Supporting Information Appendix

Progressive Increase in mtDNA 3243A>G Heteroplasmy Causes Abrupt Transcriptional Reprogramming

Martin Picard, Jiangwen Zhang, Saege Hancock, Olga Derbeneva, Ryan Golhar, Pawel Golik, Sean O'Hearn, Shawn E Levy, Prasanth Potluri, Maria Lvova, Antonio Davila, Chun Shi Lin, Juan Carlos Perin, Eric F Rappaport, Hakon Hakonarson, Ian Trounce, Vincent Procaccio and Douglas C Wallace

Table of Contents

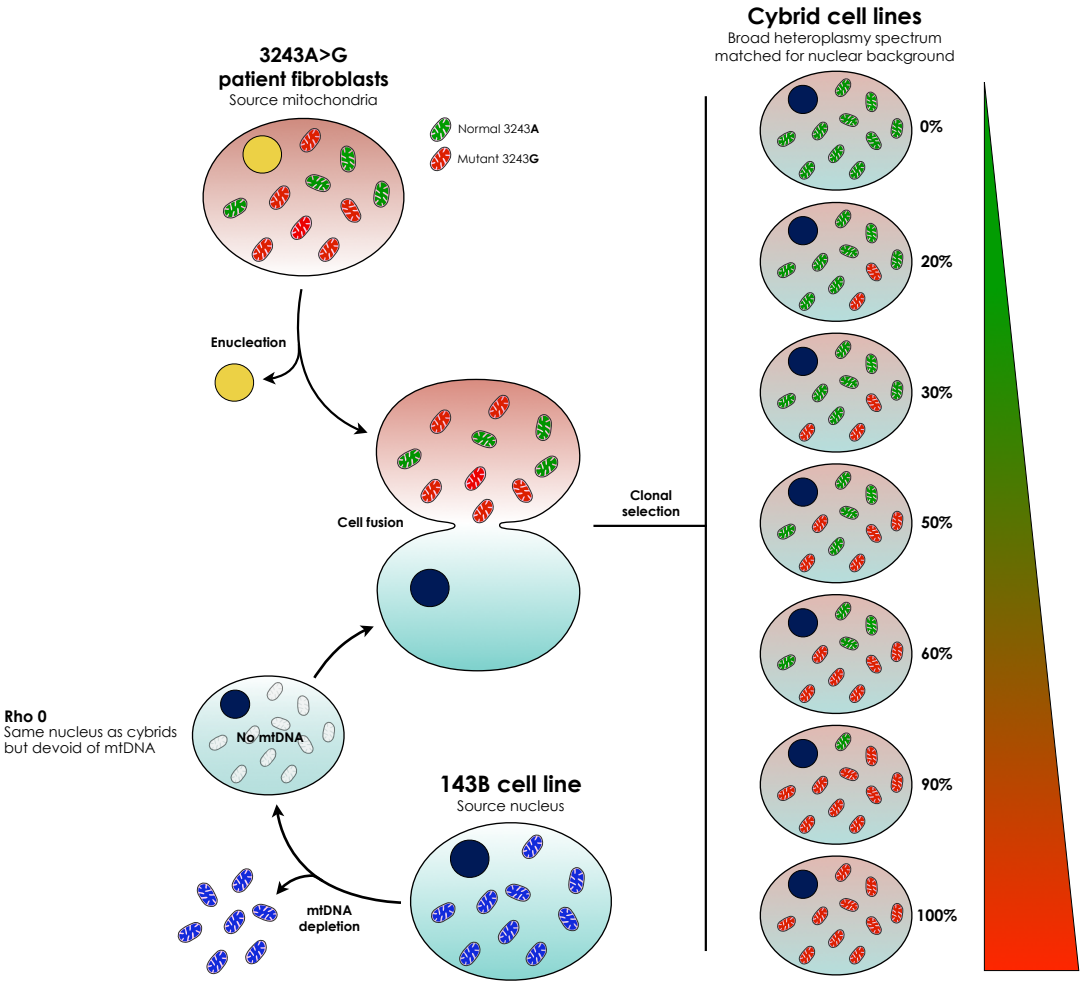


Fig. S1 Experimental paradigm for generating cybrid cell lines spanning the full spectrum of heteroplasmy for the 3243A>G point mutation. See Extended Experimental Procedures for details.

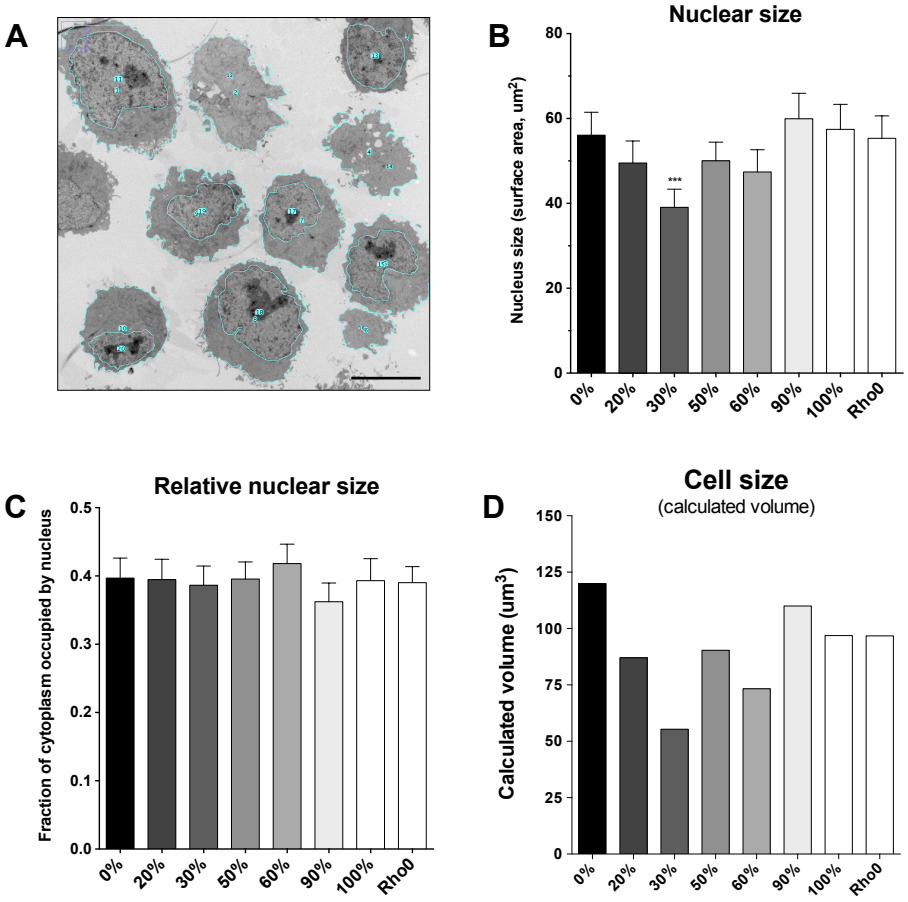
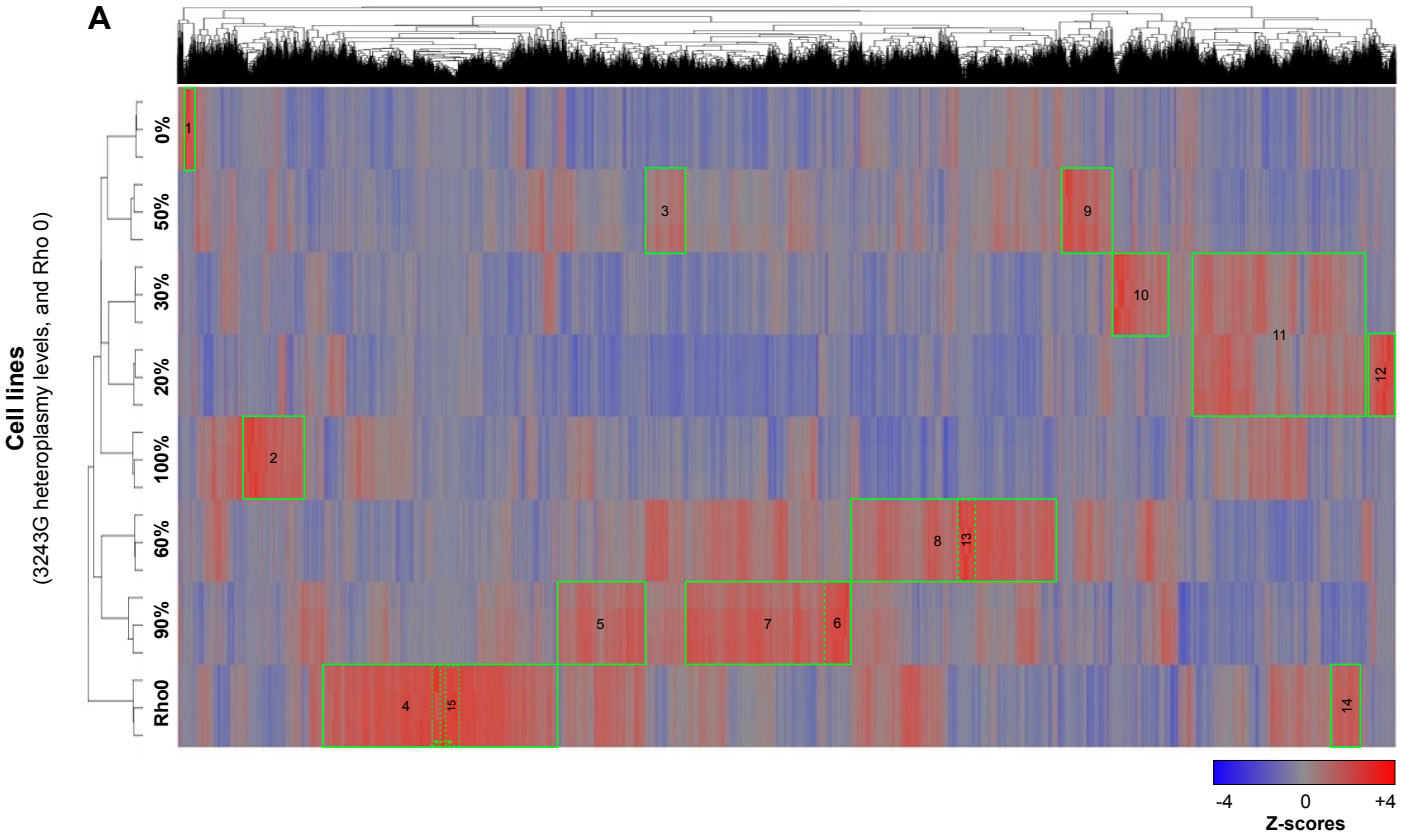


Fig. S2. Quantification of nuclear size in cell lines with 3243A>G heteroplasmy. (**A**) Cell and nucleus contour were manually traced in Image J from cell suspensions processed and imaged by electron microscopy. Scale bar = 10µm. (**B**) Mean nuclear size and (**C**) fraction of cytoplasmic space occupied by cell nucleus, with 95 % confidence interval. Mean measured cell surface area is shown in Fig. 1H, and was used to calculate (**D**) cell volume. *** : $P < 0.001$ vs 0%, n = 130-209 cells per group.

A



B

1 • alcohol transmembrane transporter activity • glycerol channel activity • chemokine receptor activity • G-protein coupled chemoattractant receptor activity • glycerol transmembrane transporter activity • polyol transmembrane transporter activity	**2** • alcohol dehydrogenase (NADP+) activity • androsterone dehydrogenase (A-specific) activity • glutamate:sodium symporter activity • anion transmembrane transporter activity • carbonate dehydratase activity • peroxiredoxin activity	**3** • structural constituent of ribosome • RNA binding • nucleotide binding • small molecule binding • isomerase activity • catalytic activity	**4** • zinc ion binding • cation binding • ion binding • metal ion binding • transition metal ion binding • DNA binding	**5** • structural constituent of ribosome • RNA binding • binding • purine ribonucleoside triphosphate binding • nucleotide binding • purine nucleotide binding
6 • hydroxyacylglutathione hydrolase activity • magnesium ion binding • fatty acid elongase activity • intramolecular oxidoreductase activity, interconverting aldoses and ketoses • purine nucleobase transmembrane transporter activity • nucleobase-containing compound transmembrane transporter activity	**7** • catalytic activity • protein binding • nucleotide binding • small molecule binding • ATP binding • purine ribonucleoside triphosphate binding	**8** • transcription cofactor activity • nucleic acid binding • DNA binding • transcription factor binding transcription factor activity • protein binding transcription factor activity • chromatin binding	**9** • binding • aminoacyl-tRNA editing activity • leucine-tRNA ligase activity • isomerase activity • purine nucleotide binding • adenyl nucleotide binding	**10** • methylenetetrahydrofolate dehydrogenase (NAD+) activity • oxidoreductase activity, acting on the CH-NH group of donors, NAD or NADP as acceptor • methenyltetrahydrofolate cyclohydrolase activity • melatonin receptor activity • symporter activity • aminopeptidase activity
11 • catalytic activity • transferase activity • small conjugating protein ligase activity • protein serine/threonine phosphatase activity • acid-amino acid ligase activity • ligase activity, forming carbon-nitrogen bonds	**12** • D-glucose transmembrane transporter activity • glucose transmembrane transporter activity • hexose transmembrane transporter activity • monosaccharide transmembrane transporter activity • integrin binding • structural constituent of eye lens	**13** • calcitonin receptor activity • chemokine receptor activity • G-protein coupled chemoattractant receptor activity • receptor activity • cytokine receptor activity • cation channel activity	**14** • nucleic acid binding • binding • DNA binding • protein transmembrane transporter activity • macromolecule transmembrane transporter activity • transferase activity, transferring phosphorus-containing groups	**15** • calcium-dependent phospholipid binding • passive transmembrane transporter activity • channel activity • gated channel activity • ion channel activity • excitatory extracellular ligand-gated ion channel activity

Fig. S3. Different heteroplasmy levels induce unique gene expression profiles. (**A**) Cluster composition analysis from unsupervised hierarchical clustering of whole-transcriptome data. Vertical elements are differentially expressed genes (n = 15,652, ANOVA model P < 0.0001) hierarchically ordered according to their similarity (Pearson correlation) in expression profiles. Horizontal elements represent each cell line hierarchically ordered as for genes, in which the RNA-Seq was performed in triplicates (see Extended Experimental Procedures for details). (**B**) Listed are the 6 most significant molecular functions as defined by gene ontology (GO) terms associated with gene clusters numbered within the heatmap in (A).

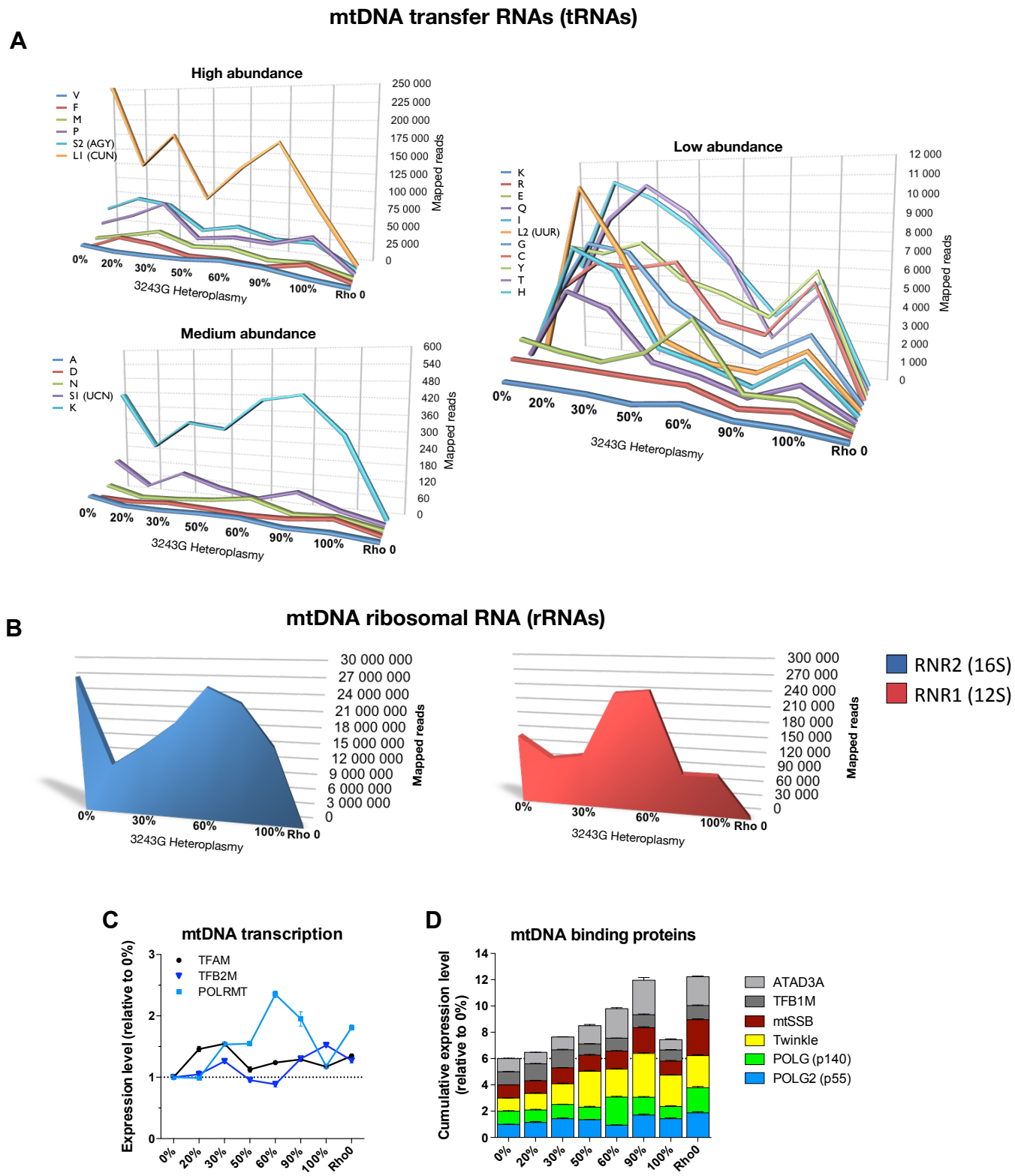


Fig. S4. Transcript levels for mtDNA-encoded tRNAs and rRNAs, and nucleoid components in response to mtDNA 3243A>G heteroplasmy. Shown are normalized mtDNA-mapped read counts normalized to total reads from RNA sequencing data of (**A**) transfer (tRNAs) and (**B**) ribosomal RNA (rRNA) genes across 3243G heteroplasmy levels. Data in (A) are plotted in three different panels to facilitate resolution of widely different transcript abundance between genes (note y axis ranges). No major light versus heavy strand effect was observed. Data are averages of measurements performed in triplicates. (**C**) Relative transcript levels for genes encoding the mtDNA transcription machinery, including transcription factor of mitochondria (TFAM), transcription factor B2 (TFB2M), and mitochondrial RNA polymerase (POLMRT); and (**D**) other mtDNA binding proteins including the ATPase Family, AAA Domain Containing 3 isoform A (ATAD3A), transcription factor B1 (TFB1M), Single-Stranded DNA Binding Protein 1 (mtSSB), Twinkle helicase, and the two subunits of the mitochondrial DNA polymerase (POLG and POLG2) (Gilkerson et al. Cold Spring Harb Perspect Biol 2013; 5(5): a011080).

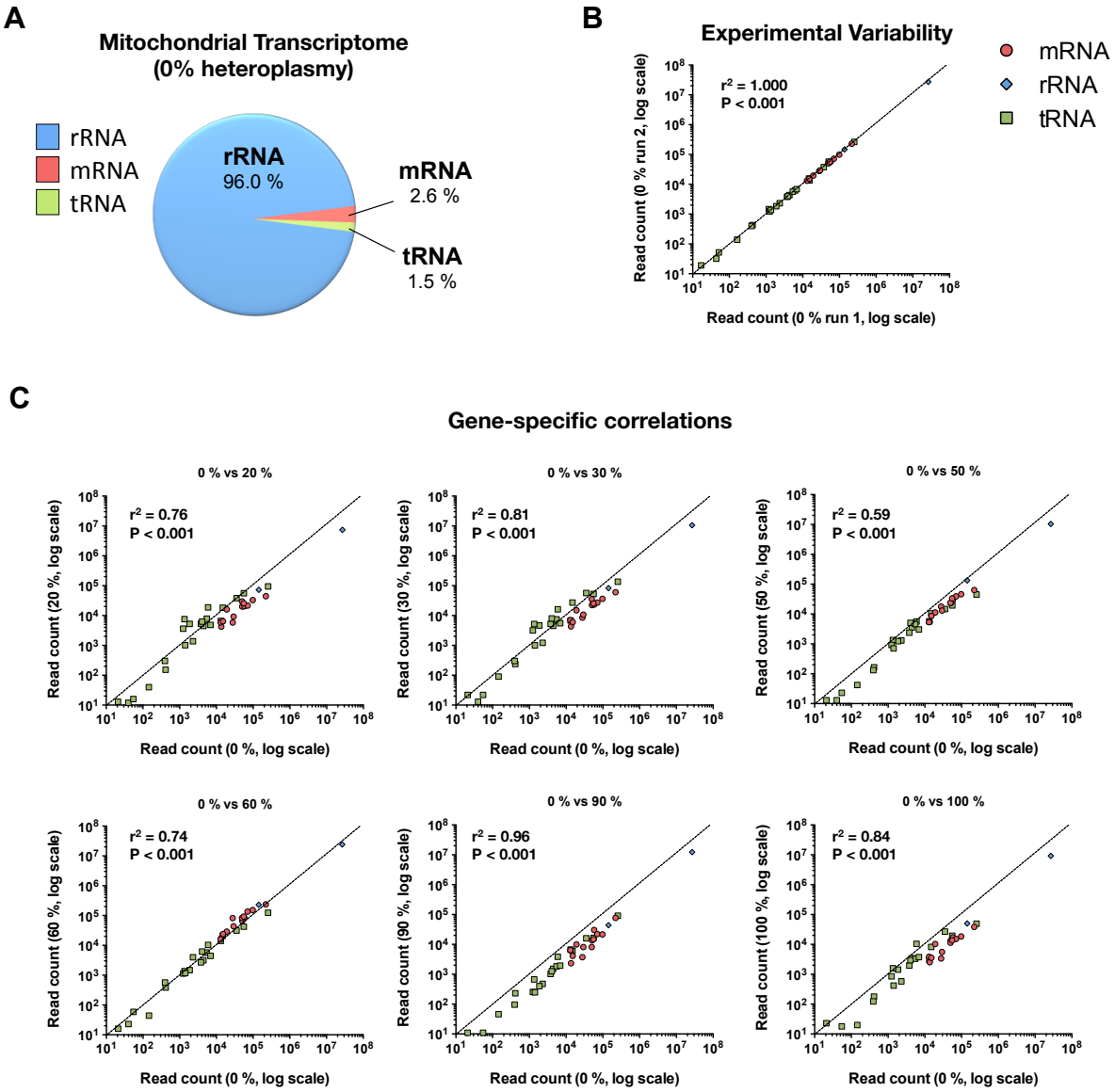


Fig. S5. Composition of the mitochondrial transcriptome and impact of 3243A>G heteroplasmy on mtDNA gene expression. (**A**) Relative abundance of ribosomal (rRNA), messenger (mRNA) and transfer (tRNAs) at baseline in the 0% heteroplasmy line. The relative abundance of transcripts for the different heteroplasmy levels are shown in main Fig. 3B. (**B**) Correlations of transcript levels on a per-gene basis, for all 37 mtDNA-encoded transcripts, measured from two different runs to assess experimental variability. Diagonal dotted line indicates an equal ratio (1:1) between runs. (**C**) Correlation as in B, but between the 0% heteroplasmy cell line and the other 3243G cell lines, revealing the upregulation of some transcripts (above dotted line), and downregulation of others (below dotted line). Data points represent the mean of three independent RNA-Seq runs for each gene.

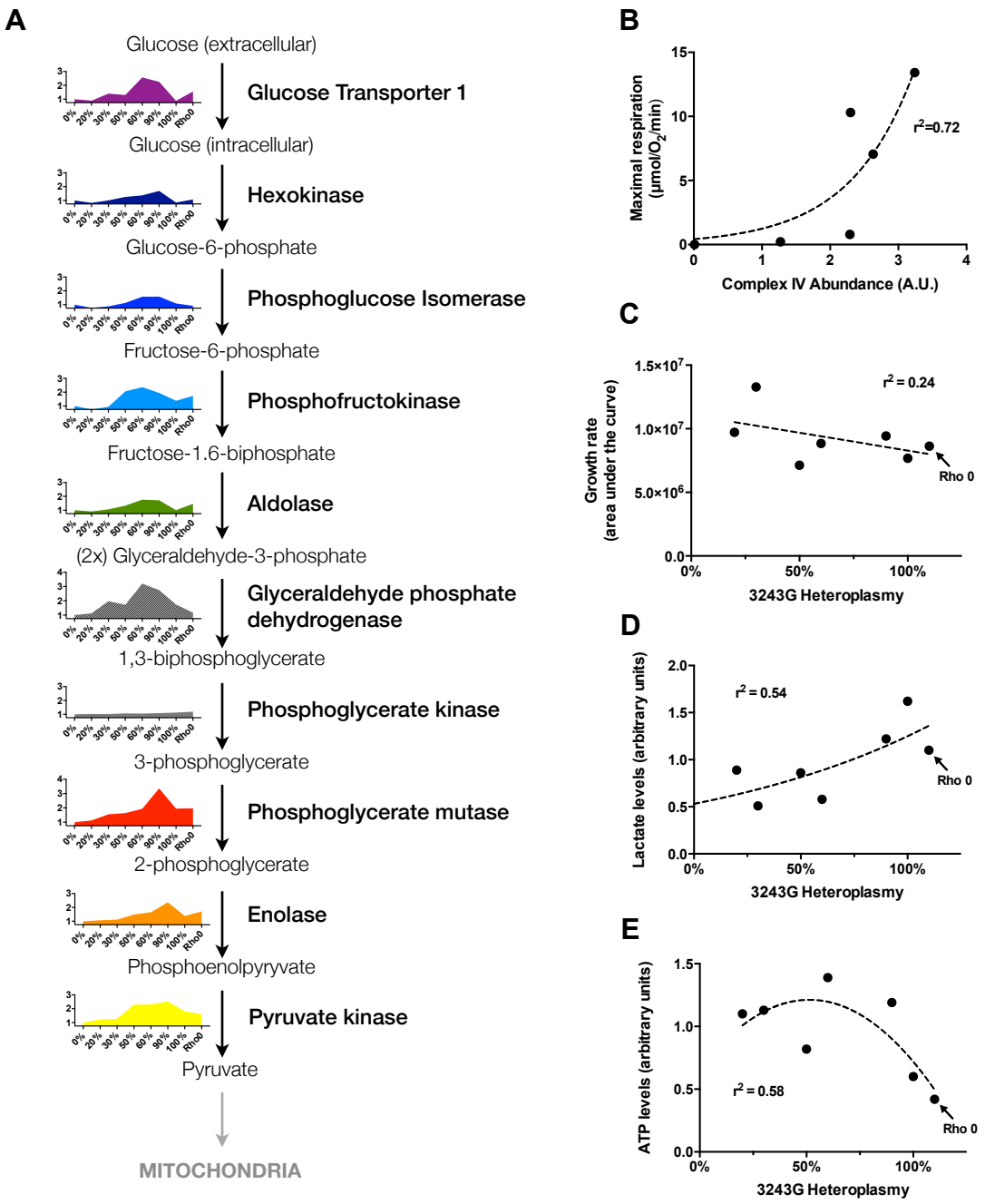


Fig. S6. Altered energy metabolism and in response to increasing 3243G heteroplasmy. (**A**) Step-by-step analysis of gene expression for enzymatic reactions of the glycolysis pathway. (**B**) Correlation (exponential) between the abundance of assembled complex IV (cytochrome c oxidase, COX) measured by Blue Native gel and maximal uncoupled respiration measured in the presence of the uncoupler CCCP. (**C**) Correlation (linear) between heteroplasmy and growth rate. (**D**) Correlation (exponential) between mtDNA heteroplasmy and lactate levels measured from the culture medium. (**E**) Correlation (polynomial second order) between heteroplasmy and ATP levels.

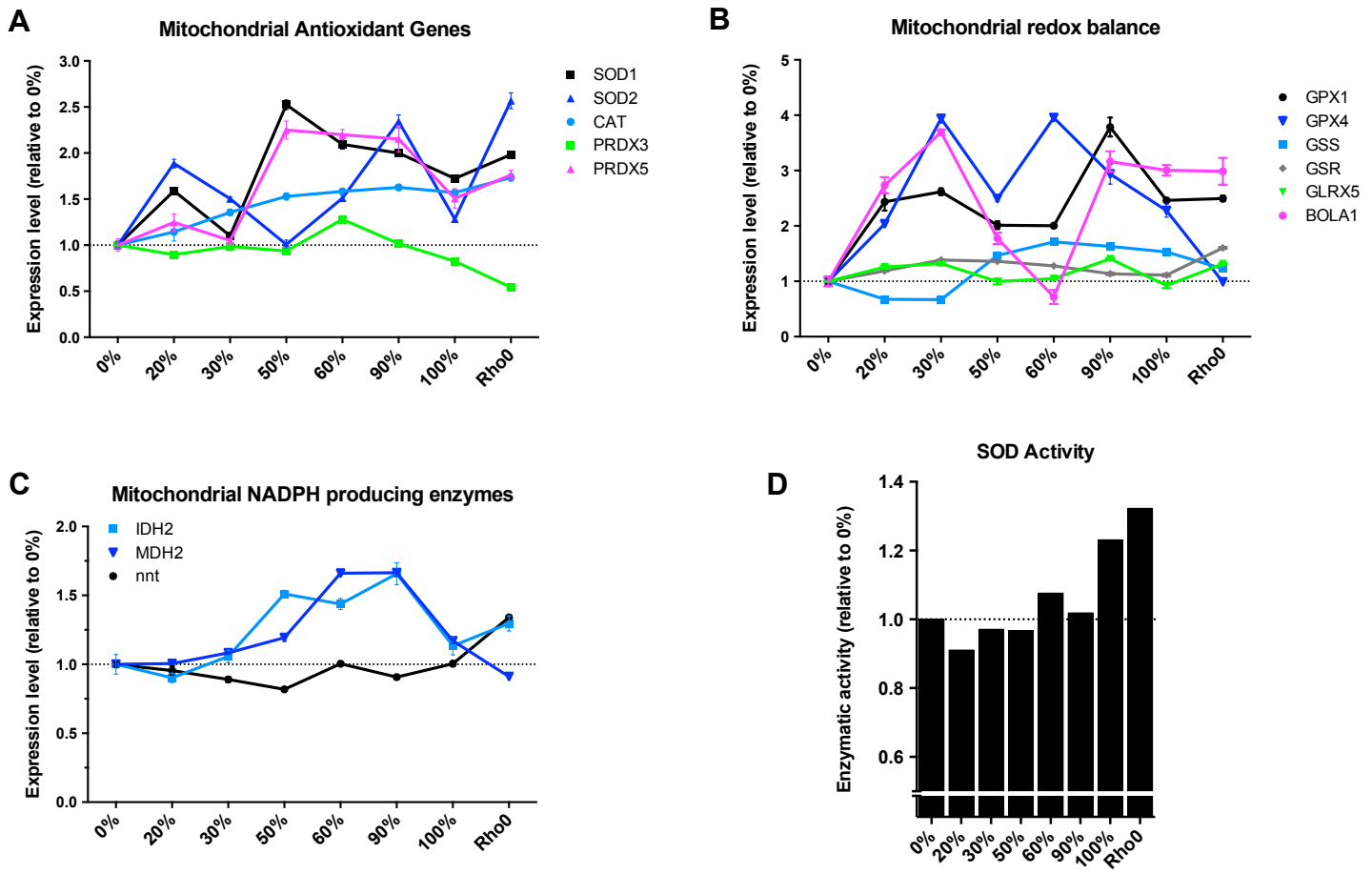


Fig. S7. Induction of antioxidant defenses and redox balance with mtDNA 3243A>G heteroplasmy. (**A**) Transcript levels for antioxidant enzymes localized to mitochondria, including SOD1 (CuZn SOD, superoxide dismutase 1), SOD2 (MnSOD, superoxide dismutase 2), CAT (catalase), PRDX3 and PRDX5 (peroxiredoxins 3 and 5). (**B**) Transcript levels for enzymes involved in the regulation of the mitochondrial redox balance, including BOLA1 (BolA-like 1), GSR (glutathione reductase, mitochondrial), GLRX5 (glutaredoxin-related protein 5, mitochondrial), GPX1 (glutathione peroxidase 1), GSS (glutathione synthetase) and GPX4 (glutathione peroxidase 4, also known as phospholipid hydroperoxide glutathione peroxidase, mitochondrial). (**C**) Transcript levels for the three mitochondrial NADPH-producing enzymes IDH2 (isocitrate dehydrogenase 2), MDH2 (malonate dehydrohenase 2), and nicotinamide nucleotide transhydrogenase (nnt). Shown are means ± S.E.M., n = 3 per group. (**D**) Total superoxide dismutase (SOD) activity measured from total cell lyzates (P for linear trend = 0.017) (see Materials and Methods for details).

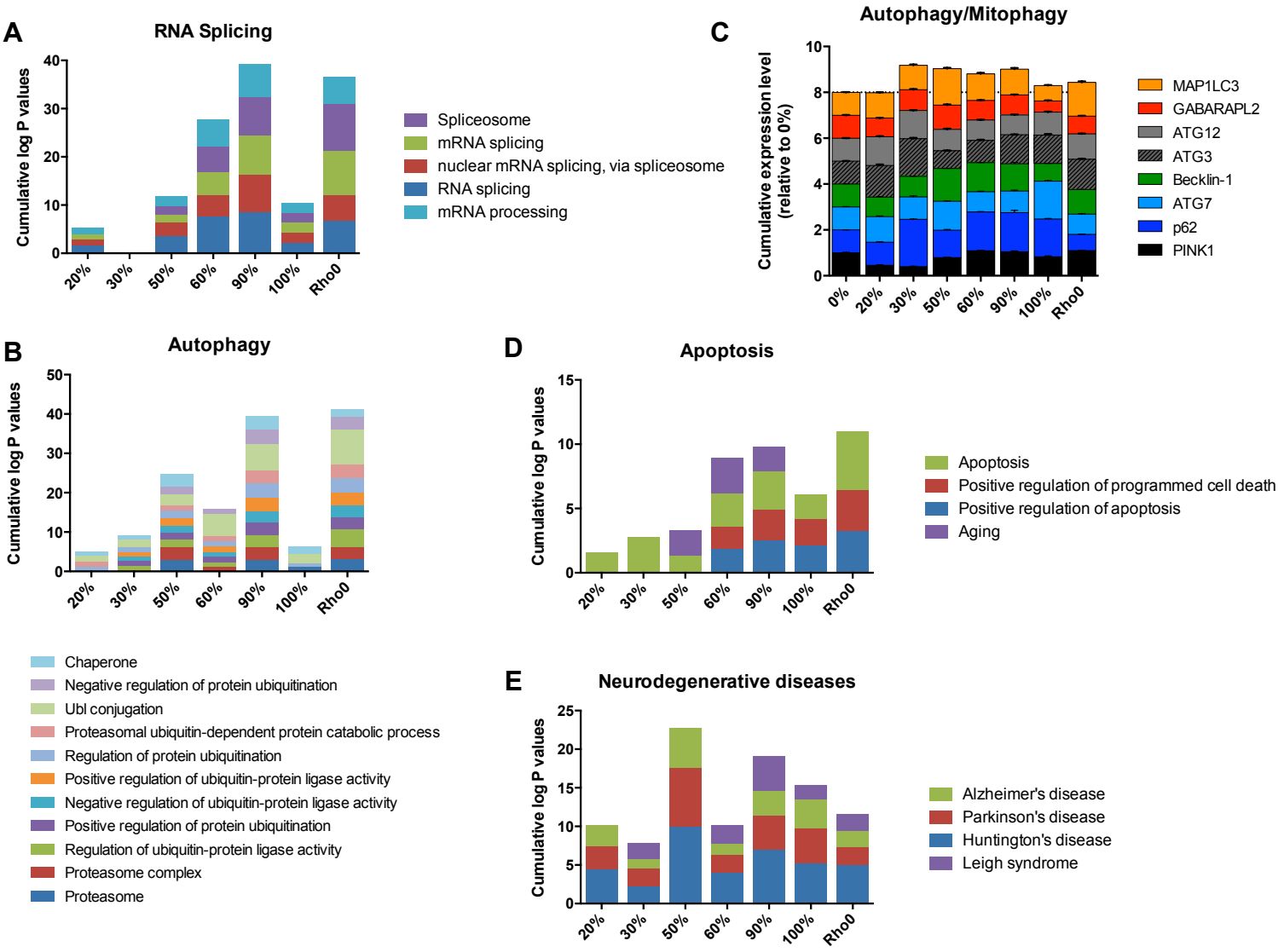


Fig. S8. Gene ontology analysis of whole-transcriptome data by 3243G heteroplasmy levels. (**A**) Enrichment log P values for gene ontology (GO) terms associated with RNA splicing; (**B**) autophagy, showing the simultaneous upregulation of both positive and negative regulators of autophagy. (**C**) Relative gene expression levels for canonical components of autophagy and mitophagy, demonstrating marginal and inconsistent induction of these genes with increasing heteroplasmy. (**D**) Enrichment log P values for gene ontology (GO) terms associated with apoptosis, and (**E**) neurodegenerative diseases.

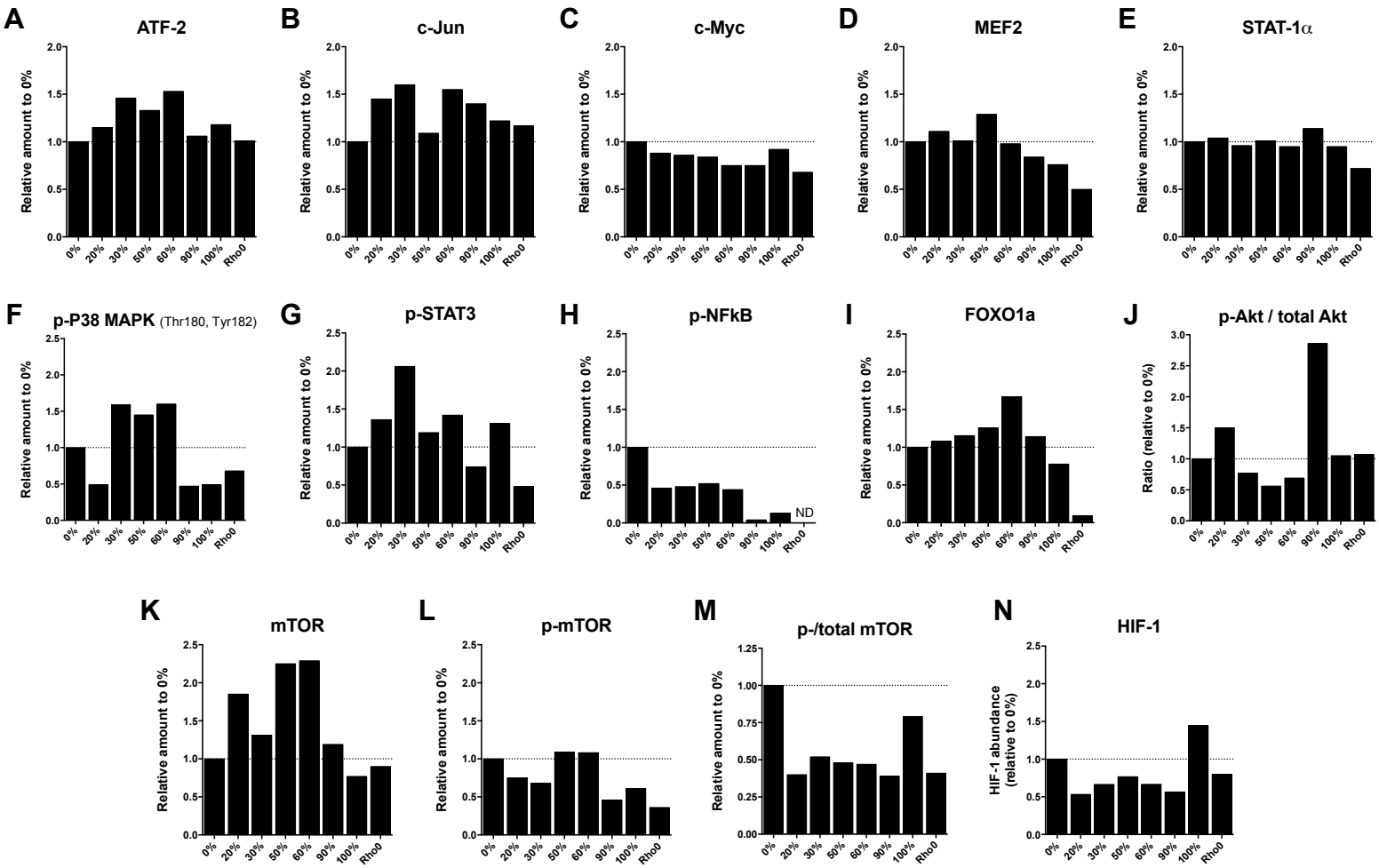


Fig. S9. Proteomic analysis of major transcription factors and key retrograde signaling molecules in response to mtDNA 3243A>G heteroplasmy. (**A**) Protein levels of ATF-2; (**B**) c-Jun; (**C**) c-Myc; (**D**) MEF2; (**E**) STAT-1α expressed relative to 0% heteroplasmy. (**F**) Protein levels of the phosphorylated form of P38 MAPK (Thr180, Tyr 182); (**G**) STAT3; (**H**) NF-κB, and (**I**) FOXO1a. (**J**) Activated (phosphorylated) form of Akt (p-Akt) relative to total Akt protein levels, (**K**) mammalian target of rapamycin (mTOR), (**L**) level of the phosphorylated form (p-mTOR), and (**M**) the ratio of phospho/total mTOR reflecting its activation status. (**N**) Abundance of total hypoxia inducible factor-1 (HIF-1). *ND = non detectable*

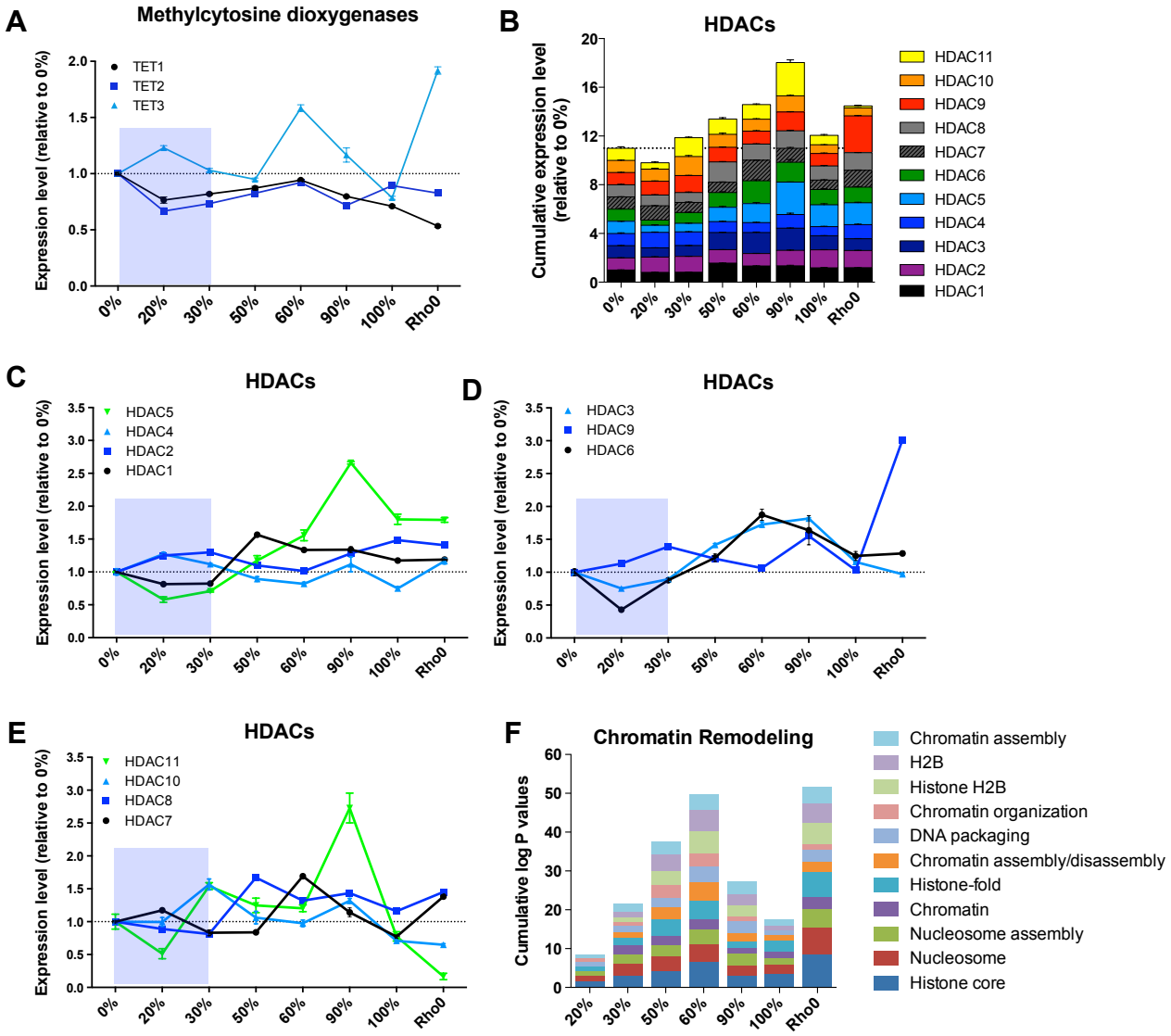


Fig. S10. Bi-phasic induction of the chromatin remodeling machinery in response to 3243A>G heteroplasmy. (**A**) Transcript levels for putative DNA de-methylation enzymes methylcytosine dioxygenases (TETs). (**B**) Cumulative expression levels of the HDACs expressed relative to 0 % heteroplasmy. (**C-E**) Transcript levels for each histone deacetylases (HDACs) 1 - 11. (**F**) Significance values relative to 0 % representing enrichment for GO terms containing differentially expressed families of genes involved in chromatin remodeling at different 3243G heteroplasmy levels.

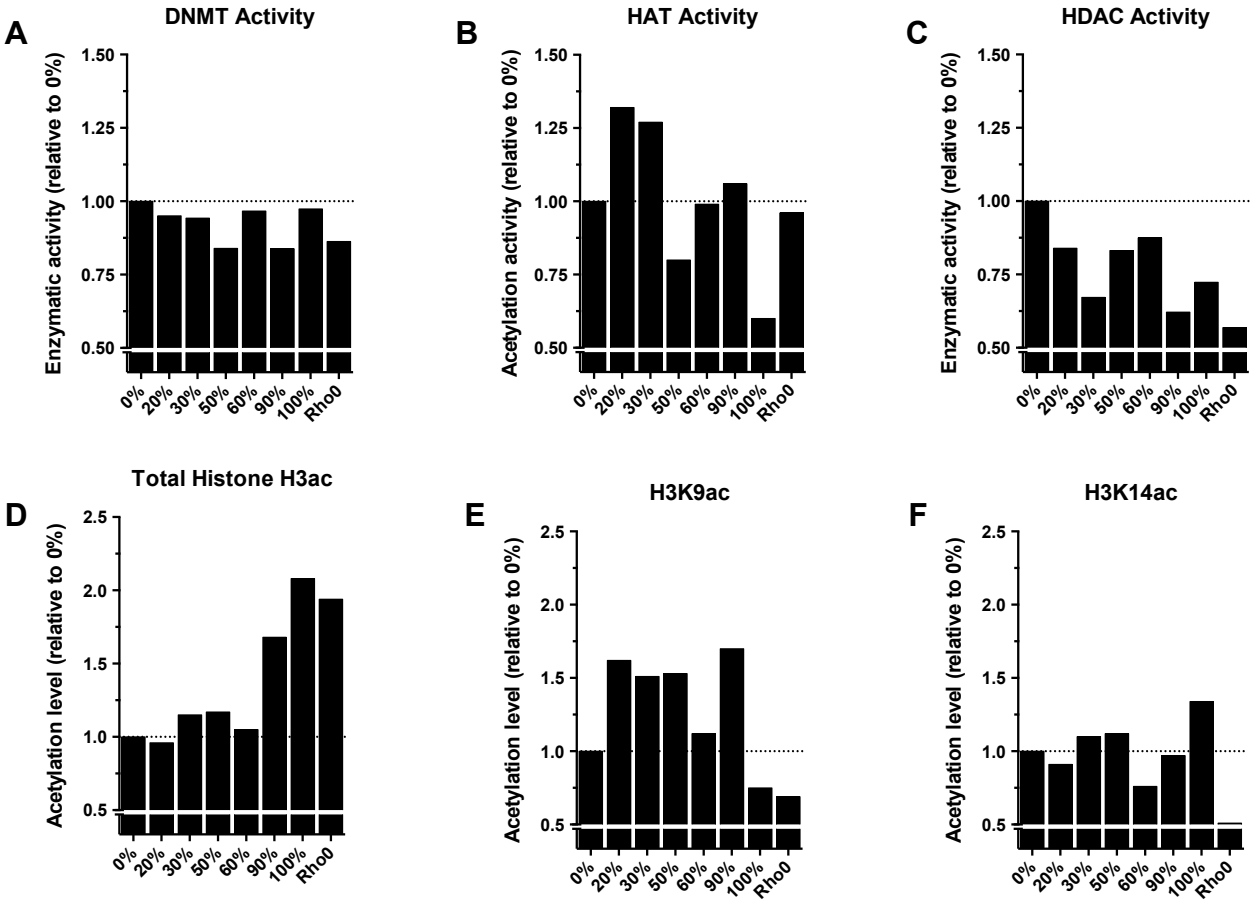


Fig. S11. Enzymatic activity of chromatin modifiers and histone 3 (H3) acetylation status in response to 3243A>G heteroplasmy. (**A**) Total DNA methyltransferase (DNMT) activity; (**B**) histone acetyltransferase (HAT) activity; and (**C**) total HDAC activity measured by enzyme-linked immunosorbent assay (ELISA, see Materials and Methods for details). (**D**) Total histone H3 acetylation levels; (**E**) histone H3 lysine 9 (H3K9); (**F**) and histone H3 lysine 14 (H3K14) acetylation relative to 0% measured by ELISA. Note that the reduction in HDAC activity is associated with an increased in total histone H3 acetylation in 90 %, 100 % and Rho 0 cell lines.

DAVID functional annotation of strongly repressed genes

Clusters	**GO terms** (# of genes)	Enrichment score	P value
1	**GPCR** (33) **Transducer** (34) **GPCR rhodopsin-like superfamily** (30) **GPCR signaling pathway** (38) **GPCR rhodopsin-like** (28) **7TM GPCR rhodopsin-like** (28)	5.47	All $P < 10^{-5}$
2	**Transmembrane region** (113) **Transmembrane** (113) **Intrinsic to membrane** (121) **Membrane** (131) **Integral to membrane** (114)	5.35	All $P < 10^{-5}$
3	**Sensory transduction** (27) **Sensory perception** (28) **Cognition** (30) **Neurological system process** (35)	4.46	All $P < 10^{-4}$

Fig. S12. Functional classification analysis of strongly repressed (>100 fold) genes at high (60-100%) 3243G heteroplasmy levels. Results from DAVID functional pathway analysis of strongly repressed genes at high heteroplasmy levels (see Fig. 5B). Shown are the three most significant clusters of biological functions enriched in this gene set. Notably, these GO clusters of genes were induced in 20 and 30% 3243G mutant cell lines (see Fig. 5B).

CELLULAR PHENOTYPE	0%	20 - 30%	50 - 90 %	100%
Energy production (Oxidative capacity)	Baseline	⟷	⟷, ↓↓	↓↓↓ (complete)
Tubular mitochondria (MTG and TEM)	Baseline	↑↑	↑↑↑	↑
Cell Size (measured by TEM)	Baseline	↓↓	↓, ⟷	⟷
Proportion mtDNA/nDNA transcipts (% mt RNA)	Baseline	↓↓	↓, ⟷	↓
mRNA abundance (%mtDNA transcripts)	Baseline	⟷, ↑	↑↑, ↓↓	↓↓
tRNA abundance (% mtDNA transcripts)	Baseline	↑↑	↓	⟷
Glycolysis (mRNA levels)	Baseline	⟷	↑↑	↑
DNMTs (mRNA levels)	Baseline	⟷, ↓	↑	↑, ⟷, ↓
ANTs (mRNA levels)	Baseline	↑	⟷, ↑↑↑	↑
Antioxidant defenses (mRNA levels)	Baseline	⟷, ↑	↑↑	⟷, ↑
Cytoplasmic/ER UPR (mRNA levels)	Baseline	⟷, ↑	↑↑↑	⟷, ↑↑

EXPRESSION PROFILES				
Growth and degradation (mTOR, lysosome)	Baseline	↓↓↓	↑↑, ⟷	⟷
7-Transmembrane GPCR	Baseline	↑	⟷, ↓↓↓	↓↓↓
Neurodegeneration	Baseline	⟷	↑↑	↑↑↑
Mitochondrial biogenesis, TCA, ETC	Baseline	⟷	↑↑	↑↑↑

Clinical Phenotypes	Normal	Autism & Diabetes	Multisystem degenerative disorder	Perinatal Leigh syndrome

Fig. S13. Summary of functional impact of the mtDNA 3243A>G mutation in 134B cybrid cell lines. Progressively increasing mtDNA 3243G mutation load causes divergent cellular phenotypes, summarized here as either unchanged (⟷), mild (↓,↑), moderate (↓↓,↑↑) or severe (↓↓↓,↑↑↑) increase or decrease in the listed parameter. MTG, mitotracker green; TEM, transmission electron microscopy; mRNA, messenger RNA; tRNA, transfer RNA; ANTs, adenine nucleotide translocator; DNMTs, DNA methyltransferases; ER, endoplasmic reticulum; UPR, unfolded protein response; mTOR, mammalian target of rapamycin; GPCR, G protein-coupled receptor; TCA, tricarboxylic acid cycle; ETC, electron transport chain.

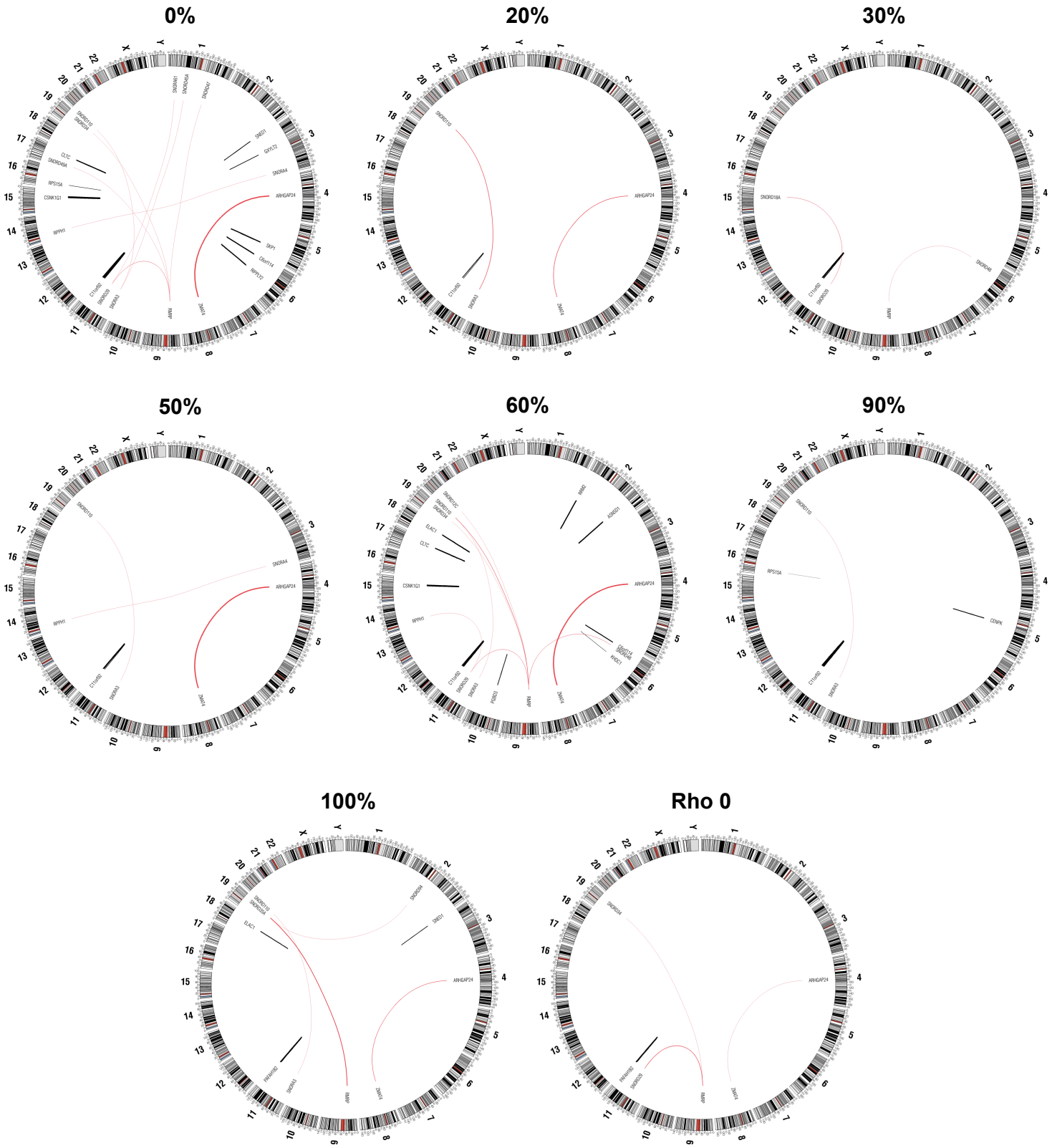


Fig. S14. Chromosomal rearrangements do not correlate with transcriptional reprogramming in cybrid cell lines. Circos plot analysis (LifeScope) showing gene fusion events obtained from RNA-Sequencing data. Line thickness denotes confidence in the fusion product. Black and red lines represent intra- and inter-chromosomal fusion products, respectively.

Extended Experimental Procedures

Generation of transmitochondrial cybrids

A mtDNA-deficient (ρ^o) cell line derivative from the osteosarcoma 143B thymidine kinase-deficient (TK^-) cell line (143B(TK^-), ATCC CRL 8303) (1) was fused to enucleated lymphoblastoid cells (1) from a patient heteroplamsic for the $tRNA^{Leu(UUR)}$ nucleotide (nt) 3243A>G mtDNA mutation (2) using polyethylene glycol (1). After being established by incubation for two to three days in complete DMEM (Gibco #10569), cybrids were selected in DMEM supplemented with 10% of dialized FBS, 110 ug/mL sodium pyruvate and 50 ug/mL 5-bromo-2'-deoxyuridine (BrdU), without uridine. The BrdU selected for the 143B(TK^-) nucleus and against the lymphoblastoid cell nucleus, and the absence of uridine selected against the mtDNA-deficient 143B(TK^-) cells while permitting growth of those 143B(TK^-) cells that had acquired the lymphoblastoid cell line's mtDNAs. Transmitochondrial cybrid colonies were isolated between 14 and 21 days and subsequently expanded in DMEM media containing 4.5 g/L (25mM) glucose and 110 mg/L sodium pyruvate (Gibco #10569) supplemented with 50 μg/ml uridine and non-essential amino acids (1:100, Gibco #11140) at 37°C in 95% air, 5 % CO_2. Clonal isolates were screened for those harboring the 143B(TK^-) nucleus yet were heteroplasmic for the 3243A>G variant. Heteroplasmy for the $tRNA^{Leu(UUR)}$ nt 3243A>G mtDNA mutation was monitored by polymerase chain reaction (PCR) amplification of a 264 nt fragment surrounding the 3243 site and digestion with HaeIII restriction which distinguishes the mutant and normal base pair (1, 3). Clones heteroplasmic for different percentages of the 3243A>G mtDNA mutation were generated by partial depletion of the cybrid clones by treatment with 5 μg/ml ethidium bromide (EtBr) for 11 days, with the level of mtDNA depletion monitored each day by qPCR (4). At the end of the EtBr treatment, the mtDNA levels reached a very low level. The cells were then plated at low density in medium lacking EtBr to facilitate the isolation of clones with varying levels of the 3243G mutant mtDNAs (5).

A series of clones were ultimately obtained harboring approximately 0, 20, 30, 50, 60, 90, and 100% 3243G mutant mtDNA. These clones have proven remarkably stable in their heteroplasmic levels under the above culture conditions. This contrasted with the reduced heteroplasmy stability that we observed when cytoplasts from the 143B(TK^-) cybrids were used to transfer the heteroplasmic mutation to SH-SY5Y neuroblastoma cells (6). Cell lines used for experiments were grown in 75 cm^2 dishes, and used when at 50-80% confluence.

Measurement of mtDNA heteroplasmy

Having established the 143B(TK^-) 3243A>G cybrids with the desired mtDNA heteroplasmy levels, the heteroplasmy levels were monitored PCR amplification of the 264 nt fragment, digestion with HaeIII (New England Biolabs #R0108), and separation of the fragments initially by agarose gel electrophoresis and densitometric analysis and subsequently by capillary electrophoresis on an Genetic Analyzer 3500 automated sequencer (Applied Biosciences). Biases in the determined 3243A>G mtDNA levels in these cybrids were corrected using standard curves (6). Genotypic analysis by this procedure prior to the RNASeq analysis gave the values listed in Materials and Methods Table 1.

Table 1

Heteroplasmy group	0 %	20 %	30 %	50 %	60 %	90 %	100 %
Clone name	CL9	DW4	DW5	DW9	DW7	DW10	CL3
RFLP % 3243G	1.09 ± 0.01	19.37 ± 2.75	33.23 ± 0.74	52.86 ± 0.09	61.35 ± 1.50	94.37 ± 0.20	100.0 ± 0.00

Mean mutation load (3243G) measured by RFLP. Data are means ± S.D. for 2-3 replicates.

Heteroplasmy levels at subsequent passages were assessed by sequencing the mtDNA using next generation semi-conductor sequencing (M&M Table 2). The entire mtDNA was first PCR amplified in two fragments using the SequalPrep Long PCR Kit (Invitrogen #A10498) with the following primers: (1) hmtL9611: 5'-TCCCACTCCTAAACACATCC-3', hmtH1405: 5'-ATCCACCTTCGACCCTTAAG-3'; and (2) hmtL1305: 5'-GTAAGCGCAAGTACCCACG-3', hmtH9819: 5'-GCCAATAATGACGTGAAGTCC-3'. Product was visualized on an agarose gel with ethidium bromide to assess amplification and product size-specificity. For library preparation the two PCR fragments were combined in equal ratios to make 200 ng of DNA, fragmented and blunt-end ligated to 3' P1 and barcoded 5' A-BC adaptors (Life Technologies) using the NEBNext Fast DNA Fragmentation & Library Prep Set for Ion Torrent (New England Biolabs #E6285S/L). Size selection was done using Agencourt AMPure XP beads (Beckman Coulter). Ready amplified libraries were quantified with the Ion Library Quantitation kit (Life Technologies #4468802). Multiple barcoded samples per run were pooled in equimolar ratios and templated Ion Sphere Particles (ISPs) were prepared and enriched on the One Touch 2 and ES systems using the Ion PGM Template OT2 200 kit (Life Technologies #4480974). The enriched templated ISPs were then loaded on an Ion 318 sequencing chip and sequenced on an Ion Personal Genome Machine (PGM) using Ion PGM Sequencing 200 v2 kit (Life Technologies). Data was analyzed using the Ion Torrent Suite v4.0.1 and NextGENe software v2.3.3 and A/G heteroplasmy levels at nucleotide position 3243 were verified. Results are shown in Table 2 below.

Table 2

Heteroplasmy group	0 %	20 %	30 %	50 %	60 %	90 %	100 %
Clone name	CL9	DW4	DW5	DW9	DW7	DW10	CL3
NGS % 3243G	0.00 (N.D.)	17.0	30.7	66.9	71.6	73.8	100.0

Mutation load (3243G) measured by next generation sequencing. N.D.: no detectable G allele.

The complete mtDNA sequence of all six cybrid clones was identical, with two exceptions. A spontaneously occurring synonymous nucleotide transition was observed at position 9494A>G (COIII, no amino-acid change: Glycine-Glycine) in the 30% (DW5, 19.0-26.5% 9494G heteroplasmy) and 50% (DW9, 3.2-4.2% 9494G heteroplasmy) 3243G heteroplasmy cell lines. The second spontaneous transition mutation was observed at 1042T>C in the 16S rRNA in the 60% (DW7, 7.6-8.4% 1042C heteroplasmy) 3243G cell line. Given the synonymous nature of the 9494A>G mutation and the low levels (<10%) of the 1042T>C mutations these changes were unlikely to have influenced the cellular phenotypes or gene expression profile.

mtDNA 3243A>G heteroplasmy levels were also estimated indirectly from RNA-Sequencing data, based on the number of transcripts observed to contain the 3243 A versus G nucleotide (M&M Table 3).

Table 3

Heteroplasmy group	0 %	20 %	30 %	50 %	60 %	90 %	100 %
Clone name	CL9	DW4	DW5	DW9	DW7	DW10	CL3
RNA-Seq % 3243G	0.12 ± 0.07	1.75 ± 0.15	15.74 ± 0.49	50.98 ± 1.04	68.74 ± 1.32	97.00 ± 1.95	97.74 ± 1.93

Mean mutation load (3243G) measured by counting allele frequency from RNA-Sequencing data. Data are means ± S.D for triplicates.

Comparison of the RNA-Seq and NGS data from samples collected at the same time suggests that there is an under-representation of the 3243G $tRNA^{Leu(UUR)}$ at the 20 and 30% mtDNA mutation in the RNA-Seq data. By contrast there is a variable representation of the 3234G mtDNA in the 50-90% cybrid by NGS. The reduced level of the 3243G mutant $tRNA^{Leu(UUR)}$ at 20-30% heteroplasmy could reflect instability, improper processing, or altered transcription of the 3243G $tRNA^{Leu(UUR)}$ when at low heteroplasmy. The variable NGS counts of the 3243A mtDNA at 50-90% 3243G heteroplasmy probably reflects non-linear amplification during library preparation. However, taken together, the NGS and RNASeq data support the heteroplasmy estimates generated by the more conventional HaeIII fragment analysis.

Western blotting and blue native electrophoresis gels

Electron transport chain (ETC) subunits Complex I 20 kDa, Complex II 30kDa, Complex III Core II, Complex IV COII, and Complex V (ATP synthase) F1α abundance were quantified by Western Blotting involving electrophoresis of 20 μg protein on a pre-cast sodium dodecyl sulfate (SDS), 10% polyacrylamide gel (SDS-PAGE). Gel separated proteins were electro-transferred onto polyvinylidene difluoride (PVDF) membranes and incubated overnight with antibody cocktail (Mitosciences #MS601). Equal protein loading between samples was verified by ponceau staining. Membranes were washed and incubated with horseradish peroxidase-conjugated secondary antibody (dilution 1:5,000). Gels were exposed to photographic films and specific band densitometry measured and quantified by normalizing to β-tubulin (Abcam #ab21058). The data is presented in Fig. 1C.

Mitochondrial respiratory complexes were separated by BN-PAGE (7). Mitochondria were isolated by differential centrifugation from 10-20 x 10^6 cells, and mitochondrial pellets equivalent to 400 mg of protein were solubilized with 800 mg of Dodecyl-b-D-Maltoside (Sigma #D4641) in 5 mM 6-aminohexanoic acid, 50 mM imidazole-HCl (pH 7.0) and 10% glycerol. Coomassie Brilliant Blue G-250 (Serva # 35050) was added to the solubilized samples at a dye/detergent ratio of 1:5 (w/w). Mitochondrial lysates equivalent to 50 ug were separated by electrophoresis on a 4-13% acrylamide gradient gel. Western blotting was performed using standard transfer conditions for 3 h. The stain from the blot was removed by washing with 100% methanol, which did not compromise the quality of the bands.

Mitochondrial respiration and ATP measurements

Mitochondrial respiration in each 3243G heteroplasmic cell line was measured on plated cells using the Oxygen Biosensor System (BD Biosciences) (8). Measurements were performed in triplicates in standard media under basal conditions, after inhibition of the F_0F_1 ATP synthase with oligimycin (state IV), and maximal respiration uncoupled with the protonophore carbonyl cyanide 4-trifluoromethoxyphenylhydrazone (FCCP). Leak mitochondrial respiration, obtained after inhibition with oligomycin, did not differ between cell lines (data not shown). Respiration expressed as pmol O_2/sec. Cellular ATP was measured with the bioluminescence assay kit CLS II (Roche, #1-699-695). ATP levels were quantified by normalizing relative fluorescence units (RFU) to total protein content.

Electron microscopy

Cybrid cells were detached with 0.25 % trypsin at 37^{o}C for 2 minutes, briefly centrifuged at 400 g for 1 minute, and immediately fixed with 2.5% glutaraldehyde, 2.0% paraformaldehyde in 0.1M sodium cacodylate buffer, pH 7.4. overnight at 4^{o}C. After subsequent buffer washes,

the samples were post-fixed in 2.0% osmium tetroxide for 1 hour at room temperature, and rinsed in distilled H_2O prior to in-bloc staining with 2% uranyl acetate. After dehydration through a graded ethanol series, each cell pellet was infiltrated and embedded in EMbed-812 (Electron Microscopy Sciences, Fort Washington, PA). Thin sections were stained with uranyl acetate and lead citrate, and examined on a JEOL 1010 electron microscope fitted with a Hamamatsu digital camera and AMT Advantage image capture software.

For quantification of cell and nuclear size, electron micrographs collected at 2,500x were measured. Cell and nuclear contours were manually traced by two different investigators blinded to group identity. From the dimension-calibrated micrographs the surface area (in μm^2) was calculated using Image J (NIH, rsbweb.nih.gov/ij/). Values for each cell and its corresponding nucleus (when visible) were exported and analyzed using Microsoft Excel and GraphPad Prism 6. Results from the two sets of data were pooled, and means with 95% and 99% confidence intervals (C.I.) calculated. Between 130 - 209 cells, and 98 - 133 nuclei were measured for each cell line. Statistical differences were evaluated using ANOVA with Dunnett's multiple comparison post hoc test. To normalize mtDNA content to cell content, cell volume was then calculated assuming spherical shape (cells in suspension), using the following formula: $V = 4/3\ \pi\ r^3$ (Fig. S2D) with the radius derived from mean cell surface area shown in Fig. 1H.

Cell and nuclear size varied in parallel as a function of 3243G heteroplasmy, with proportions of nuclear to total cell volume – termed karyoplasmic ratio (9) remaining constant (Fig. S2C). Nuclear/total cellular ratio is a conserved property across phylogenetically lower eukaryotic organisms such as in the budding yeast (10). Accordingly, the mtDNA mutation 3243G does not affect this conserved parameter.

Confocal microscopy

For confocal microscopy, 1 x 10^4 cells were grown for 36 hours on individual 35 mm glass-bottom culture dishes (MatTek, #P35GC-1.5-14.C). Cells were incubated in 125 nM Mitotracker Green TM (Molecular Probes, #7514) for 45 minutes at 37°C and then washed twice in media. Images were acquired using a laser scanning confocal microscope (Zeiss LSM 710) with a PlanApo 63x/1.40 oil immersion objective, optovar 1.25x and digital zoom of 1, 1024x1024 image size (final pixel size = 0.132 x 0.132 μm). Live cells were imaged in a temperature-controlled, humidified and 5% CO_2 chamber. Excitation settings were: 458 laser at 1.2 % power, master gain = 872, digital gain = 1, digital offset = -13.89, with optical filter 488-620 and pinhole size of 49 μm. Images were sequentially acquired with the Zen 2010 software at the following settings: 1024 x 1024 pixels, pixel dwell time = 25.2 μsec.

For analyzing mitochondrial morphology, images were processed with Image J (Wayne Rasband, NIH) version 1.42q. Raw images files were processed using an automated macro available from the authors upon request. Briefly, images were convolved using the following matrix [-1 -1 -1 -1 -1\n-1 -1 -1 -1 -1\n-1 -1 24 -1 -1\n-1 -1 -1 -1 -1\n-1 -1 -1 -1 -1\n], thresholded (117-225). Objects < 0.15 μm^2 representing background noise or out-of-focus mitochondria were discarded. Aspect ratio and roundness for each mitochondrion was exported to Microsoft Excel and Graphpad Prism 6.0 for analysis. An average of 5855 mitochondria were analyzed per heteroplasmy group. Data is expressed as means ± 95 % confidence intervals (C.I.), and significance was determined based on 99% C.I. of the mean.

RNA sequencing – Library preparation

Total RNA was extracted with Trizol and depleted of cytosolic rRNA (RiboMinus, Life Technologies #10837-08). RNA was quality checked and quantified on a Bioanalyzer 2100, RNA 6000 Nano kit (Eukaryote Total RNA Nano, Agilent Techonolgies) and Qubit 2.0 fluofometer RNA assay kit (Molecular Probes #Q32852). Five hundred ng of RNA was used for cDNA library preparation using (Ambion, Total RNA-Seq kit #4445374). Samples were barcoded such that they could be pooled and run on the same sequencing lanes (see below).

RNA sequencing – Sequencing and data processing

RNA from all eight experimental groups (trans-micotochondrial cybrid cell lines 0, 20, 30, 50, 60, 90 and 100 % 3243G heteroplasmy plus the ρ^o parental cell line) were sequenced on the ABI SOLiD 5500 platform. Barcoded triplicates were sequenced on the same slide in different lanes according to the table below. All sequencing was performed using paired-end chemistry of 50 (forward) x 35 (reverse) base pairs.

Sample	Replicate 1	Replicate 2	Replicate 3
0 %	Lane 2	Lane 4	Lane 6
20 %	Lane 4	Lane 5	Lane 6
30 %	Lane 4	Lane 5	Lane 6
50 %	Lane 4	Lane 5	Lane 6
60 %	Lane 4	Lane 5	Lane 6
90 %	Lane 4	Lane 5	Lane 6
100 %	Lane 4	Lane 5	Lane 6
Rho 0	Lane 4	Lane 5	Lane 6

Sequenced reads were mapped with LifeScope 2.5.1 using the Whole-Transcriptome Paired-End mapping pipeline against the 1000 Genomes Build 37 Decoy 5 reference (11) with default parameters. Mapped Reads were further filtered using Picard tools [http://picard.sourceforge.net/] post-mapping to retain only those reads in which both read pairs from each

paired-end fragment were properly paired resulting on average 114 million reads per replicate.

RNA sequencing – Data analysis

RNA sequencing data was further analyzed using Partek Genomics Suite™ software version 6.12.0109. Out of a total of 22,449 uniquely annotated genes for which transcripts were detected, 15,652 were differentially expressed (ANOVA model, $P < 0.0001$). Data was normalized and expressed as reads per kilobase per million reads (RPKM). Normalizing data using R package DESeq yielded similar results. Principal component analysis was performed in Partek Genomics Suite using default settings, and unsupervised hierarchical clustering was performed using the distance method with eulcidean parameter. Data was also analyzed in LifeScope™ Genomic Analysis Solutions (Life Technologies) using the gene fusion module to detect intra- and inter-chromosomal rearrangements (Fig S14) (12).

The Database for Annotation, Visualization and Integrated Discovery v6.7 (DAVID, (13)) was used to derive functional significance from highly repressed genes (n = 211) at 60-100% 3243G mtDNA levels (Figs. 4B and S12).

Pathway analysis

To quantify alterations in gene expression between cell lines with different mtDNA heteroplasmy, the sample with 0% 3243G mtDNA was set as baseline. The fold change in gene expression between all other heteroplasmy levels was calculated. All genes were ranked from most up-regulated to most down-regulated for each cell line with different mtDNA heteroplasmy level. Gene Set Enrichment Analysis (GSEA) was conducted (14) with the gene sets from the Molecular Signatures Database (MolSigDB v4.0). Gene sets with false discovery rate (FDR) less than 0.1 were selected for further analysis. The activity of affected transcription factors and pathways was inferred from its associated P value, and was positively correlated with minus logarithm transformed p.value, i.e. –log10(P value). In addition to GSEA analysis, hypergeometric distributions were used to evaluate the significance of enriched biological processes using the Gene-Ontology (GO) categories from differentially regulated gene lists. Only biological processes with significant enrichment ($P < 0.01$) were selected, and visualized by heatmap.

Mitochondrial DNA copy number

MtDNA copy number was measured by quantitative real-time PCR (qRT-PCR). Total genomic DNA was extracted and purified (DNeasy, Qiagen #69506) from all cell lines at four non-sequential passages, and diluted to 2.5 ng/μl. The ratio of mitochondrial to nuclear DNA (mtDNA/nDNA) was quantified by the ΔΔCt method using the following primer pairs for nDNA: β-2-microglobulin forward [TCCTCTCCCGCTCTGCACCC] (3'-5') and reverse [GGCGGGCCACCAAGGAGAAC]; and mtNDA: ND1 forward [GCAGAGACCAACCGAACCCCC] and reverse [GGGCCTGCGGCGTATTCGAT]. The PCR mixture contained 1x iTaq™ Universal SYBR® Green Supermix (BioRad #172-5120), 400 nM primers and 5 ng of template DNA. No mtDNA could be detected from the ρ^o cell line.

Chromatin remodeling and enzymatic assays

All assays were performed on non-denatured nuclear fractions isolated using the Nuclear Extract Kit (Activ Motif #40010) and measurements performed in triplicates. Total acetylated histone H3, H3 lysine 9 (H3.K9), and lysine (H3.K14) were measured from 100 μg of nuclear extract protein, using the PathScan Acetylated Histone H3 Sandwich ELISA Kit (Cell signaling technology #7232). Histone acetyltransferase (HAT) activity assay was measured using a colorimetric kit (Abcam #65352) from 50 μg of nuclear extract proteins. Histone deacetylase (HDAC) activity was measured using a colorimetric kit (Abcam #56210) from 20 μg non-denatured nuclear extract proteins, and specific activity assessed using the HDAC inhibitor trichostatin A. DNA methyltransferase (DNMT) activity was measured from 10 μg of nuclear extract using total DNMT (DNMT1, 3A, 3B) activity assay (Active motif #55006). Superoxide dismutase (SOD) activity was measured from 20 μg of whole cell lysate using the SOD detection kit (Cell Technology, #CSOD 100-2 colorimetric).

Transcription factors and retrograde signaling

The abundance and activation state of key transcription factors were measured by a combination of ELISA and Western blotting. The mitogen activated protein kinase (MAPK) pathway was probed by quantifying the abundance of ATF-2, c-Jun, c-Myc, MEF-2, and STAT-1α from 20 μg of nuclear extract using the TransAM MAPK Family Transcription Factor kit (Active Motif #47296). The signaling protein Akt1 and its phosphorylated form p-Akt1 (Ser 473), p-MEK1 (Ser217/221), p-p-38 MAPK (Thr180/Tyr182), p-STAT3 (Tyr705) and p-NF-κB were quantified from 10 μg of nuclear extract using PathScan® Signaling Nodes Multi Target Sandwich ELISA Kit (Cell Signaling Tecnology #7272). Myc (c-Myc) was quantified from 10 μg of nuclear extract using the TransAM c-Myc kit (Active Motif #43396). FOXO1a (Abcam #ab12161), mTOR (#ab2732) and p-mTOR (#ab109268) were quantified by Western immunoblotting and normalized to β-actin (Sigma #A2228) as loading control.

References

1. Trounce IA, Kim YL, Jun AS, & Wallace DC (1996) Assessment of mitochondrial oxidative phosphorylation in patient muscle biopsies, lymphoblasts, and transmitochondrial cell lines. *Methods Enzymol.* 264:484-509.

2. Heddi A, Stepien G, Benke PJ, & Wallace DC (1999) Coordinate induction of energy gene expression in tissues of mitochondrial disease patients. *J. Biol. Chem.* 274(33):22968-22976.

3. Goto Y, Nonaka I, & Horai S (1990) A mutation in the $tRNA^{Leu(UUR)}$ gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* 348(6302):651-653.

4. Wiseman A & Attardi G (1978) Reversible tenfold reduction in mitochondria DNA content of human cells treated with ethidium bromide. *Mol. Gen. Genet.* 167(1):51-63.

5. Procaccio V*, et al.* (2006) Detection of low levels of the m.3243A>G mutation in mitochondrial DNA in blood derived from diabetic patients. *Mol. Diagn. Ther.* 10(6):381-389.

6. Desquiret-Dumas V*, et al.* (2012) Metabolically induced heteroplasmy shifting and l-arginine treatment reduce the energetic defect in a neuronal-like model of MELAS. *Biochim. Biophys. Acta.* 1822(6):1019-1029.

7. Wittig I, Karas M, & Schagger H (2007) High resolution clear native electrophoresis for in-gel functional assays and fluorescence studies of membrane protein complexes. *Mol. Cell. Proteomics* 6(7):1215-1225.

8. Wang W, Upshaw L, Strong DM, Robertson RP, & Reems J (2005) Increased oxygen consumption rates in response to high glucose detected by a novel oxygen biosensor system in non-human primate and human islets. *J Endocrinol* 185(3):445-455.

9. Gregory TR (2005) Genome Size Evolution in Animals. *The Evolution of the Genome*, ed Gregory TR (Elsevier Academic Press, London), pp 3-87 (Chapter One).

10. Jorgensen P*, et al.* (2007) The size of the nucleus increases as yeast cells grow. *Mol. Biol. Cell* 18(9):3523-3532.

11. Abecasis GR*, et al.* (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56-65.

12. Berger MF*, et al.* (2010) Integrative analysis of the melanoma transcriptome. *Genome Res.* 20(4):413-427.

13. Huang da W, Sherman BT, & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4(1):44-57.

14. Subramanian A*, et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102(43):15545-15550.