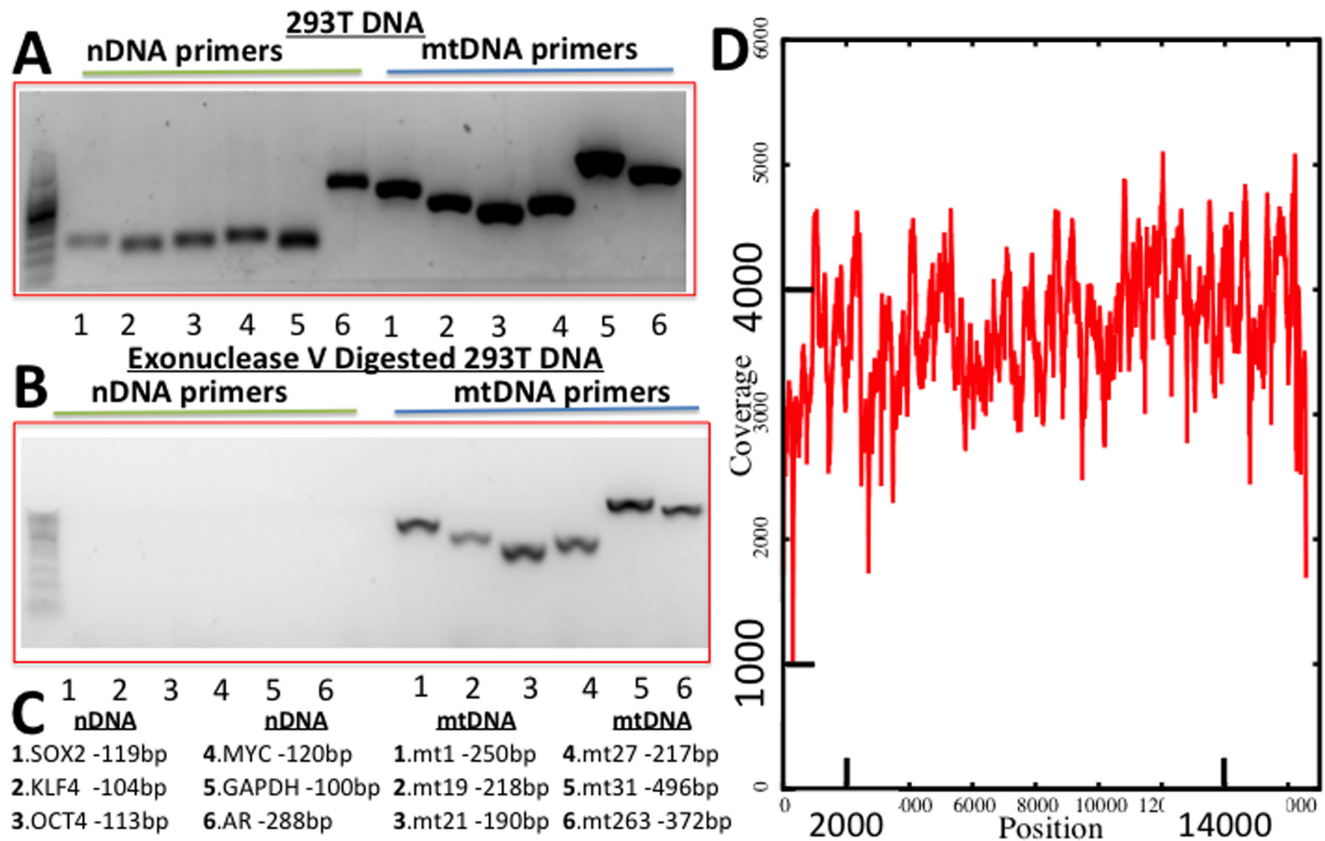


Figure 1. Performance of Mseek. (A) PCR products run on a 2% Agarose gel using primers for six nDNA genes (OCT4, MUC, KLF4, SOX2, GAPDH and AR, Supplementary Table ST1) and five regions of mtDNA Supplementary (Supplementary Table ST2) before exonuclease digestion. (B) After digestion, the nDNA bands disappear. (C) Sizes of expected PCR products. (D) Deep sequencing, read depth (y-axis) versus position on mtDNA (x-axis) demonstrating uniformity of coverage. 1.23 million mtDNA reads and 50 000 nDNA reads implying >98% pure mtDNA.

penicillin/streptomycin (Gibco, Grand Island, NY, USA). All cells were sub-cultured or collected using 0.05% trypsin-EDTA (Gibco) and maintained at 37°C and 90% humidity in a 5% CO₂ incubator. For selection of cells carrying the neomycin selection marker, 750 µg/ml G418 (Gibco) was added to the culture medium for 2 weeks. Control cells not carrying the resistance marker were used to verify cell death by G418. Clonal isolation of tumor cells was performed by serial dilution into 96-well plates and visual examination of wells for single cells. Single cells were then expanded for an additional 28–30 population doublings, expanding into larger tissue culture plates as necessary.

Fluorescence-activated cell sorting

Cultured cells were collected in phosphate buffered saline (Gibco) at a density 5×10^6 cells/ml and passed through a 40 µm filter. Cells were sorted using the BD FACS Aria II cell sorter (Becton-Dickinson, Mountain View, CA, USA), using the 488 nm laser. Sorting was performed in a sterile BSL2+ biosafety cabinet. FACSDiva Version 6.1.2 software (Becton-Dickinson) was used for analysis.

Lentiviral GFP expression

To ectopically express GFP, we used a NSPI-derived lentiviral vector that drives GFP expression with a constitutive

PGK promoter and contains a neomycin selection marker (33). High titer lentiviral production and infection were carried out as previously described (33).

Mixing experiments

Two strategies were used for selecting GFP-positive cells in the mixing experiments.

In the first strategy, HCC1806 and U2OS cells were transduced with a high titer of lentivirus to constitutively express GFP. Four days after infection, GFP-positive cells were isolated by utilizing the BD FACSaria II cell and FACSDiva Version software. GFP-positive cells were placed back in culture. An examination of several hundred cells showed that all were GFP positive, suggesting that fluorescence-activated cell sorting (FACS) was highly efficient and there were no GFP-negative cells.

Pairs of cell lines (one GFP-positive and the other GFP-negative) were mixed and co-cultured in 150 mm tissue culture dishes. In one group of experiments, MDA-MB-157 was co-cultured with HCC1806-GFP (ratio of 10:1 ensured by counting cells). In a second group of experiments, A382 cells were co-cultured with U2OS-GFP cells. The cells were allowed to grow for 4 weeks and were sub-cultured 1:4 or 1:5 when they reached near confluency, which was about every four days. The culture was monitored visually to en-

sure both GFP-positive and GFP-negative cells existed in culture.

Live GFP-positive and GFP-negative cells were separately gated and sterile sorted using the BD FACSAria II cell and FACSDiva software according to GFP status. Sorted GFP-positive cells were placed back into culture and allowed to expand for a week or two, to obtain 5×10^6 cells in a 150 mm plate, giving $\approx 4.5 \times 10^6$ of unlabeled cells and 5×10^5 of GFP-labeled cells, which was then processed for mtDNA sequencing. The purity of the cells after sorting was further verified by visual examination under the microscope.

The second strategy was similar to the above with the following changes. (i) After the infection with the high titer GFP lentivirus with neomycin resistance, GFP-positive cells were selected for neomycin resistance. (ii) These cells were co-cultured with marker-negative cells for 4 weeks, GFP-negative cells were specifically selected against using the neomycin selection marker as an alternative to cell sorting. Visual examination was used to further verify the purity of the GFP-positive population. As a control, unmixed cells lacking the neomycin resistant marker failed to form any colonies in the presence of the antibiotic, G418. In this way, only GFP-positive/neomycin resistant cells were collected for mtDNA isolation. MDA-MB-157-GFP mixed with IMR90 and WI-38-GFP mixed with IMR90 were processed by this strategy.

Analyses

The sequencing data is generated as fastq format files. These were processed using a pipeline developed by us for whole-exome analysis called MiST which is described elsewhere (34). In brief, the sequences were filtered for quality (sequences with >10 consecutive nucleotides with $Q < 20$ were eliminated) and mapped to the reference mitochondrial genome (accession NC_012920 from Genbank). Identical reads were identified as being clonal and were considered only once, irrespective of the number of copies, toward variant calling. A variant call was made only if there were at least three non-clonal reads carrying the variant, at least 10 nt away from the ends, and a minimum coverage of 10 was required at the variant.

Variants occurring on reads predominantly on one strand ($>80\%$) of the mtDNA were excluded to further reduce errors, based on our prior experience (34). The error rate in Miseq and Hiseq reads are usually <1 in a 1000 (phred score $Q > 30$), requiring at least three non-clonal reads reduces the error rate to well under one in a million. Nuclear contamination was estimated using sequences that map to repeat elements such as Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs), which only occur in nDNA. This enables reliable estimation of the level of nDNA contamination, which ranged from 0.5 to 1.5%.

The annotation of variants was determined using mtDNA annotations from MITOMAP13. Common single nucleotide polymorphisms (SNPs) and haplotype indicators were identified from dbSNP14. Various programs that annotate the effect of variants on protein function were tested. We eliminated programs that indicated com-

mon SNPs in mtDNA proteins were deleterious, *Mutation Assessor* (35) performed the best under this test, we used it to assess the impact of mtDNA mutations on protein function.

Mutation Assessor uses conservation of structure across orthologues to identify mutations in the DNA (and consequent changes in amino-acids) with potentially deleterious effects. The mutations are rated *high*, *medium*, *low* or *neutral* based on their impact on protein function. We highlight the *high* and *medium* impact mutations in our graphs, as they might affect mitochondrial function.

Custom code was developed for simulations and the plots were created using Gnuplot and R.

RESULTS

Mseek: an efficient method to isolate and sequence mtDNA

To efficiently purify mtDNA, we sought to take advantage of the difference in topology between nDNA and mtDNA using an exonuclease to digest the linear nDNA, while leaving intact the circular mtDNA. Total DNA was extracted from HEK 293T cells and digested with exonuclease V or left undigested. To determine the efficiency of digestion, sequences specific to nDNA and mtDNA were PCR amplified using appropriate primers (Supplementary Tables ST1 and ST2). As expected, in the undigested samples of total DNA we detected both nDNA and mtDNA (Figure 1A). In sharp contrast, in the samples treated with exonuclease V we only detected mtDNA (Figure 1B). The lengths of the expected PCR products are shown in Figure 1C.

Using this approach, mtDNA was prepared and sequenced on the Illumina MiSeq platform. Out of a total of 3.05 million 100 nt reads, 1.233 million mapped to the mitochondrial genome and 50 000 ($<2\%$) mapped to the nDNA. The rest of the reads were adapter dimers, formed during library preparation by the ligation of adapters to each other. Over 98% of the mappable reads were derived from mtDNA with an average coverage $>3000\times$ (Figure 1D). More than 50 distinct samples were processed similarly to consistently obtain high purity mtDNA. This approach, designated Mseek (Supplementary Figure S4), provides a means of unmatched efficiency in accurately sequencing the mtDNA contained within a population of cells.

Our comparisons of Mseek to other kits on the market show that Mseek substantially outperforms all of them in terms of the purity of mtDNA and yield (Supplementary material and Figure S5). As currently implemented, a limitation of Mseek is its requirement of at least 4 μg of intact total DNA, which is a big improvement over the 50 μg of total DNA that were needed for the initial experiments. Further improvements of sample preparation may reduce the input amount required, but are beyond the scope of this paper. For $<4 \mu\text{g}$ of total DNA we recommend using long-range PCR amplification with mtDNA-specific primers after exonuclease V treatment, which depletes nDNA to minimize distortions arising from Numts. The primers for long-range PCR are specified in Supplementary Table ST2 and enzyme used for amplification is from Clontech (Advantage Genomic LA Polymerase Mix catalog # S4775).

The DNA needs to be handled gently, without excessive centrifugation, to avoid shearing the circular mtDNA.

However, any linearization of mtDNA is very unlikely to be biased against a specific mtDNA, ensuring that Mseek gives an accurate representation of mtDNA heteroplasmy. While multiple rounds of treatment with exonuclease V may be needed to achieve high purity, a single treatment usually achieves around 80% purity. This is sufficient to avoid errors arising from Numts, which amount to <10 copies of mtDNA in the reference human genome (Supplementary Figures S2 and S3), enabling mtDNA sequencing at 50× depth with ≈20 000 50 bp reads.

Thus, Mseek can be used to call rare variants to any level of sensitivity, only limited by the depth of sequencing. Most methods will not allow this, for example, in PCR-amplicon sequencing, sensitivity does not necessarily increase with depth of sequencing as errors introduced during amplification cannot be corrected by greater sequencing depth. This is a valuable feature, that distinguishes it from other techniques, enabling the tracking of rare variants. Duplex sequencing, which uses adaptors with random tags to track clones in order to reduce the errors introduced during sample preparation, should be used in conjunction with Mseek to identify and study rare variants (with frequencies <0.1%) (36).

Ubiquity of heteroplasmy

Clonally derived cell lines would be expected to have identical nDNA and mtDNA. There are two possible reasons for this, either (i) a fitness advantage for a haplotype or (ii) allelic drift due to stochastic processes (37), both of which lead to the selection of a single nuclear genome and mtDNA homoplasmy.

To study clonality in mtDNA, we applied Mseek to 30 samples including four human PBMCs and human cell lines derived from human diploid fibroblasts (501T), glioma (A382) and breast carcinoma (HCC1806 and MDA-MB-157). Repeat content of the sequences was computationally identified to estimate nDNA contamination, which ranged from 0.5 to 5%; confirming the specificity of Mseek. Importantly, because of this high degree of mtDNA purity, we were able to multiplex all 30 samples in a single MiSeq run, with average coverage of >50×. We show the coverage at various positions in Supplementary Tables ST4 and ST5.

The sequences were analyzed for variants using MiST (34). Variants with a frequency of 1 indicate homoplasmic mtDNA. Frequencies <1 imply the co-existence of multiple haplotypes. Strikingly, in both cell lines and human blood-derived mtDNA, we observed variants occurring in the 0.1–0.9 frequency range (Figure 2), indicating the presence of multiple haplotypes. Most mutations were transitions (Supplementary Table ST3), suggesting that the mutations most likely arise from errors introduced by polymerase- γ activity instead of oxidative stress (36). The program *Mutation Assessor* (35) was used to label the variants as *high*, *medium*, *low* or *neutral* signifying their predicted impact on protein function. Cell lines and human PBMCs did not exhibit mutations of putative high effect at a high frequency (>5%), consistent with the expectation that functioning cells should have functional mitochondria.

Each sample had unique, distinguishing mutations, ranging in frequency from 0.36 to 1.0. There were a number of

variants unique to each of the human PBMC samples (ranging in number from 5 to 15) and each of the cell lines (ranging in number from 5 to 21). Our findings hold for cell lines derived from a variety of tissues, suggesting they are of general applicability. It was not possible to distinguish cell lines (normal and cancer) from human blood-derived mtDNA, as they exhibited similar properties with respect to the occurrences of deleterious mutations and the degree of heteroplasmy.

Stability of heteroplasmy in cell lines

The results above indicate that heteroplasmy exists within a cell population but questions remain about the nature of heteroplasmy in individual cells. A mixture of homoplasmic cells with different haplotypes can appear to be heteroplasmic. In order to establish heteroplasmy in individual cells, we placed the severest possible bottleneck on the population by deriving colonies from single cells, utilizing MDA-MB-157 and U20S breast carcinoma and osteosarcoma lines respectively (Figure 3). In each of the derived colonies (four colonies per cell line), variants from the original lines existed in the derived colonies at approximately the same frequencies as in the original colonies. The sharing of mutations between the original and derived colonies implies the diversity in mtDNA is present in individual cells. The preservation of the frequencies between the original and derived colonies further indicates that this heteroplasmy is uniform across cells in the original line (Figure 3). Low-frequency mutations (frequency <2%) in the original colony do disappear in the derived colonies, suggesting that the rarer variants might be present in a subset of cells, unlike the more common variants, and might be transient.

A simple model of mtDNA genetics assumes random assortment of mtDNA haplotypes between daughter cells upon cell division, followed by multiplication of mitochondria. This model would predict drift toward homoplasmy, as seen in our simulation of this process (Figure 4) and by others (37). The rate of drift in haplotype frequencies is a function of the number of mtDNA molecules per cell and the original frequency of the haplotypes (Figure 4). After many passages, irrespective of the original mtDNA distribution, the likelihood of two randomly selected cells having the same heteroplasmic mix would be extremely low, which is at odds with the stable and uniform heteroplasmy that we observed in the clonally derived cell lines. This suggests the existence of an active mechanism to counteract this drift.

Exchange of mtDNA between cells within a population is the simplest explanation for the uniformity of heteroplasmy and its stability. Exchange can counteract the effects of drift by bringing the haplotype distribution closer to the average across the population. Balancing selection (38), which arises from an active selection of haplotypes through interactions between them or segregated replication of groups with fixed composition (39), could also conceivably explain the lack of drift. However, these possibilities can be discounted because most variants are neutral and specific to each cell line, suggesting the selection needs to be different for each cell line without an obvious selective pressure. In addition, the co-culturing experiments that we discuss below suggest that the

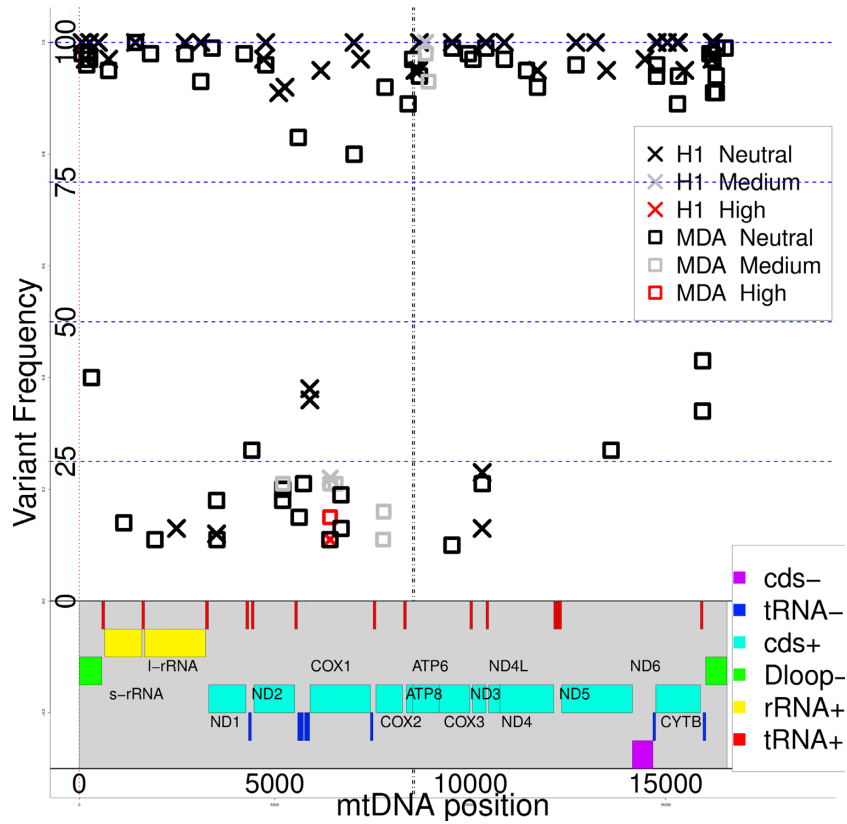


Figure 2. Variants in a cell line (MDA-MB-157, squares) and a human (PBMC, crosses) compared to the reference mtDNA sequence. Mutation frequency (y-axis) versus position on mtDNA (x-axis). Genes are colored bands at bottom of graph (+, - represent the strand, cds is coding sequence). *Neutral, low, medium* and *high* are the effect of the mutation on protein function (*Mutation Assessor* (35)). Except for the D-loop, most of the mtDNA codes for a transcript, with a few gaps. The 45 nt long overlapping region between ATP8 and ATP6 is marked by black vertical lines. Mutation frequencies between 0 and 1 arise from the co-existence of multiple mtDNA haplotypes. Heteroplasmy at a cellular level is demonstrated in Figure 3. There are no stark differences between the human and cell-line derived mtDNA.

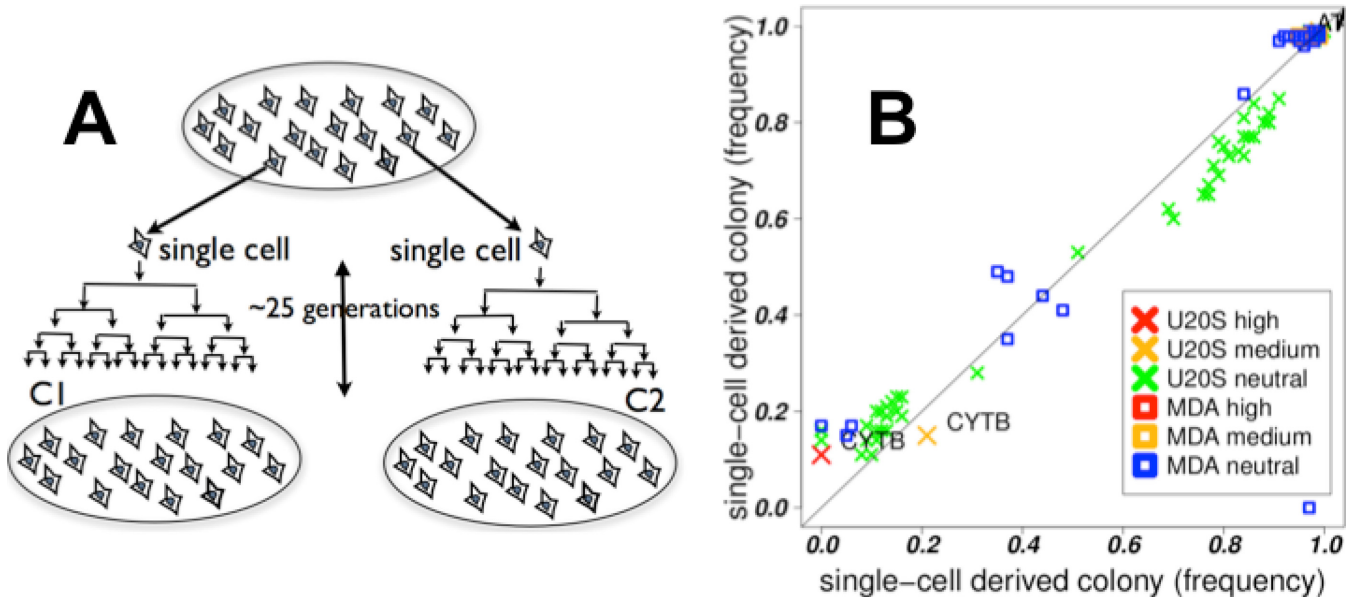


Figure 3. Stability of heteroplasmy. (A) C1 and C2 are colonies derived from single cells in the original colony, placing a severe bottleneck on the mtDNA, and then passaged ≈ 25 times. (B) Mutation frequencies in C1 (x-axis) versus C2 (y-axis) for two cell lines, U20S and MDA-MB-157. The mutations mostly lie along the diagonal; the heteroplasmic mix of mutations in the derived colonies are similar to each other. This implies the heteroplasmic mix exists at the single-cell level and is stable over many divisions. A drift in frequencies is expected with random assortment of mtDNA haplotypes (simulations, Figure 4). The stability of the frequencies implies active mechanisms to counteract the drift, such as the exchange of mtDNA between cells.

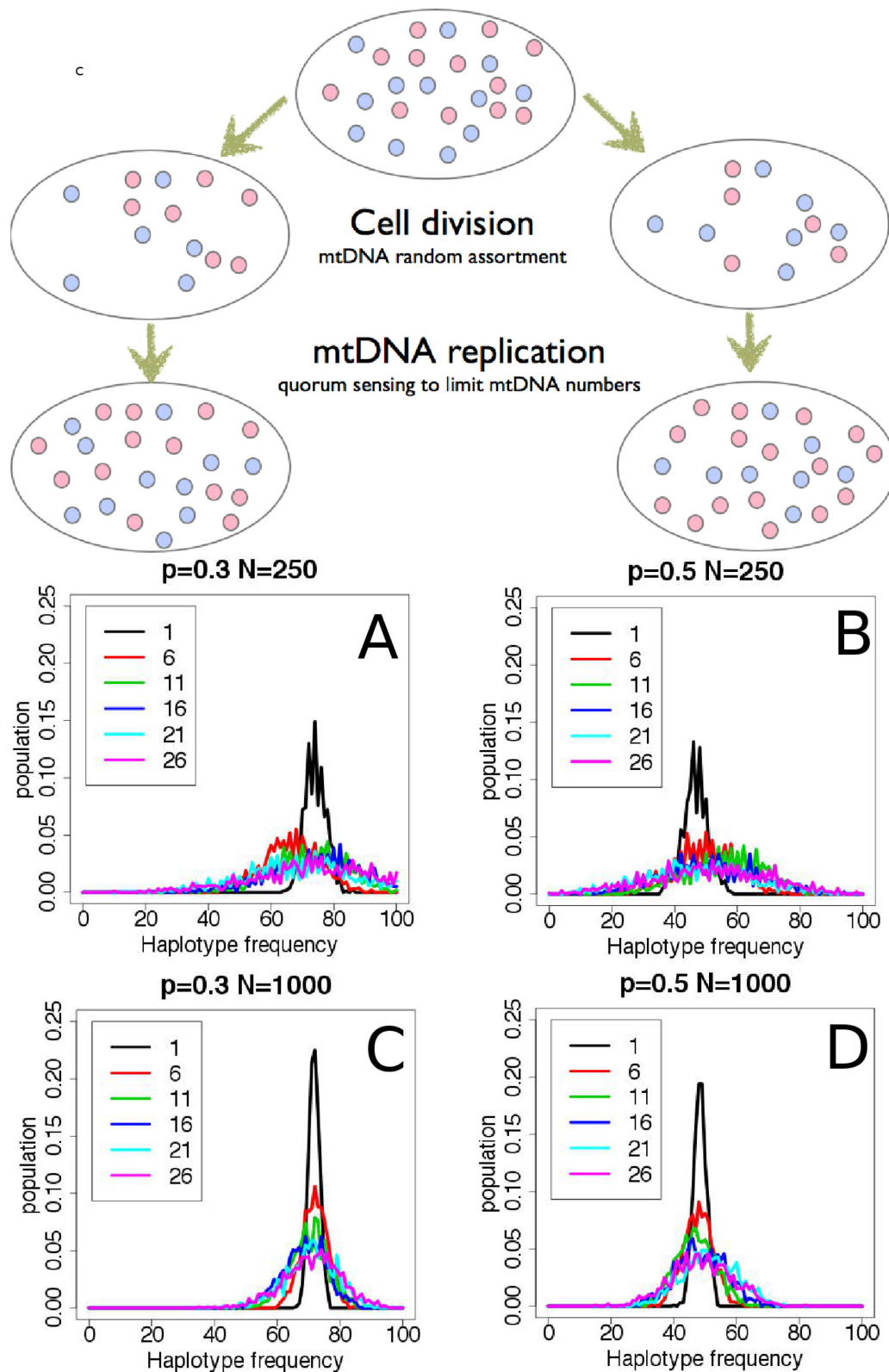


Figure 4. Simulations of mtDNA replication. Each cell contains a mixture of mtDNA haplotypes, here we consider only two species, red and blue, for the sake of simplicity. At cell division, the mitochondria assort randomly between the daughter cells and divide (with mtDNA replication) until a quorum of mitochondria, specific to each tissue-type, is reached. Quorum sensing is not well-understood, but the number of mitochondria per cell is tissue-specific (8). The graphs show simulations of the evolution of heteroplasmy over time, based on this model. The distribution of haplotype frequencies spreads over generations (drift), implying two randomly selected cells from a later generation are less likely to have the same mtDNA haplotype distribution. The plots show the distribution of frequencies after 1, 6, 11, 16, 21 and 26 divisions. The starting number of mitochondria per cell (N) is 250 in the upper panels (A, B) and 1000 in the lower panels (C, D). One of the alleles occurs with a frequency (P) of 0.3 in the left panels (A, C) and $P = 0.5$ in the right panels (B, D). The drift is slower and the distributions narrower with larger N and smaller deviations of P from 0.5.

composition of mtDNA haplotypes can change under appropriate conditions.

Experimental demonstration of mtDNA exchange between cells

In order to test the ability of mtDNA to transfer between cells, we co-cultured cell lines with distinct mtDNA heteroplasmy signatures. Since we were not sure which kinds of cells would allow transfer and if contact between cells were important, we tested mixtures of a pair of untransformed cell lines from embryo lung fibroblasts (IMR90 and WI-38) as well as four cancer cell lines (MDA-MB-157, U2OS, A382 and HCC1806).

One cell line in each of the co-cultured pair was labeled with constitutively expressed GFP, with 10-fold more non-GFP cells than GFP. After co-culturing for 4 weeks, the GFP labeled cells were isolated (either sorting by FACS or by using Neomycin resistance) and placed in culture again for up to 2 weeks, to obtain 10 million cells, which were then prepared for mtDNA sequencing. The details are given in the methods section. There were four experiments,

1. HCC1806-GFP mixed with MDA-MB-157 (both breast cancer) (FACS)
2. U2OS-GFP mixed with A382 (both cancer cells) (FACS)
3. MDA-MB-157-GFP (breast cancer) mixed with IMR90 (normal fibroblast) (Neomycin resistance)
4. WI-38-GFP mixed with IMR90 (normal fibroblast cells) (Neomycin resistance)

In two cases (WI-38-GFP, IMR90 shown in Figure 5 and Supplementary Table ST4) and (HCC1806-GFP, MDA-MB-157 shown in Supplementary Table ST5), we detected variants private to the non-GFP-labelled cell line in the co-cultured partner cell line, suggesting the transfer of mtDNA between the cell lines. The private variants were transferred to varying degrees, ranging from thorough mixing to no transfer, arguing against the results arising from errors in sorting or cytoplasmic/nuclear exchange between cells. Additionally, the purity of the sorted cells, based on FACS, excludes nuclear exchange as an explanation for the findings. The lack of transfer in the two pairs, (i) MDA-MB-157-GFP, IMR90 (Supplementary Table ST6) and (ii) U2OS-GFP, A382 (Supplementary Table ST7), further suggests that the transfer results are not artifacts. In all cases, the cultures were visually inspected to ensure that both cell lines were thriving.

DISCUSSION

Sensitive detection of heteroplasmy is important as its variability may have clinical significance, as a biomarker and in disease progression (40). Analyses of heteroplasmy are confounded by Numts, highlighted by a study that used whole-genome data from the TCGA to infer that deleterious mtDNA mutations are more common in cancer cells compared to normal tissue (15). In contrast, findings of low mutations rates in tumor mtDNA from a colorectal cancer study (14) are more in line with our findings that cancer cell lines do not exhibit higher rates of deleterious mutations

compared to normal cells from human tissues. A study of mtDNA from twins showed that Numts can influence the ascertainment of heteroplasmy, highlighting the need for further investigation of Numts (25).

Mseek provides the means to purify and deeply sequence mtDNA and determine heteroplasmy accurately by eliminating Numts and PCR-related biases. The sensitivity of Mseek increases with sequencing depth and can be made more accurate by using Duplex sequencing (36). Using longer, paired-end reads can overcome some of the problems arising from Numts. However, the ability to sequence mtDNA at a much lower cost strongly favors Mseek. As currently implemented, a limitation of Mseek is its requirement of at least 4 μ g of intact total DNA. Using Mseek we have demonstrated that cells from a wide range of cell lines and human samples exhibit heteroplasmy, in accord with results from several studies (14,19).

Heteroplasmy might be an essential feature of mtDNA, seemingly providing a *fingerprint* that can identify cells. A larger survey is needed to understand the resolution of this fingerprint and its ability to distinguish cellular origins. Such a fingerprint could be used to identify contamination of cell cultures, which might be more convenient than using short tandem repeat profiling (41). We found that mtDNAs from transformed human cell lines and primary human lymphocytes are similar with respect to the distributions of densities and frequencies of mutations (benign and deleterious ones).

By studying mtDNA from colonies derived from single cells, we showed that heteroplasmy is stable at the single-cell level, which is surprising in light of (i) the higher rates of mutation in mtDNA (42), which should increase the diversity of mtDNA, and (ii) drift, which should lead to homoplasmy in ~ 70 generations (14,37). Another study has determined the stability of heteroplasmy using PCR to track a particular mutation (A3243G), they proposed multiplication of segregating units with fixed mixtures of haplotypes to explain this stability (39). Intercellular exchange of mtDNA is the simplest explanation for this stability, on the basis of our cell-line data (Figure 3) in conjunction with simulations (Figure 4) and co-culturing experiments (Supplementary Tables ST4, ST5 and Figure 5).

The transfer of mtDNA most likely occurs through the transfer of either mitochondria-derived vesicles (43) or mitochondrial organelles. We cannot exclude transfer of free mtDNA, though the presence of DNA in the cytoplasm can trigger an innate immune response (44), that could lead to cell death. To ensure that the results are not artifacts, we utilized powerful biological selection including FACS selection of GFP-positive cells or a selectable resistance marker in one co-cultured cell line to ensure elimination of the other co-cultured line. For example, control experiments established that there was no detectable survival of cells lacking the selectable marker (less than one cell in a million). We also inspected the cells daily to make sure that the colonies were growing stably and the cultures were not taken over by one cell line. The Mseek data also provides evidence that the results are not artifacts, as the transfer of variants is not uniform, some variants show significant transfer while others show very little or none. Additionally, certain pairs of cell lines do not show mtDNA transfer despite being co-

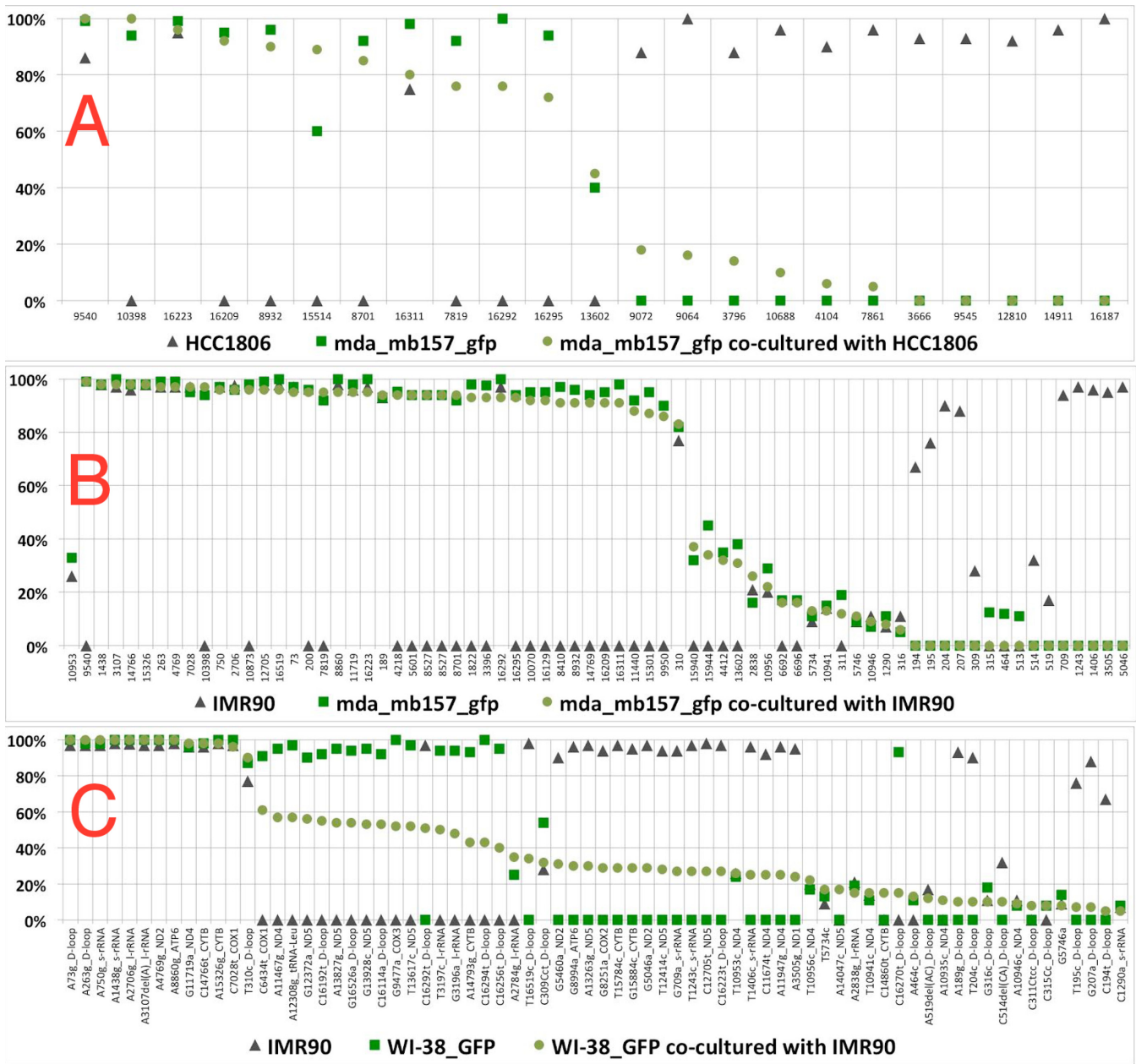


Figure 5. mtDNA transfer between co-cultured cells. The variants specific to each of the two cell lines before co-culturing (green squares for one and black triangles for the other) and their frequencies (y-axis) are depicted in the three plots. In panel A, after the co-culturing, the GFP-positive cells were selected by FACS sorting, in panels B and C the GFP-negative cells were killed using a neomycin selection marker which was incorporated with the GFP. In all cases, the remaining GFP-positive cells were regrown in culture and mtDNA variants sequenced (green circles). The variants on the x-axis are organized by the frequency of the green circles in descending order, the names of the variants identify their position, the alleles and the host gene. (A) Two cancer cell lines, MDA-MB-157 (GFP labeled, green squares) and HCC1806 (black triangles), were co-cultured. Transfer of variants from HCC1806 to MDA-MB-157 cells is observed, as the green circles are between the green square and black triangles for the most part. (B) A cancer cell line, MDA-MB-157 (GFP labeled, green squares), and a normal fibroblast cell line, IMR90 (black triangles), were co-cultured. No transfer of variants from IMR90 to MDAMB157 cells is observed, the green circles are roughly coincident with the green squares. (C) Two normal fibroblast cell lines, WI-38 (GFP labeled, green squares) and IMR90 (black triangles), were co-cultured. Transfer of variants from IMR90 to WI-38 cells is observed, as the green circles are intermediate between the black triangles and the green squares.

cultured for over 4 weeks, further suggesting that the data is not an artifact and contact between cells might be needed to facilitate organelle or vesicle transfer. It is possible that co-culturing cell-line pairs that exhibit transfer for a sufficiently long time will lead to a thorough mixing of mtDNA between the cells. We can exclude the possibility of allelic drift, as the cell lines passaged on their own exhibit stable allele frequencies over multiple generations.

Intercellular transfer of mtDNA has been seen in other instances. Horizontal transfer of genetic material between species of yeasts has been shown (45). Organelle transfer between cells through microtubule formation is increasingly a focus of many studies (46). Exchange of mtDNA between mitochondria within a cell is facilitated through networks created by fusion, mediated by the mitofusins *Mfn1* and *Mfn2* (47). This intracellular exchange is essential for functional mitochondria; knocking out the fusins causes muscles to atrophy through the accumulation of deleterious mtDNA mutations (47). *In vivo*, exchanges of mitochondria between cells have also been demonstrated in the rejuvenation of cells with damaged mitochondria by transfer of functional mitochondria from mesenchymal stem cells (48). Rejuvenation of cells containing damaged mtDNA by transfer of functional mtDNA from neighboring cells in culture has also been observed (49).

The present study is the first explicit demonstration of mtDNA transfer between healthy human cells in culture. Prior studies have shown transfer from cells with functional mtDNA into ones with non-functioning mtDNA using protein markers (48,49). The exchange of mtDNA between cells may help to explain its stability over the lifetime of an organism, which can be inferred from the relative paucity of age-related disorders originating from somatic mtDNA mutations. This exchange may also explain the stability of mtDNA over generations, which can be inferred from the distinct geographic signatures in human mtDNA (50). The stability of mtDNA against deleterious mutations could be enhanced by a coupling between replication and transcription (51), ensuring the depletion of non-functional mtDNA by inefficiencies in replication.

The sequencing of mtDNA in cell lines allows us to understand the nature of mtDNA variability and its maintenance in cell populations. Somatic mutations in mtDNA could play a role in various human disorders and in aging, especially when the transfer of mtDNA between cells is impeded. Thus, mechanisms involved in mtDNA transfer might be fruitful targets for therapeutic intervention. The transfer of functional mtDNA into diseased cells could be used to treat disorders arising from mtDNA defects. There is great value in surveying large populations in order to establish the normal range of heteroplasmy for use in genome-wide association studies (GWAS). By making mtDNA sequencing economical, Mseek enables large-scale studies of heteroplasmy for GWAS applications and clinical monitoring of mtDNA in tissues.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

We thank Brian Brown, Eirini Papapetrou, Viviana Simon, Benjamin tenOever, Avinash Waghray, Sunniva Bjorklund and Vessela Kristensen for helpful discussions and numerous suggestions.

FUNDING

National Institutes of Health (NIH) [1R21HG007394-01 to R.S., A.J., L.L.]; Tisch Cancer Institute (TCI) [to J.S.]; A developmental award from the TCI and a pilot grant from the Venture Capital Research Funding Program of Children's Environmental Health Center (CEHC) at Mount Sinai funded several experiments; National Cancer Institute (NCI) [to E.K.B.]; Breast Cancer Research foundation [to S.A.A.]. Funding for open access charge: NIH [1R21HG007394-01].

Conflict of interest statement. A.J. and R.S. have filed for a patent on Mseek and founded Girihlet Inc. to provide wider access to Mseek. Swapna Gone and Masoud Toloue are employees of a commercial enterprise that designs and produces kits for next generation sequencing.

REFERENCES

- Wallace,D.C. (1982) Structure and evolution of organelle genomes. *Microbiol. Rev.*, **46**, 208–240.
- Wallace,D.C. (1999) Mitochondrial diseases in man and mouse. *Science*, **283**, 1482–1488.
- Liu,X., Kim,C.N., Yang,J., Jemmerson,R. and Wang,X. (1996) Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome c. *Cell*, **86**, 147–157.
- Newmeyer,D.D. and Ferguson-Miller,S. (2003) Mitochondria: releasing power for life and unleashing the machineries of death. *Cell*, **112**, 481–490.
- Westermann,B. and Neupert,W. (2003) 'Omics' of the mitochondrion. *Nat. Biotechnol.*, **21**, 239–240.
- Brandon,M.C., Lott,M.T., Nguyen,K.C., Spolim,S., Navathe,S.B., Baldi,P. and Wallace,D.C. (2005) MITOMAP: a human mitochondrial genome database—2004 update. *Nucleic Acids Res.*, **33**, D611–D613.
- Legros,F., Malka,F., Frachon,P., Lombes,A. and Rojo,M. (2004) Organization and dynamics of human mitochondrial DNA. *J. Cell Sci.*, **117**, 2653–2662.
- Fernandez-Vizarrá,E., Enriquez,J.A., Perez-Martos,A., Montoya,J. and Fernandez-Silva,P. (2011) Tissue-specific differences in mitochondrial activity and biogenesis. *Mitochondrion*, **11**, 207–213.
- Wallace,D.C. (1994) Mitochondrial DNA mutations in diseases of energy metabolism. *J. Bioenerg. Biomembr.*, **26**, 241–250.
- DiMauro,S. and Schon,E.A. (2001) Mitochondrial DNA mutations in human disease. *Am. J. Med. Genet.*, **106**, 18–26.
- Lebon,S., Chol,M., Benit,P., Mugnier,C., Chretien,D., Giurgea,I., Kern,I., Girardin,E., Hertz-Pannier,L., de Lonlay,P. *et al.* (2003) Recurrent de novo mitochondrial DNA mutations in respiratory chain deficiency. *J. Med. Genet.*, **40**, 896–899.
- Taylor,R.W. and Turnbull,D.M. (2005) Mitochondrial DNA mutations in human disease. *Nat. Rev. Genet.*, **6**, 389–402.
- Greaves,L.C. and Taylor,R.W. (2006) Mitochondrial DNA mutations in human disease. *IUBMB life*, **58**, 143–151.
- He,Y., Wu,J., Dressman,D.C., Iacobuzio-Donahue,C., Markowitz,S.D., Velculescu,V.E., Diaz Luis,A.J., Kinzler,K.W., Vogelstein,B. *et al.* (2010) Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature*, **464**, 610–614.
- Larman,T.C., DePalma,S.R., Hadjipanayis,A.G., Protopopov,A., Zhang,J., Gabriel,S.B., Chin,L., Seidman,C.E., Kucherlapati,R. and Seidman,J.G. (2012) Spectrum of somatic mitochondrial mutations in five cancers. *Proc. Natl Acad. Sci. U.S.A.*, **109**, 14087–14091.
- Goto,H., Dickins,B., Afgan,E., Paul,I.M., Taylor,J., Makova,K.D. and Nekrutenko,A. (2011) Dynamics of mitochondrial heteroplasmy

- in three families investigated via a repeatable re-sequencing study. *Genome Biol.*, **12**, R59.
17. Avital, G., Buchshtav, M., Zhidkov, I., Feder, J.T., Dadon, S., Rubin, E., Glass, D., Spector, T.D. and Mishmar, D. (2012) Mitochondrial DNA heteroplasmy in diabetes and normal adults: role of acquired and inherited mutational patterns in twins. *Hum. Mol. Genet.*, **21**, 4214–4224.
 18. Ye, K., Lu, J., Ma, F., Keinan, A. and Gu, Z. (2014) Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc. Natl Acad. Sci. U.S.A.*, **111**, 10654–10659.
 19. Diroma, M.A., Calabrese, C., Simone, D., Santorsola, M., Calabrese, F.M., Gasparre, G. and Attimonelli, M. (2014) Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. *BMC Genomics*, **15**, S2.
 20. Birsoy, K., Possemato, R., Lorbeer, F.K., Bayraktar, E.C., Thiru, P., Yucel, B., Wang, T., Chen, W.W., Clish, C.B. and Sabatini, D.M. (2014) Metabolic determinants of cancer cell sensitivity to glucose limitation and biguanides. *Nature*, **508**, 108–112.
 21. Woischnik, M. and Moraes, C.T. (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.*, **12**, 885–893.
 22. Lascaro, D., Castellana, S., Gasparre, G., Romeo, G., Saccone, C. and Attimonelli, M. (2008) The RHNumtS compilation: features and bioinformatics approaches to locate and quantify Human NumtS. *BMC Genomics*, **9**, 267.
 23. Calabrese, F.M., Simone, D. and Attimonelli, M. (2012) Primates and mouse NumtS in the UCSC genome browser. *BMC Bioinformatics*, **13**, S15.
 24. Calabrese, C., Simone, D., Diroma, M.A., Santorsola, M., Gutt, C., Gasparre, G., Picardi, E., Pesole, G. and Attimonelli, M. (2014) MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinformatics*, **30**, 3115–3117.
 25. Bouhhal, Y., Martinez, S., Gong, H., Dumas, K. and Shieh, J.T.C. (2013) Twin mitochondrial sequence analysis. *Mol. Genet. Genomic Med.*, **1**, 174–186.
 26. Vallone, P.M. (2012) Capillary electrophoresis of an 11-plex mtDNA coding region SNP single base extension assay for discrimination of the most common Caucasian HV1/HV2 mitotype. *Methods Mol. Biol.*, **830**, 159–167.
 27. Paull, D., Emmanuele, V., Weiss, K.A., Treff, N., Stewart, L., Hua, H., Zimmer, M., Kahler, D.J., Golland, R.S., Noggle, S.A. et al. (2012) Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature*, **493**, 632–637.
 28. Maricic, T., Whitten, M. and Paabo, S. (2010) Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One*, **5**, e14004.
 29. Lang, B.F. and Burger, G. (2007) Purification of mitochondrial and plastid DNA. *Nat. Protoc.*, **2**, 652–660.
 30. Ameer, A., Stewart, J.B., Freyer, C., Hagstrom, E., Ingman, M., Larsson, N.-G. and Gyllenstein, U. (2011) Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. *PLoS Genet.*, **7**, e1002028.
 31. Picardi, E. and Pesole, G. (2012) Mitochondrial genomes gleaned from human whole-exome sequencing. *Nat. Methods*, **9**, 523–524.
 32. McKernan, K.J., Spangler, J., Zhang, L., Tadigotla, V., McLaughlin, S., Warner, J., Zare, A. and Boles, R.G. (2014) Expanded genetic codes in next generation sequencing enable decontamination and mitochondrial enrichment. *PLoS One*, **9**, e96492.
 33. Benson, E.K., Mungamuri, S.K., Attie, O., Kracikova, M., Sachidanandam, R., Manfredi, J.J. and Aaronson, S.A. (2014) p53-dependent gene repression through p21 is mediated by recruitment of E2F4 repression complexes. *Oncogene*, **33**, 3959–3969.
 34. Subramanian, S., Pierro, V.D., Shah, H., Jayaprakash, A.D., Weisberger, I., Shim, J., George, A., Gelb, B.D. and Sachidanandam, R. (2013) MiST: a new approach to variant detection in deep sequencing datasets. *Nucleic Acids Res.*, **41**, e154.
 35. Reva, B., Antipin, Y. and Sander, C. (2011) Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.*, **39**, e118.
 36. Schmitt, M.W., Kennedy, S.R., Salk, J.J., Fox, E.J., Hiatt, J.B. and Loeb, L.A. (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl Acad. Sci. U.S.A.*, **109**, 14508–14513.
 37. Collier, H.A., Khrapko, K., Bodyak, N.D., Nekhaeva, E., Herrero-Jimenez, P. and Thilly, W.G. (2001) High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nat. Genet.*, **28**, 147–150.
 38. Connallon, T. and Clark, A.G. (2014) Balancing selection in species with separate sexes: insights from fisher's geometric model. *Genetics*, **197**, 991–1006.
 39. Raap, A.K., Jahangir Tafrechi, R.S., van de Rijke, F.M., Pyle, A., Whlby, C., Szuhai, K., Ravelli, R.B.G., de Coo, R.F.M., Rajasimha, H.K., Nilsson, M. et al. (2012) Non-random mtDNA segregation patterns indicate a metastable heteroplasmic segregation unit in m.3243A>G hybrid cells. *PLoS One*, **7**, e52080.
 40. Wallace, D.C. and Chalkia, D. (2013) Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb. Perspect. Biol.*, **5**, a021220.
 41. Marx, V. (2014) Cell-line authentication demystified. *Nat. Methods*, **11**, 483–488.
 42. Dowling, D.K., Friberg, U. and Lindell, J. (2008) Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol. Evol.*, **23**, 546–554.
 43. Sugiura, A., McLelland, G.-L., Fon, E.A. and McBride, H.M. (2014) A new pathway for mitochondrial quality control: mitochondrial-derived vesicles. *EMBO J.*, **33**, 2142–2156.
 44. Zhang, Q., Raoof, M., Chen, Y., Sumi, Y., Sursal, T., Junger, W., Brohi, K., Itagaki, K. and Hauser, C.J. (2010) Circulating mitochondrial DAMPs cause inflammatory responses to injury. *Nature*, **464**, 104–107.
 45. Marinoni, G., Manuel, M., Petersen, R.F., Hvidtfeldt, J., Sulo, P. and Pikur, J. (1999) Horizontal transfer of genetic material among *Saccharomyces* yeasts. *J. Bacteriol.*, **181**, 6488–6496.
 46. Rogers, R.S. and Bhattacharya, J. (2013) When cells become organelle donors. *Physiology*, **28**, 414–422.
 47. Chen, H., Vermulst, M., Wang, Y.E., Chomyn, A., Prolla, T.A., McCaffery, J.M. and Chan, D.C. (2010) Mitochondrial fusion is required for mtDNA stability in skeletal muscle and tolerance of mtDNA mutations. *Cell*, **141**, 280–289.
 48. Islam, M.N., Das, S.R., Emin, M.T., Wei, M., Sun, L., Westphalen, K., Rowlands, D.J., Quadri, S.K., Bhattacharya, S. and Bhattacharya, J. (2012) Mitochondrial transfer from bone-marrow-derived stromal cells to pulmonary alveoli protects against acute lung injury. *Nat. Med.*, **18**, 759–765.
 49. Spees, J.L., Olson, S.D., Whitney, M.J. and Prockop, D.J. (2006) Mitochondrial transfer between cells can rescue aerobic respiration. *Proc. Natl Acad. Sci. U.S.A.*, **103**, 1283–1288.
 50. Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. and Barbujani, G. (2000) Geographic Patterns of mtDNA Diversity in Europe. *Am. J. Hum. Genet.*, **66**, 262–278.
 51. Clayton, D.A. (2000) Transcription and replication of mitochondrial DNA. *Hum. Reprod.*, **15**(Suppl. 2), 11–17.