PhD Summary

Unravelling Biased Segregation of Mitochondrial DNA

Daniel Jeremy Ives

Point mutations, deletions, and duplications of mitochondrial DNA (mtDNA) can cause mitochondrial disease and are associated with Parkinson's disease and ageing. Cells from patients with mitochondrial disease often contain a mixture of mutant and wild type mtDNA, a state termed heteroplasmy. A heteroplasmic cell can maintain normal respiratory function until the proportion of mutant mtDNA reaches a critical threshold, above which respiratory decline is exhibited. The study of pathological mtDNA mutations has been aided by the development of cybrid cell lines. A cybrid is created by fusion of an enucleated cell to an immortal cell depleted of its mtDNA. Cybrids enable the study of mtDNA mutations in a controlled nuclear background. Previous studies have found that the proportion of mutant mtDNA can increase or decrease depending on the nuclear background in a phenomenon termed biased segregation. In the case of the m.3243A>G point mutation, which can cause mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes, the proportion of mutant mtDNA was found to increase in a sub-population of 143B osteosarcoma cells, whereas it decreased in a sub-population of A549 adenocarcinoma cells.

The NCBI Gene Expression Omnibus is a database for global gene expression data and includes datasets for cell lines studied in the context of biased mtDNA segregation. These datasets were curated and utilised for global expression analysis in an effort to identify a molecular basis for biased segregation of mtDNA. Comparison of datasets from A549 cells with datasets from cell lines known to select at least one form of mutant mtDNA revealed enrichment for transcripts linked to the endoplasmic reticulum (ER) stress response in A549 cells. Stimulation of the ER stress response with 2-deoxy-D-glucose (2DG) reduced the proportion of mutant mtDNA in multiple clones and populations of A549 cybrid cells, whereas stable heteroplasmy persisted in replica cells cultured without 2DG.

A549 cybrids do not represent a homogenous population as only a fraction select WT mtDNA. Detailed analysis of datasets from A549 cells revealed they could be divided into two sub-populations characterised by differential expression of PPARGC1A, a key regulator of mitochondrial biogenesis. Moreover, PPARGC1A expression was associated with genesets linked to the ER stress response.

Mapping of the ER stress response to an unbiased architecture of the cell derived from NCBI GEO revealed enrichment for the stimulation of innate immunity and the interferon-response. However, interferon treatment inhibited the spontaneous reduction in the proportion of mutant mtDNA in an A549 cybrid clone. Global expression analysis of interferon treated primary human hepatocytes revealed marked down-regulation of PPARGC1A with interferon treatment. Taken together, these results suggest that a 2DG-stimulated ER stress response can promote segregation to wild-type mtDNA in adenocarcinoma cells; however this behaviour is contingent on other factors, which might include elevated PPARGC1A expression.

Declaration

I declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. No part of this dissertation has been submitted for qualification at another university. The work described was carried out at the Medical Research Council Mitochondrial Biology Unit under the supervision of Dr. Ian J. Holt, during the period of October 2009 to December 2012, and under the supervision of Dr. Mike P. Murphy from January 2013 to June 2013.

Daniel Jeremy Ives

June 2013

Acknowledgements

First and foremost, I want to acknowledge Catherine Nezich, who was integral to the success of this project. I want to thank Antonella Spinazzola, who made me feel at home, Mike Murphy, who ensured the safe passage of my project and Michael Minczuk, for his advice and generosity. I am indebted to Janet Luck and Steven Drinkwater, who went beyond the call of duty to resolve issues of an administrative and technical nature. I am further indebted to Yvonne Vallis at the MRC Laboratory of Molecular Biology, who briefly hosted me and ensured the safe passage of my project. I want to acknowledge Lawrence Kazak, who unknowingly served as a nuclear powered role model, Jiuya He, whose good nature and wisdom were unsurpassed, Helen Cooper, for her Finnish companionship and Stuart Wood, who is the nicest guy I'm likely to meet. Thanks are also due to Gloria Brea and Sara Vidoni, who really looked out for me during my 'Ghost' period. I am in debt to Andrew Raine, Merlin Hartley and Ashley Hartley, who provided ongoing computing support and expertise. I want to thank Payam Gammage for his sense of humour, Bastiaan Leerkotte for his sense of humour and Christopher Powell for his sense of humour. I also want to thank my friend Nick Gowler, for the crazy holidays and essential advice, which preserved my spirit and were a tonic to the realities of my project. Thank you to Dr David Hartley, who previously kept me on the scientific path and to whom I will always be indebted. Thank you also to my parents, for their unwavering support. Last and foremost, I want to acknowledge my supervisor, Dr Ian Holt, who gave me the freedom to follow my own path. The success of this project is built upon his previous work, critical technical and intellectual input, and careful steering of my energies. To all those that have helped me, but have not been acknowledged by name, thank you.

Statement of length

This dissertation is 19170 words in length and meets the requirements of the Biology Degree Committee to not exceed a limit of 60000 words.

Unravelling Biased Segregation of Mitochondrial DNA

Contents

1 Introduction 7
2 Materials and methods 21
2.1 Primers 21
2.2 Primer design 21
2.3 Probes and plasmids 21
2.4 Reagents 21
2.5 Solutions 22
2.6 Vessels 22
2.7 Cell lines 22
2.8 Maintenance of cells 23
2.9 Freezing cells 23
2.10 Creation of a ρ^0 HEK293T cell line. 24
2.11 Creation of a HEK293T m.3243A>G cybrid cell line 24
2.12 Bacterial transformation 26
2.13 DNA minipreps 26
2.14 Cell transfection 27
2.15 A549 m.3243A>G cybrid sub cloning 27
2.16 Clone C4 sub-clone treatment regime 27
2.17 A549 m.3243A>G cybrid sub-clone treatment regime 28
2.18 Low throughput DNA extraction 28
2.19 Medium throughput DNA extraction 29
2.20 High throughput DNA extraction 29
2.21 DNA processing for pyrosequencing 29
2.22 Last-cycle Polymerase Chain Reaction 30
2.23 Pyrosequencing 30
2.24 Southern blotting 31
2.25 Microarray data curation and compression 32
2.26 Babelomics 4.0 32
2.27 GenePattern 33
2.28 Connectivity map 34
2.29 ProfileChaser 35
2.30 Unbiased biological process mapping 35
2.31 Enrichment map 35
2.32 Statistical analysis 36
3 Results 37
3.1 Isolation of a ρ^0 HEK293T cell line. 37
3.2 Creation of the HEK293T m.3243A>G cybrid cell line 38
3.3 Curation of microarray datasets for A549, HEK293, HeLa, NT2 and fibroblast cell lines from NCBI GEO. 41
3.4 Differential expression analysis 42
3.5 Gene set enrichment analysis 46
3.6 Heteroplasmic spread and a low level of spontaneous biased segregation in HEK293T m.3243A>G cybrid sub-clones 49

3.7 Intermittent amino acid starvation is not associated with biased segregation in HEK293T m.3243A>G cybrid sub-clones and pooled populations.....52
3.8 Intermittent 2DG treatment is not associated with biased mtDNA segregation in HEK293T m.3243A>G cybrid sub-clones and pooled populations.....55
3.9 ER stress has no appreciable effect on the segregation of m.3243A>G mtDNA in HEK cybrids without transgenic Parkin. 58
3.10 Intermittent 2DG treatment induces a decrease in the proportion of mutant mtDNA in A549 m.3243A>G cybrid sub-clones and pooled populations that is not observed in the absence of treatment. 59
3.11 Unbiased biological process mapping 63
3.12 IFNα/γ treatment has no appreciable effect on the segregation of m.3243A>G mtDNA in HEK cybrids expressing transgenic Parkin. 65
3.13 The only effect of intermittent IFNα/γ treatment is to inhibit the segregation to WT mtDNA in a single A549 m.3243A>G cybrid sub-clone.....66
3.14 Intermittent combined 2DG/IFNα/γ treatment may permit spontaneous biased segregation in an A549 m.3243A>G cybrid sub-clone that previously maintained stable heteroplasmy with intermittent IFNα/γ treatment. 69
3.15 Enrichment map.....72
3.16 Connectivity map, ProfileChaser, UBP and leading-edge analysis of A549↑ versus A549↓ datasets.....76
3.17 The hypersensitivity of PPARGC1A expression to amino acid starvation in the A549 cell line.....83
4 Discussion 85
4.1 PPARGC1A is a candidate for influencing mtDNA segregation 86
4.2 CMAP and ProfileChaser suggest the reciprocal co-regulation of PPARGC1A expression and autophagy 87
4.3 Modelling the effects of autophagic flux on mtDNA segregation 87
4.4 ROS response and glutathione metabolism 90
4.5 MtDNA turnover 91
4.6 ER stress response and biased mtDNA segregation.....94
4.7 ER stress and the interferon-response.....97
5 References.....104
6 Appendix.....120

Abbreviations

143B: 143B osteosarcoma

Δψm: Membrane potential

2DG: 2-deoxy-D-glucose

A549: A549 adenocarcinoma

ADP: Adenosine diphosphate

ATP: Adenosine triphosphate

CMAP: Connectivity map

bp: Base pair

D-loop: Displacement loop

DMEM: Dulbecco's modified eagle's medium

DNA: Deoxyribonucleic acid

DOX: Doxycycline hyclate

eIF2α: Eukaryotic translation initiation factor 2

ERV: Endogenous retrovirus

EB: Ethidium bromide

ET: Electron tomography

ETC: Electron transport chain

FAD: Flavin adenine dinucleotide

FBS: Fetal bovine serum

GEO: Gene expression omnibus

GSEA: Gene set enrichment analysis

HEK293T: Flp-In T-REx human embryonic kidney 293

HeLa: Henrietta Lacks

IFNα/γ: Interferon α and γ

IFN-ß: Interferon β

MELAS: Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes

MtDNA: Mitochondrial DNA

mTERF: Mitochondrial transcription termination factor

NAD^+: Nicotinamide adenine dinucleotide

NCR: Non-coding region

mRNA: Messenger RNA

NDV: Newcastle disease virus

NES: Normalised enrichment score

PCR: Polymerase chain reaction

ROS: Reactive oxygen species

RNAi: RNA interference

rRNA: Ribosomal RNA

tRNA: Transfer RNA

UBP: Unbiased biological process

WT: Wild type

Glossary

ρ^0 cell: A cell devoid of mtDNA

A549↑: A549 microarray datasets in which expression of PPARGC1A is high.

A549↓: A549 microarray datasets in which expression of PPARGC1A is low.

Clone C4: HEK293T m.3243A>G cybrid Clone C4

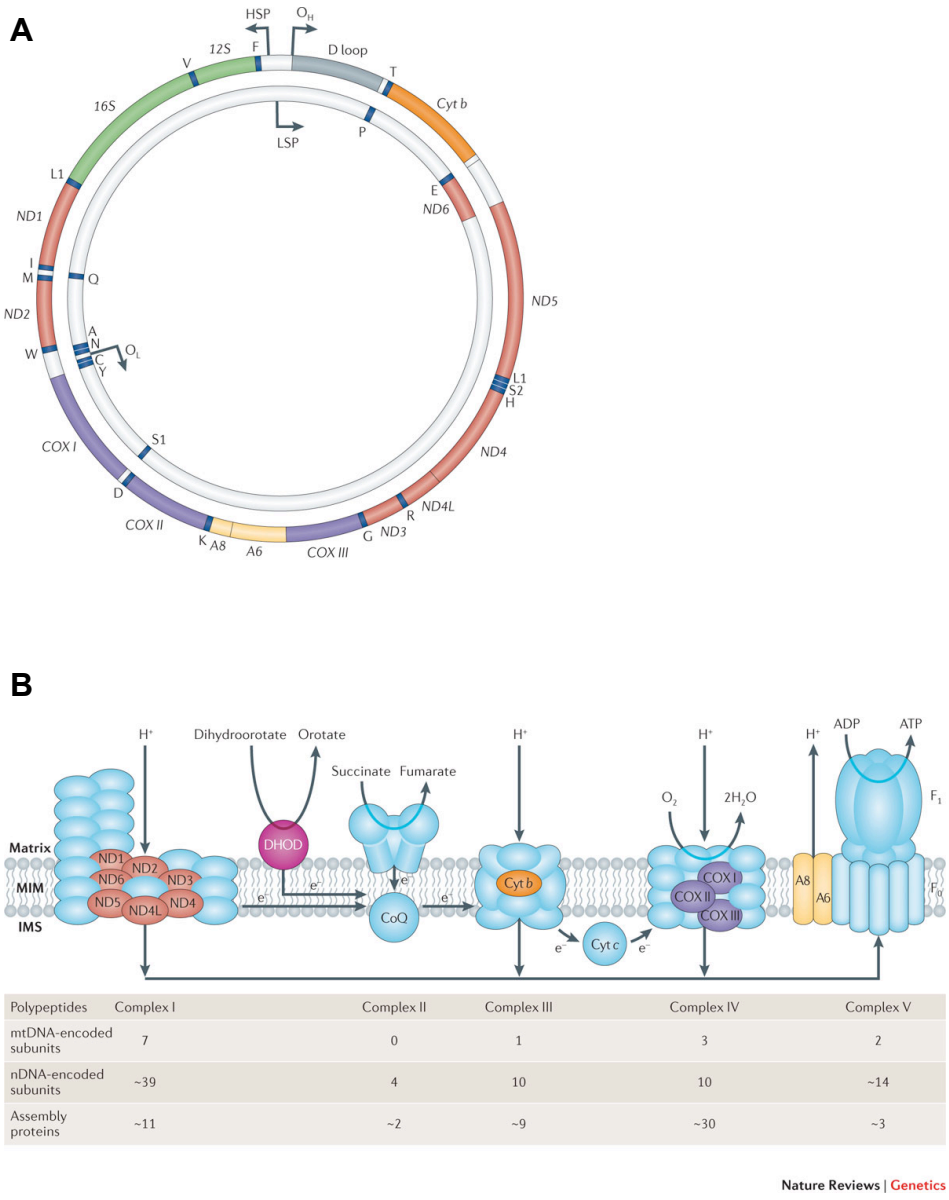
HepG2: Human hepatoma cell line

HT1080: Human fibrosarcoma cell line

1 Introduction

Mitochondria are membrane-enclosed cellular organelles whose primary function is adenosine triphosphate (ATP) synthesis via oxidative phosphorylation. Mitochondria generate ATP using an electron transport chain (ETC) that couples electron flow to proton pumping, establishing a proton-motive force that drives production of ATP. The ETC consists of four protein complexes (Complex I-IV), which through sequential redox reactions undergo conformational changes that pump protons across the inner mitochondrial membrane into the intermembrane space. NAD^+ and FAD are reduced in the matrix by the tricarboxylic acid cycle and donate electrons to Complexes I and II respectively, whilst molecular oxygen serves as the terminal electron acceptor at Complex IV. The proton gradient generated by Complex I-IV is utilised by the F-ATP synthase (Complex V), which couples dissipation of the gradient to phosphorylation of adenosine diphosphate (ADP) to generate ATP (Figure 1.1 B). The quantity of ATP generated for each glucose substrate by aerobic oxidative phosphorylation is 19-fold higher than anaerobic glycolysis.

Mitochondria are dynamic organelles that undergo continual cycles of fission and fusion in proliferating cells[1]. Fission is mediated in part by Drp1[2] and Fis1[3] and occurs at points of endoplasmic reticulum (ER) contact[4]. Fusion is mediated in part by Opa1[5] and Mfn1/2[6,7]. Mitochondrial morphology is linked to multiple cellular processes including the G_1-S transition of the cell cycle[8] and autophagy[9], whose stimulation converts mitochondria from isolated fragments to a 'hyper-fused' network.



Polypeptides	Complex I	Complex II	Complex III	Complex IV	Complex V
mtDNA-encoded subunits	7	0	1	3	2
nDNA-encoded subunits	~39	4	10	10	~14
Assembly proteins	~11	~2	~9	~30	~3

Figure 1.1. A schematic of human mtDNA (A) and the ETC (B) (Schon et al., 2012).

This network may contribute to energy distribution, with mitochondrial branches acting as power transmitting 'cables'[10]. In contrast, apoptotic stimuli promote conversion of mitochondria to isolated fragments[11].

According to the endosymbiotic theory, mitochondria are descended from α-proteobacteria engulfed by a eukaryotic progenitor and stably maintained as an endosymbioant[12]. Subsequently, most genes were transferred from α-proteobacteria to the nucleus of the host, now expressed in the cytoplasm and imported in to the mitochondria via protein translocases of the outer and inner mitochondrial

membranes[13,14]. Despite the transfer, mitochondria stubbornly maintain an independent genome of 16569 base pairs (bp) termed mitochondrial DNA (mtDNA)[15]. MtDNA encodes 13 proteins critical to the functioning of the ETC, aswell as 22 tRNAs and two rRNAs required for their expression in mitochondria (Figure 1.1 A). MtDNA and nuclear encoded proteins combine to form 'mosaic' respiratory chain complexes (Figure 1.1 B). MtDNA is inherited exclusively from the mother in humans[16] and most multicellular organisms[17]. Maternal mtDNA inheritance has been proposed to optimise 'Mitonuclear match' of respiratory chain subunits, minimising mtDNA heteroplasmy (a mixture of mtDNA haplotypes) to enable selection of mtDNA haplotypes that combine optimally with nuclear-encoded subunits to maximise respiratory function[18].

Considerable resources are required to maintain a second genetic system for the synthesis of a mere 13 proteins, suggesting there is a strong selection pressure to maintain these genes in mitochondria. One hypothesis is that mtDNA is an evolutionary relic. Mitochondrial DNA may be maintained because some mitochondrial proteins are too hydrophobic for import into the organelle. This is supported by the fact that all genes encoded by mtDNA are hydrophobic, cytochrome *c* oxidase subunit I and cytochrome *b*, which have never transferred to the nucleus during evolution are highly hydrophobic[19] and that mitochondrial import of apocytochrome *b* expressed in the cytoplasm is limited by the number of hydrophobic transmembrane helices in yeast[20]. An alternative hypothesis is that mtDNA is retained for evolutionary advantage. Mitochondrial DNA may serve as a long-term redox damage sensor, with cumulative mutation serving as a proxy for bioenergetic and genomic competence in cell lineages[21]. At a critical mutation threshold, mtDNA is proposed to initiate retrograde signalling to the nucleus in order to trigger anterograde protective and cell death pathways. This serves to select the energetically 'fittest' cells and remove cells that compromise the survival of the organism.

Mitochondrial DNA replication proceeds via a mechanism, distinct from bacterial chromosome and nuclear DNA replication, in which leading strand synthesis proceeds two-thirds of the way around the molecule before lagging strand synthesis initiates[22]. During the delay between initiation of leading- and lagging- strand synthesis, it was

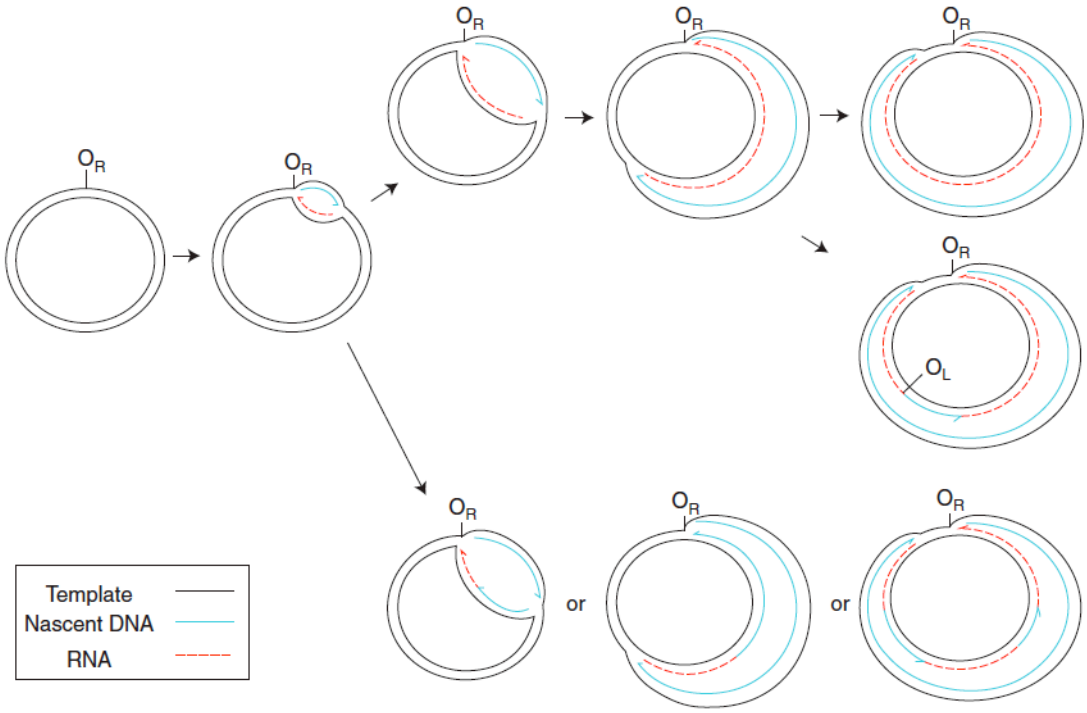


Figure 1.2. A mtDNA replication model. Replication initiates at the origin of replication (O_R) and leading strand synthesis proceeds with concurrent incorporation of RNA on the lagging strand. At some point lagging strand synthesis initiates (frequently O_L) and RNA is replaced or converted to DNA (Holt and Reyes., 2012).

long-believed that the lagging strand remained single stranded[23], yet later analyses of replication intermediates detected only duplex molecules[24-26]. Treatment of replication intermediates with RNaseH, which degrades RNA/DNA hybrid, generated extensive single stranded regions indicating that mtDNA replication intermediates contain long stretches of RNA/DNA hybrid. Further analysis of replication intermediates suggested that during leading strand synthesis, RNA derived from preformed transcripts, is incorporated throughout the lagging strand and is later replaced by DNA (Figure 1.2) [26,27]. However, a minority of mtDNA replication intermediates are resistant to both RNase H and single strand nuclease and can be enriched under specific experimental conditions[28,29]. This suggests that an alternative replication mechanism generating duplex DNA may operate in some circumstances.

Many pathological rearrangements of mtDNA contain a deletion junction with a residual direct repeat, which may form via replication slippage[30]. Such

rearrangements are thankfully rare in human populations and RNA incorporation on the lagging strand template may be important to minimise their occurrence[31], since exposing direct repeats as single stranded DNA increases the frequency of deletion in bacterial plasmids[32]. Alternatively, incorporation of RNA could reduce the risk of DNA damage imposed by the formation of single stranded DNA required during conventional DNA replication, by obviating the need to loop out the lagging strand for Okazaki fragment synthesis[31]. Moreover, without RNA on the lagging strand, repairing single-strand DNA breaks would be problematic. Lastly, incorporated RNA might inhibit mitochondrial transcription on actively replicating mitochondrial DNA molecules to prevent the deleterious consequences of collisions between replication and transcription complexes[33].

Human cells in culture each maintain 1000-5000 copies of mtDNA packaged into nucleoprotein complexes called nucleoids. Earlier reports determined that each nuceloid contains 2-10 copies of mtDNA[34,35,36], but a recent study utilising super-resolution microscopy revealed that the nucleoid frequently contains only a single copy of mtDNA (mean of ~1.4 mtDNA molecules per nucleoid)[37]. Nucleoid proteins are encoded by the nucleus and protect mtDNA from reactive oxygen species (ROS) generated by the ETC, regulate mtDNA expression and ensure faithful replication and maintenance of mtDNA[38]. A major nucleoid component is the DNA-binding protein Tfam, which according to a report is expressed at levels sufficient to coat the entire mitochondrial genome[39]. Tfam expression correlates with mtDNA copy number[40], and is required for both mtDNA transcription and replication[41].

Another protein associated with mitochondrial nucleoids is ATAD3. ATAD3A, ATAD3B and ATAD3C are three adjacent nuclear genes that presumably arose from gene duplication events and specify very similar proteins, although expression of ATAD3C has not been observed. Several findings suggest that ATAD3A and ATAD3B might be involved in nucleoid formation or segregation. The N-terminal region of ATAD3B has a marked preference for binding synthetic displacement loops (D-loops), similar to those found in a triple stranded segment of the major noncoding region (NCR) of mtDNA[40]. Both proteins localise to a sub-population of nucleoids and simultaneous silencing of both genes reduces the number of protein dependent mtDNA multimers, whilst having a modest impact on mtDNA copy number[42].

A549 adenocarcinoma (A549) cells express no ATAD3B and lower levels of ATAD3A than 143B osteosarcoma (143B) cells and have larger and fewer nucleoids[43]. Furthermore, ATAD3 had been observed to form 'filaments' between nucleoids[43].

MtDNA is dependent on the nucleus for its maintenance and replication; therefore mitochondrial disease may arise from mutations in either the mitochondrial or the nuclear genomes[44]. Point mutations, deletion and duplication of the mtDNA have been reported[45-47]. Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) is one of the most prominent mitochondrial diseases, frequently caused by an adenine to guanine point mutation in the mitochondrial $tRNA^{Leu(UUR)}$ gene at bp 3243[48]. Mutation at m.3243 reduces $tRNA^{Leu(UUR)}$ aminoacylation[49], causing a mitochondrial translation defect that manifests in patients as severe respiratory chain activity deficiency, with complexes I and IV the most affected[50]. MtDNA deletions may occur *de novo*[51] or as a consequence of primary mutations in nuclear genes including POLG, TWINKLE and ANT1, which cause autosomal dominant progressive external ophthalmoplegia[52]. Nuclear mutations may also cause gross mtDNA depletion, respiratory chain subunit deficiency or defects in iron or coenzyme Q metabolism[52].

Cells from patients with mtDNA mutations causing mitochondrial disease often contain a mixture of mutant and wild type (WT) mtDNA, a state termed heteroplasmy. A heteroplasmic cell can maintain normal ETC function until the proportion of mutant mtDNA exceeds a critical threshold. Above this threshold, ETC activity rapidly declines, which can be identified by reduced cytochrome *c* oxidase activity in muscle[53-55]. In myotubes from patients with myoclonus epilepsy and ragged red fibers (MERFF), carrying an adenine to guanine mutation at nucleotide position 8344 in the $tRNA^{Lys}$ gene, COX activity remained at normal levels up to 85% mutant mtDNA[53], above which a marked decline was exhibited.

Mutation of mtDNA is associated with Parkinson's disease (PD), an age-associated neurodegenerative disease characterised by degeneration and accumulation of COX deficient dopaminergic neurons in the substantia nigra[56]. High levels of deleted mtDNA accumulate in COX-deficient substantia nigra neurons of aged controls, and

this phenomenon is amplified in individuals with PD[57,58]. The deletions are variable between cells, but individual cells contain a specific mtDNA rearrangement, indicating a primary mutational event followed by clonal expansion[57]. Patients with idiopathic PD have complex I deficiency[59], and several nuclear mutations associated with familial PD reside in proteins that are targeted to the mitochondrion or interact with the organelle. Two of these proteins, Parkin and Pink[60-62], are thought to underpin a mitochondrial quality control system. Pink1 is rapidly degraded in functional mitochondria, but accumulates on the outer membrane of mitochondria with deficient membrane potential (Δψm), where it recruits Parkin to mediate their removal by autophagy[63,64].

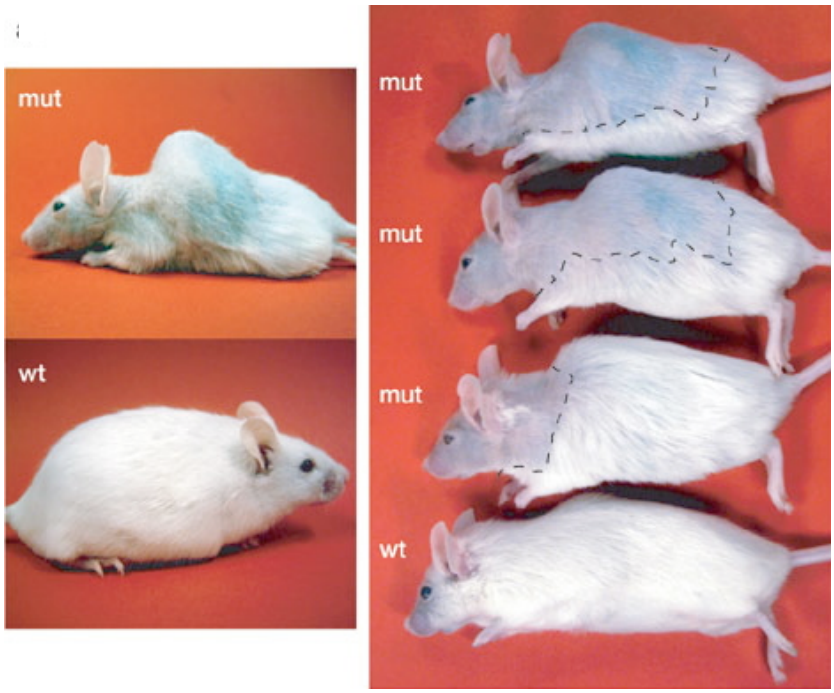


Figure 1.3. Aging-related phenotypes observed in mtDNA-mutator mice. MtDNA mutator-mice (mut) exhibit reduced body size, kyphosis, reduced hair density and different stages of alopecia (delineated by dashed line) in comparison with littermate WT mice (wt) at the age of 40–45 weeks (Trifunovic et al., 2004).

Therefore, loss of function mutations in either gene might be expected to lead to the accumulation of dysfunctional mitochondria, including those harbouring mutant mtDNA.

MtDNA mutations accumulate during ageing in humans[57,58,65], monkeys[66] and mice[67]. A causative link between mtDNA mutations and ageing was established with the creation of the mtDNA 'Mutator mouse'. The Mutator mouse is homozygous for a defective mtDNA polymerase[68] (*Polg*$^{mut/mut}$), which causes ~2500-fold higher levels of random mtDNA point mutations according to a random mutation capture assay[72], as well as increased amounts of a linear 11kb fragment of mtDNA[68,69]. This fragment was later shown to be a product of stalled replication[70]. The Mutator mouse exhibits a median lifespan of 11 months compared to 26 months for WT littermates, together with premature onset of ageing-related phenotypes including alopecia, kyphosis, osteoporosis, anemia and reduced fertility[68] (Figure 1.3). A mtDNA 'Deletor mouse', carrying predominant amounts of mtDNA with a ~5kb deletion with some mtDNA duplications, also exhibits kyphosis and reduced life span[71]. However, mice heterozygous for the defective mtDNA polymerase (*Polg*$^{+/mut}$) do not exhibit a reduction in lifespan despite ~500 times higher levels of random mtDNA mutation than age-matched WT mice[72]. Therefore, the relationship between random mutations and ageing is not linear. Although this study was unable to detect large deletions that accumulate in *Polg*$^{mut/mut}$ mice[71], transgenic mice expressing mutant *twinkle* have an organism-wide increase in the accumulation of large deletions and have a normal lifespan[73].

It is unlikely that the threshold for which random mtDNA mutations become limiting for lifespan (~2500-fold increase) will be reached in WT mice. However, time dependent selection of specific pathogenic mtDNA mutants may be the primary cause of aging in WT and *Polg*$^{+/mut}$ mice, whilst the ~2500-fold higher mutation load in *Polg*$^{mut/mut}$ mice may be sufficient to short-circuit this process through increased *de novo* generation of pathogenic mtDNA mutants. Alternatively, the highly-elevated random mtDNA mutation levels observed in *Polg*$^{mut/mut}$ mice may impact aging indirectly by effecting processes in early life such as development.

The study of mtDNA mutations has been aided by the development of cybrid cell lines. A cybrid is created by the fusion of an enucleated cell to a cell depleted of its mtDNA (ρ^0 cell)[74]. By enucleating cells with mutant mtDNA from mitochondrial disease patients, it is possible to transfer mutant mtDNA to a control nuclear background. If such cybrids are respiratory deficient, and the deficiency correlates with mutant load, the defect can be confidently attributed to the mtDNA mutation. An unexpected finding from the cybrid system was that the proportion of mutant mtDNA could change during continuous culture, a phenomenon termed biased segregation. This was first observed with an increase in the proportion of deleted mtDNA in cybrids with a HeLa nuclear background[75] and an increase in the m.3243A>G mtDNA mutation in cybrids with an osteosarcoma nuclear background[76].

Later cybrid studies found that the proportion of m.3243A>G mtDNA could decrease in another nuclear background, indicating that the nucleus influences the direction of biased segregation

.

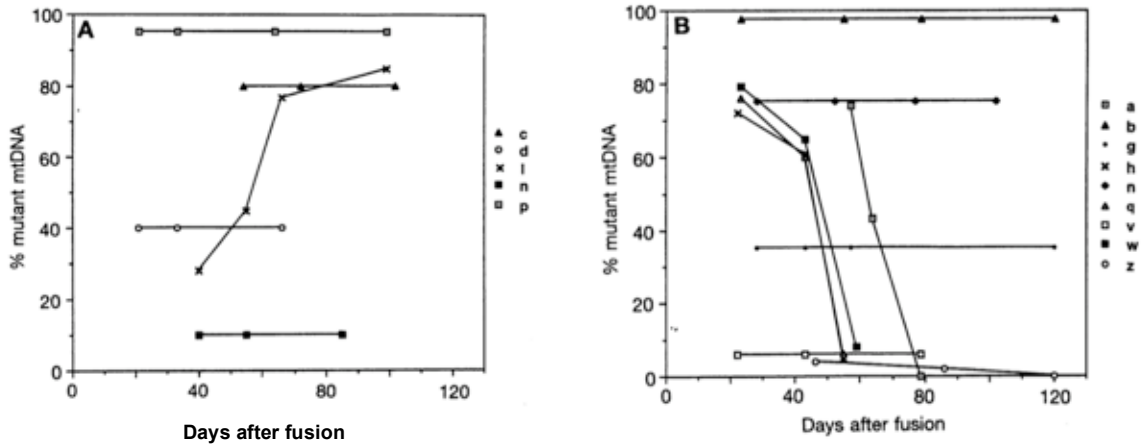


Figure 1.4. Biased segregation in a sub-population of cybrid clones. The heteroplasmy of 143B m.3243A>G cybrid (**A**) and A549 m.3243A>G cybrid (**B**) clones is plotted with respect to time (Dunbar et al., 1995).

In cell lines heteroplasmic for the MELAS m.3243A>G point mutation, the proportion of mutant mtDNA was found to increase in 1/33-143B m.3243A>G cybrid clones and decrease in 5/25-A549 m.3243A>G clones[77] (Figure 1.4 A/B).

Moreover, in cybrids heteroplasmic for partially duplicated mtDNA, rapid loss of partially duplicated mtDNA is observed in the A549 nuclear background, whilst maintenance or further rearrangement of duplicated mtDNA to generate partially triplicated mtDNA is observed in the 143B nuclear background[78].

These results indicate that the A549 nuclear background can reduce the proportion of multiple types of mutant mtDNA. Conversely, the 143B nuclear background can increase the proportion of multiple types of mutant mtDNA and generate further rearrangements. Characterisation of A549 and 143B cells revealed that A549 cells express no ATAD3B, lower levels of ATAD3A and have fewer and larger nucleoids than 143B cells[43]. However, the relevance of these differences to biased segregation was not established. Table 1.1 provides a summary of the results from studies investigating changes in the proportion of mutant mtDNA and mtDNA copy number for different types of nuclear backgrounds and mtDNA mutation.

mtDNA mutation	+ve selection	-ve selection	Depletion
A3243G	HOS[28,29], NT2[35]	ADC[29]	NT2[37]
Deletion	HeLa [23] Lymphoblasts [30,31] Fibroblasts [36]		Fibroblasts[36]
Duplication	HOS[32]	ADC[32]	HOS[32]
T8993G			HOS[35], ADC[35]

Table 1.1. Biased segregation and depletion of mtDNA in different cell lines. Cell lines with abbreviated names include ADC (A549 adenocarcinoma), HeLa (Henrietta Lacks), HOS (143B osteosarcoma) and NT2 (Neuron committed teratocarcinoma), (Holt., 2010).

Cybrids provide a convenient model for the study of mitochondrial disease because they can be maintained indefinetly in cell culture, but the relevance of the cybrid model to mitochondrial disease patients remains open to question. This is because cybrids are derived from aneuploid cancer-cell lines, and are routinely maintained in non-physiological cell-culture conditions[74].

Early studies of mtDNA segregation in the germ-line of heteroplasmic mice found that it was random[79,80]. However, investigation of a range of tissues in adult mice carrying a mixture of the mtDNA haplotypes BALB and NZB revealed biased segregation in several tissues; BALB mtDNA was selected in hematopoietic cells, whereas in kidney and liver there was selection for the NZB haplotype[81].

Thus, nuclear background influences biased segregation *in vivo*, as well as in cell-culture models. Furthermore, nuclear background can also influence the phenotypic outcome of mtDNA mutations. Mutator or deletor mice with the B6J nuclear background do not exhibit the greying and alopecia observed in the C57BL/6 nuclear background, whilst Mutator or deletor mice with either background exhibit kyphosis and a shortened life span[69].

DRP1 and FIS1, whose protein products promote mitochondrial fission, were the first nuclear genes to be linked to biased segregation. Drp1 and Fis1 knockdown in human cells heteroplasmic for the m.3243A>G mutation increased the proportion of mutant mtDNA by 15% and 11% respectively, suggesting that mitochondrial fission favours WT mtDNA[82]. PARK2, whose protein product Parkin is mutated in familial early onset PD, was also found to participate in biased segregation.

Parkin expression in 143B cells decreased the proportion of mtDNA mutated in the cytochrome *c* oxidase gene[83]. However, Parkin expression did not decrease the proportion of mtDNA mutated in the cytochrome *b* gene, suggesting the ability of Parkin to affect biased segregation is haplotype-specific. The outer mitochondrial membrane GTPase GIMAP3 was identified as a critical regulator of biased segregation in leukocytes from mice heteroplasmic for the neutral mtDNA haplotypes BALB and NZB, by genome wide linkage scan of mice with and without the capacity for selection of the BALB haplotype[84].

The human ortholog of GIMAP3, GIMAP5, also localises to the outer mitochondrial membrane[85], suggesting functional conservation. The mechanism by which GIMAP3 regulates biased segregation remains unknown, but it was suggested that involvement in mitochondrial peptide export would allow the cell to recognise particular mtDNA

haplotypes[84]. Together, these findings indicate that genes with a link to mitochondria participate in biased segregation.

A mechanism predicated on the idea of 'selfish' DNA has been proposed to explain selection of mutant mtDNA. In yeast, the WT 75kb circular mitochondrial genome (ρ^+) undergoes massive deletion at high frequency, giving rise to respiratory-deficient mutant genomes (ρ^-). Matings between ρ^+ and some ρ^- mutants can give rise to >95% ρ^- zygotic diploid clones[86]. The ρ^- mutant is said to be >95% suppressive or 'hypersuppresive' since it severely suppresses the ρ^+ genome, supposedly by monopolising the replication or segregation apparatus.

Hypersuppressive ρ^- mutants contain tandem repeats of a 300bp stretch called the *rep* sequence, which is believed to serve as a replication origin[86] and could provide replicative advantage to ρ^- mtDNA. The segregation advantage of hypersuppresive ρ^- mtDNA is abolished by mutations in the Holliday junction resolvase CCE1/MGT1[87], a key component of the DNA recombination machinery in yeast mitochondria. This implies that replicative advantage of ρ^- mtDNA could be mediated by a form of DNA replication utilising homologous recombination.

In humans, the amplified region in partially duplicated and triplicated mtDNA includes the NCR, which contains one or more origins of replication. Thus, the maintenance or increase in the proportion of duplicated mtDNA in the 143B nuclear background, and its displacement by partially triplicated mtDNA, suggest that additional origins confer a direct replicative advantage[78]. Although the m.3243A>G mutation is located far from all the proposed origins of replication, a replication pause site maps to the 3243-region due to the binding of the mitochondrial transcription termination factor (mTERF)[88]. The m.3243A>G mutation decreases mTERF-binding affinity *in vitro*, suggesting that the mutation alleviates replication pausing *in vivo*, and further supporting replicative advantage as a physical basis for biased mtDNA segregation.

In mice, selection of NZB mtDNA in kidney and liver, and selection of BALB mtDNA in hematopoietic tissue[81], has been linked to ROS production. ROS

production is higher and respiratory capacity is lower for NZB mtDNA compared to BALB mtDNA. Cells containing NZB mtDNA also maintain a higher copy number than BALB mtDNA due to ROS induced mitochondrial biogenesis[89], although the mechanism is unknown. However, ROS production has been mechanistically linked to mitochondrial biogenesis in yeast. ROS induces mitochondrial biogenesis by oxidising NTG1. Oxidised NTG1 introduces a double strand break at the mtDNA replication origin, which serves to initiate rolling circle mtDNA replication by the homologous DNA pairing protein Mhr1[90]. NTG1 does not have a homologue in animals, but an unrelated protein might provide a functional substitute.

The study of biased mtDNA segregation in a greater variety of contexts would help to generate increased insight. One approach would be to express a gene linked to biased segregation, such as PARK2[83], across a panel of mtDNA mutant and nuclear backgrounds to isolate common features of those backgrounds in which Parkin expression is sufficient to induce biased mtDNA segregation. However, current methods for creation of cybrid cell lines are laborious and have a high failure rate.

An alternative approach would be to create a facility to modulate expression of a panel of candidate genes in a control mtDNA mutant haplotype and nuclear background to determine their effect on biased segregation. RNA interference (RNAi) is a method that can enable this approach and has been used to modulate candidate gene expression in previously established cybrid cell lines[82]. However, there are two limitations to RNAi. Firstly, although the method has been used to up-regulate expression of selected genes[91,92], it is restricted to gene silencing in most cases. Secondly, persistent knockdown is frequently difficult to achieve[93].

The Flp-In T-REx 293 cell line (HEK293T), a derivative of the human 293 embryonic kidney cell line, is based on a system that allows inducible expression of a candidate gene from a specific transcriptionally active genomic locus[94]. Not only does this cell line provide the facility to increase and decrease candidate gene expression, selection of cells that repress or remove the transgene is avoided because the transgene is dormant during creation of the transgenic cell lines.

By enabling the transgenic expression of a range of gene products in a heteroplasmic cybrid, a HEK293T m.3243A>G cybrid (HEK cybrid) would provide the opportunity to explore the genetic contribution to biased segregation. Prior to this study, attempts to create a ρ^0 variant of the HEK293T cell line for creation of heteroplasmic cybrids were unsuccessful (Spelbrink and Holt, personal communications). Overcoming this technical barrier would enable the expression of a panel of candidate genes with plausible links to biased segregation, including genes involved in mitochondrial nucleoid formation, network morphology, mitochondrial biogenesis and mitophagy.

An alternative approach to the study of biased segregation would be to identify global differences in gene expression between nuclear backgrounds associated with an increase or decrease of mutant mtDNA with time. However, identifying which differences in gene expression are related to biased segregation remains problematic. Helpfully, recent advances in bioinformatics have enabled the study of global gene expression differences at the functional level[95], allowing changes in underlying biological processes to be recognised more readily. Furthermore, the emergence of public databases containing thousands of global gene expression datasets such as NCBI gene expression omnibus (GEO), and the free availability of powerful bioinformatics software, has largely removed the resource barrier to global gene expression analysis.

The overarching aim of this study was to identify genes, genesets or cellular processes associated with selection of WT mtDNA, and to validate them by manipulation of gene expression to induce biased segregation of mtDNA in at least a sub-population of heteroplasmic cybrids. This aim was divided into a series of subsidiary aims;

1. The identification of genes, genesets or cellular processes associated with selection of WT mtDNA by curation and bioinformatic analysis of global gene expression datasets.
2. To investigate the capacity of chemicals and drugs impacting cellular processes identified in aim ‘1.’ to induce selection of WT mtDNA.
3. To create cell lines carrying both m.3243A>G mtDNA and inducible transgenes in order to test the ability of specific genes to induce selection of WT mtDNA in at least a sub-population of heteroplasmic cybrids.

2 Materials and methods

2.1 Primers

NT2F: TCAACAATAGGGTTTACGACCTCG
NT2R: AGGGGGGTTCATAGTAGAAGAGCG
DigControlF: CTTAAAACTCAAAGGACCTGGC
DigControlR: GTTTGGCTAAGGTTGTCTGGTAG

2.2 Primer design

Primers were designed using the MacVector software package.

2.3 Probes and plasmids

D-loop (H1) and 18S rRNA probes were kindly gifted (Aurelio Reyes).
D-loop (H1) probe (16241-141) primers;
Forward - TTACAGTCAAATCCCTTCTCGT
Reverse - GGATGAGGCAGGAATCAAAGACG
18S rRNA probe primers;
Forward - GTTGGTGGAGCGATTTGTCT
Reverse - GGCCTCACTAAACCATCCAA
SF-Park plasmid was kindly gifted (Catherine Nezich).

2.4 Reagents

TaKaRa LA TaqTM (without Mg^{2+}), 125 Units (TaKaRa)
Phusion High-Fidelity DNA Polymerase (Finnzyme/Thermo)
Doxycycline hyclate (Sigma)
2-deoxy-D-glucose (Sigma)
Interferon-α protein (Abcam)

Interferon-γ, human (Roche)

Dulbecco's modified eagle's medium (DMEM) (+) Na pyruvate, (-) glucose, (-) amino acids (Dundee Cell)

Fetal bovine serum (Invitrogen)

Tetracycline free fetal bovine serum (Thermo Fisher)

Dialysed serum (Dundee Cell)

Uridine (Sigma)

2.5 Solutions

Lysis buffer; 75 mM NaCl, 50 mM EDTA, 20 mM hepes-NaOH pH 7.4, 0.5% SDS, 0.2 mg/ml proteinase K.

1X SSC; 15 mM $Na_3C_6H_5O_7$, 0.15 M NaCl, pH 7.0.

SOC; 2.0% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 20 mM glucose, pH 7.0.

1X TBE; 89 mM tris, 89 mM boric acid, 2 mM EDTA, pH 8.0.

2.6 Vessels

6-well cell-culture-treated plates (Corning)

24-well cell-culture-treated plates (Corning)

96-well cell-culture-treated plates (Corning)

100 mm cell-culture dishes (BD Falcon).

Nunc biobanking and cell-culture cryogenic tubes (Thermo Scientific).

2.7 Cell lines

The Flp-In T-REx 293 cell line (Invitrogen) is a derivative of the human 293 embryonic kidney cell line[96] containing two stably and independently-integrated plasmids. The first plasmid introduces a single FRT recombination site into a transcriptionally active but unmapped genomic locus whilst stably expressing the

lacZ-zeocin fusion gene from a constitutive promoter. The second plasmid expresses the Tet repressor gene and a blasticidin resistance cassette from a constitutive promoter. Integration of pCDNA5 expression plasmid into the FRT site allows inducible gene expression by addition of DOX, which de-represses gene expression by binding and inhibiting the action of the Tet repressor on an upstream promoter.

A549 m.3243A>G cybrids were generated previously by fusion of A549 ρ^0 cells to enucleated myoblasts heteroplasmic for the m.3243A>G mutation[77].

2.8 Maintenance of cells

HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) (Invitrogen) supplemented with 10% Fetal Bovine Serum (FBS), 15 μg/ml blasticidin and 100 μg/ml zeocin (standard medium) at 37°C and 5% CO_2. ρ^0 HEK293T cells were maintained in standard medium supplemented with 50 μg/ml uridine (ρ^0 medium). HEK cybrid cells were isolated in DMEM supplemented with 10% Dialysed Serum (DS), 15 μg/ml blasticidin and 100 μg/ml zeocin (isolation medium). A549 m3243A-G cybrid cells were maintained in DMEM supplemented with 10% tetracycline free FBS (Tet- FBS).

2.9 Freezing cells

Cells were grown to 70% confluency in a 100 mm cell-culture dish, trypsinised, centrifuged at 1100 g for two minutes, resuspended in 2 ml cell stock solution (50% DMEM, 40% Tet- FBS, 10% DMSO) and 1 ml transferred to each of two cryotubes. Cryotubes were immediately transferred to a 'Mr Frosty' (Thermo) and placed at -80°C overnight. Cryotubes were then transferred to a -80°C freezer or liquid nitrogen for long term storage.

2.10 Creation of a ρ^0 HEK293T cell line.

Isolation of a ρ^0 variant of the HEK293T cell line is an essential prerequisite for creation of a heteroplasmic cybrid derivative[74]. Incubation of culture medium with low concentrations of ethidium bromide (EB) has previously been used to isolate ρ^0 variants of 143B[97], GM701[98], HeLa[75], A549[99] and Namalwa[100] cell lines. To isolate a ρ^0 variant of the HEK293T cell line, cells were grown for an extended period of time in standard medium supplemented with 50 ng/ml EB and 50 μg/ml uridine.

EB treated HEK293T cells were tested for mtDNA depletion by growth on galactose medium. Galactose medium contained glucose-free DMEM, 1 mM (0.5 mg/l) pyruvate and 5 mM (0.9 mg/ml) filter-sterilised D-(+)-galactose (Sigma). Cells were further tested for total mtDNA depletion by Southern blot of total DNA with D-loop (H1) probe. Prior to Southern blot, total DNA was extracted by isoproponal precipitation, digested overnight with *Pvu*II at 37°C, heat denatured at 70°C for 10 minutes and separated on a 1.0% TBE agarose gel at room temperature for 3 hours at 100 volts.

2.11 Creation of a HEK293T m.3243A>G cybrid cell line.

Cybrids are generated by fusion of an enucleated cell (cytoplast) to a ρ^0 cell[74]. The cytoplast and ρ^0 cell serve as mitochondrial and nuclear donors respectively. Cytoplasts were isolated from both A549 m.3243A>G cybrid (B2) and 143B m.3243A>G cybrid (206) cells by cytoskeletal disruption with cytochalasin B (Calbiochem) and centrifugation using the ficoll gradient technique[74]. A continuous ficoll gradient was prepared in 11 x 60 mm Beckman ultra-clear centrifuge tubes (Beckman Coulter). First, 50% w/w Ficoll PM 400 (Sigma) stock was autoclaved. Ficoll solutions were then prepared in DMEM containing 10 μg/ml cytochalasin B. 0.92 ml 25% ficoll solution was pipetted into each of two centrifuge tubes. 1.18 ml 12.5% ficoll solution was then carefully pipetted on top of the 25% ficoll layer. Tubes were capped and transferred to a Biocomp gradient station (Biocomp), the '12-25% Ficoll' program selected, and the gradient stored at 37°C in a CO_2 incubator.

~1.3x10^7 A549 m.3243A>G cybrid (B2) and 143B m.3243A>G cybrid (206) cells were each resuspended in 0.9 ml 12.5% ficoll solution, layered on the ficoll gradient and DMEM layered on the cells to bring the contents within 5 mm of the top of the tube. A SW60 Ti 99E 1802 rotor (Beckman Coulter) and Beckman optima LE-80K ultracentrifuge were pre-warmed to 37°C and run at 99976 g for one hour prior to centrifugation. The centrifuge tubes were balanced and transferred to the rotor for centrifugation at 99976 g for one hour at 37°C. Sterile 1000μl filter tips (Starlab) were modified by sterile blade to increase the bore size and used to collect cytoplast and nucleated cell fractions from the tubes (upper and lower bands respectively). Cells were washed with DMEM, pelleted for five minutes at 1100 g, the supernatant was removed and the procedure was repeated once more, leaving 50-100 μl of supernatant over the pellet. ~1.3x10^7 ρ^0 HEK293T cells were harvested, resuspended at 5x10^6 cells/ml in ρ^0 medium and layered on the pellet.

Cytoplasts were fused to ρ^0 HEK293T cells with fusogen polyethylene-glycol-1500 (PEG1500). Cells were pelleted for three minutes at 1100 g, supernatant was carefully aspirated, 0.1 ml of sterile 50% PEG1500 (BDH) solution was added and cells were slowly suspended. With suspension near complete, the pellet was carefully suspended for 30 seconds then left to stand for 30 seconds. 10 ml of ρ^0 medium was added, gently mixed and the solution was plated at various dilutions (8/10, 1/10, 1/20, 1/40) on 100 mm cell-culture dishes. HEK293T m.3243A>G cybrid cells were isolated by prolonged incubation in isolation medium, which replaced ρ^0 medium 24 hours after the plating. Isolation medium lacked uridine, but contained blasticidin and zeocin.

Blasticidin and zeocin select against nucleated A549 m.3243A>G cybrid (B2) and 143B m.3243A>G cybrid (206) cells, whereas HEK293T m.3243A>G cybrid or ρ^0 HEK293T cells carry blasticidin and zeocin resistance cassettes and so are resistant to the drugs. The absence of uridine selects against ρ^0 HEK293T cells, which are uridine auxotrophic due to the inhibition of the mitochondrial enzyme dihydroorotate dehydrogenase, a crucial enzyme in pyrimidine biosynthesis and DNA replication whose activity requires a functional ETC[101].

m.3243A>G heteroplasmy in HEK cybrid clones was quantified by restriction fragment length polymorphism (RFLP) utilizing last-cycle PCR. The m.3243A>G mutation creates an *Apa*I cleavage site, hence, amplification of a region of mtDNA encompassing this site by PCR, incorporation of radioactive marker into the PCR product and digestion by *Apa*I can be used to quantify the percentage of mutant and WT mtDNA.

MtDNA depletion was determined by Southern blot of total DNA with the D-loop (H1) and 18S rRNA probes. The major mtDNA fragment at 16.5 kb was used exclusively for quantification. Prior to Southern blot, 3.0 µg of total DNA was digested with 10U of *Pvu*II in a 30 µl total volume. Reaction was separated on a 0.7% TBE agarose gel at room temperature for 20 hours at 50 volts, and transferred to a nylon membrane.

2.12 Bacterial transformation

Competent *Escherichia coli* (DH5α) were transformed by heat shock[102]. 10 ng purified plasmid was incubated with 50 µl of competent cells on ice for 30 minutes, transferred to a water bath at 42°C for 45 seconds, then transferred to ice for two minutes. 250 µl of 37°C pre-warmed Super Optimal broth with Catabolite repression (SOC) medium was added and the cells were placed horizontally in a 37°C incubator on a shaker set to 225 rpm for one hour. 300 µl of cell mixture was spread on a Liquid Broth (LB)-agar plate containing 100 µg/ml ampicillin and incubated overnight at 37°C.

2.13 DNA minipreps

Transformed DH5α colonies were picked using an inoculation loop, transferred to 5 ml of LB with 100 µg/ml ampicillin, and incubated overnight in a 37°C incubator on a shaker set to 225 rpm. Plasmid DNA was isolated using a QIAprep spin miniprep kit (Qiagen) according to the manufacturers instructions.

2.14 Cell transfection

Clone C4 was co-transfected with 150 ng of SF-Park plasmid and 1350 ng of pOG44 plasmid (Invitrogen) using Lipofectamine 2000 (Invitrogen). Transfected sub-clones were selected by culture in DMEM supplemented with 10% Tet- FBS, 15 μg/ml blasticidin, 100 μg/ml hygromycin and 10 ng/ml DOX (standard culture conditions).

2.15 A549 m.3243A>G cybrid sub-cloning

A549 m.3243A>G cybrid cells were grown to confluency in a 100 mm cell-culture dish. Cells were trypsinised and counted twice with a Countess automated cell counter (Invitrogen). The mean cell number was adjusted to $1x10^5$ cells/ml in 3 ml of media. Cells were sequentially diluted 10-fold in 3 ml of media from $1x10^5$ cells/ml to $1x10^2$ cells/ml. Cells were diluted 10-fold from $1x10^2$ cells/ml to $1x\ 10^1$ cells/ml in 20 ml of media, then diluted two or four fold in 20 ml of media to five cells/ml and 2.5 cells/ml respectively. 200μl/well of diluted cells were transferred to a 96 well plate by multi-channel pipette (Thermo Fisher) to give one cell/well or 0.5 cells/well.

2.16 Clone C4 sub-clone treatment regime

Amino acid starvation, 2DG, or IFNα and IFNγ (IFNα/γ), in combination with 10 ng/ml DOX treatment, was initiated three days following transfection of Clone C4. Transfected sub-clones were subjected to one, two or three days amino acid starvation, 24 hours 10 mM 2DG treatment or six hours 1000 U/ml IFNα/γ treatment followed by recovery in standard culture conditions. Sub-clones were identified visually by microscope, isolated by pipette and expanded in 24-well cell-culture plates in 1 ml culture medium. At confluence, cells were disrupted by pipette and samples were collected for DNA extraction. 2.5×10^4 cells were seeded for further passage. After two days, cells were subjected to a new round of treatment. Sub-clones that exhibited less than a 15% change in heteroplasmy over a minimum of four time points or 30 days were discarded.

2.17 A549 m.3243A>G cybrid sub-clone treatment regime

$1x10^4$ or $5x10^3$ cells were seeded for treatment or no treatment respectively in 24-well cell-culture dishes. After two days, cells seeded for treatment were subjected to 2DG, IFNα/γ, or combined 2DG and IFNα/γ treatment for 24 hours, six hours and 24 hours respectively in 0.5 ml of medium, followed by recovery in the absence of treatment in 1.0 ml of standard culture conditions medium. At confluence, cell samples were collected by trypsinisation and seeded for a new round of treatment. Sub-clones exhibiting less than a 15% change from initial heteroplasmy over a minimum of six time points or 26 days were discarded.

2.18 Low-throughput DNA extraction

Cells were washed once with PBS, incubated for three minutes at room temperature in lysis buffer (section 2.5), then for 30 minutes at 4°C and for 10 minutes at 50°C. An equal volume of phenol was added to the tube, which was then mixed on a rotator for 15 minutes at room temperature. The tube was spun at 6800 g for 15 minutes at room temperature, aqueous phase was transferred to a new tube, an equal volume of chloroform isoamyl was added and the tube was mixed on a rotator for 15 minutes at room temperature. The tube was spun at 6800 g for 15 minutes at 4°C, the aqueous phase transferred to a new tube, an equal volume of isoproponal and 175 mM NaCl added and the DNA precipitated for 10 minutes at 4°C.

The tube was spun at 6800 g for 15 minutes at 4°C, the supernatant was discarded, the pellet was transferred in 1 ml of 70% EtOH to a new tube and spun at 13000 g for five minutes at 4°C. The DNA pellet was air dried and resuspended in 50 µl EB-buffer (Qiagen).

2.19 Medium-throughput DNA extraction

Cell suspension was transferred to a 1.5 ml tube and spun at 6128 g for one minute at room temperature. Supernatant was aspirated, leaving ~50 µl above the pellet. 300 µl of Cell lysis solution (Qiagen) and 100 µl of Protein precipitation solution (Qiagen) were added and the pellet was suspended for 5-10 seconds by pipette. Lysed cells were spun at 15932 g for seven minutes at room temperature. Supernatant was transferred to a new 1.5 ml with 500 µl isopropanol, inverted 50 times and spun at 15932 g for five minutes at room temperature. Supernatant was removed, 500 µl of 70% EtOH added, and the tube inverted three times then spun at 15932 g for five minutes at room temperature. Supernatant was removed and the pellet was air-dried until translucent. The translucent DNA pellet was resuspended in 30 µl EB-buffer.

2.20 High-throughput DNA extraction

All steps were carried out in accordance with manufacturers protocols. Materials included;

Vac-man laboratory vacuum manifold (Promega)
Vacuum pump (Welch)
Wizard SV genomic DNA purification system (Promega)
Eppendorf multipette plus pipette (Fisher Scientific)
10 ml Combitips pipette tips (Fisher Scientific)

2.21 DNA processing for pyrosequencing

DNA concentration was measured on a NanoDrop 8000 spectrophotometer (NanoDrop/Thermo) and adjusted to 10 ng/µl with Nuclease-free water (Life Technologies) in a 96 Well PCR segmented plate (Axygen) capped with PCR strip caps (Axygen), or a MicroAmp optical 96-well reaction plate (Applied Biosystems/Life Technologies) capped with MicroAmp optical 8-cap strips (Applied Biosystems/Life Technologies).

2.22 Last-cycle Polymerase Chain Reaction

100 ng of template was amplified by PCR with 1U of LA-Taq polymerase, 100 nM of primers NT2F and NT2R, 1x Mg free Taq buffer, 2.5 mM $MgCl_2$ and 16 μM dNTPs in a 50 μl total volume. Reactions were subjected to 22 cycles of 95°C for 50 seconds, 60°C for 30 seconds and 72°C for 45 seconds, followed by addition of 1:10 dilution [^{32}P] dCTP and a last cycle of 95°C for three minutes, 60°C for one minute and 72°C for 10 minutes. 2.5 μl of PCR product was digested with 10U *Apa*I for one hour in 20 μl total volume. Reaction volume was separated on a 6% non-denaturing polyacrylamide gel at room temperature for 90 minutes at 150 volts. Last-cycle PCR of sequence spanning a native *Apa*I site with the primers DigControlF and DigControlR was used as a positive control to ensure complete digestion.

2.23 Pyrosequencing

Pyrosequencing was used for the high-throughput quantification of heteroplasmy at the m.3243 position. Briefly, biotinylated PCR product was generated using primers spanning the m.3243 position (F1 and R1). PCR product was then bound to Streptavidin sepharose beads (GE Healthcare) and heat denatured to release the non-biotinylated strand. Sequencing primer (S1) was added and bound to the biotinylated strand to allow pyrosequencing at the m.3243G position. Release of pyrophosphate by incorporation of nucleotide triphosphate to the 3'-end of the sequencing primer allows identification of nucleotide sequence; pyrophosphate is converted by sulfurylase to ATP, which drives the production of light by luciferase, detected as a peak on the pyrogram. PSQ 96 MA 2.1 software was used for generation of pyrograms, pyrogram quality control and automatic allelic quantification. Paul Hawtin (Qiagen) assisted with primer design.

m.3243A>G heteroplasmy of DNA samples from five cybrid cell lines with between 63.5% and 90.2% mutant mtDNA was measured, both by pyrosequencing and last-cycle PCR, to validate pyrosequencing for heteroplasmy quantification (see Appendix, Figure 6.1). DNA samples from HEK293T and ρ^0 HEK293T cells were also measured by pyrosequencing to determine the accuracy of heteroplasmy

measurment in samples with 0% mutant mtDNA or no mtDNA respectively. A standard deviation range of 0.0006-0.0464, or a 0.06-4.64% change in heteroplasmy, was obtained across 359 samples measured in triplicate by pyrosequencing. All steps were carried out in accordance with the manufacturers protocols. Materials included;

PSQ 96MA pyrosequencer (Biotage/Qiagen)
Pyromark vacuum prep workstation (Biotage/Qiagen)
Pyromark vacuum prep filter probe (100) (Qiagen)
Dri-block DB2D (Techne)
Pyromark gold Q96 reagents (5x96) (Qiagen)
Pyromark Q96 plate low (100) (Qiagen)
Pyromark Q96 cartridge (3) (Qiagen)
Pyromark binding buffer (200 ml) (Qiagen)
Pyromark denaturation solution. (500 ml) (Qiagen)
Pyromark wash buffer (200 ml) (Qiagen)
Pyromark annealing buffer (250 ml) (Qiagen)
Pyromark PCR kit (800) (Qiagen)
PCR with biotinylated forward primer (Sigma);
F1: [Btn]TTATACCCACACCCACCCAAGAA (HPLC purified)
R1: GCGATTAGAATGGGTACAATGAGG
S1: ATGCGATTACCGGGC

2.24 Southern blotting

TBE agarose gels were incubated in depurination solution (250 mM HCl) for 20 minutes at room temperature, incubated with denaturation solution (0.6 M NaCl, 0.2 M NaOH) for five minutes, then 20 minutes with new solution, incubated with neutralisation solution (0.6 M NaCl, 0.24 M Tris-HCl pH 7.4) for five minutes, then 20 minutes with new solution, and dry blotted to a nylon membrane with 3MM paper and paper towels overnight. DNA was cross-linked to the membrane by 120 mJ/cm^2 of ultra-violet (UV) light exposure.

50 ng of DNA probe in 10 µl of ddH_2O was denatured at 95°C for three minutes and transferred to ice. A Ready-to-go DNA-labeled bead (GE Healthcare) was resuspended in 35 µl of nuclease-free water (Life Technologies) and transferred to ice. 3 µl of ^{32}P α-dCTP and the re-suspended Ready-to-go DNA-labeled bead were then added to the probe. The mixture was then incubated at 37°C for 30 minutes, 99°C for five minutes, and transferred to ice water at 4°C for three minutes.

The cross-linked membrane was incubated with hybridisation buffer (0.25 M Na_2HPO_4, 7% SDS) for 30 minutes at 65°C in a hybridisation tube, rolling, then again for 30 minutes with new buffer. ^{32}P α-dCTP labeled probe was then added to the tube, which was incubated overnight, rolling. The membrane was then washed three times with 1X SSC for 20 minutes, followed by a single wash with 1X SSC 0.1% SDS for 20 minutes at 65°C, rolling. The membrane was wrapped in cling film and exposed to phosphor screens for 2-16 hours. The screens were visualised on a Typhoon™ 9410 variable mode imager (GE Healthcare).

2.25 Microarray data curation and compression

CEL files were downloaded from the National Centre for Biotechnical Information Gene Expression Omnibus (NCBI GEO), which is available online at http://www.ncbi.nlm.nih.gov/geo/. Curated datasets were compressed using the application Yemuzip with the 'PC compatible' option selected. Yemuzip is available to download online at http://www.yellowmug.com/yemuzip/.

2.26 Babelomics 4.0

Babelomics is a platform for the analysis of transcriptomic, proteomic and genomic data[103], which is accessible online at http://babelomics.bioinfo.cipf.es/.

The 'Upload data' tab was selected for upload of compressed datasets. Data was normalised by selection of the 'Processing' tab, followed by the 'Normalise >

Expression > One-channel > Affymetrix' thread. On the 'Affy normalisation' page, no parameters were modified in the 'Analysis' box.

Differential expression of the normalised data was performed by selection of the 'Expression' tab, followed by the 'Differential expression > Class comparison' thread. On the 'Differential expression: class comparison' page, 'T-test', 'Limma' and 'Fold-change' options were sequentially specified in the 'Select test' box. No other parameters were modified.

Genes enriched by differential expression analysis of A549 cells versus cell lines with a known segregation bias toward mutant mtDNA were filtered for those genes with a link to mitochondria based on the domain expertise of lab members, MitoCarta, an inventory of 1013 human genes with evidence of mitochondrial protein localisation[104] or those genes with a link to mitochondria based on hypothesis. Filtered genes were assigned a score of three if differential expression was identified as significant by Fold-change, T-test and Limma methodologies in a biased analysis. Genes were assigned a score of two or one if identified as significant by only two or one of these methodologies respectively. Scores were multiplied by a factor of three if differential expression was identified as significant by Fold-change, T-test and Limma methodologies in an unbiased analysis. Scores were multiplied by a factor of two or one if identified as significant by only two or one of these methodologies respectively.

2.27 GenePattern

Genepattern is a platform for gene-expression analysis, proteomics, single nucleotide polymorphism analysis, flow-cytometry and RNA-sequencing analysis[105], which is accessible online at http://genepattern.broadinstitute.org/gp/pages/login.jsf.

The 'ExpressionFileCreator' module was selected for the upload and normalisation of compressed datasets. No parameters were modified.

The ‘ComparativeMarkerSelection’ module was selected for the differential expression analysis of normalised data. No parameters were modified.

The ‘Comparitive-MarkerSelectionVisualiser’ module was selected for the extraction of differential expression data generated by the ‘ComparativeMarkerSelection’ module.

The ‘ExtractComparitiveMarkerResults’ module was selected for the transformation of differential expression data for GSEA. The ‘statistic’ parameter was set to FDR(BH) and ‘min’ and ‘max’ parameters set to 0 and 0.05 respectively.

The GSEA module was selected for Gene Set Enrichment Analysis (GSEA)[95] analysis of transformed data. The chip platform ‘HG_U133_Plus_2.chip’ was selected and no parameters were modified.

‘javaGSEA desktop application’ software was used for the leading-edge analysis of genesets enriched using GSEA. The 20 most up and down-regulated genesets were each used to determine the top 20 highest scoring up-regulated and down-regulated leading-edge genes respectively, from each of the C2 (curated), C3 (motif), C4 (computational) and C5 (Gene Ontology) geneset collections.

2.28 Connectivity map

Connectivity map (CMAP) is a bioinformatic tool that uses pattern-matching algorithms to identify Food and Drug Administration (FDA) approved small molecules that induce global transcriptional changes that are most similar to those identified between two experimental conditions[106,107]. CMAP is accessible online at http://www.broadinstitute.org/cmap/. A list of the 200, 100, 50 and 30 most differentially expressed genes extracted from output of the ‘ComparitiveMarker-SelectionVisualiser’ module of GenePattern was used as input for CMAP.

2.29 ProfileChaser

ProfileChaser is a bioinformatic tool that searches NCBI GEO for experiments whose differential expression looks similar to a query experiment[108]. ProfileChaser is accessible online at http://profilechaser.stanford.edu/. Log_2-transformed gene expression data was extracted as an expression matrix from the output of the 'ExpressionFileCreator' module of GenePattern. Probe Identification Descriptions (ID's) of the Affymetrix human genome U133 plus 2.0 array were converted to Entrez gene IDs using the AILUN server, which is accessible online at http://ailun.stanford.edu/. The transformed expression matrix used as input for ProfileChaser.

2.30 Unbiased biological process mapping

Genesets that were enriched in GSEA analysis were mapped to 115 unbiased biological processes (UBPs) derived from an unbiased architecture of the cell[109]. Only UBPs in which the mapped genesets were consistently up or down-regulated were considered for further analysis. The median or mean normalised enrichment score (NES) of mapped genesets was used as a UBP enrichment score.

2.31 Enrichment map

Enrichment map is a Cytoscape plugin for functional enrichment visualisation[110], which is accessible online at http://www.baderlab.org/Software/EnrichmentMap. Using the GenePattern platform, log_2-transformed gene expression data was extracted as an expression matrix from the output of the 'ExpressionFileCreator' module and converted with the 'CollapseDataset' module to be used as input in the 'Expression tab' of the Enrichment Map plugin. Other files required as input were obtained from the output of GSEA analysis.

2.32 Statistical analysis

Statistical analysis of group differences was examined using the BINOMDIST function of Microsoft excel. Differences were considered significant at the 95% confidence level ($P < 0.05$).

3 Results

3.1 Isolation of a ρ^0 HEK293T cell line.

To derive a ρ^0 variant of the HEK293T cell line, cells were grown for an extended period of time in medium supplemented with 50 ng/ml of ethidium bromide (EB). After 95 days of treatment, cells were tested for mtDNA depletion by growth on galactose medium. In medium containing galactose in place of glucose, cells are forced to rely almost exclusively on oxidative phosphorylation to produce ATP[111]. Cells depleted of mtDNA are deficient for oxidative phosphorylation and therefore lose viability in galactose medium. EB treated cells exhibited extensive cell death on galactose medium (data not shown). The procedure was repeated after 116 days of EB treatment, with the same outcome. These results were consistent with mtDNA depletion, but did not preclude maintenance of some residual mtDNA in the cells.

To establish the extent of mtDNA depletion, total DNA was extracted from EB treated and untreated HEK293T cells and screened by Southern hybridisation after digestion with *Pvu*II and fractionation by agarose gel electrophoresis. Using a radiolabelled amplified fragment of human mtDNA as a probe, bands of ~16 kb and 650 bp corresponding to mtDNA and D-loop (7S DNA) respectively, were detected in untreated HEK293T cells, but not in cells treated for 95 or 116 days with EB (Figure 3.1.1). This result suggested that prolonged EB treatment had successfully depleted HEK293T cells of their mtDNA. This was confirmed when the putative ρ^0 cells were grown in the absence of EB for a period of several months and still found to lack any detectable mtDNA (Dr Jiuya He, personal communication).

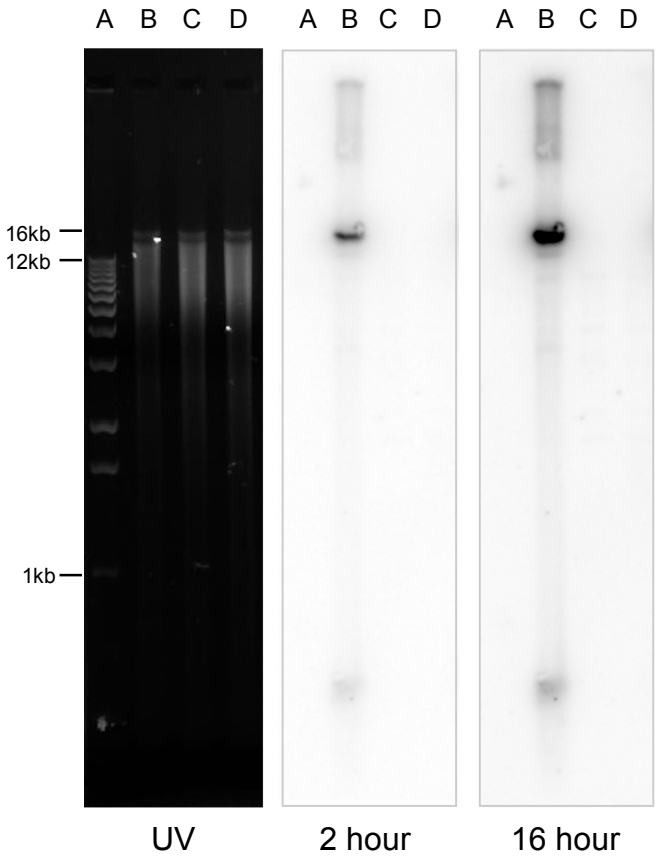


Figure 3.1.1. Total depletion of mtDNA in HEK293T cells. Total DNA was extracted from cells, digested with *Pvu*II and imaged by exposure to ultra-violet (UV) light, or blotted and probed for mtDNA and exposed to a phosphor screen for two hours (2 hour) or 16 hours (16 hour) respectively. (**A**) 1 kb DNA Ladder. (**B**) Un-treated HEK293T cells total DNA. (**C**) EB treated HEK293T cells total DNA at 95days. (**D**) EB treated HEK293T total DNA at 116 days.

3.2 Creation of the HEK293T m.3243A>G cybrid cell line

The creation of a HEK293T m.3243A>G cybrid cell line (HEK cybrid) was an essential prerequisite for implementation of a plan to determine the effect of candidate gene overexpression on biased segregation of mtDNA.

At the first attempt, seven HEK cybrid clones were isolated, six from the fusion of A549 m.3243A>G cytoplasts with ρ^0 HEK293T cells, the seventh where the mitochondrial donor was a 143B m.3243A>G cytoplast.

The heteroplasmy level of each clone was determined by last-cycle PCR and mtDNA copy number was determined by Southern blotting. Cybrid clones exhibited a range of heteroplasmy, from 68.4% (C4) to 90.2% (B2) mutant with a mean heteroplasmy of 77.4% mutant (Figure 3.2.1).

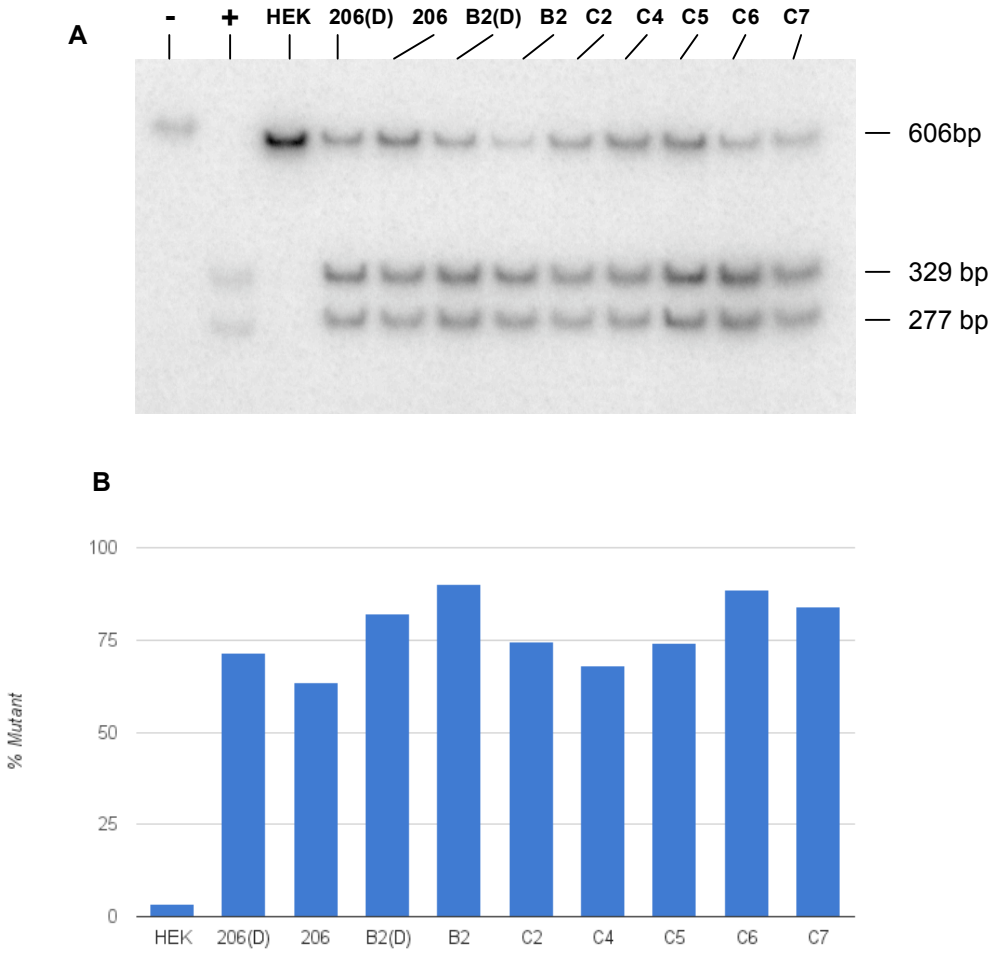


Figure 3.2.1. Heteroplasmy of HEK293T m.3243A>G cybrid clones. (**A/B**) Last-cycle PCR was used to quantify heteroplasmy in seven HEK cybrid clones (206, B2, C2, C4-7), A549 m.3243A>G cybrid (B2(D)), 143B m.3243A>G cybrid (206(D)) and parental HEK293T (HEK) cells. Mutant mtDNA generates two bands at 329 bp and 277 bp whilst WT mtDNA generates a single band of 606 bp. PCR of sequence spanning a native *Apa*I site in mtDNA followed by incubation in the presence (+) or absence (-) of *Apa*I was used as digestion control.

All clones contained mtDNA (Figure 3.2.2). Heteroplasmy levels were similar to those of mitochondrial donor cybrids at the time of cytoplast generation, which were 71.5% and 82.2% for A549 m.3243A>G and 143B m.3243A>G cybrids respectively.

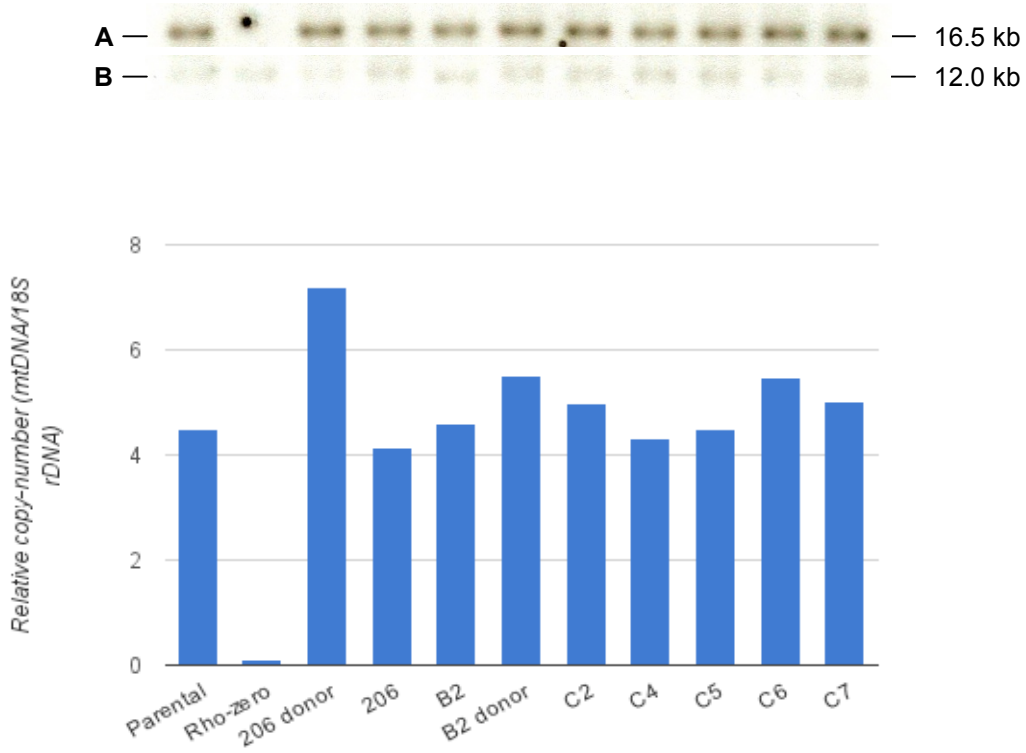


Figure 3.2.2. The repopulation of HEK293T cells with m.3243A>G mtDNA. Seven HEK293T m.3243A>G cybrid clones (206, B2, C2, C4-7) isolated from the cybridisation of ρ^0 HEK293T cells to cytoplasts derived from a A549 m.3243A>G cybrid (B2 donor) or 143B m.3243A>G cybrid (206 donor). Mitochondrial DNA copy number was measured by the ratio of the major ~16 kb mtDNA fragment (A) to 12 kb 18S rDNA (B).

3.3 Curation of microarray datasets for A549, HEK293, HeLa, NT2 and fibroblast cell lines from NCBI GEO.

NCBI GEO is a database repository for global gene expression data. Control datasets from studies submitting data to NCBI GEO can be curated and utilised for global expression analysis on a scale unfeasible to a single laboratory, in a so-called 'meta-analysis'. A meta-analysis helps to overcome false positives associated with an elevated false discovery rate in global expression data[112], which for a microarray data consists of tens of thousands of intensity measurements, as well as false positives or outliers associated with a particular study. Furthermore, large numbers of datasets allow the reliable determination of expression differences for genes with a low dynamic range of mRNA expression, where interpretation of differences between conditions would otherwise be compromised due to the increased influence of noise. Of critical importance to this study, several cell types studied in the context of biased segregation have also been widely utilised for global expression analysis by microarray.

Those cell types with both a known segregation bias and datasets in the NCBI GEO include A549 (31 datasets), fibroblast (34 datasets), HeLa (29 datasets) and NT2 (3 datasets). However, with the exception of fibroblasts, datasets correspond to the parental cell lines, whereas studies of biased segregation were performed in cybrid derivatives, which have been found to exhibit some differences in gene expression[113]. Therefore, a caveat in the use of microarray datasets to study segregation bias is that they do not necessarily represent cybrid global gene expression profiles.

An argument frequently leveled against the use of microarray data is that differences at the mRNA level correlate poorly with differences at the protein and functional level, and therefore provide limited insight into changes in cellular phenomena[114]. This is because absolute mRNA level does not distinguish mRNA synthesis and degradation, or capture protein synthesis, degradation or post-translational modification. All of these processes contribute to gene expression and are controlled by gene regulatory events[115,116]. Changes at the functional level that are undetectable to microarray include translational control by selective expression of ribosomal

proteins[117], and a switch to internal-ribosomal-entry-site dependent initiation under conditions of global translational down-regulation[118]. However, it has been found that on average, at least 40% of the variance in protein level is determined by mRNA level[119]. In addition, cellular changes at the functional level create 'gene-expression signatures' at the transcriptional level that can be detected using bioinformatic tools[106,107]. Therefore, microarray data still presents an opportunity to generate insight into many cellular phenomena and processes.

3.4 Differential expression analysis

Differential expression analysis can be used to determine which genes exhibit the greatest difference in expression between two conditions, and are therefore most correlated with a particular condition. Prior to differential expression analysis, datasets for each cell line were filtered for those genes with a link to mitochondria based on the cumulative domain expertise of group members (see Appendix, Table 6.1). This included genes linked to mitochondrial biogenesis, autophagy, dynamics, nucleoid and PD associated genes. Datasets for a single cell line were then clustered on the Babelomics 4.0 platform to allow selective removal of a minority of datasets, visualised as outliers on a dendrogram (data not shown), which exhibited expression differences in genes linked to mitochondria compared to the majority of datasets. It was hypothesised that inclusion of these datasets could obscure differences in gene expression between cell lines that may be linked to biased segregation.

During this procedure it was observed that A549 datasets could be divided into two sub-populations characterised by differential expression of the gene PPARGC1A (Figure 3.4.1), whose protein product PGC-1α is a regulator of mitochondrial biogenesis and function[120]. This was surprising since other genes with a link to mitochondria did not show such a high degree of polarised expression, and all datasets were taken from cells cultured in identical conditions. However, because the significance of polarised PPARGC1A expression was not known, A549 datasets were pooled irrespective of PPARGC1A expression for comparison to other cell lines.

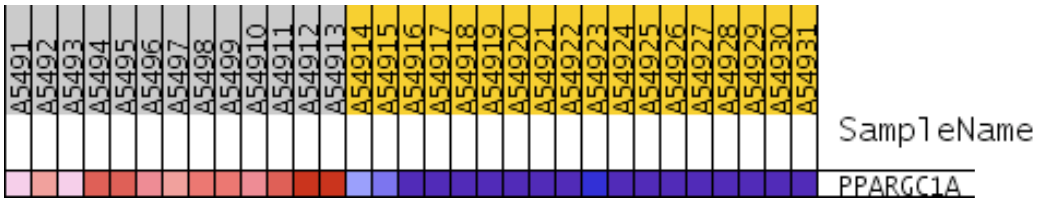


Figure 3.4.1. PPARGC1A expression is polarised in A549 datasets. Datasets were taken from A549 cells cultured in equivalent culture conditions. PPARGC1A up and down-regulation is proportional to the intensity of red and blue respectively.

Using multiple cell lines and many datasets is a much more powerful means of identifying differences between cells that can or cannot select WT mtDNA than a simple pairwise comparison (e.g. A549 cybrid versus 143B cybrid). Nevertheless, a weakness of the analysis was the availability of only one cell line known to select WT mtDNA, and the tens of thousands of gene products in the screen inevitably meant that many of the differences were unrelated to mtDNA segregation.

One approach to parsing the data is to restrict the analysis to genes with a link to mitochondrial function, although because of the complex integrated circuitry of the cell, this is a much larger geneset than the ~1500 proteins of the mitochondrial proteome. Although this approach limits understanding of biased segregation to the context of previously characterised genes, it may help to exclude genes unrelated to the phenomenon, and therefore provide increased insight. Alternatively, an unbiased analysis offers the possibility of identifying genes with a previously unrecognised link to mitochondria, but prioritises genes based on differences in mRNA expression alone, which may create bias towards genes with a high dynamic range of mRNA expression.

Both biased and unbiased approaches were pursued, since insight can be generated from both. Genes with a link to mitochondria based on cumulative domain expertise of lab members (see Appendix, Table 6.1), genes from MitoCarta, an inventory of 1013 human genes with evidence of mitochondrial protein localisation[104] and any other gene with a link to mitochondria based on hypothesis were used for biased analysis.

Three methods are commonly used to determine the significance of differential expression. The simplest method is to measure the fold-change difference in gene expression between cell lines and impose an arbitrary threshold of significance. Alternatively, the t-test and limma methods take both the magnitude of differential expression and the variability of expression values into account, in order to reduce the influence of noise. Despite this, the fold-change method generates the most reproducible results between experiments[121,122]. The MicroArray Quality Control (MAQC) group, established to develop guidelines for microarray data analysis, recommends the combining of fold-change and probabilistic t-test and limma methods to detect changes in gene expression between datasets, and this recommendation was followed here.

A549 datasets were sequentially compared to HeLa, NT2 and fibroblast datasets, because limitations of the Babelomics 4.0 platform prevented comparison of A549 to all the other cell lines grouped together. Both unbiased and biased analysis of each comparison was performed. Ranks were awarded to each gene based on the magnitude of differential expression according to the fold-change, t-test and limma methods, before combining these ranks to generate a compound score (section 2.25). The sum of compound scores, from sequential comparisons of A549 datasets to HeLa, NT2 and fibroblast datasets, was used to measure association of gene expression with the A549 cell line.

Extracting insight from unbiased analysis proved challenging. Nonetheless, 48 genes that exhibited differential expression in biased analysis also exhibited significant differential expression in at least one comparison in unbiased analysis, defined as the top 5% of all differentially expressed genes (Table 3.4.1). The most up-regulated gene across comparisons was MCT4, which mediates lactate and ketone body transport across the plasma membrane[123]. The most downregulated gene was HK2, which catalyses the first step in glycolysis. MCT4 up-regulation could increase the availability of respiratory substrates besides glucose, whilst HK2 down-regulation suggests an increased dependence on aerobic respiration. Both changes suggest an increased dependence upon mitochondrial function, which could aid the discrimination and removal of dysfunctional mitochondria containing mutant mtDNA,

which in turn could promote biased mtDNA segregation. SQSTM1 up-regulation and SNCA down-regulation might also be relevant to biased segregation of mtDNA because SQSTM1 is required for Parkin-induced mitochondrial clustering[124] and SNCA down-regulation induces mitochondrial fusion[125].

GENE	A549 v FIB	A549 v HELA	A549 v NT2	TOTAL
MCT4	**3**	**9**	**9**	**21**
SNAP25	9	3	9	21
NRG1	3	9	3	15
DAPK1	3	9	2	14
UCHL1	2	9	2	13
IGFBP1	3	3	6	12
SRC	3	6	3	12
LRRK2	2	3	6	11
RND1	3	6	2	11
RAI1	2	2	6	10
BCLXL	3	3	3	9
DNM1	3	3	3	9
DGKD	3	2	3	8
SQSTM1	**2**	**2**	**3**	**7**
SIRT3	2	2	2	6
LRPPRC	3	0	2	5
HSP75	2	0	2	4
ULK2	-2	3	3	4
NBR1	-2	2	3	3
SCARB2	-3	0	3	0
Drp1	2	0	-3	-1
MAP1B	-2	3	-2	-1
PKP4	0	2	-3	-1
FEN1	3	-2	-3	-2
MYH10	-3	3	-3	-3
SGK1	-2	-3	2	-3
C9orf86	2	-3	-3	-4
SNCA	**2**	**-3**	**-3**	**-4**
COL16A1	-3	-2	0	-5
GTPBP5	0	-6	1	-5
NUP155	2	-6	-1	-5
CCNB1	2	-2	-6	-6
ZMYM4	-2	-2	-2	-6
ZYG11B	-3	0	-3	-6
ATPIF1	-2	-2	-3	-7
C13orf27	-2	-3	-2	-7
FBXL12	-2	-2	-3	-7
FDX1L	-2	-2	-3	-7
LEPRE1	-3	-2	-2	-7
GLUT3	-2	0	-6	-8
MAP4K4	-3	-2	-3	-8
MCT1	-2	-3	-3	-8
SYT11	-6	0	-3	-9
MOSC1	2	-9	-3	-10
FYN	-6	-3	-3	-12
RPS6KA2	-9	-3	0	-12
HK2	**-3**	**-9**	**-3**	**-15**

Table 3.4.1. Genes that are up or down-regulated in A549 cells compared with three other cell types. A549 datasets were compared to datasets from cell lines with a known segregation bias towards mutant mtDNA including fibroblast (FIB), HeLa (HELA) and NT2 cells. Cumulative scores were calculated from individual scores for each of the three comparisons (TOTAL). Genes that may be relevant to biased segregation shown in bold typeface and highlighted green.

3.5 Geneset enrichment analysis

Differential expression analysis can generate a list of genes whose expression is correlated with a particular condition, but extracting insight from this list is often problematic. Furthermore, differential expression of a single gene may not have significance at the functional level. Gene set enrichment analysis (GSEA) overcomes these problems by comparing differential expression of sets of genes, or 'genesets', defined by function or previously published knowledge[95]. GSEA leverages previously published knowledge to provide functional insight and can detect differences at the functional level that are not evident from expression changes of individual genes at the mRNA level. This is achieved through the use of a cumulative score derived from the combined scores of individual genes in a geneset. GSEA utilises gene sets from the Molecular Signatures Database (MSigDB), which are divided into five major collections: positional, curated, motif, computational and Gene Ontology (GO) defined gene sets. Furthermore, GSEA provides the facility to upload user-defined genesets, allowing incorporation of unpublished knowledge into the analysis. GSEA also includes a tool called 'Leading-edge analysis', which searches for those genes that feature most frequently in the enriched genesets. This allows a user to isolate the genes most likely to control the functional differences observed between two conditions.

GSEA is a module within the GenePattern genomic analysis platform, which provides the capability to compare A549 datasets to datasets from multiple cell lines grouped together. Exploiting this capability, A549 datasets were compared to datasets from cell lines with a known segregation bias towards mutant mtDNA including fibroblast, HeLa and NT2 cells. Furthermore, A549 datasets were compared to a group of datasets taken from the NCI60 study, which generated global expression data for a panel of 60 cancer cell lines including the A549 adenocarcinoma cell line[126]. This comparison provided an opportunity to discover gene expression patterns unique to the A549 cell line. Although A549 is the only cell line with an established segregation bias within the group, learning what is unique about the A549 cell line may provide clues to the nature of biased segregation.

Analogous to differential expression analysis, GSEA generates a list of ranked markers, but in GSEA the markers are not genes but genesets, and are ranked by normalised enrichment score (NES) instead of fold-change difference or P-value. Although the transition to genesets provides increased insight, higher organisation of these genesets is lacking, which obscures an overall interpretation of the enrichment. One approach is to organise the genesets based on subjective understanding of the phenomena under study. However, this approach serves to introduce a bias, but the scope for bias is bounded if applied to an unbiased list of enriched genesets.

Enriched genesets were organised into categories whose annotations may be linked to biased mtDNA segregation, based on personal domain expertise. Comparing A549 datasets to a group of datasets from cell lines with a known segregation bias towards mutant mtDNA revealed up-regulation of genesets with a link to mitochondria, metabolism and response to stress, and down-regulation of genesets with a link to cell growth (Figure 3.5.1 B). Comparison of A549 datasets in which PPARGC1A is expressed at high levels (A549↑) to those datasets in which it is expressed at low levels (A549↓) revealed a similar profile, suggesting that the A549↑ subset of datasets was responsible for enrichment of features in the multiple-cell-line comparison (Figure 3.5.1 A).

Supporting this hypothesis, substitution of A549↑ for A549 in the multiple-cell-line comparison augmented the enrichment of features in the original profile (Figure 3.5.1 C). Comparison of A549↑ datasets to a group of datasets taken from cell lines of the NCI60 study (with exception of A549) caused further enrichment of features in this profile (Figure 3.5.1 D). Taken together, this suggests that features associated with the A549 cell line, and by extension, a reduction in the proportion of mutant mtDNA, are enriched in the A549↑ subset of datasets and are unique to the A549 cell line.

Amino acid starvation of HEK293T cells was found to reproducibly boost respiratory chain activity and reduce growth (Dr. A. Spinazzola, personal communication). Genesets associated with both these features are enriched in the profile of A549↑ datasets (Figure 3.5.1 A/C/D).

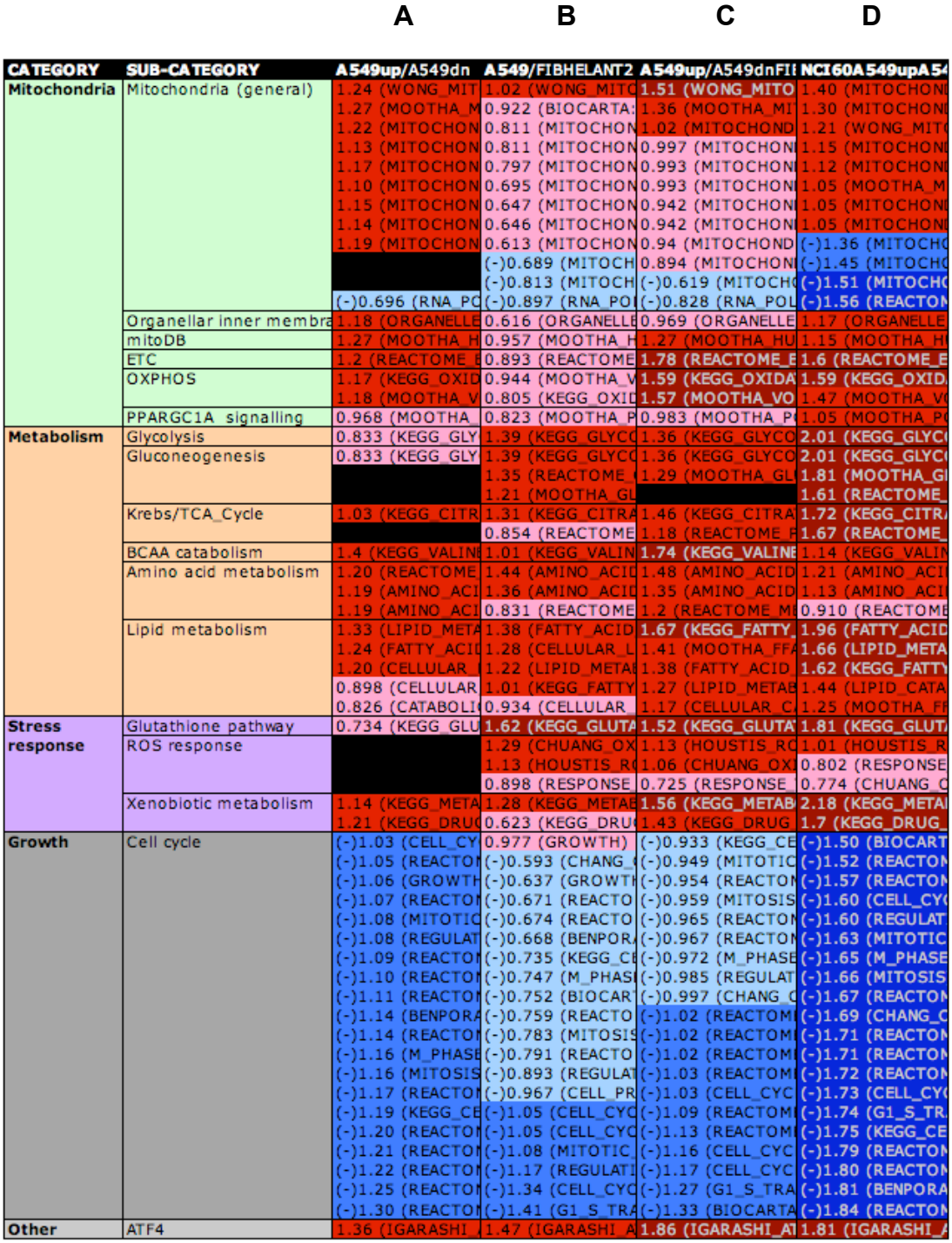


Figure 3.5.1. **Organisation of selected genesets from multiple comparisons**. Genesets with a possible link to biased segregation based on subjective understanding were organised into categories and sub-categories for comparisons A549↑ versus A549↓ (**A**), A549 versus cell lines with known segregation bias towards mutant mtDNA (**B**), A549↑ versus A549↓ and cell lines with a known segregation bias towards mutant mtDNA (**C**) and A549↑ versus remaining NCI60 cell lines (**D**). The strength of colour represents the amplitude of up-regulation (red) and down-regulation (blue) of geneset expression according to NES.

Amino acid starvation causes accumulation of uncharged tRNAs and unfolded proteins in the endoplasmic reticulum, leading to ER stress. Uncharged tRNAs activate protein kinase Gcn2[127], which phosphorylates serine 51 on the α-subunit of eukaryotic translation initiation factor 2 (eIF2α)[128]. Phosphorylation of eIF2α reduces global translation yet induces expression of transcription factor ATF4. ATF4 stimulates expression of genes involved in amino acid import, glutathione biosynthesis and resistance to oxidative stress[128]. Genesets with these annotations are all enriched in the A549↑ profile (Figure 3.5.1 A/C/D). Moreover, a geneset corresponding to the targets of ATF4 is one of the most enriched genesets across comparisons (Figure 3.5.1, category-Other). Collectively, this data suggests that the cells with an A549↑ profile are exhibiting an ER stress response.

3.6 Heteroplasmic spread and a low level of spontaneous biased segregation in HEK293T m.3243A>G cybrid sub-clones

Heterogeneity has been observed between single cells[129], clonal cell lines[77] and pooled populations (section 3.4, Figure 3.4.1) cultured under identical conditions. This heterogeneity influences the capacity for biased segregation[77], so can be exploited by studying the effect of candidate interventions across multiple clonal cell lines. This increases the probability that interventions may stimulate biased segregation, which may otherwise be masked by the study of pooled populations. However, it is first necessary to characterise the heteroplasmy level of clonal cell lines over time under standard growth conditions. This provides a baseline, which must be exceeded before an intervention can be considered to have had an impact on the levels of mutant and WT mtDNA.

Transfection of HEK cybrids with empty vector (plasmid pCDNA5) can be used to isolate clonal lines for determination of heteroplasmy levels over time in the absence of any intervention. HEK293T m.3243A>G cybrid Clone C4 (Clone C4) with 68% mutant mtDNA was transfected with pCDNA5 and sub-clones isolated in 'standard culture conditions' (section 2.14), to determine whether cell heterogeneity or the transfection and drug selection process influences heteroplasmy levels or the maintenance of heteroplasmy over time.

Any cells remaining after the isolation of sub-clones were pooled to create 'pooled populations'. Sub-clones and pooled populations were maintained in culture for at least 28 days or the time required to collect four consecutive DNA samples. 55 sub-clones and seven-pooled populations were isolated from two transfections. Pyrosequencing was used to determine heteroplasmy, with a standard deviation range of 0.0006-0.0464, or 0.06-4.64% change in heteroplasmy, obtained across 359 samples measured in triplicate. Sub-clones and pooled populations exhibited a heteroplasmic spread, ranging from 28% to 90% mutant mtDNA with a mean heteroplasmy of 65.6% mutant. Dividing the initial heteroplasmy level of sub-clones into 3% intervals, the modal interval was 69-72% mutant mtDNA (Figure 3.6.1). One of 55 (1.82%) sub-clones exhibited a spontaneous decrease in the proportion of mutant mtDNA from 50% mutant to 9% over 93 days (Figure 3.6.2 C). Remaining sub-clones and pooled populations maintained stable heteroplasmy, defined as less than a 15% change in the proportion of mutant mtDNA over 28 days or 4 consecutive time points in HEK293T m.3243A>G cybrids (Figure 3.6.2 A/B).

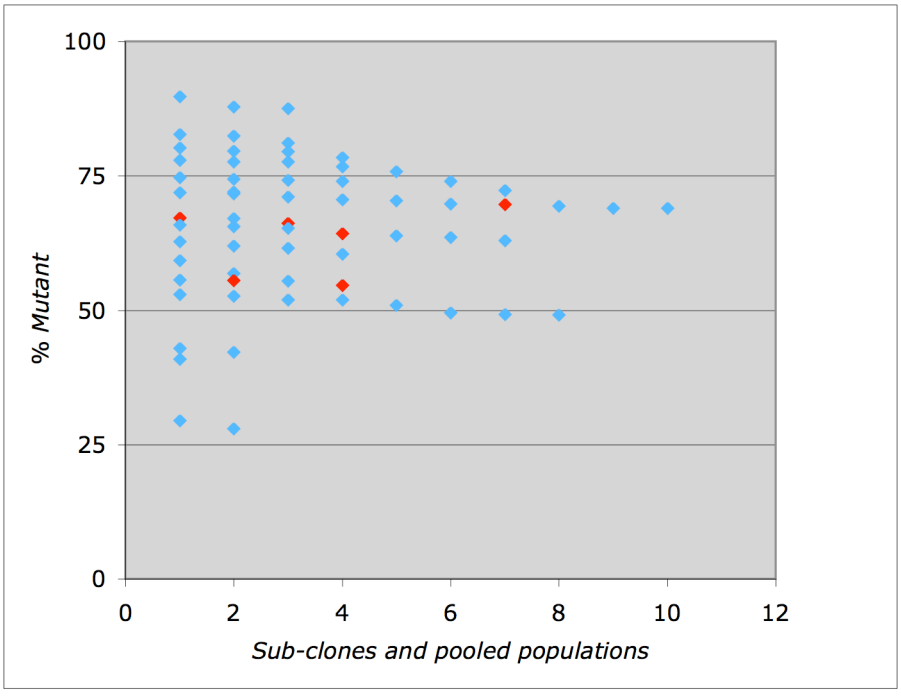
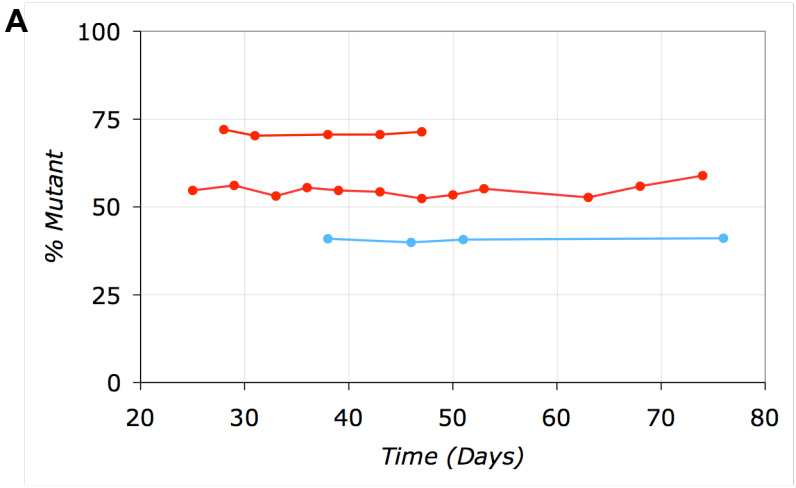
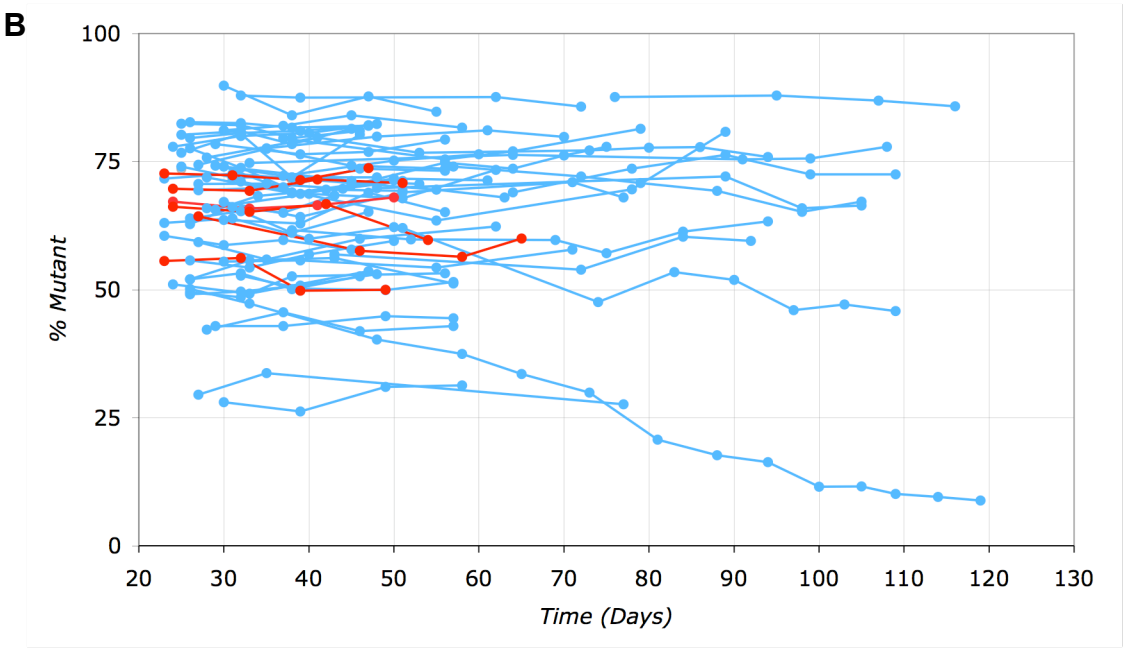


Figure 3.6.1. Heteroplasmic spread from 28.0% to 89.8% mutant mtDNA. Clone C4 was transfected with plasmid pCDNA5 to generate 55 sub-clones (blue) and seven pooled populations (red) with a variable initial heteroplasmy level.


A
100
75
50
25
0
% Mutant
20
30
40
50
60
70
80
Time (Days)


B
100
75
50
25
0
% Mutant
20
30
40
50
60
70
80
90
100
110
120
130
Time (Days)

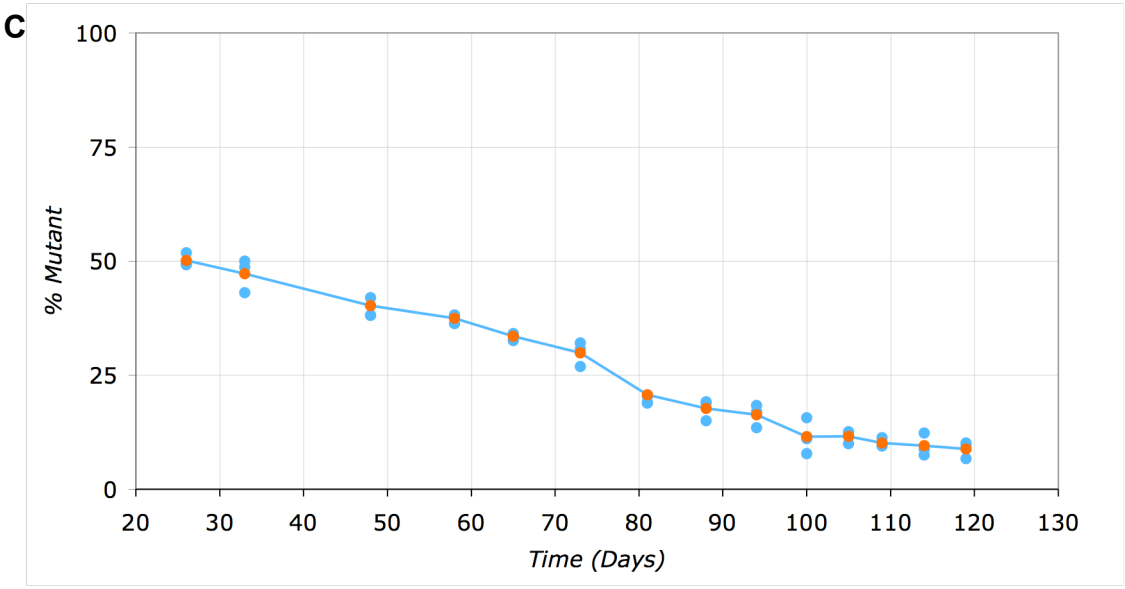

C
100
75
50
25
0
% Mutant
20
30
40
50
60
70
80
90
100
110
120
130
Time (Days)

Figure 3.6.2. Heteroplasmy is stable in the vast majority of HEK cybrids with m.3243A>G mtDNA. (**A/B**) 55 HEK cybrid sub-clones (blue) and seven pooled populations (red) isolated from two transfections of Clone C4 were maintained in standard culture conditions for up to 119 days, following transfection with plasmid pCDNA5 conferring resistance to hygromycin. DNA was harvested and the proportion of mutant mtDNA determined by pyrosequencing (•/•). (**C**) A spontaneous decrease in the proportion of mutant mtDNA was exhibited in a single sub-clone. Single (•) and mean (•) heteroplasmy measurements are plotted for each time point.

This result demonstrates that stable heteroplasmy is the norm in HEK293T m.3243A>G cybrids. Nevertheless, clonal cell line heterogeneity or the transfection and drug selection process can influence both heteroplasmy levels and the proportion of mutant mtDNA over time. Furthermore, there appears to be an upper limit of ~90% mutant mtDNA for HEK293T m.3243A>G cybrids in standard culture conditions, which has not been observed for 143B m.3243A>G or A549 m.3243A>G cybrids. Stable heteroplasmy and a low frequency of spontaneous change in the proportion of mutant mtDNA make the HEK cybrid a suitable model to study the effect of candidate interventions. However, the considerable heterogeneity in heteroplasmy and low level of spontaneous change in the proportion of mutant mtDNA in clonal cell lines suggest that is essential to study the effect of candidate interventions on biased mtDNA segregation in multiple clonal cell lines.

3.7 Intermittent amino acid starvation is not associated with biased segregation in HEK293T m.3243A>G cybrid sub-clones and pooled populations.

To test the hypothesis that a decrease in the proportion of mutant mtDNA is associated with the cellular response to ER stress, HEK cybrid sub-clones were subjected to cycles of amino acid starvation followed by recovery in standard culture conditions. To isolate sub-clones, Clone C4 was transfected with doxycycline hyclate (DOX) inducible Parkin overexpression plasmid ‘SF-Park’.

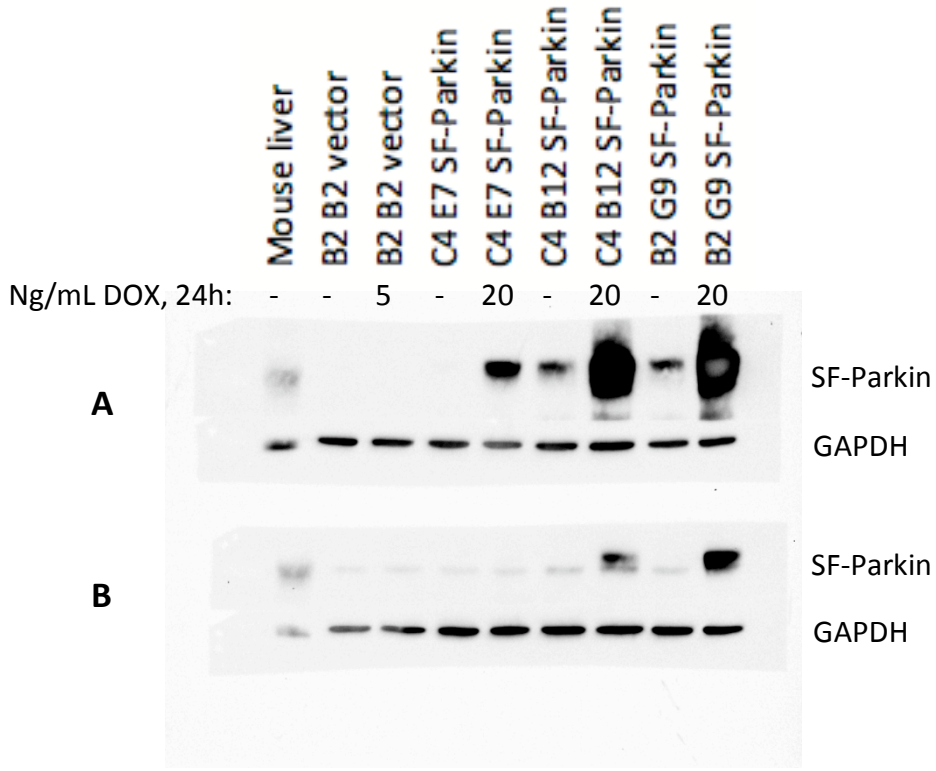
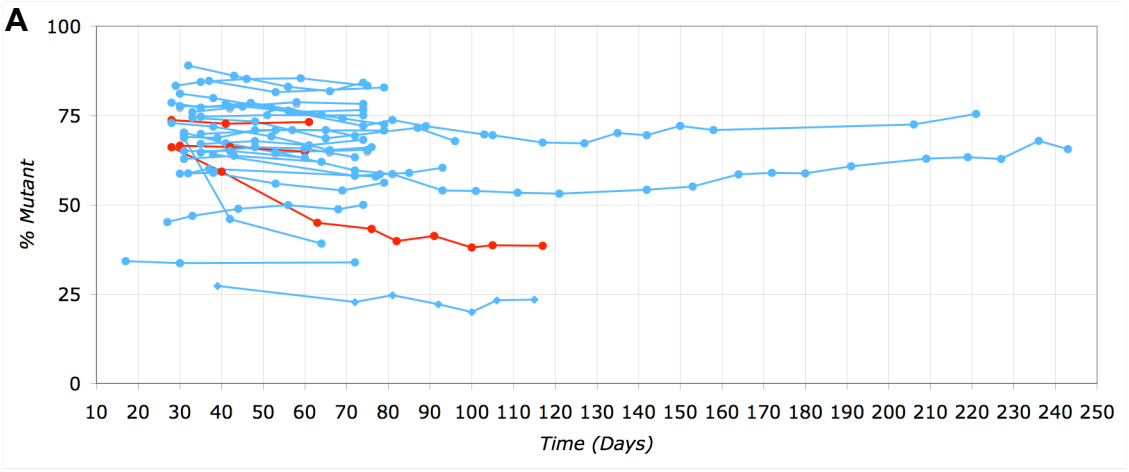
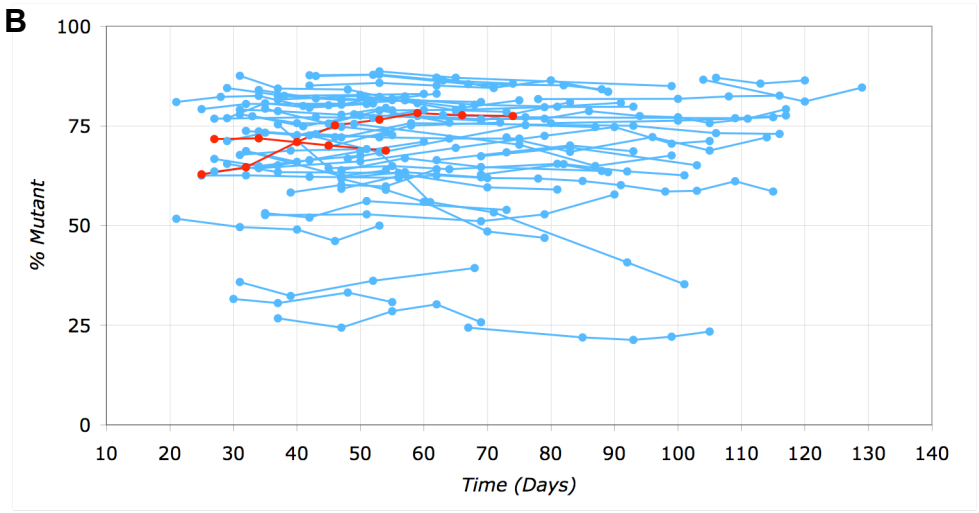
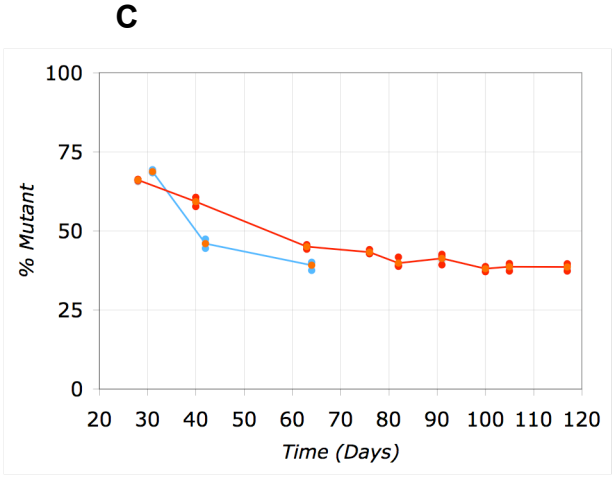


Figure 3.7.1. Parkin expression in mouse liver and HEK293T m.3243A>G cybrid clones. 20μg of protein from total cell lysate of HEK cybrid clones transfected with Parkin overexpression vector SF-Park (C4-E7 SF-Parkin, C4 B12 SF-Parkin, B2-G9 SF-Parkin) or pCDNA5 (B2 B2 vector), which were cultured in the presence or absence of DOX for 24 hours, was harvested for western blot. 20μg of protein from mouse liver, which expresses high levels of Parkin, was harvested for western blot as a positive control. High (**A**) and low (**B**) exposures of a single blot are shown (Catherine Nezich). GAPDH is used as a loading control.

Sub-clones were induced with DOX, because Parkin expression was previously found to contribute to a decrease in the proportion of mutant mtDNA in the 143B-COXICA65 cybrid cell line[83], but is not sufficient to stimulate this change in the HEK cybrid cell line (Catherine Nezich, personnel communication). Transgenic expression of Parkin induced by DOX has previously been demonstrated in the HEK cybrid cell line (Catherine Nezich, Figure 3.7.1). Three of 86 (3.48%) HEK cybrid sub-clones exhibited a progressive decrease in the proportion of mutant mtDNA (Figures 3.7.2 C/D). This was higher than the frequency of segregation in cells continuously supplied with amino acids (1.82%) but not statistically significant ($P = 0.206$).


A
100
75
50
25
0
% Mutant
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250
Time (Days)


B
100
75
50
25
0
% Mutant
10 20 30 40 50 60 70 80 90 100 110 120 130 140
Time (Days)


C
100
75
50
25
0
% Mutant
20 30 40 50 60 70 80 90 100 110 120
Time (Days)

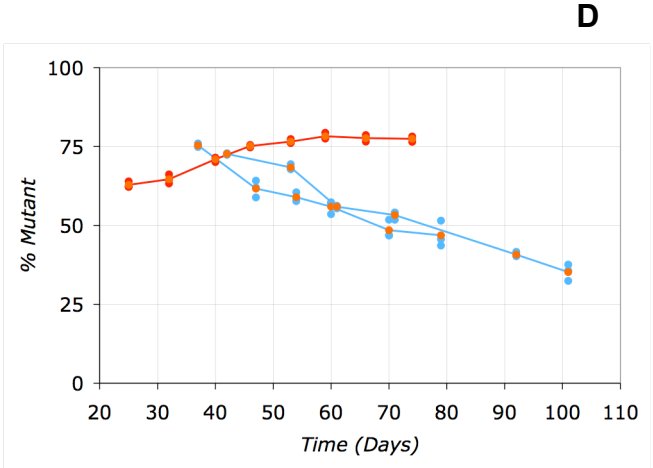
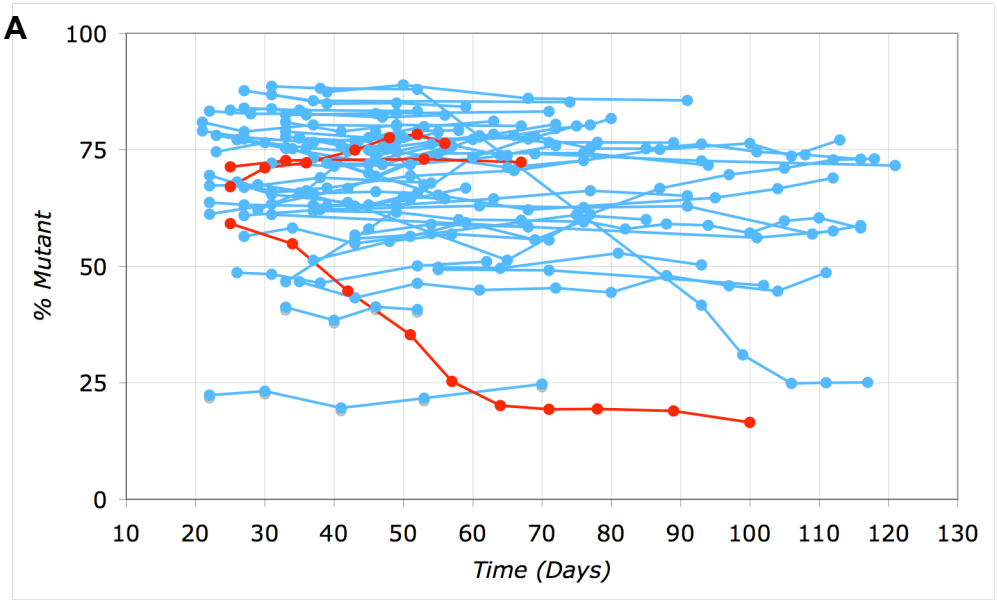

D
100
75
50
25
0
% Mutant
20 30 40 50 60 70 80 90 100 110
Time (Days)

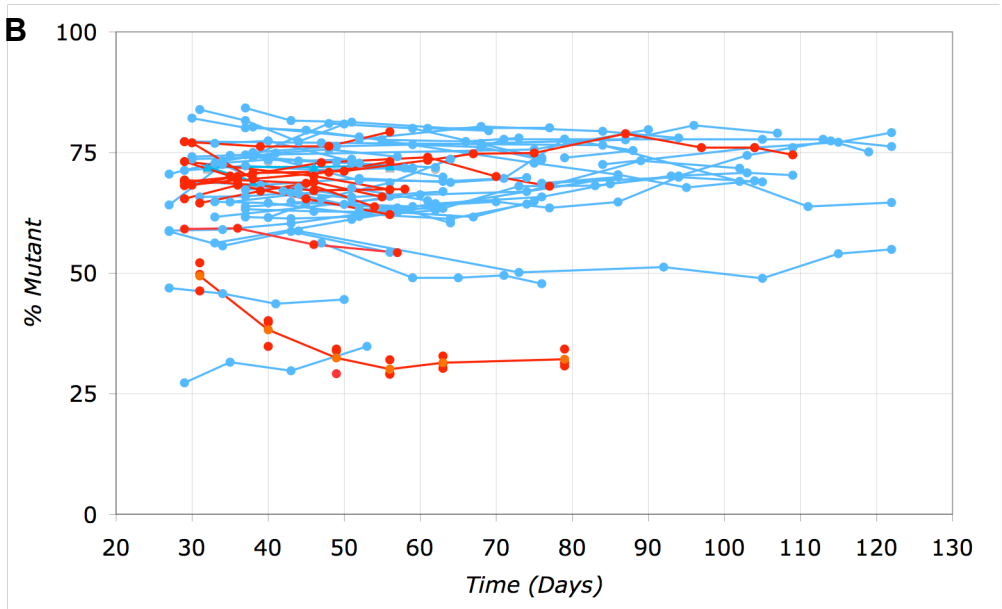
Figure 3.7.2. Heteroplasmy is stable in the vast majority of HEK293T m.3243A>G cybrids expressing Parkin and subjected to intermittent amino acid starvation. (**A/B**) 86 HEK cybrid sub-clones (blue) and five pooled populations (red) isolated from two transfections of Clone C4 were subjected to cycles of amino acid starvation followed by recovery in standard culture conditions for up to 243 or 129 days after transfection with plasmid SF-Park, expressing Parkin and conferring resistance to hygromycin. DNA was harvested and the proportion of mutant mtDNA determined by pyrosequencing (•/•). (**C/D**) Segregation to WT mtDNA was exhibited in three sub-clones and opposing segregation of mutant mtDNA in single pooled populations with intermittent amino acid starvation. Single (•/•) and mean (•) heteroplasmy measurements are plotted for each time point.

Segregation occurred in two of five pooled populations (40%) subjected to intermittent amino acid deprivation, but in opposite directions (Figure 3.7.2 C/D). Although stable heteroplasmy was maintained in all seven pooled populations cultured under standard conditions, the number of samples is too small to draw firm conclusions. Remaining sub-clones and pooled populations maintained stable heteroplasmy (Figure 3.7.2 A/B). The findings suggest that intermittent amino acid starvation and elevated Parkin expression may increase the proportion of segregation occurring. However, this treatment does not favour mutant or WT mtDNA.

3.8 Intermittent 2DG treatment is not associated with biased mtDNA segregation in HEK293T m.3243A>G cybrid sub-clones and pooled populations.

The ER stress response is engaged by amino acid starvation, but other features of amino acid starvation may be involved in changes in the proportion of mutant mtDNA. To determine whether the ER stress response is associated with biased mtDNA segregation, the alternative ER stressor 2-deoxy-D-glucose (2DG) was used in surrogate. Clone C4 was transfected with SF-Park plasmid and induced to express Parkin, whilst sub-clones were subjected to cycles of 2DG treatment followed by recovery in standard culture conditions.


A
100
75
50
25
0
% Mutant
10 20 30 40 50 60 70 80 90 100 110 120 130
Time (Days)


B
100
75
50
25
0
% Mutant
20 30 40 50 60 70 80 90 100 110 120 130
Time (Days)

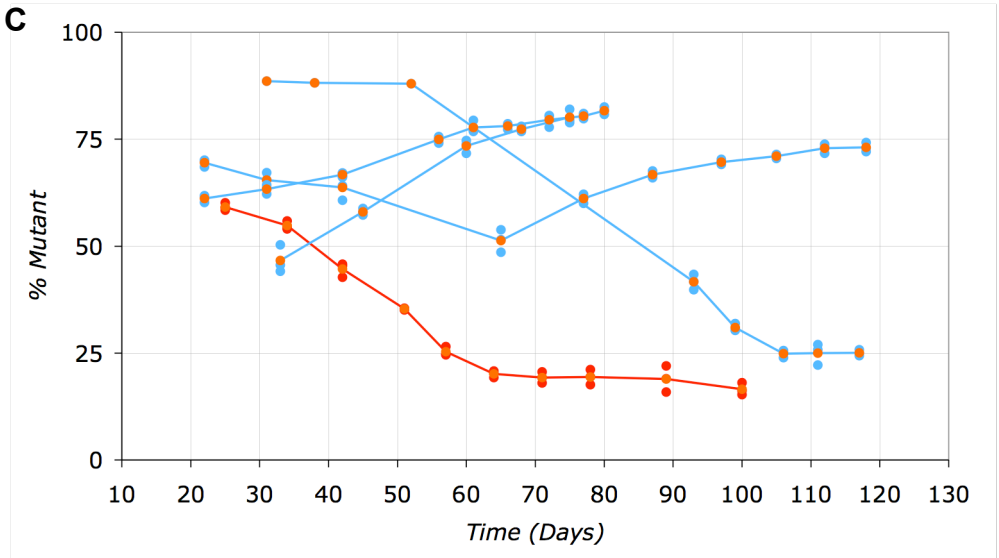

C
100
75
50
25
0
% Mutant
10 20 30 40 50 60 70 80 90 100 110 120 130
Time (Days)

Figure 3.8.1. Heteroplasmy is stable in the vast majority of HEK293T m.3243A>G cybrids expressing Parkin and subjected to intermittent 2DG treatment. (**A/B**) 86 HEK cybrid sub-clones (blue) and 15 pooled populations (red) isolated from two transfections of Clone C4 were subjected to cycles of 24 hours 2DG treatment followed by recovery in standard culture conditions for up to 121 or 122 days after transfection with plasmid SF-Park, expressing Parkin and conferring resistance to hygromycin. DNA was harvested and the proportion of mutant mtDNA was determined by pyrosequencing (•/•). (**B/C**) Two sub-clones and two pooled populations exhibited a reduction in the proportion of mutant mtDNA, two sub-clones exhibited an increase in the proportion of mutant mtDNA, and a single sub-clone segregated to WT mtDNA initially, but the segregation bias reversed and heteroplasmy returned to initial levels. Single (•/•) and mean (•) heteroplasmy measurements are plotted for each time point.

Two of 86 (2.33%) 2DG-treated sub-clones exhibited a decrease in the proportion of mutant mtDNA (Figure 3.8.1 B/C). This was only marginally higher than the frequency of segregation in the absence of 2DG treatment (1.82%) and was not statistically significant ($P = 0.465$). Two of 15-pooled populations (13.3%) exhibited a decrease in the proportion of mutant mtDNA, whereas seven pooled populations not exposed to 2DG have already been shown to maintain stable heteroplasmy (Figure 3.8.1 A/B). Furthermore, two of 86 (2.33%) sub-clones treated with 2DG exhibited an increase in the proportion of mutant mtDNA, which does not occur spontaneously (Figure 3.8.1 C). One of the sub-clones that segregated to WT mtDNA initially, reversed its segregation bias after 65 days, and after a further 32 days it had returned to its initial level of ~70% mutant mtDNA (Figure 3.8.1 C).

Together with the results from intermittent amino acid starvation of HEK cybrids expressing Parkin, this suggests that the ER stress response and elevated Parkin expression might increase the proportion of segregation occurring, but does not introduce a bias for mutant or WT mtDNA. Amino acid starvation and 2DG treatment of sub-clones and pooled populations has been observed to stimulate both an increase and decrease in the proportion of mutant mtDNA, suggesting that both types of behaviour are variations of the same underlying phenomena. The transition from a decrease to an increase in mutant mtDNA in a single sub-clone with 2DG treatment supports this idea.

3.9 ER stress has no appreciable effect on the segregation of m.3243A>G mtDNA in HEK cybrids without transgenic Parkin.

To determine whether transgenic Parkin contributed to the segregation of m.3243A>G mtDNA in HEK cybrids subjected to intermittent 2DG treatment, Clone C4 was transfected with an empty vector (pCDNA5). After establishing sub-clones and pooled populations, the cell lines were subjected to intermittent cycles of 2DG treatment followed by recovery in standard culture conditions.

One of 41 (2.44%) sub-clones isolated from a single transfection exhibited an increase in the proportion of mutant mtDNA (Figure 3.9.1). This was similar to the two of 86 (2.33%) sub-clones expressing Parkin that exhibited an increase in the proportion of mutant mtDNA (P = 0.380). Remaining sub-clones and pooled populations exhibited stable heteroplasmy.

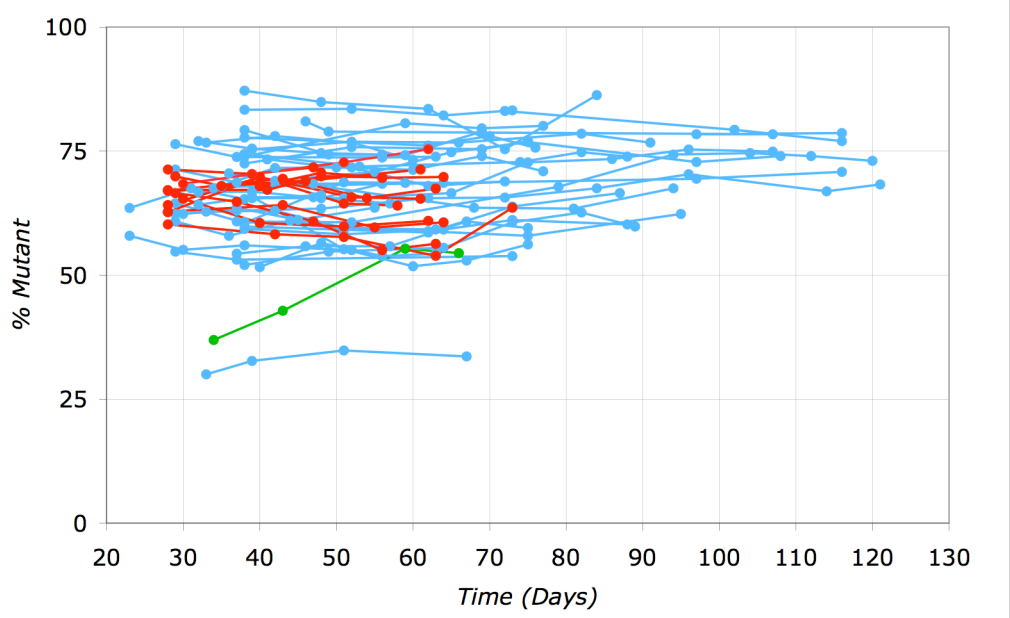


Figure 3.9.1. Intermittent 2DG treatment does not significantly change the proportion of mutant mtDNA in HEK293T cybrid sub-clones and pooled populations. HEK cybrids were subjected to cycles of 24 hours 2DG treatment followed by recovery in standard culture conditions for up to 121 days, after transfection with plasmid pCDNA5 conferring resistance to hygromycin. DNA was harvested and the proportion of mutant mtDNA determined by pyrosequencing (•/•/•). Of 41 sub-clones (blue/green) and 12-pooled populations (red), a single sub-clone exhibited an increase in the proportion of mutant mtDNA (green).

Although in experiments with 2DG, two of 98 clones carrying the Parkin transgene exhibited segregation to WT mtDNA and none of 41 clones with only endogenous Parkin did likewise (Figure 3.9.1), this cannot be taken as evidence that elevated Parkin expression favours selection of WT mtDNA in HEK cybrids. Not only is the number of clones sampled too small, one of 55 (1.82%) sub-clones without transgenic Parkin or 2DG treatment segregated to WT mtDNA. Thus, the spontaneous rate of segregation is unchanged by elevated Parkin expression or 2DG treatment of HEK cells.

3.10 Intermittent 2DG treatment induces a decrease in the proportion of mutant mtDNA in A549 m.3243A>G cybrid sub-clones and pooled populations that is not observed in the absence of treatment.

In a previous study, 20% of A549 m.3243A>G cybrid clones were found to spontaneously segregate to WT mtDNA following cybridisation[77]. This compares with 1.82% of HEK cybrids recorded in this study. A549 cells have also been found to express high levels of Parkin compared to HEK293 and 143B cell lines (Catherine Nezich, personal communication,), which has been found to contribute to a reduction in mutant mtDNA[83]. This suggests A549 cells have a much greater propensity to segregate mutant and WT mtDNA. Thus, if 2DG treatment of HEK cells was responsible for the apparent small increase in the frequency of mtDNA segregation, then this effect should be more evident in A549 cells. That is, intermittent 2DG treatment of A549 m.3243A>G cybrids (A549 cybrids) would be expected to promote a decrease in the proportion of mutant mtDNA in a larger proportion of sub-clones than in HEK cybrids. To test this hypothesis, limiting dilution was used to isolate 16 sub-clones and a pooled population from the A549 cybrid cell line. The sub-clones and the pooled population exhibited a range of initial heteroplasmy, with a maximum of 100.0% and minimum of 10.2% mutant mtDNA (Figure 3.10.1 A/B). They were subjected to intermittent cycles of 2DG treatment followed by recovery in standard culture conditions, whilst untreated duplicates of each sub-clone and the pooled population were cultured in parallel for the purposes of comparison. In a first experiment performed by my colleague Catherine Nezich, one of 16 sub-clones (6.25%) and the pooled population exhibited segregation to WT mtDNA dependent

on intermittent 2DG treatment (Figure 3.10.1 A). One other sub-clone exhibited segregation to WT mtDNA both with and without 2DG treatment, although this was accelerated with 2DG treatment at earlier time points (Figure 3.10.1 B). Remaining sub-clones exhibited stable heteroplasmy irrespective of 2DG treatment, defined as less than a 15% change in the proportion of mutant mtDNA over 26 days or 6 consecutive time points in A549 m.3243A>G cybrids (Figure 3.10.1 A/B). In a repeat experiment of mine, 13 sub-clones and a pooled population were monitored. Sub-clones and the pooled population exhibited a range of initial heteroplasmy with a maximum of 95.0% and a minimum of 34.2% mutant mtDNA (Figure 3.10.2 A/B). Two of the 13 sub-clones (15.4%) and the pooled population segregated to WT mtDNA, dependent on intermittent 2DG treatment (Figure 3.10.2 B). Remaining sub-clones exhibited stable heteroplasmy (Figure 3.10.2 A/B).

Combining the data from the two experiments, the mean change in heteroplasmy for clones exhibiting stable heteroplasmy (defined above) irrespective of 2DG treatment, was 0.7%, whilst the mean change in heteroplasmy for clones and pooled populations exhibiting a change in heteroplasmy dependent on 2DG treatment was 21.6%, which is statistically significant by the Welch's T-test ($p=0.0056$). The frequency of 2DG-induced segregation to WT mtDNA in A549 cybrid sub-clones was 10.3% (3/29) and 100% in pooled populations (2/2). Although HEK cybrid sub-clones and pooled populations transfected with empty vector (pCDNA5) were not monitored concurrently in both the presence and absence of intermittent 2DG treatment, the absence of segregation with 2DG treatment in 41 sub-clones and 12 pooled populations implies a maximum frequency of less than 2.5% (1/40) and less than 8.3% (1/12) respectively for 2DG induced segregation to WT-mtDNA. These frequencies are at least four-fold and 12-fold lower than those for A549 cybrid sub-clones or pooled populations respectively. Moreover, the frequency of spontaneous segregation to WT mtDNA in HEK293 m.3243A>G cybrid sub-clones, of 1.8 % (1/55), was almost two-fold lower than that observed among A549 cybrid sub-clones (3.5%, 1/29). Thus, the data support the hypothesis that A549 cybrids have a greater propensity to segregate to WT mtDNA than HEK293T cybrids. Moreover, it supports the contention that ER stress, in the form of intermittent 2DG treatment, promotes segregation of mutant and WT mtDNA in otherwise stable heteroplasmic cell lines.

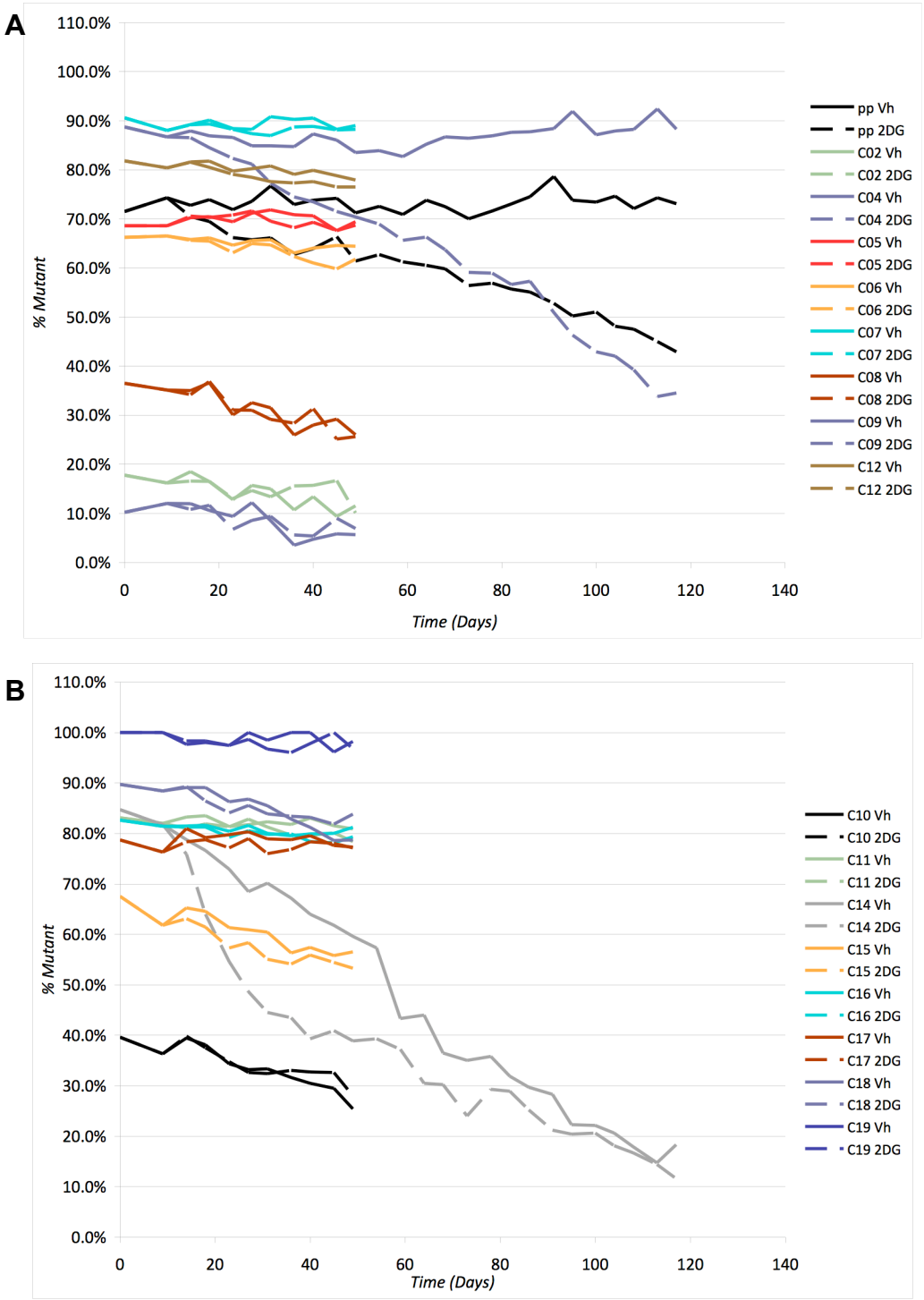


Figure 3.10.1. 2DG treatment decreases the proportion of mutant mtDNA in an A549 m.3243A>G cybrid sub-clone and pooled population. (**A/B**) 16 A549 cybrid sub-clones and a pooled population were subjected to cycles of 24 hours 2DG treatment followed by recovery in standard culture conditions for up to 117 days, whilst duplicates were maintained in the absence of treatment. DNA was harvested and the proportion of mutant mtDNA was determined by pyrosequencing.

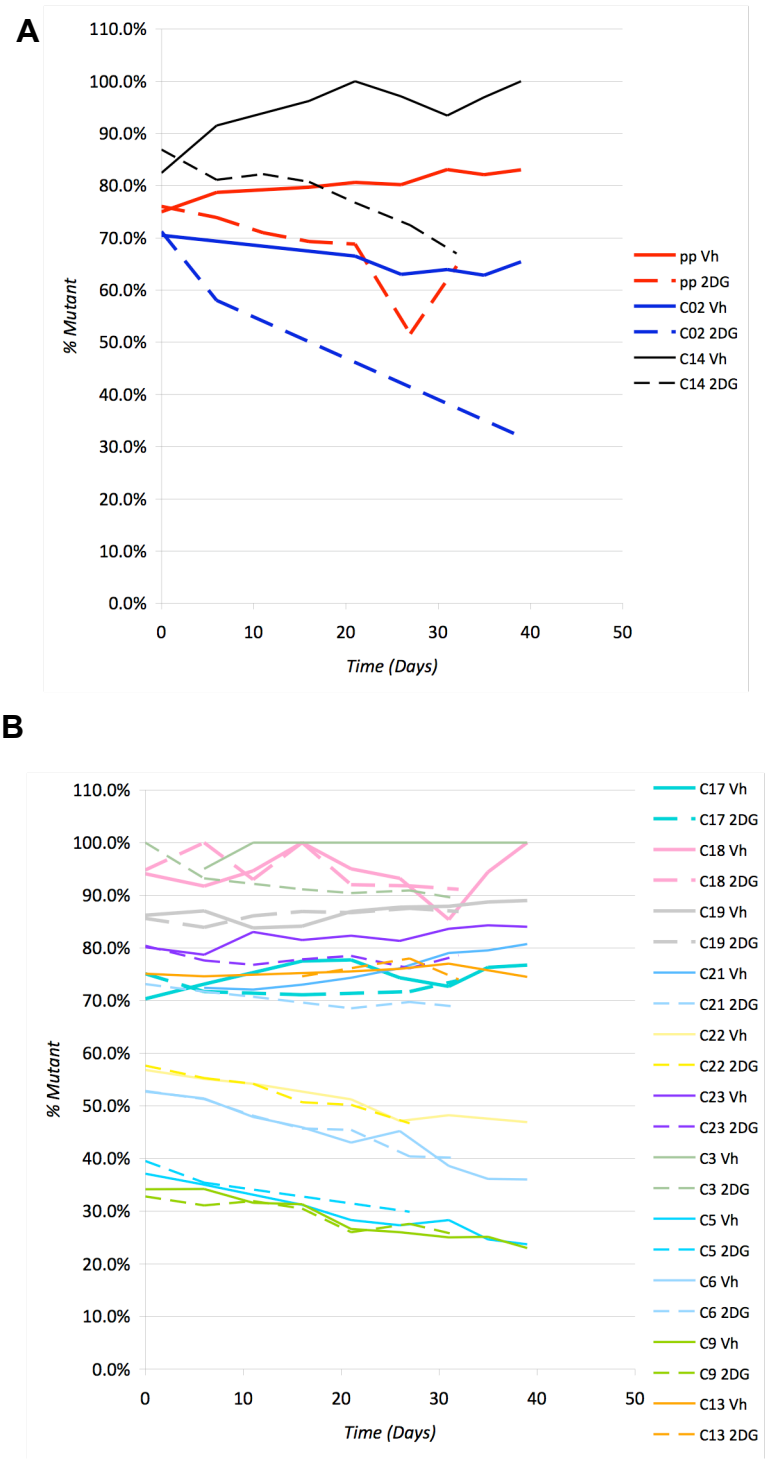


Figure 3.10.2. 2DG treatment decreases the proportion of mutant mtDNA in two A549 m.3243A>G cybrid sub-clones and a pooled population. (**A/B**) 13 A549 cybrid sub-clones and a pooled population were subjected to cycles of 24 hours 2DG treatment followed by recovery in standard culture conditions for up to 39 days whilst duplicates were maintained without treatment. DNA was harvested and the proportion of mutant mtDNA determined by pyrosequencing.

3.11 Unbiased biological process mapping

NCBI GEO contains 947654 datasets (23rd June 2013) corresponding to genetic, pharmaceutical and environmental perturbations of the cell. If a geneset achieves significant regulation across the NCBI GEO database, as evidenced by significant correlation among its genes during perturbation of the cell, then it may represent an underlying functional process[109]. Based on this hypothesis, J. Chang created an unbiased architecture of the cell composed of 115 unbiased biological processes (UBPs), which each contain genesets that share a similar transcriptional profile.

The UBP architecture serves as an unbiased higher organising principle for genesets, and therefore offers a route to increased insight from a list of enriched genesets identified by GSEA. The annotation of genesets from enriched UBPs may also provide insight beyond the annotation of the mapped genesets. Moreover, since UBPs are derived from the cellular responses to genetic, pharmaceutical and environmental perturbations, mapping the gene expression response of unknown phenomena to UBPs can provide a route to genetic, pharmaceutical and environmental perturbations, in isolation or combination, that induce the same phenomena. This method could overcome the main technical hurdle to the development of combinatorial interventions that are not approachable by traditional screening methods due to combinatorial explosion.

To identify functional processes stimulated by 2DG treatment, and therefore associated with a decrease in the proportion of mutant mtDNA, GSEA of datasets from fibrosarcoma (HT1080) cells treated with 2DG versus untreated cells was performed, followed by mapping of enriched genesets to UBPs. Multiple UBPs with immunological annotations were enriched with 2DG treatment (23, 28 and 16), aswell as a UBP consistent with the ER stress response (UBP15) (Figure 3.11.1 2DG).

Inspection of genesets from enriched UBPs revealed ‘Interferon-response’ as a common annotation in both UBP23 and UBP28. Moreover, the identity of multiple genesets within the enriched UBP profile is consistent with an interferon, or downstream response.

2DG

UBP	Geneset	NES
15	AMUNDSON_RESPONSE_TO_ARSENITE	1.370609
15	GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS_RED_UP	1.5052528
15	HELLER_SILENCED_BY_METHYLATION_DN	1.6916149
15	KRIGE_AMINO_ACID_DEPRIVATION	1.6690961
15	ZHAN_MULTIPLE_MYELOMA_CD1_VS_CD2_UP	1.8842561
23	DER_IFN_BETA_RESPONSE_UP	1.264482
23	DER_IFN_GAMMA_RESPONSE_UP	1.478955
23	KIM_LRRC3B_TARGETS	1.07655
23	KRASNOSELSKAYA_ILF3_TARGETS_UP	1.239814
23	LIANG_SILENCED_BY_METHYLATION_2	0.878895
23	MAHADEVAN_RESPONSE_TO_MP470_UP	1.5053672
23	MISSIAGLIA_REGULATED_BY_METHYLATION_UP	1.454222
23	SANA_TNF_SIGNALING_UP	1.5460292
23	SEITZ_NEOPLASTIC_TRANSFORMATION_BY_8P_DELETION_UP	0.957863
23	TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_10D_UP	1.248571
23	TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_8D_UP	0.905044
23	ZHAN_MULTIPLE_MYELOMA_LB_DN	1.322458
23	ZHU_CMV_8_HR_UP	1.368012
23	ZHU_CMV_ALL_UP	1.112327
28	DER_IFN_ALPHA_RESPONSE_UP	1.492909
28	GRANDVAUX_IFN_RESPONSE_NOT_VIA_IRF3	
28	RADAEVA_RESPONSE_TO_IFNA1_UP	1.098855
28	UROSEVIC_RESPONSE_TO_IMIQUIMOD	1.231243
16	BILD_HRAS_ONCOGENIC_SIGNATURE	1.237727
16	FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS_UP	
16	HINATA_NFKB_TARGETS_FIBROBLAST_UP	1.032761
16	HINATA_NFKB_TARGETS_KERATINOCYTE_UP	1.119782
16	KIM_WT1_TARGETS_UP	1.003472
16	SENESE_HDAC1_AND_HDAC2_TARGETS_UP	1.11256
16	TIAN_TNF_SIGNALING_VIA_NFKB	1.113881
71	KEGG_COMPLEMENT_AND_COAGULATION_CASCADES	1.345246
71	LIU_VAV3_PROSTATE_CARCINOGENESIS_UP	1.315439
71	TARTE_PLASMA_CELL_VS_PLASMABLAST_UP	1.244642
71	VART_KSHV_INFECTION_ANGIOGENIC_MARKERS_UP	1.0576
71	WANG_ESOPHAGUS_CANCER_VS_NORMAL_UP	1.162583
71	ZHANG_ANTIVIRAL_RESPONSE_TO_RIBAVIRIN_UP	0.809041
71	module_107	1.050309
71	module_172	0.995695
71	module_209	1.067206
101	module_274	0.699442
101	CATION_CHANNEL_ACTIVITY	0.825217
101	CATION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	1.313688
101	CELL_CELL_SIGNALING	1.043674
101	GATED_CHANNEL_ACTIVITY	0.903554
101	ION_CHANNEL_ACTIVITY	0.907237
101	ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	1.234466
101	METAL_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	1.180119
101	NEUROLOGICAL_SYSTEM_PROCESS	0.893782
101	SUBSTRATE_SPECIFIC_CHANNEL_ACTIVITY	0.837515
101	SUBSTRATE_SPECIFIC_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	1.414688
101	SUBSTRATE_SPECIFIC_TRANSPORTER_ACTIVITY	1.430744
101	SYNAPTIC_TRANSMISSION	1.051538
101	TRANSMEMBRANE_TRANSPORTER_ACTIVITY	1.404913
101	TRANSMISSION_OF_NERVE_IMPULSE	0.946199
101	VOLTAGE_GATED_CATION_CHANNEL_ACTIVITY	0.741454
101	VOLTAGE_GATED_CHANNEL_ACTIVITY	0.837589

NoAA

UBP	Geneset	NES
23	DER_IFN_BETA_RESPONSE_UP	
23	DER_IFN_GAMMA_RESPONSE_UP	
23	KIM_LRRC3B_TARGETS	
23	KRASNOSELSKAYA_ILF3_TARGETS_UP	
23	LIANG_SILENCED_BY_METHYLATION_2	
23	MAHADEVAN_RESPONSE_TO_MP470_UP	
23	MISSIAGLIA_REGULATED_BY_METHYLATION_UP	1.4230852
23	SANA_TNF_SIGNALING_UP	
23	SEITZ_NEOPLASTIC_TRANSFORMATION_BY_8P_DELETION_UP	
23	TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_10D_UP	
23	TAKEDA_TARGETS_OF_NUP98_HOXA9_FUSION_8D_UP	
23	ZHAN_MULTIPLE_MYELOMA_LB_DN	
23	ZHU_CMV_8_HR_UP	
23	ZHU_CMV_ALL_UP	
15	AMUNDSON_RESPONSE_TO_ARSENITE	1.3894634
15	GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS_RED_UP	
15	HELLER_SILENCED_BY_METHYLATION_DN	1.5374563
15	KRIGE_AMINO_ACID_DEPRIVATION	1.3852336
15	ZHAN_MULTIPLE_MYELOMA_CD1_VS_CD2_UP	
98	BASSO_CD40_SIGNALING_UP	
98	DIAZ_CHRONIC_MEYLOGENOUS_LEUKEMIA_DN	
98	DUNNE_TARGETS_OF_AML1_MTG8_FUSION_UP	
98	HADDAD_T_LYMPHOCYTE_AND_NK_PROGENITOR_DN	
98	HELLER_SILENCED_BY_METHYLATION_UP	0.9784917
98	LINDSTEDT_DENDRITIC_CELL_MATURATION_A	
98	REACTOME_SIGNALING_IN_IMMUNE_SYSTEM	1.2004755
98	SMID_BREAST_CANCER_LUMINAL_B_DN	0.6168267
98	TONKS_TARGETS_OF_RUNX1_RUNX1T1_FUSION_HSC_DN	
98	WILENSKY_RESPONSE_TO_DARAPLADIB	
98	module_165	
98	module_177	
98	module_27	1.3761657
98	module_64	1.3382909
98	module_76	
98	DEFENSE_RESPONSE	1.190741
98	IMMUNE_RESPONSE	1.6715913
98	IMMUNE_SYSTEM_PROCESS	1.3516976
16	BILD_HRAS_ONCOGENIC_SIGNATURE	1.2106909
16	FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS_UP	1.3558673
16	HINATA_NFKB_TARGETS_FIBROBLAST_UP	
16	HINATA_NFKB_TARGETS_KERATINOCYTE_UP	
16	KIM_WT1_TARGETS_UP	1.3061521
16	SENESE_HDAC1_AND_HDAC2_TARGETS_UP	1.1931876
16	TIAN_TNF_SIGNALING_VIA_NFKB	
91	GEISS_RESPONSE_TO_DSRNA_UP	
91	GRAHAM_CML_QUIESCENT_VS_NORMAL_DIVIDING_UP	
91	GRAHAM_NORMAL_QUIESCENT_VS_NORMAL_DIVIDING_UP	
91	KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	1.2467966
91	REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES	
91	RUTELLA_RESPONSE_TO_CSF2RB_AND_IL4_UP	0.6922505
91	TSAI_DNAJB4_TARGETS_UP	
91	module_108	
91	module_263	
91	CHEMOKINE_ACTIVITY	
91	CHEMOKINE_RECEPTOR_BINDING	
91	INFLAMMATORY_RESPONSE	

Figure 3.11.1. UBPs with immunological annotations are enriched with 2DG treatment or amino acid starvation. Genesets annotated with the 'Interferon-response' (Yellow), genesets for signalling pathways consistent with an interferon-response (Orange), genesets for metal ion transporters (Green) associated with an interferon-response and other pathways associated with a response to pathogens (blue) are highlighted in enriched UBPs. The strength of colour represents the amplitude of geneset up-regulation (red) according to NES.

Interferon is a cytokine secreted in response to bacterial or viral infection that inhibits the infection of neighboring cells by binding to interferon receptors at the cell surface to activate cell-autonomous immunity[130]. Cell autonomous immunity is mediated in part by the 'Resistome'[131], a set of genes most stimulated by interferon, but also stimulated by TNFα (UBP16, UBP23) or extracellular pathogens themselves

(UBP16). One feature of the resistome is the up-regulation of metal ion transporters that promote the efflux of metal ions from the cytosol to the extracellular space. This helps to combat pathogens, whose enzymes are more dependent upon these ions (UBP101). The UBP profile is consistent with the finding that 2DG has antiviral activity towards a variety of enveloped viruses[132].

GSEA of datasets from amino acid starved human hepatoma (HepG2) compared to cells with sufficient amino acids, followed by mapping of enriched genesets to UBPs, revealed up-regulation of UBP23, containing two genesets annotated with the 'Interferon-response' (Figure 3.11.1 NoAA). This suggests that the capacity of 2DG to stimulate an ER stress response is also the activity responsible for the subsequent activation of an interferon-response.

Interferon-inducible genes are members of a group identified as essential to Parkin mediated mitophagy by an siRNA screen[133]. This group of genes is enriched with both 2DG treatment of HT1080 cells and A549 cells, compared to other NCI60 cell lines (data not shown). Interferon has also been found to reduce somatic mutation of mtDNA in liver tissue from chronic viral hepatitis patients[134]. It is tempting to speculate that stimulation of Parkin mediated mitophagy by a 2DG-induced interferon-response stimulates segregation to WT mtDNA in HEK and A549 cybrids. To test this hypothesis, Clone C4 was transfected with SF-Park plasmid and induced to express Parkin, whilst sub-clones were subjected to cycles of combined Interferon α and Interferon γ treatment (IFNα/γ) followed by recovery in standard culture conditions.

3.12 IFNα/γ treatment has no appreciable effect on the segregation of m.3243A>G mtDNA in HEK cybrids expressing transgenic Parkin.

One of 36 (2.78%) sub-clones exhibited a decrease in the proportion of mutant mtDNA with intermittent IFNα/γ treatment followed by recovery in standard culture conditions (Figure 3.12.1), which was elevated relative to the spontaneous frequency (1.82%) but not statistically significant ($P = 0.483$). Remaining sub-clones and pooled populations exhibited stable heteroplasmy (Figure 3.12.1).

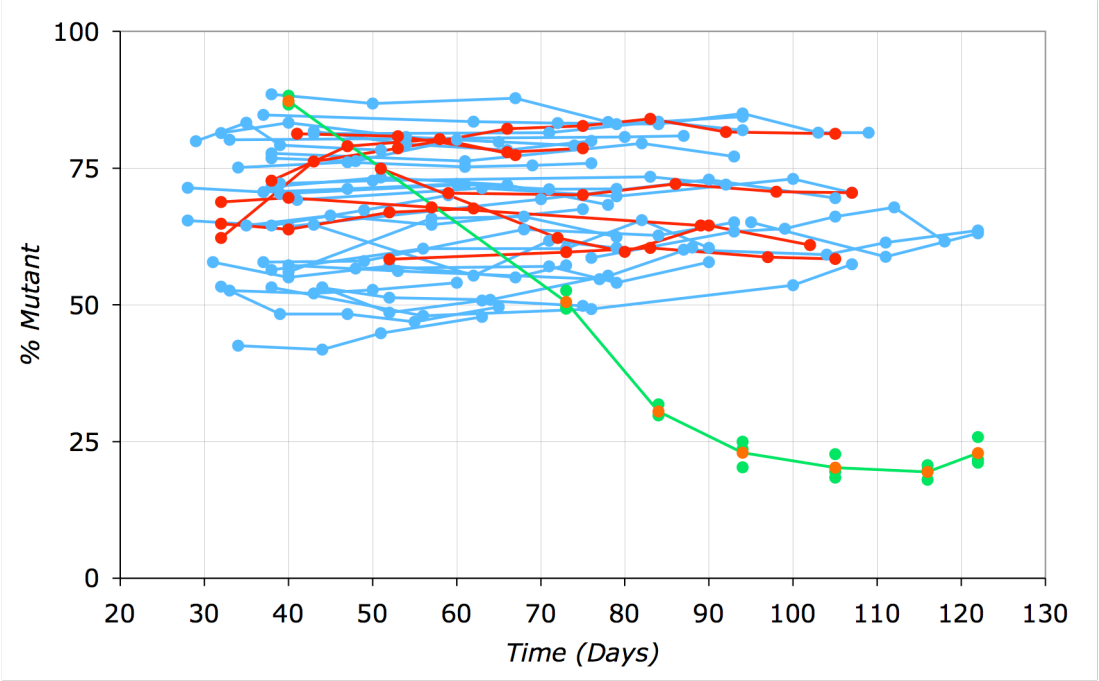


Figure 3.12.1. Intermittent IFNα/γ treatment does not change the proportion of mutant mtDNA in HEK293T cybrid sub-clones and pooled populations. HEK cybrids were subjected to cycles of 6 hours IFNα/γ treatment followed by growth in standard culture conditions for up to 122 days following transfection with plasmid SF-Park, expressing Parkin and conferring resistance to hygromycin. DNA was harvested and the proportion of mutant mtDNA was determined by pyrosequencing (•/•/•). Of 36 sub-clones (blue/green) and eight-pooled populations (red), a single sub-clone exhibited a decrease in the proportion of mutant mtDNA (green). Single (•/•/•) and mean (•) heteroplasmy measurements are plotted for each time point.

This result suggests that intermittent IFNα/γ treatment and Parkin expression do not affect the capacity to decrease the proportion of mutant mtDNA in sub-clones and pooled populations.

3.13 The only effect of intermittent IFNα/γ treatment is to inhibit the segregation to WT mtDNA in a single A549 m.3243A>G cybrid sub-clone.

The data here support the hypothesis that A549 cybrids have a greater propensity to segregate to WT mtDNA than HEK293T cybrids.

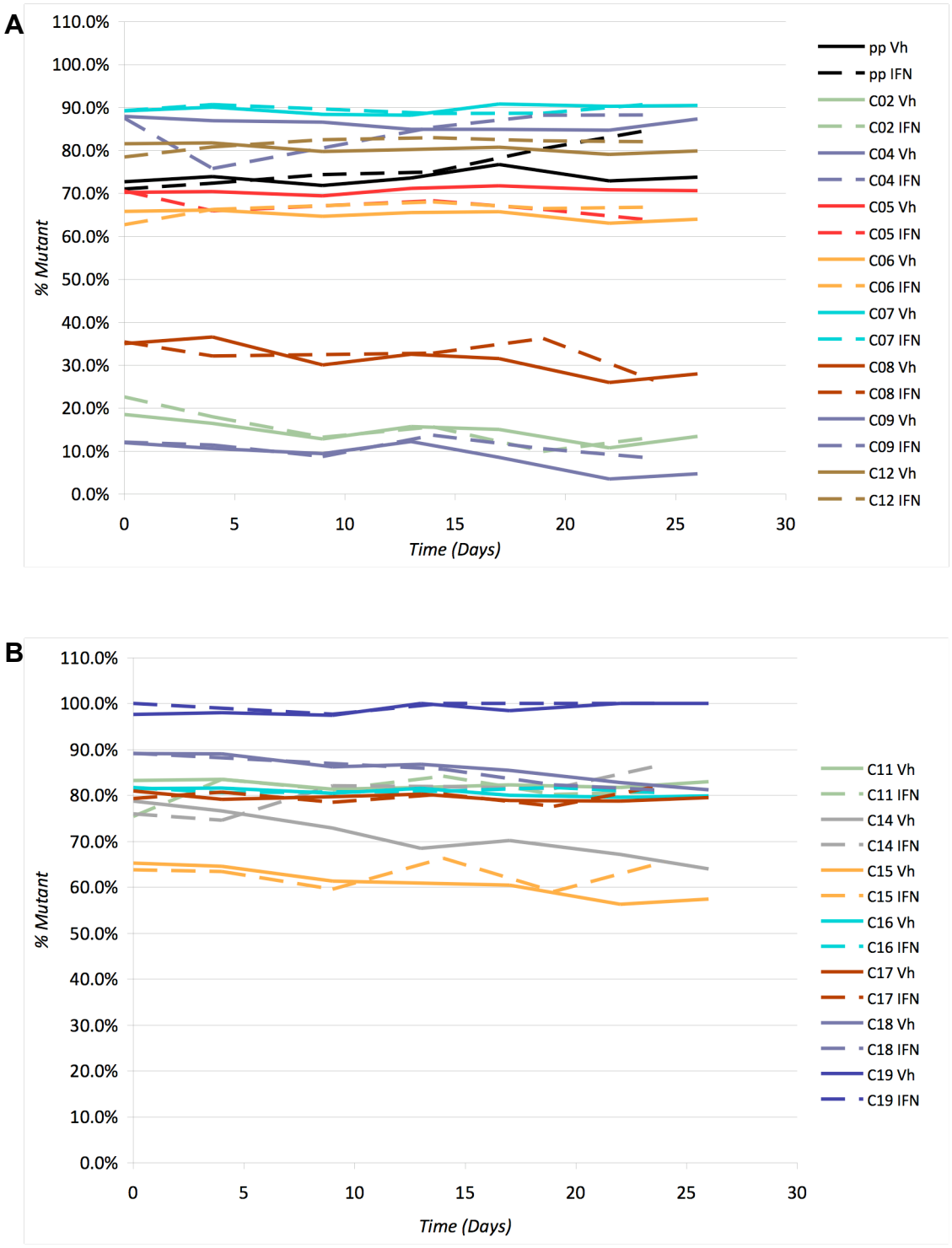


Figure 3.13.1. IFNα/γ treatment does not influence the proportion of mutant mtDNA. (**A/B**) A549 cybrid sub-clones and a pooled population were subjected to cycles of 6 hours IFNα/γ treatment followed by recovery in standard culture conditions for up to 26 days whilst duplicates were maintained in absence of treatment. DNA was harvested and the proportion of mutant mtDNA determined by pyrosequencing.

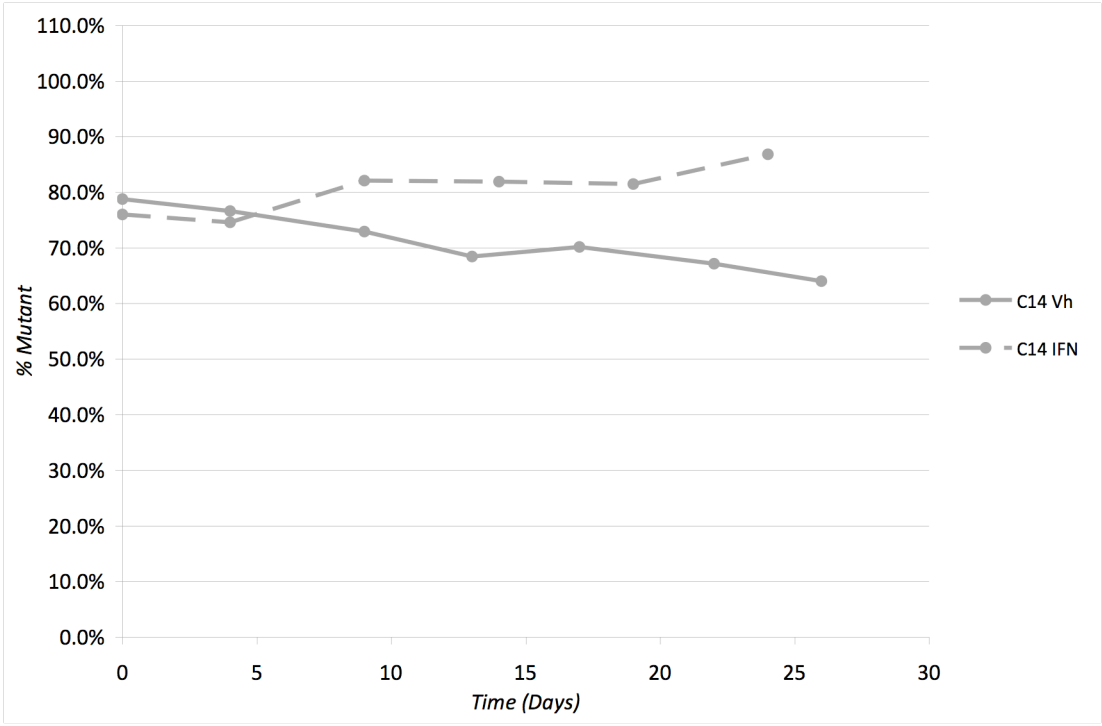
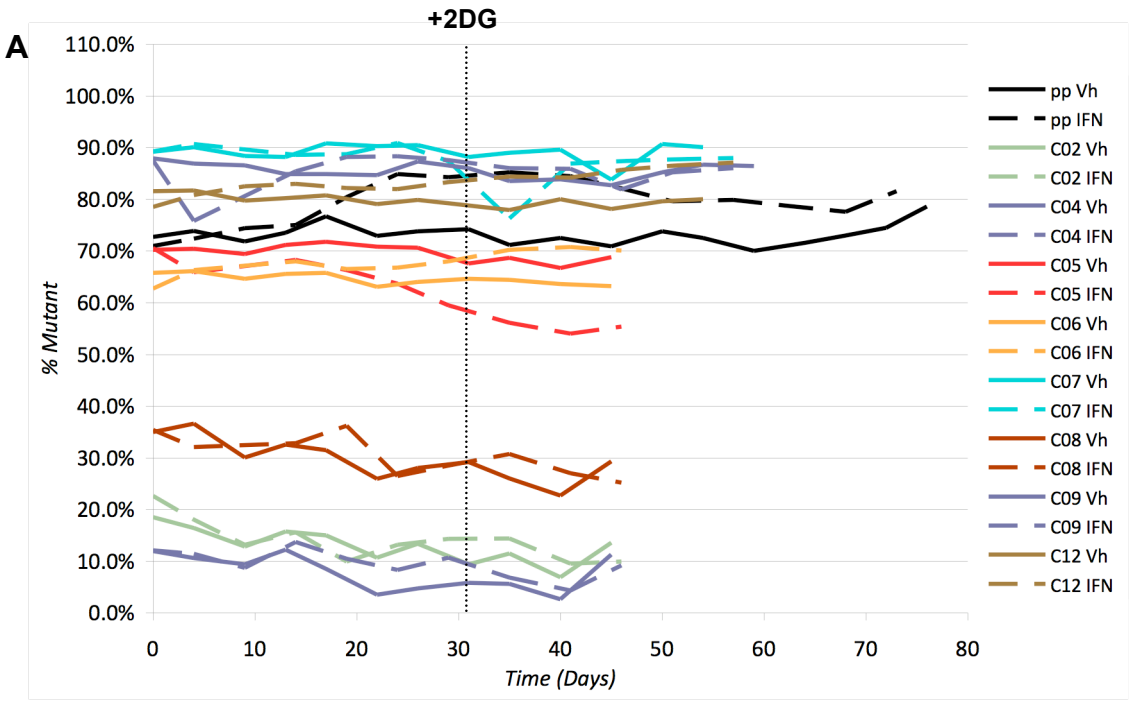


Figure 3.13.2. IFNα/γ treatment inhibits a spontaneous decrease in the proportion of mutant mtDNA. Of 16 sub-clones and a pooled population, a single sub-clone subjected to IFNα/γ treatment (C14 IFN) prevented a spontaneous decrease in the proportion of mutant mtDNA observed in the absence of treatment (C14 Vh). Heteroplasmy values were confirmed by triplicate measurement.

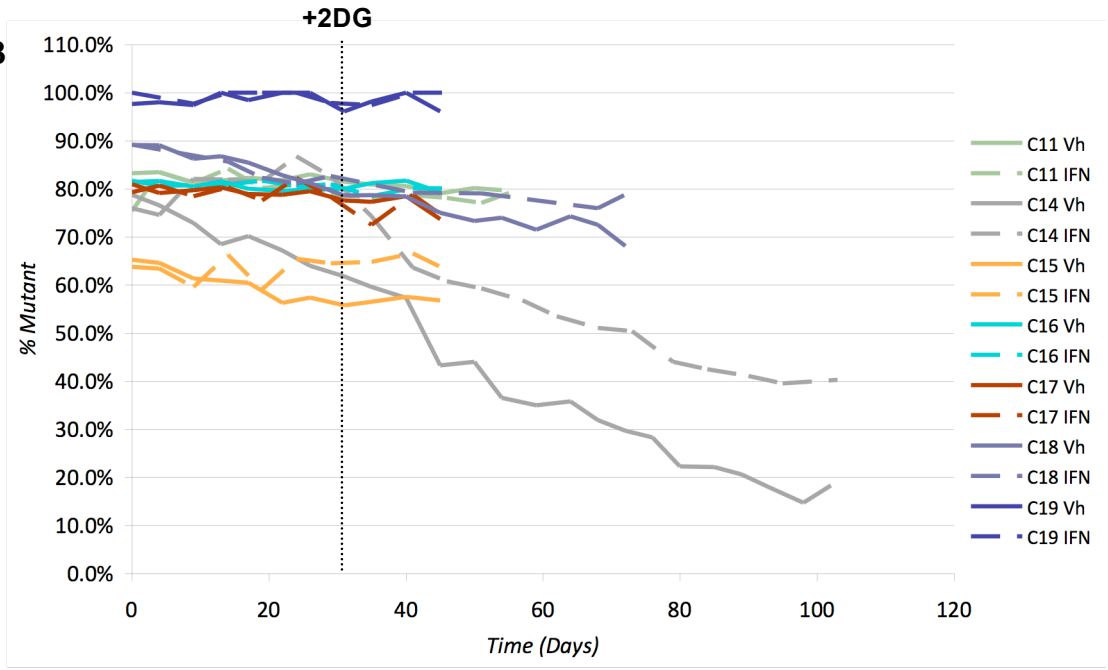
Therefore, the A549 cell line would be expected to have a greater propensity to segregate to WT mtDNA with IFNα/γ treatment, should this treatment be capable of stimulating biased segregation. Based on this hypothesis, A549 cybrid sub-clones and a pooled population were subjected to cycles of six hours IFNα/γ treatment followed by recovery in standard culture conditions. Surprisingly, IFNα/γ treatment of 15 sub-clones and a pooled population not only failed to stimulate segregation to WT mtDNA (Figure 3.13.1 A/B), but prevented the spontaneous segregation to WT mtDNA observed without treatment in a single sub-clone (Figure 3.13.2).

3.14 Intermittent combined 2DG/IFNα/γ treatment may permit spontaneous biased segregation in an A549 m.3243A>G cybrid sub-clone that previously maintained stable heteroplasmy with intermittent IFNα/γ treatment.

Stimulation of the interferon-response, as suggested by UBP analysis of 2DG-treated HT1080 cells, represents only one facet of the cellular response to 2DG. In isolation this may not be sufficient to stimulate a decrease in the proportion of mutant mtDNA. However, combining IFNα/γ with 2DG treatment may augment the ability of 2DG to stimulate segregation to WT mtDNA. To test this hypothesis, 15 sub-clones and a pooled population of A549 cybrids, previously subjected to cycles of six hours IFNα/γ treatment followed by recovery in standard culture conditions for 29 days, were transitioned to cycles of 24 hours combined 2DG/IFNα/γ treatment followed by recovery in standard culture conditions. Combined 2DG/IFNα/γ treatment failed to stimulate segregation to WT mtDNA (Figure 3.14.1 A/B) but may have permitted spontaneous segregation to WT mtDNA in a single sub-clone that was also observed in the absence of treatment (Figure 3.14.1 C).



B


+2DG
110.0%
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
% Mutant
0
20
40
60
80
100
120
Time (Days)
C11 Vh
C11 IFN
C14 Vh
C14 IFN
C15 Vh
C15 IFN
C16 Vh
C16 IFN
C17 Vh
C17 IFN
C18 Vh
C18 IFN
C19 Vh
C19 IFN

C

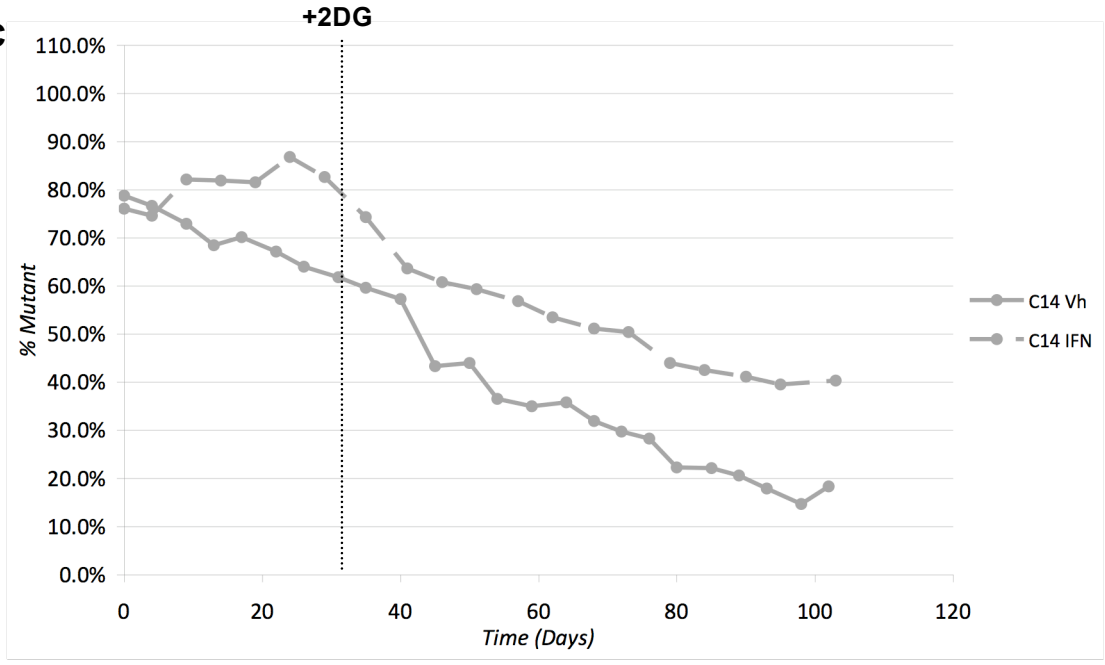
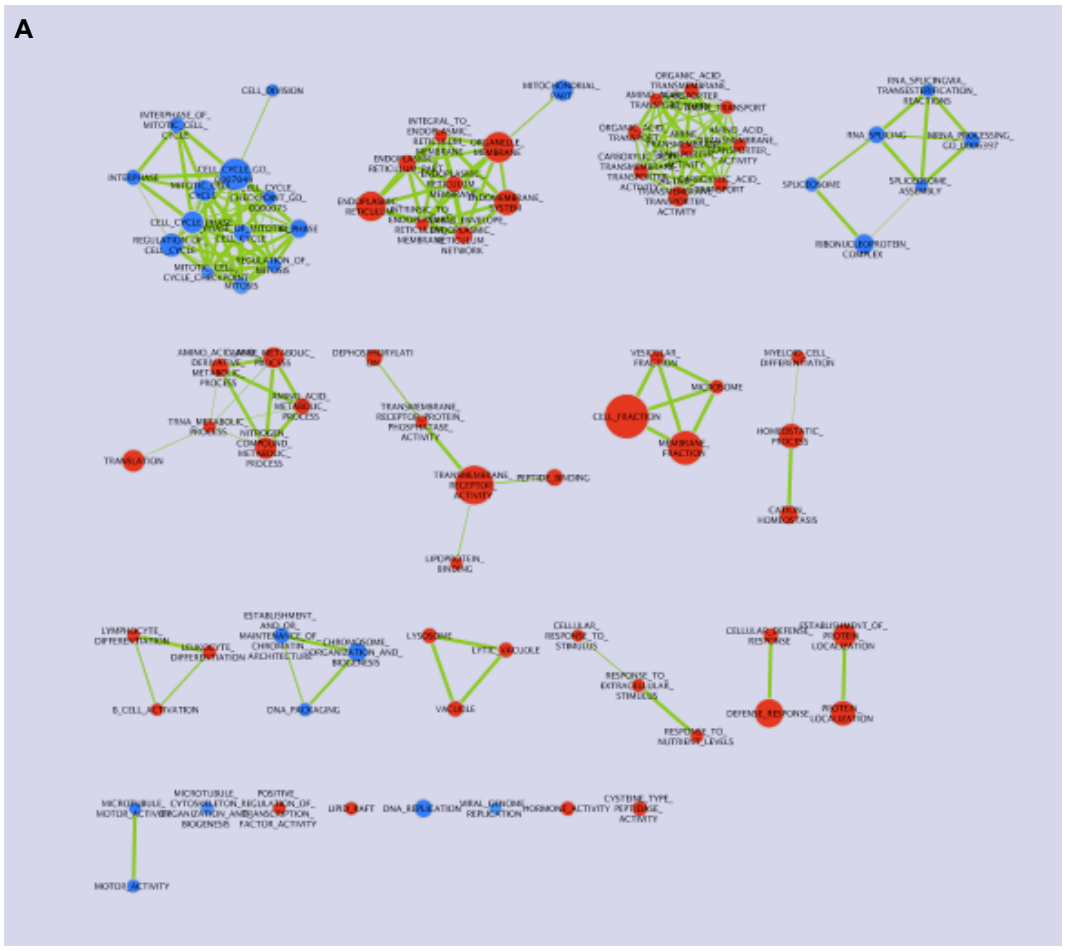

+2DG
110.0%
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
% Mutant
0
20
40
60
80
100
120
Time (Days)
C14 Vh
C14 IFN

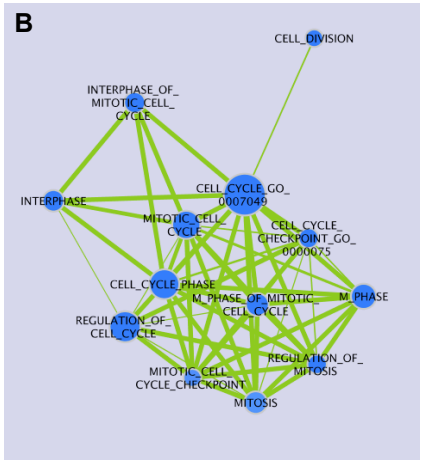
Figure 3.14.1. Combined 2DG/IFNα/γ treatment does not induce a decrease in the proportion of mutant mtDNA. (**A/B**) Following 29 days of cycles of 6 hours IFNα/γ treatment followed by recovery in standard culture conditions, A549 cybrid sub-clones and a pooled population were transitioned to cycles of 24 hours combined 2DG/IFNα/γ treatment followed by recovery in standard culture conditions for up to 74 days. Duplicates were maintained in the absence of treatment. DNA was harvested and the proportion of mutant mtDNA was determined by pyrosequencing. (**C**) Of 15 sub-clones and a pooled population, a single sub-clone exhibited segregation to WT mtDNA following the transition, which was also observed in the absence of treatment. Heteroplasmy values (•) were confirmed by triplicate measurement.

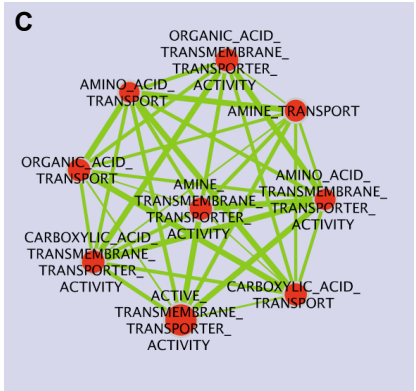
3.15 Enrichment map

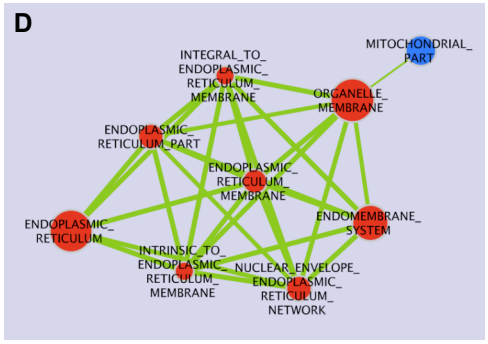
Enrichment map is a bioinformatic tool that organises genesets enriched from GSEA analysis into a network, where each geneset is a node and edges represent the overlap between genesets[110]. An automated network layout group's genesets with common genes into clusters, aiding the identification of functional 'themes'. To identify functional processes stimulated by 2DG treatment, which stimulates segregation to WT mtDNA in a sub-population of A549 cybrid sub-clones and pooled populations, GSEA of 2DG treated versus control HT1080 cells was performed, followed by analysis of enriched genesets with Enrichment Map. GO genesets formed multiple clusters (Figure 3.15.1 A). Major functional themes included the inhibition of cell cycle progression (Figure 3.15.1 B), amino acid transporter up-regulation (Figure 3.15.1 C), endoplasmic reticulum biogenesis (Figure 3.15.1 D) and inhibition of splicesome assembly (Figure 3.15 E). Clusters with themes of innate immunity (Figure 3.15.1 F) and lysosome biogenesis (Figure 3.15.1 G) are less prominent but potentially more significant to biased mtDNA segregation. Intriguingly, the geneset 'MITOCHONDRIAL _PART' formed part of a major cluster with the functional theme 'endoplasmic reticulum biogenesis', but showed marked down-regulation in contrast to other members of the cluster (Figure 3.15.1 D).

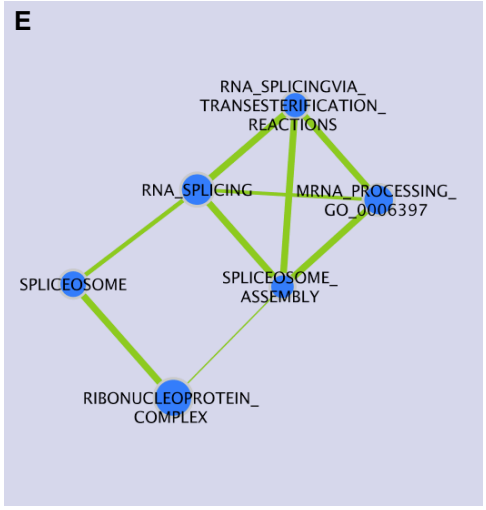
Curated genesets formed multiple clusters (Figure 3.15.1 H), some with themes identical or similar to those previously identified, including inhibition of the cell cycle (Figure 3.15.1 I), lysosome biogenesis (Figure 3.15.1 K) and amino acid or small molecule transporter up-regulation (Figure 3.15.1 M). Inhibition of transcription was a theme not previously identified (Figure 3.15.1 J), whilst inhibition of mitochondrial gene expression (Figure 3.15.1 L) confirmed a theme less evident with GO genesets. Cis-regulatory-motif-defined genesets formed a single major cluster composed of genesets regulated by the transcription factor E2F (Figure 3.15.1 N), whose down-regulation is consistent with inhibition of cell cycle progression. Computationally defined genesets formed clusters whose themes were not identified (Figure 3.15.1 O). A caveat of Enrichment Map is that the number of genesets in an enriched cluster is a combination of the genesets enriched with a particular condition and the redundancy of those genesets.

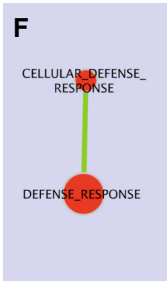

A

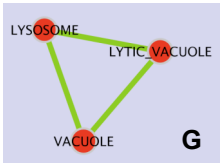

B
CELL_DIVISION
INTERPHASE_OF_MITOTIC_CELL_CYCLE
INTERPHASE
CELL_CYCLE_GO_0007049
MITOTIC_CELL_CYCLE
CELL_CYCLE_CHECKPOINT_GO_0000075
CELL_CYCLE_PHASE
M_PHASE_OF_MITOTIC_CELL_CYCLE
M_PHASE
REGULATION_OF_CELL_CYCLE
REGULATION_OF_MITOSIS
MITOTIC_CELL_CYCLE_CHECKPOINT
MITOSIS

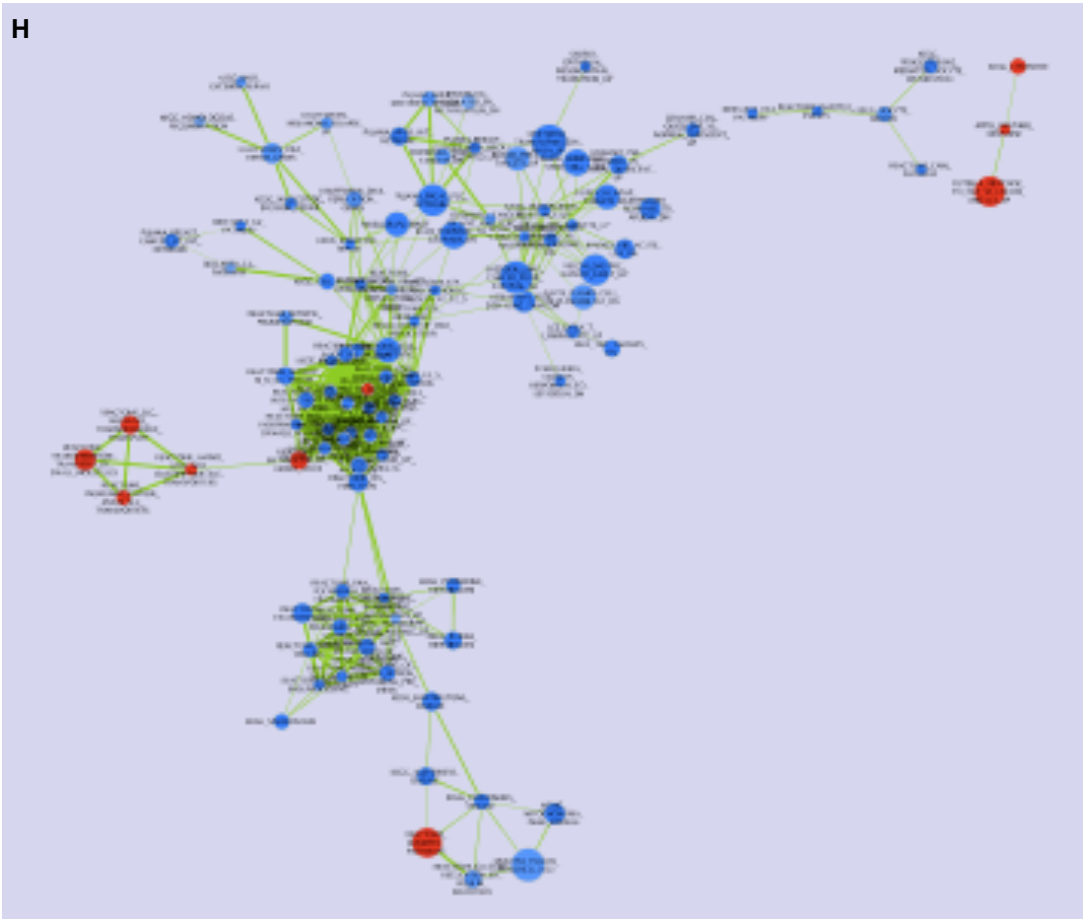

C
ORGANIC_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
AMINO_ACID_TRANSPORT
AMINE_TRANSPORT
ORGANIC_ACID_TRANSPORT
AMINE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
AMINO_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
CARBOXYLIC_ACID_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
ACTIVE_TRANSMEMBRANE_TRANSPORTER_ACTIVITY
CARBOXYLIC_ACID_TRANSPORT

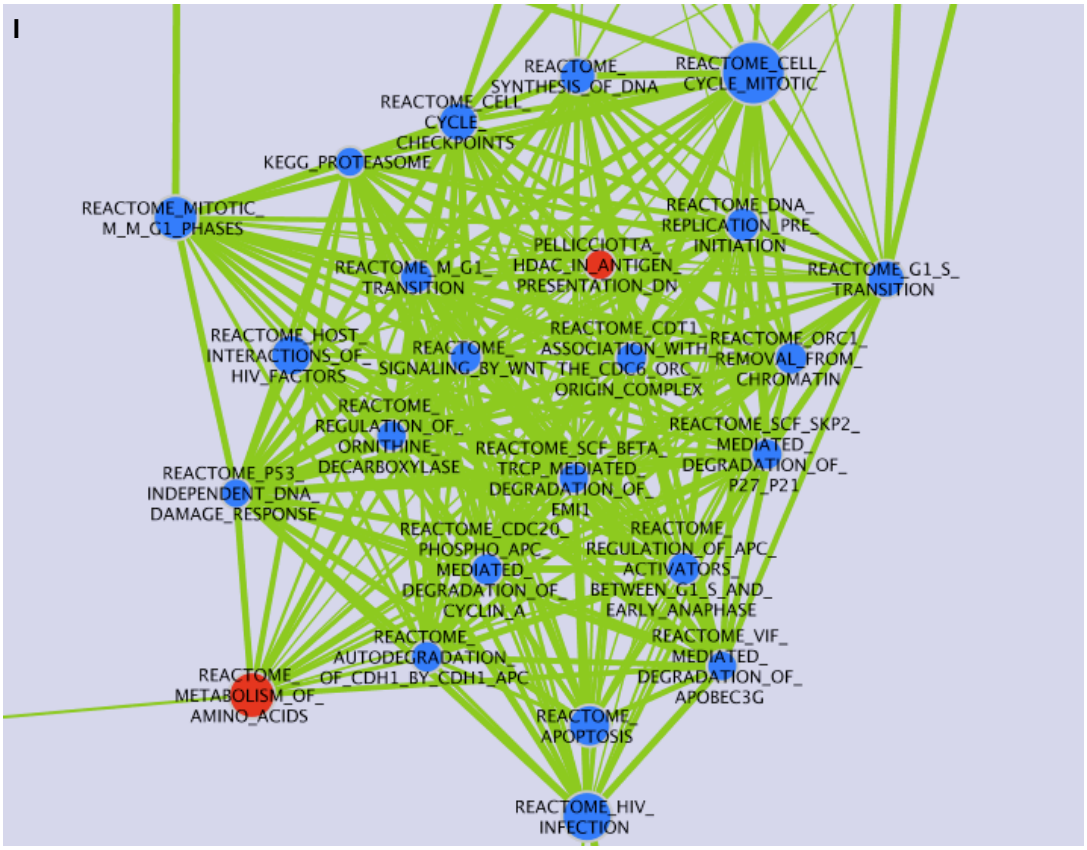

D
MITOCHONDRIAL_PART
INTEGRAL_TO_ENDOPLASMIC_RETICULUM_MEMBRANE
ORGANELLE_MEMBRANE
ENDOPLASMIC_RETICULUM_PART
ENDOPLASMIC_RETICULUM_MEMBRANE
ENDOPLASMIC_RETICULUM
ENDOMEMBRANE_SYSTEM
INTRINSIC_TO_ENDOPLASMIC_RETICULUM_MEMBRANE
NUCLEAR_ENVELOPE_ENDOPLASMIC_RETICULUM_NETWORK

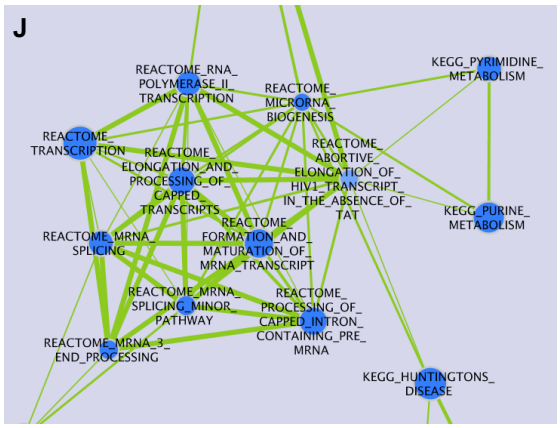

E
RNA_SPLICINGVIA_TRANSESTERIFICATION_REACTIONS
RNA_SPLICING
MRNA_PROCESSING_GO_0006397
SPLICEOSOME
SPLICEOSOME_ASSEMBLY
RIBONUCLEOPROTEIN_COMPLEX

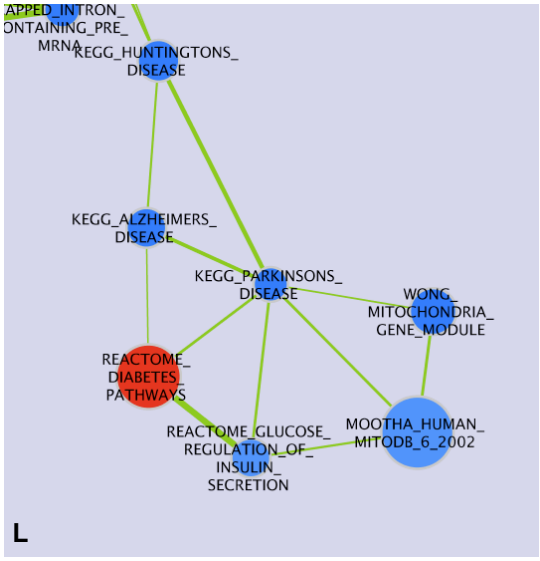

F
CELLULAR_DEFENSE_RESPONSE
DEFENSE_RESPONSE

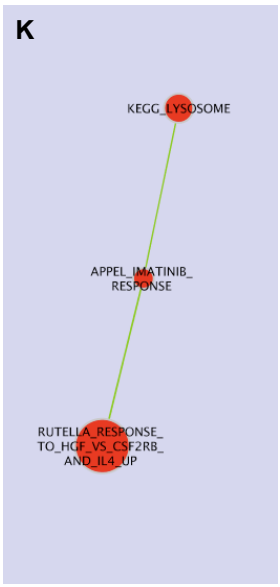

LYSOSOME
LYTIC_VACUOLE
VACUOLE
G

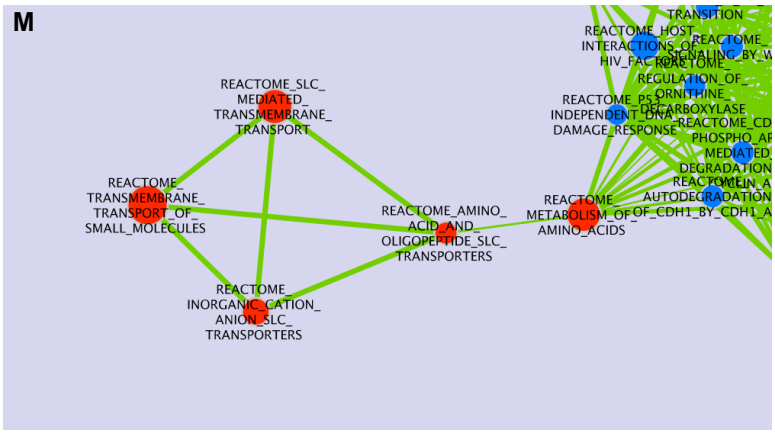

H

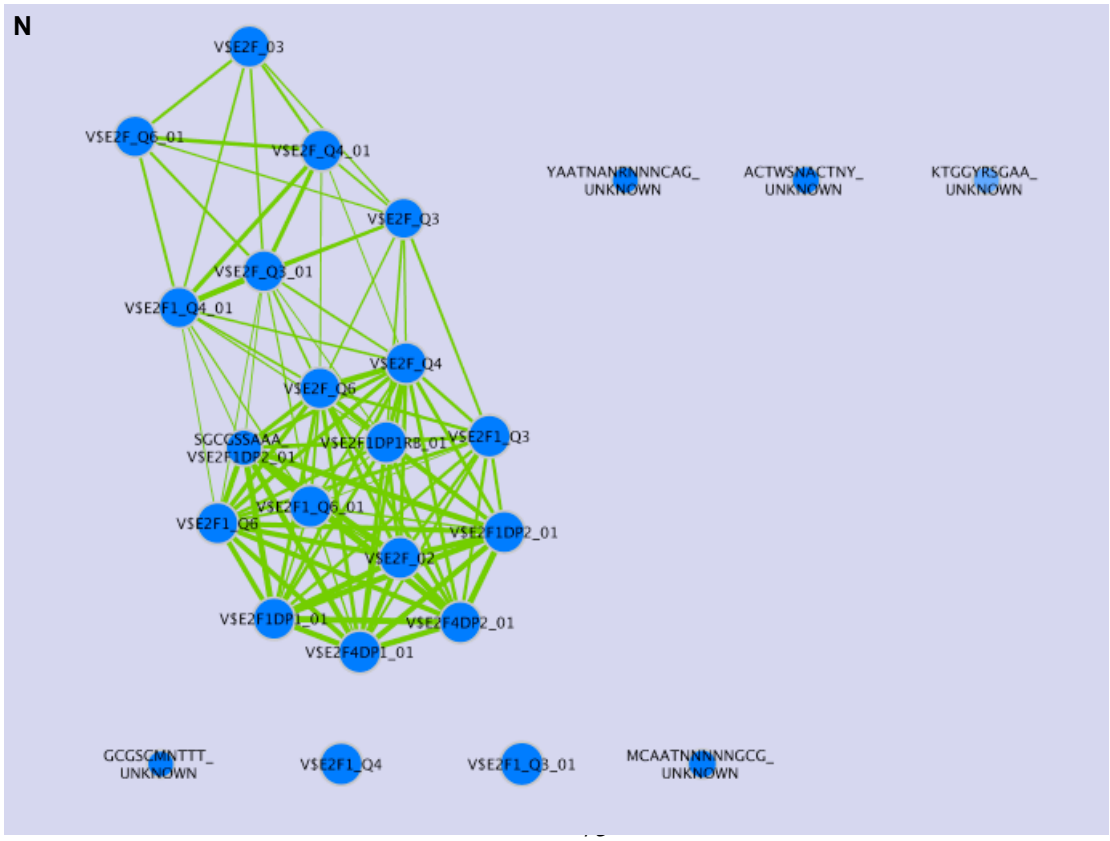

I
REACTOME_ SYNTHESIS_OF_DNA
REACTOME_CELL_ CYCLE_MITOTIC
REACTOME_CELL_ CYCLE_ CHECKPOINTS
KEGG_PROTEASOME
REACTOME_MITOTIC_ M_M_G1_PHASES
REACTOME_DNA_ REPLICATION_PRE_ INITIATION
PELLICCIOTTA_ HDAC_IN_ANTIGEN_ PRESENTATION_DN
REACTOME_M_G1_ TRANSITION
REACTOME_G1_S_ TRANSITION
REACTOME_HOST_ INTERACTIONS_OF_ HIV_FACTORS
REACTOME_ SIGNALING_BY_WNT
REACTOME_CDT1_ ASSOCIATION_WITH_ THE_CDC6_ORC_ ORIGIN_COMPLEX
REACTOME_ORC1_ REMOVAL_FROM_ CHROMATIN
REACTOME_ REGULATION_OF_ ORNITHINE_ DECARBOXYLASE
REACTOME_SCF_BETA_ TRCP_MEDIATED_ DEGRADATION_OF_ EMI1
REACTOME_SCF_SKP2_ MEDIATED_ DEGRADATION_OF_ P27_P21
REACTOME_P53_ INDEPENDENT_DNA_ DAMAGE_RESPONSE
REACTOME_CDC20_ PHOSPHO_APC_ MEDIATED_ DEGRADATION_OF_ CYCLIN_A
REACTOME_ REGULATION_OF_APC_ ACTIVATORS_ BETWEEN_G1_S_AND_ EARLY_ANAPHASE
REACTOME_ AUTODEGRADATION_ OF_CDH1_BY_CDH1_APC
REACTOME_VIF_ MEDIATED_ DEGRADATION_OF_ APOBEC3G
REACTOME_ METABOLISM_OF_ AMINO_ACIDS
REACTOME_ APOPTOSIS
REACTOME_HIV_ INFECTION


J
REACTOME_RNA_POLYMERASE_II_TRANSCRIPTION
REACTOME_MICRORNA_BIOGENESIS
KEGG_PYRIMIDINE_METABOLISM
REACTOME_TRANSCRIPTION
REACTOME_ELONGATION_AND_PROCESSING_OF_CAPPED_TRANSCRIPTS
REACTOME_ABORTIVE_ELONGATION_OF_HIV1_TRANSCRIPT_IN_THE_ABSENCE_OF_TAT
KEGG_PURINE_METABOLISM
REACTOME_MRNA_SPLICING
REACTOME_FORMATION_AND_MATURATION_OF_MRNA_TRANSCRIPT
REACTOME_MRNA_SPLICING_MINOR_PATHWAY
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA
REACTOME_MRNA_3_END_PROCESSING
KEGG_HUNTINGTONS_DISEASE


APPED_INTRON_
ONTAINING_PRE_
MRNA
KEGG_HUNTINGTONS_DISEASE
KEGG_ALZHEIMERS_DISEASE
KEGG_PARKINSONS_DISEASE
WONG_MITOCHONDRIA_GENE_MODULE
REACTOME_DIABETES_PATHWAYS
REACTOME_GLUCOSE_REGULATION_OF_INSULIN_SECRETION
MOOTHA_HUMAN_MITODB_6_2002
L


K
KEGG_LYSOSOME
APPEL_IMATINIB_RESPONSE
RUTELLA_RESPONSE_TO_HGF_VS_CSF2RB_AND_IL4_UP


M
REACTOME_SLC_MEDIATED_TRANSMEMBRANE_TRANSPORT
REACTOME_TRANSMEMBRANE_TRANSPORT_OF_SMALL_MOLECULES
REACTOME_AMINO_ACID_AND_OLIGOPEPTIDE_SLC_TRANSPORTERS
REACTOME_INORGANIC_CATION_ANION_SLC_TRANSPORTERS
REACTOME_METABOLISM_OF_AMINO_ACIDS
REACTOME_DECARBOXYLASE_
REACTOME_REGULATION_OF_
REACTOME_HOST_


N
VSE2F_Q3
VSE2F_Q6_01
VSE2F_Q4_01
YAATNANRNNNCAG_UNKNOWN
ACTWSNACTNY_UNKNOWN
KTGGYRSGAA_UNKNOWN
VSE2F_Q3
VSE2F_Q3_01
VSE2F1_Q4_01
VSE2F_Q4
VSE2F_Q6
SGCGSSAAA_VSE2F1DP2_01
VSE2F1DP1RB_01
VSE2F1_Q3
VSE2F1_Q6_01
VSE2F1_Q6
VSE2F1DP2_01
VSE2F_Q2
VSE2F1DP1_01
VSE2F4DP2_01
VSE2F4DP1_01
GCGSCMNTTT_UNKNOWN
VSE2F1_Q4
VSE2F1_Q3_01
MCAATNNNNNGCG_UNKNOWN

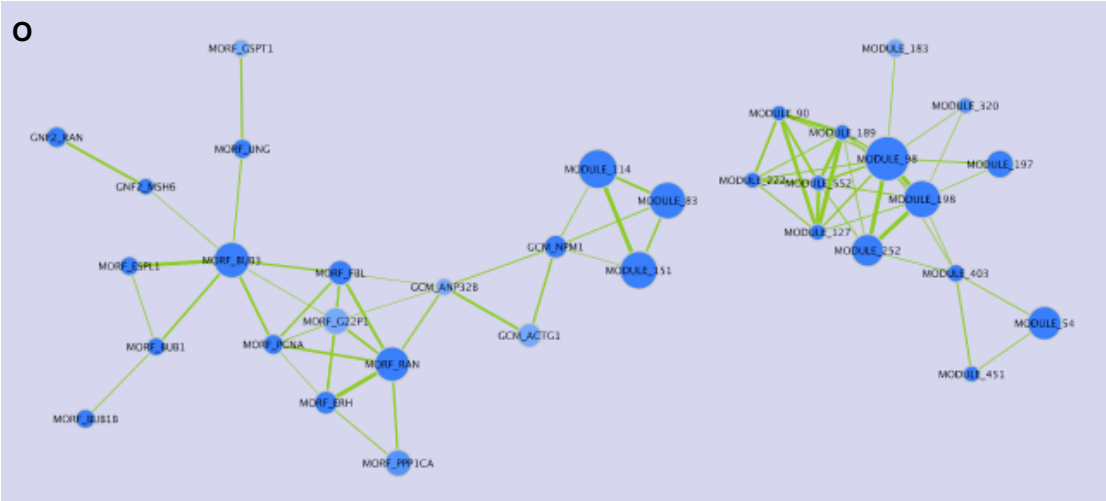


Figure 3.15.1. Enrichment map analysis of 2DG treated versus non-treated HT1080 cells. Nodes represent the genesets, the strength of edges indicate overlap between the genesets and the strength of the colour represents the amplitude of up-regulation (red) or down-regulation (blue) of geneset expression. An automated network layout groups functionally related genesets into clusters.

Therefore, prominence of an enriched cluster may be a partial artifact of redundancy amongst its constituent genesets, which can be recognised by strong edges between nodes in a cluster. An example of high geneset redundancy, with thick edges between nodes in a cluster, is enrichment for genesets showing inhibition of cell cycle progression (Figure 3.15.1 I). In contrast, an example of low geneset redundancy, with thin edges between nodes in a cluster, is enrichment for genesets showing inhibition of mitochondrial gene expression (Figure 3.15.1 L).

3.16 Connectivity map, ProfileChaser, UBP and leading-edge analysis of A549↑ versus A549↓ datasets.

In a previous study, 20% of A549 cybrid clones spontaneously segregated to WT mtDNA following cybridisation[77] and here I discovered that A549 microarray datasets derived from cells cultured in equivalent conditions can be divided into two sub-populations characterised by differential expression of PPARGC1A (section 3.4, Figure 3.4.1). Given the well-recognised role of PGC-1α in mitochondrial biogenesis,

it is tempting to speculate that one of the sub-populations, but not the other, represent a cellular state that permits segregation to WT mtDNA. GSEA analysis of A549↑ versus A549↓ datasets revealed a profile consistent with response to ER stress in A549↑ (section 3.5, Figure 3.5.1 A). However, efforts to recapitulate this profile by treatment of A549 cybrid sub-clones with the ER stressor 2DG only stimulated segregation to WT mtDNA in a fraction of sub-clones (3/29 or 10.3%). If we make the assumption that 2DG stimulates similar levels of ER stress in all sub-clones, then the response to ER stress provides only a partial explanation for segregation to WT mtDNA.

Connectivity map (CMAP) is a bioinformatic tool that identifies Food and Drug Administration (FDA) approved drugs that induce global gene expression differences similar to those observed between two experimental conditions[106,107]. CMAP was used to identify candidate drugs for neuroblastoma and hepatocellular carcinoma that were subsequently validated in cell-culture[135,136].

CMAP							
Rank	cmap name and cell line	mean	n	enrichment	p	specificity	% non-null
1	wortmannin - MCF7	0.499	10	0.741	0	0.061	80
2	LY-294002 - MCF7	0.458	34	0.605	0	0.0197	70
3	trichostatin A - MCF7	0.33	92	0.432	0	0.6019	66
4	trichostatin A - HL60	0.274	34	0.421	0	0.2638	64
5	LY-294002 - PC3	0.518	12	0.688	0.00002	0.035	83
6	lidocaine - PC3	-0.732	2	-0.989	0.00026	0	100
7	0173570-0000 - PC3	0.573	4	0.845	0.00086	0.0323	100
8	seneciphylline - MCF7	-0.681	2	-0.98	0.00091	0.0054	100
9	aciclovir - PC3	-0.669	2	-0.977	0.00111	0	100
10	ivermectin - PC3	0.673	2	0.97	0.00151	0.0111	100
11	thioridazine - MCF7	0.362	11	0.531	0.00202	0.2372	63
12	niclosamide - PC3	0.639	2	0.96	0.00266	0.0419	100
13	lynestrenol - PC3	0.714	2	0.96	0.00266	0.0062	100
14	sulfametoxydiazine - MC	-0.623	2	-0.965	0.00274	0	100
15	aztreonam - PC3	-0.656	2	-0.965	0.0028	0	100
16	triamterene - PC3	0.635	2	0.959	0.00292	0	100
17	fluphenazine - PC3	0.553	3	0.883	0.00322	0.0667	100
18	5707885 - MCF7	0.579	2	0.957	0.00324	0.0101	100
19	Prestwick-1100 - PC3	0.584	2	0.955	0.00366	0	100
20	disopyramide - MCF7	-0.725	2	-0.959	0.00368	0.0319	100
21	0175029-0000 - MCF7	0.599	2	0.949	0.00483	0.0739	100
22	dehydrocholic acid - PC3	-0.594	2	-0.952	0.00489	0	100
23	tobramycin - MCF7	-0.603	2	-0.949	0.00567	0.0199	100
24	hydroflumethiazide - MCF	0.57	2	0.94	0.00704	0.0155	100
25	luteolin - MCF7	0.586	2	0.938	0.00744	0.1181	100
26	gossypol - PC3	0.627	2	0.938	0.00748	0.046	100
27	resveratrol - PC3	0.625	2	0.936	0.00785	0.0957	100
28	thiamazole - MCF7	-0.275	3	-0.84	0.00817	0.0134	66
29	edrophonium chloride - M	-0.654	2	-0.935	0.00885	0.0057	100
30	PF-00562151-00 - PC3	0.366	4	0.737	0.00937	0.0069	75

Table 3.16.1. CMAP analysis of A549↑ versus A549↓ datasets. The output of CMAP analysis is ranked by connectivity score. Blue rows indicate a common annotation.

ProfileChaser				
GEO	Title	Subset 1 vs 2	Score	q-value
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	4 h vs 72 h	0.581675	0.0059
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	0.5 h vs 72 h	0.579398	0.0059
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	1 h vs 72 h	0.577963	0.006
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	8 h vs 72 h	0.568768	0.0067
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	8 h vs 24 h	0.534525	0.0099
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	control vs 72 h	0.529879	0.0103
GDS709	Enterocyte differentiation time course	8 d vs 2 d	0.528616	0.0104
GDS709	Enterocyte differentiation time course	post-proliferati	0.528616	0.0104
GDS709	Enterocyte differentiation time course	15 d vs 2 d	0.517471	0.011
GDS709	Enterocyte differentiation time course	differentiated v	0.517471	0.011
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	4 h vs 24 h	0.507409	0.0125
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	0.5 h vs 24 h	0.490815	0.0146
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	1 h vs 24 h	0.489177	0.0149
GDS3706	Benzo[a]pyrene diol epoxide effect on lung WI-	1 uM vs control	0.48764	0.0152
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	8 h vs 16 h	0.485761	0.0153
GDS2852	Interleukin 13 effect on bronchial cell line: time	24 h vs 0 h	0.468271	0.0172
GDS3029	Small-cell lung carcinoma cell lines of varying se	Bcl-2 copy ge 3	0.466865	0.0173
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	16 h vs 72 h	0.462083	0.0183
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	4 h vs 16 h	0.460106	0.0186
GDS3706	Benzo[a]pyrene diol epoxide effect on lung WI-	1 uM vs 0.1 uM	0.455366	0.0194
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	1 h vs 16 h	0.447183	0.0211
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	0.5 h vs 16 h	0.446364	0.0214
GDS2604	Asbestos effect on epithelial and mesothelial lu	epithelial vs m	0.444779	0.0215
GDS2852	Interleukin 13 effect on bronchial cell line: time	24 h vs 4 h	0.427684	0.0249
GDS2445	Polycomb group protein depletion effect on emb	EED depletion v	0.42663	0.0251
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	control vs 24 h	0.425748	0.0252
GDS2604	Asbestos effect on epithelial and mesothelial lu	A549 vs Met5A	0.424899	0.0253
GDS2445	Polycomb group protein depletion effect on emb	EED depletion v	0.422484	0.0258
GDS2341	Type I and Type II interferons effect on lung ep	24 h vs 6 h	0.421798	0.0258
GDS3710	TGF-beta-induced epithelial-mesenchymal trans	16 h vs 24 h	0.419098	0.0266

Table 3.16.2. ProfileChaser analysis of A549↑ versus A549↓ datasets. The output of ProfileChaser analysis is ranked by confidence score. Blue rows indicate the common annotation of 'TGF-β signalling' whilst the green row indicates the unique annotation of 'Interferon signalling'.

Moreover, the annotation of an enriched drug may provide insight into the functional changes accompanying mRNA expression differences, allowing deduction of functional and post-transcriptional changes from transcriptional information. CMAP requires a list of those genes most differentially expressed between two conditions as input, with a 'signature' of between ten and five hundred genes found to perform well[106,107]. Using a signature of the 100 most differentially expressed genes from A549↑ versus A549↓ datasets as input revealed enrichment for autophagy inhibitors LY-294002 and wortmannin (Table 3.16.1). This suggests that A549↑ represents a state of reduced autophagic flux.

ProfileChaser is a bioinformatic tool that allows the identification of any genetic, pharmaceutical or environmental perturbation submitted to the NCBI GEO database that induces global gene expression differences similar to those identified between two experimental conditions[108].

ProfileChaser is more powerful than CMAP, since its output is not restricted to FDA approved drugs. However, successful application of the tool has yet to be reported in the literature. Analysis of A549↑ versus A549↓ datasets revealed enrichment for early versus late TGF-β response (Table 3.16.2), or more broadly, reduced TGF-β signalling. Intriguingly, there was also weak enrichment for a 24-hour versus six-hour interferon-response.

A major limitation of CMAP and ProfileChaser is that they try to match global expression changes between two conditions to single interventions, which is analogous to representing a complex image with a single pixel. This restricts insight into underlying biological events, and provides a poor platform to design targeted interventions. To characterise biological changes at higher resolution, enriched genesets from GSEA of A549↑ versus A549↓ datasets were mapped to 115 UBPs (Figure 3.16.1). This revealed up-regulation in A549↑ for UBPs associated with peroxisome biogenesis (UBP 102), lipid metabolism (UBP 51), aerobic respiration and the mitochondrial respiratory chain (UBPs 86 and 103). Furthermore, UBPs consistent with hypoxia were up-regulated (UBPs 11 and 12).

To determine if PPARGC1A expression is responsible for these biological changes, leading-edge analysis of A549↑ versus A549↓ datasets was performed. Leading-edge analysis identifies genes that occur with the highest frequency in enriched genesets, which may represent the 'core' of a geneset that drives the enrichment signal[95]. For enriched genesets upregulated in A549↑ versus A549↓ datasets, PPARGC1A was the first and third highest scoring leading-edge gene using the C3 (motif) and C5 (GO) geneset collections respectively (Table 3.16.3). Intriguingly, ribosomal protein subunit encoding genes were the highest scorers in leading-edge analysis of C2 (curated gene sets) and C4 (computational gene sets), consistent with the finding that ribosome content is proportional to mitochondrial mass (Francisco J. Iborra, personnel communication). Moreover, only two genes are more differentially expressed than PPARGC1A (data not shown). The first is CEACAM6, which encodes a glycoprotein involved in cell adhesion[137] and the second is COX7C, which encodes a subunit of Complex IV.

UBP	Geneset	NES
102	KEGG_PEROXISOME	1.260615
102	KEGG_PYRUVATE_METABOLISM	
102	KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	1.399005
102	COENZYME_METABOLIC_PROCESS	
102	COFACTOR_METABOLIC_PROCESS	
102	MICROBODY	
102	PEROXISOME	
9	CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL_UP	1.323971
9	CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_UP	1.3655555
9	COLDREN_GEFITINIB_RESISTANCE_DN	1.3230242
9	ONDER_CDH1_TARGETS_2_DN	1.3730222
9	module_180	1.2608261
9	module_342	1.3434447
51	ACEVEDO_LIVER_CANCER_DN	1.0941697
51	CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_DN	1.1095341
51	HOSHIDA_LIVER_CANCER_SUBCLASS_S3	1.3381625
51	HSIAO_LIVER_SPECIFIC_GENES	1.4264753
51	KEGG_PROPANOATE_METABOLISM	
51	REACTOME_METABOLISM_OF_LIPIDS_AND_LIPOPROTEINS	1.1409743
51	RODWELL_AGING_KIDNEY_NO_BLOOD_DN	1.2670969
51	WOO_LIVER_CANCER_RECURRENCE_DN	1.3693125
51	CAR_HPX	1.1001761
51	CAR_IGFBP1	
51	module_40	1.3634125
51	LIPID_METABOLIC_PROCESS	1.3310381
86	module_43	1.2033381
86	AEROBIC_RESPIRATION	
86	CELLULAR_RESPIRATION	
103	module_25	
103	MITOCHONDRIAL_RESPIRATORY_CHAIN	1.1483629
103	MITOCHONDRIAL_RESPIRATORY_CHAIN_COMPLEX_I	
103	NADH_DEHYDROGENASE_COMPLEX	
103	RESPIRATORY_CHAIN_COMPLEX_I	
61	GNF2_CEBPA	
61	GNF2_GSTM1	
61	GNF2_HPN	1.124367
61	GNF2_HPX	1.2358568
61	GNF2_LCAT	1.1490189
61	GNF2_TST	1.0962669
61	module_24	1.0265934
25	CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_5	1.1982253
25	GOZGIT_ESR1_TARGETS_DN	1.0722685
25	MASSARWEH_TAMOXIFEN_RESISTANCE_UP	0.8290771
87	KAAB_HEART_ATRIUM_VS_VENTRICLE_DN	1.0340654
87	module_212	1.0942689
87	CARBOXYLIC_ACID_METABOLIC_PROCESS	1.2164121
87	ELECTRON_CARRIER_ACTIVITY	1.079375
87	ORGANIC_ACID_METABOLIC_PROCESS	1.2320741
87	OXIDOREDUCTASE_ACTIVITY	1.1839894
76	TAKAO_RESPONSE_TO_UVB_RADIATION_UP	0.9549236
76	module_273	1.1787311
76	module_363	
81	module_23	1.2319226
81	module_55	1.0335251
81	module_60	0.5888658
81	module_88	1.0801307
30	KEGG_CITRATE_CYCLE_TCA_CYCLE	1.0264286
30	MOOTHA_TCA	
30	REACTOME_CITRIC_ACID_CYCLE	
30	REACTOME_PYRUVATE_METABOLISM_AND_TCA_CYCLE	
88	CATABOLIC_PROCESS	0.825727
88	CELLULAR_CATABOLIC_PROCESS	0.8976114
11	ELVIDGE_HIF1A_AND_HIF2A_TARGETS_DN	0.8284903
11	ELVIDGE_HIF1A_TARGETS_DN	0.9520851
11	ELVIDGE_HYPOXIA_BY_DMOG_UP	0.7502653
11	ELVIDGE_HYPOXIA_UP	0.5943185
11	MENSE_HYPOXIA_UP	0.7871208
67	module_100	0.7238065
67	module_11	0.7970342
67	module_137	0.7951529
67	module_66	0.7238065

UBP	Geneset	NES
12	ELVIDGE_HIF1A_AND_HIF2A_TARGETS_UP	-1.4461676
12	ELVIDGE_HIF1A_TARGETS_UP	-1.383734
12	ELVIDGE_HYPOXIA_BY_DMOG_DN	-1.2392923
29	REACTOME_CENTROSOME_MATURATION	-1.0453959
29	REACTOME_G2_M_TRANSITION	-1.0633756
29	REACTOME_LOSS_OF_NLP_FROM_MITOTIC_CENTROSOMES	-1.0462898
29	MICROTUBULE_CYTOSKELETON	-0.9636095
29	ORGANELLE_ORGANIZATION_AND_BIOGENESIS	-0.7126891
36	KEGG_NUCLEOTIDE_EXCISION_REPAIR	
36	KEGG_PYRIMIDINE_METABOLISM	-1.1484662
36	REACTOME_GLOBAL_GENOMIC_NER	
36	REACTOME_NUCLEOTIDE_EXCISION_REPAIR	
36	REACTOME_REMOVAL_OF_THE_FLAP_INTERMEDIATE	
36	REACTOME_TRANSCRIPTION_COUPLED_NER	
36	RESPONSE_TO_ENDOGENOUS_STIMULUS	-0.9201689
97	EXTRACELLULAR_REGION	-0.9087879
97	EXTRACELLULAR_REGION_PART	-1.0659416
16	BILD_HRAS_ONCOGENIC_SIGNATURE	-0.6116415
16	FULCHER_INFLAMMATORY_RESPONSE_LECTIN_VS_LPS_UP	-0.6551353
16	HINATA_NFKB_TARGETS_FIBROBLAST_UP	-1.1804017
16	HINATA_NFKB_TARGETS_KERATINOCYTE_UP	
16	KIM_WT1_TARGETS_UP	-0.9532169
16	SENESE_HDAC1_AND_HDAC2_TARGETS_UP	-0.8074681
16	TIAN_TNF_SIGNALING_VIA_NFKB	
111	BROWNE_HCMV_INFECTION_14HR_UP	-0.7897912
111	DAIRKEE_CANCER_PRONE_RESPONSE_BPA	
111	DOUGLAS_BMI1_TARGETS_UP	-1.0140567
111	KEGG_RNA_DEGRADATION	
111	LASTOWSKA_NEUROBLASTOMA_COPY_NUMBER_UP	-1.0983465
111	REACTOME_METABOLISM_OF_MRNA	
111	REACTOME_MRNA_3_END_PROCESSING	
111	REACTOME_TRANSCRIPTION	-0.6559594
111	VANHARANTA_UTERINE_FIBROID_WITH_7Q_DELETION_UP	-0.9137575
111	CACGTG_V_MYC_Q2	
111	V_MYCMAX_01	
111	MORF_TERF1	
111	DNA_BINDING	-0.6414542
111	DOUBLE_STRANDED_DNA_BINDING	
111	HELICASE_ACTIVITY	
111	NUCLEOLAR_PART	
111	NUCLEOPLASM_PART	-0.5363946
111	NUCLEOTIDYLTRANSFERASE_ACTIVITY	
111	RIBOSOME_BIOGENESIS_AND_ASSEMBLY	
111	RNA_SPLICINGVIA_TRANSESTERIFICATION_REACTIONS	
111	RRNA_PROCESSING	
96	BERTUCCI_MEDULLARY_VS_DUCTAL_BREAST_CANCER_DN	-0.9765186
96	LIU_PROSTATE_CANCER_DN	-0.6779419
96	WANG_SMARCE1_TARGETS_UP	-0.5032705
96	module_122	**-0.6971589**
96	EXTRACELLULAR_MATRIX	-1.0866461
96	EXTRACELLULAR_MATRIX_PART	
96	PROTEINACEOUS_EXTRACELLULAR_MATRIX	-1.0866461
110	RNA_BIOSYNTHETIC_PROCESS	-0.6754842
110	TRANSCRIPTION	-0.7150585
110	TRANSCRIPTION_DNA_DEPENDENT	-0.6754839
105	GEORGES_CELL_CYCLE_MIR192_TARGETS	-1.082759
105	SLEBOS_HEAD_AND_NECK_CANCER_WITH_HPV_UP	
105	CHROMOSOME_SEGREGATION	
105	CHROMOSOMEPERICENTRIC_REGION	
105	MITOTIC_SISTER_CHROMATID_SEGREGATION	
105	SISTER_CHROMATID_SEGREGATION	
109	BENPORATH_OCT4_TARGETS	-0.6821848
109	GTGCCTT_MIR_506	
109	REGULATION_OF_CELLULAR_METABOLIC_PROCESS	-0.6472895
109	REGULATION_OF_METABOLIC_PROCESS	-0.6670688

Figure 3.16.1. UBPs enriched in A549↑ versus A549↓ datasets. UBPs correlated with PPARGC1A expression in A549↑ versus A549↓ datasets and IFNα treated primary human hepatocytes are highlighted blue. The strength of colour represents the amplitude of geneset up-regulation (red) and down-regulation (blue) according to NES.

GENESET COLLECTION							
C2 (Curated)				C3 (Motif)			
NAME	NES	LEADING-EDGE	FREQ	NAME	NES	LEADING EDGE	FREQ
VECCHI_GASTRI	1.577	RPS17	6	V$AR_Q6	1.31	PPARGC1A	9
DOANE_BREAST_	1.476	RPS15	6	V$COUP_01	1.2	CACNB2	5
HSIAO_LIVER_SI	1.426	RPS13	6	V$HNF4_01_B	1.15	ZHX2	4
STEIN_ESRRA_TA	1.416	RPS27	6	V$HNF4_01	1.14	HOXA5	4
KEGG_VALINE_LE	1.399	RPS29	6	V$ERR1_Q2	1.12	ESRRA	4
TARTE_PLASMA_	1.391	RPS20	6	TGACCTTG_V	1.12	PDZK1	4
REACTOME_TRA	1.384	RPS21	6	V$TATA_01	1.12	SEMA3B	4
REACTOME_VIRA	1.383	RPS23	6	TAAWWATAG_	1.11	NR2F2	4
SENGUPTA_NASC	1.374	RPS24	6	GGCNRNWCT	1.08	NR2F1	4
ONDER_CDH1_TA	1.373	RPS9	6	RYAAAKNNNN	1.07	PCDH9	4
WOO_LIVER_CA	1.369	RPS6	6	V$HNF1_Q6	1.06	USH1C	4
CHARAFE_BREAS	1.366	RPS4X	6	V$STAT5A_02	1.05	BCL11B	4
IGARASHI_ATF4_	1.356	RPS15A	6	GAANYNYGAC	1.05	SMOC1	3
REACTOME_PEPT	1.353	RPL17	5	V$HNF4_Q6	1.02	RARB	3
KEGG_RIBOSOM	1.352	RPL14	5	RRAGTTGT_U	1.01	PELO	3
REACTOME_REG	1.352	RPL13	5	V$HNF4_DR1	1.01	ERBB3	3
REACTOME_FOR	1.35	RPL15	5	V$XBP1_01 (	1	CHCHD7	3
WAMUNYOKOLI_	1.346	RPL11	5	V$HNF3B_01	0.99	NR2C2	3
HOSHIDA_LIVER	1.338	RPL12	5	V$AR_Q2	0.97	NDRG2	3
STEIN_ESRRA_TA	1.33	RPS14	5	V$RORA1_01	0.97	LGALS4	3
CHARAFE_BREAS	1.324	CEACAM6	5	V$HNF1_01	0.96	RNF5	3

GENESET COLLECTION							
C4 (Computational)				C5 (Gene ontology)			
NAME	NES	LEADING EDGE	FREQ	NAME	NES	LEADING EDGE	FREQ
GCM_TPT1	1.377	RPS24	8	NITROGEN_C	1.4	BDH2	9
MODULE_40	1.363	COX7C	8	LIPID_METAB	1.33	ALDH6A1	8
MODULE_342	1.343	RPS23	7	OXIDOREDUC	1.27	PPARGC1A	8
MODULE_426	1.331	RPS6	7	LIPID_BINDI	1.26	HAO1	8
MORF_TPT1	1.281	RPL32	7	FATTY_ACID_	1.24	GATM	7
MODULE_180	1.261	RPL31	7	ORGANIC_AC	1.23	AKR1C2	7
GNF2_EIF3S6	1.252	RPL9	7	MITOCHOND	1.22	NDUFS4	7
MODULE_139	1.248	RPL10A	7	CARBOXYLIC	1.22	ACADM	7
GNF2_TPT1	1.248	RPL13A	7	ENERGY_DER	1.21	UQCRH	7
MORF_EIF3S6	1.246	RPL17	6	AMINE_METAB	1.21	CPT1A	7
MODULE_152	1.239	RPL14	6	CELLULAR_LI	1.2	SDHD	7
MODULE_243	1.237	RPL11	6	STRUCTURAL	1.2	OXA1L	6
GNF2_HPX	1.236	RPS17	6	AMINO_ACID	1.19	SLC7A7	6
MODULE_23	1.232	RPS14	6	MITOCHOND	1.19	GOT1	6
MODULE_83	1.226	RPS9	6	AMINO_ACID	1.19	DDAH1	6
MORF_NPM1	1.213	RPL28	6	OXIDOREDUC	1.18	ASL	6
MODULE_62	1.213	RPL29	6	ORGANELLE_I	1.18	SLC7A2	6
MODULE_42	1.209	TPT1	6	MITOCHOND	1.17	PPARA	6
MODULE_43	1.203	EEF2	6	MONOCARBO	1.17	UQCRC1	6
GCM_NPM1	1.193	RPL3	6	GENERATION_	1.16	NDUFS1	6
GNF2_ST13	1.189	RPL4	6	MITOCHOND	1.15	UQCRB	6

Table 3.16.3. PPARGC1A is enriched in leading-edge analysis of A549↑ versus A549↓ datasets. Leading-edge genes (LEADING EDGE) are ranked by frequency of occurrence (FREQ) in enriched up-regulated genesets from C2 (curated), C3 (motif), C4 (computational) and C5 (GO) geneset collections. PPARGC1A is highlighted blue and ribosomal proteins in green.

GENESET COLLECTION							
C2 (Curated)				C3 (Motif)			
NAME	NES	LEADING-EDGE	FREQ	NAME	NES	LEADING EDGE	FREQ
KEGG_DRUG_ME	-1.89	CYP3A4	8	V$PPARG_01	-1.6	EIFC2	6
REACTOME_PHA	-1.79	CYP3A7	7	TCCCCAC,MI	-1.6	ETV1	6
REACTOME_GLU	-1.77	CYP3A43	7	TACGGGT,MI	-1.4	HOXB5	6
REACTOME_BIO	-1.77	UGT2B15	7	V$STAT1_01	-1.4	CHD4	4
KEGG_METABOLI	-1.74	ADH4	6	ATCMNTCCG	-1.4	PPARGC1A	4
KEGG_FATTY_AC	-1.73	CYP2E1	6	CAATGCA,MI	-1.4	HOXD10	3
REACTOME_PHA	-1.72	CYP2B6	6	GGCAGAC,MI	-1.4	GRIN2B	3
THUM_MIR21_TA	-1.7	CYP2C8	6	GTGGTGA,MI	-1.4	RCOR2	3
ST_WNT_CA2_C	-1.68	ADH1C	6	V$HOX13_01	-1.3	ESRRA	3
ICHIBA_GRAFT_	-1.67	ADH1A	6	YAATNANRNN	-1.3	ASXL1	3
REACTOME_CYT	-1.66	UGT2B17	6	GGCNNMSMY	-1.3	GAS7	3
KEGG_RETINOL_	-1.65	CYP2A6	6	RGTTAMWNA	-1.3	KIAA1407	3
IIZUKA_LIVER_C	-1.65	ADH7	6	V$MAX_01	-1.3	CLN3	3
BIOCARTA_CHR	-1.64	CYP4A11	6	RRAGTTGT_U	-1.3	ARHGEF2	3
HSIAO_LIVER_S	-1.64	MGST2	6	V$AP4_Q6	-1.2	PCDHA10	3
KEGG_STEROID_	-1.64	CYP2D6	5	V$AMEF2_Q6	-1.2	SLC1A7	3
REACTOME_SYN	-1.63	CYP2A13	5	V$AP4_Q6_0	-1.2	ATP6V1A	3
WILLIAMS_ESR2	-1.63	GSTM1	5	V$MYCMAX_0	-1.2	CNNM1	3
KORKOLA_YOLK	-1.63	GSTM4	5	V$PPARA_01	-1.2	SYT3	3
KEGG_VALINE_LE	-1.62	GSTO2	5	V$ARNT_01	-1.2	DMD	3
TANG_SENESCEN	-1.62	GSTA3	5	V$CACBINDI	-1.2	BCL11B	3

C4 (Computational)				C5 (Gene ontology)			
NAME	NES	LEADING EDGE	FREQ	NAME	NES	LEADING EDGE	FREQ
MODULE_218	-1.78	EHHADH	6	FATTY_ACID_	-1.7	HAO2	6
MODULE_294	-1.76	PTGIS	5	ENERGY_RES	-1.6	PPARGC1A	4
MODULE_221	-1.74	ACADSB	5	MONOCARBO	-1.6	CROT	4
MODULE_368	-1.72	ACADM	5	LIPID_HOME	-1.6	MLYCD	4
MODULE_184	-1.72	MGA	4	HYDROGEN_I	-1.6	PPARA	4
MODULE_286	-1.72	PLEKHA1	4	LIPID_CATAB	-1.6	ALDH5A1	4
MODULE_343	-1.68	EVI2B	4	ACTIN_FILAM	-1.6	ABCD2	4
MODULE_305	-1.67	ZNF266	4	LIPOPROTEIN	-1.6	AGXT	3
MODULE_156	-1.67	SLC10A1	4	HORMONE_M	-1.6	APOA4	3
MODULE_240	-1.66	SLC16A2	4	CARBOHYDRA	-1.5	ACOT12	3
MODULE_43	-1.63	AP2A1	4	MICROTUBUL	-1.5	AKR1D1	3
GNF2_PRDX2	-1.62	DPYD	4	MOTOR_ACT	-1.5	GCK	3
MODULE_135	-1.6	CYP3A4	3	THIOLESTER_	-1.5	EHHADH	2
MODULE_313	-1.58	CYP3A7	3	AMINO_ACID	-1.5	IDE	2
MODULE_186	-1.57	SLC6A1	3	RESPONSE_T	-1.5	AGPS	2
MODULE_547	-1.56	SLC6A4	3	ENDOSOME_T	-1.5	ASMTL	2
CAR_IGFBP1	-1.53	IVD	3	FATTY_ACID_	-1.5	MYO6	2
GNF2_PCAF	-1.51	CYP2C8	3	MICROBODY	-1.5	KIF5C	2
MODULE_162	-1.5	CPS1	3	PEROXISOME	-1.5	PECI	2
MORF_RAB6A	-1.49	SRD5A2	3	OXYGEN_BIN	-1.5	BBS4	2
CAR_HPX	-1.48	ACO1	3	GLUTATHION	-1.5	ACADSB	2

Table 3.16.4. PPARGC1A is enriched in leading-edge analysis of IFN-α treated primary-human-hepatocyte datasets compared to untreated datasets. Leading-edge genes are ranked by frequency of occurrence in enriched down-regulated genesets from C2 (curated), C3 (motif), C4 (computational) and C5 (GO) geneset collections. PPARGC1A is highlighted blue.

Neither gene product is known to be associated with regulation of gene expression. In contrast, PPARGC1A encodes PGC-1α, which has been recognised as a regulator of mitochondrial biogenesis and function[120].

It is noteworthy that PPARGC1A is a leading-edge downregulated gene in response to IFNα treatment, based on analysis of enriched genesets within six hours of IFNα treatment of primary human hepatocytes (Table 3.16.4). Thus decreased PPARGC1A expression might explain why IFNα (in combination with IFNγ) inhibited the spontaneous segregation to WT mtDNA in an A549 cybrid sub-clone (section 3.13, Figure 3.13.2). Moreover, UBPs annotated with peroxisome biogenesis, lipid metabolism, aerobic respiration, mitochondrial respiratory chain and hypoxia, which are up-regulated in A549↑ versus A549↓ datasets, are down-regulated in IFNα treated primary human hepatocytes, suggesting their activity may be stimulated by PPARGC1A expression and could influence segregation to WT mtDNA (Figure 3.16.1). PPARGC1A is also a leading-edge gene from enriched genesets upregulated in A549 datasets versus a group of datasets from cell lines with a known segregation bias towards mutant mtDNA (data not shown). Together, these findings suggest that PPARGC1A expression, directly or through its stimulation of specific UBPs may be associated with a decrease in the proportion of mutant mtDNA.

Enrichment Map analysis of A549↑ versus A549↓ datasets was performed but did not reveal any results of apparent significance.

3.17 The hypersensitivity of PPARGC1A expression to amino acid starvation in the A549 cell line.

Although PPARGC1A expression is associated with the biological changes observed in A549↑ versus A549↓ datasets, the cause of its differential expression in equivalent culture conditions remains unknown. Measurement of PPARGC1A mRNA levels by quantitative PCR (qPCR) in multiple cell lines with amino acid starvation revealed that PPARGC1A expression is markedly up-regulated in A549 cells compared to 143B or HEK293T cells.

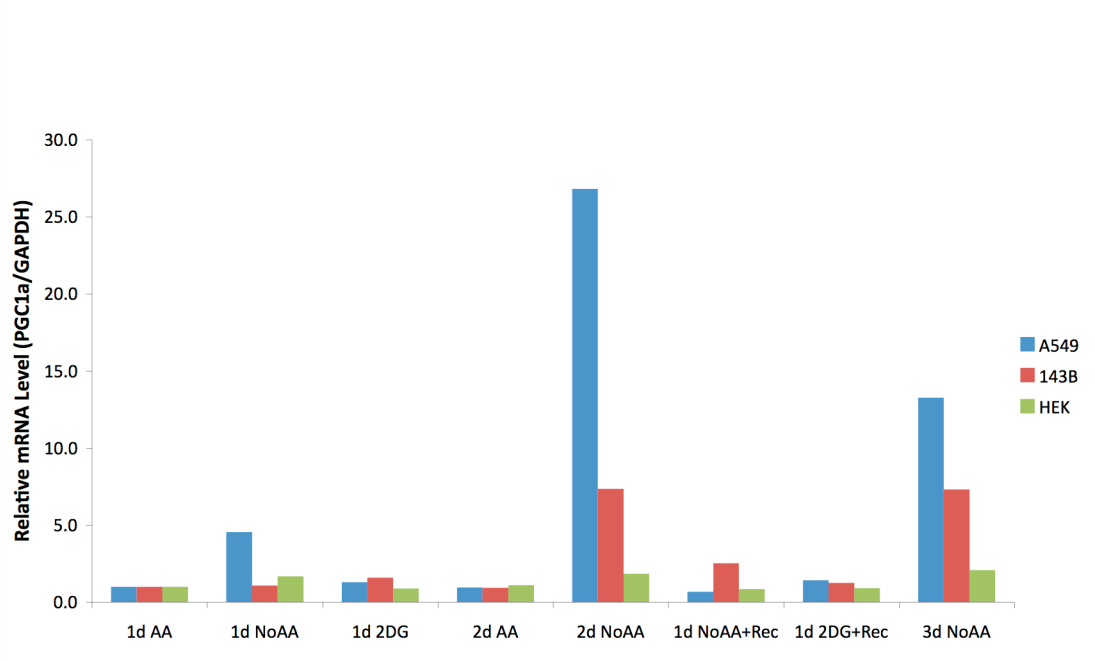


Figure 3.17.1. Amino acid starvation increases PPARGC1A expression in A549 and 143B cells. PPARGC1A mRNA expression levels were determined by qPCR in A549 (blue), 143B (red) and HEK293T (green) cells cultured in the absence of amino acids for one (1d NoAA), two (2d NoAA) or three (3d NoAA) days, in the presence of amino acids for one (1d AA) or two (2d AA) days or in the absence of amino acids for one day followed by one day recovery in the presence of amino acids (1d NoAA+Rec). n = 1 experiment. (Catherine Nezich).

After two days starvation, PPARGC1A was up-regulated 27-fold in A549 cells compared to 7.4 or 1.9-fold in 143B and HEK293T cells respectively (Catherine Nezich, Figure 3.17.1). Surprisingly, 24-hour treatment with 2DG had negligible effect on PPARGC1A expression, suggesting up-regulation is unrelated to stimulation of the ER stress response.

The hypersensitivity of PPARGC1A expression to amino acid availability suggests that its differential expression in equivalent culture conditions could be related to differences in amino acid homeostasis between pooled populations. Indeed, the 27-fold up-regulation of PPARGC1A with amino acid starvation is comparable to 11-fold up-regulation in A549↑ versus A549↓ datasets. However, this conclusion is based on results from a single experiment and therefore must be taken with caution until the same result is reproduced.

4 Discussion

It is 22 years since Hayashi's group reported the first instance of replicative advantage for pathological mutant mtDNA in human cells[75] and 18 since the selective advantage of mutant mtDNA was first attributed to the host nuclear DNA genotype[77]. In the subsequent two decades, replicative advantage has been linked to additional origins[78], or altered replication pausing[88], in the latter case implicating a specific protein (mTERF). However, only three proteins have been shown to alter the levels of mutant and WT mtDNA in human cells, Drp1, Fis1 and Parkin[82,83], which are all involved in mitochondrial dynamics and turnover. In mice, one gene, GIMAP3, has been identified that co-segregates with the tendency to select one of two mitochondrial genotypes in hematopoietic cells[84].

The mechanism is currently obscure and may not apply to pathological mtDNA variants. Hence, there remains an urgent need to understand the molecular basis of mtDNA segregation to aid the design of interventions that can restrict or reverse the selection of mutant mtDNA.

The aim of this study was to use bioinformatic tools to identify differences between cells that select mutant or WT mtDNA, and to demonstrate one or more of these differences was relevant to mtDNA segregation. Initially, the plan was to identify 'lead-genes' and express them in cells carrying mixtures of mutant and WT mtDNA to determine their effect on mtDNA heteroplasmy. The HEK293T m.3243A>G cybrid cell line was created for this purpose. It was deemed suitable for the study of candidate interventions and specifically candidate genes, because it allows the incorporation of any gene at a fixed locus, where its expression is under the control of a doxycycline-inducible promoter. However, there are two fundamental requirements for mtDNA segregation bias towards WT mtDNA, first that the gene expression profile is appropriate to segregation favouring WT over mutant mtDNA, second that the cell is able to segregate mtDNA at all, given that many cells in culture maintain a constant level of heteroplasmy for indefinite periods. The majority of HEK293T cybrids displayed stable heteroplasmy and analysis of 55 clones indicated that spontaneous mtDNA segregation occurred at very low frequency (~2%). These characteristics were suitable in the sense that there was plenty of scope for an

intervention to produce a highly significant effect. However, HEK293T cybrids clearly tend towards persistent heteroplasmy despite candidate interventions. This problem could be circumvented for chemical interventions by utilising A549 m.3243A>G cybrid cells, which had previously been shown to segregate spontaneously to WT mtDNA at a much higher frequency (20%)[77].

4.1 PPARGC1A is a candidate for influencing mtDNA segregation

Many cybrids maintain stable heteroplasmy indefinitely. Thus one or more key factors for mtDNA segregation cannot be expressed sufficiently (in an active form) in all clones. A549 cybrids have a greater propensity to segregate to WT mtDNA than other cell lines and A549 cells cultured in equivalent conditions can be divided into two sub-populations characterised by differential expression of PPARGC1A (section 3.4, Figure 3.4.1). This suggested to me that high levels of PPARGC1A might be needed to permit mtDNA segregation to occur. *A priori*, PPARGC1A was a credible candidate as it stimulates mitochondrial biogenesis[120], which in turn must impact on mitochondrial turnover, a process that has been implicated in biased mtDNA segregation[82]. Moreover, GSEA analysis identified PPARGC1A as a leading-edge gene distinguishing A549 cells from those that select mutant mtDNA (section 3.16, Figure 3.16.1). Towards the end of this study, PPARGC1A was also identified as a downregulated leading-edge gene by GSEA analysis of IFNα treated versus untreated HepG2 cells, which in combination with IFNγ, inhibited a spontaneous decrease in the proportion of mutant mtDNA in an A549 m.3243A>G cybrid sub-clone (section 3.13, Figure 3.13.2). Lastly, a 2.7 kb viral RNA interacts with complex I to stabilise $\Delta\psi m$ for ATP production[138], suggesting that down-regulation of PPARGC1A could be part of a cellular defence response to minimise ATP production by mitochondria, which could otherwise be hijacked for viral replication.

On the basis of leading-edge analysis (section 3.16), a HEK293T cybrid cell line was generated that carried a PPARGC1A transgene (data not shown). The intention was to assess the effect of elevated PPARGC1A expression on m.3243A>G segregation in HEK293T cells.

However, GSEA analysis of A549 datasets versus a group of datasets from cell lines with known segregation bias towards mutant mtDNA highlighted pathways and processes, rather than individual genes, which might influence mtDNA segregation bias. This reduced the impact of the PPARGC1A findings, although it remains a worthwhile experiment in the future.

4.2 CMAP and ProfileChaser suggest the reciprocal co-regulation of PPARGC1A expression and autophagy

CMAP analysis of A549↑ versus A549↓ datasets revealed the greatest enrichment for autophagy inhibitors LY-294002 and Wortmannin (section 3.16, Table 3.16.1), consistent with a state of reduced autophagic flux in A549↑. Enrichment for two autophagy inhibitors increases confidence that this process is specifically and significantly affected. This appears somewhat paradoxical as PGC-1α stimulates mitochondrial biogenesis, and it is difficult to see how this would reduce the requirement for 'mitophagy' in particular, or autophagy generally. Nevertheless, it has been shown that stimulation of autophagy with rapamycin in human skeletal muscle decreases PPARGC1A expression[139], which is also consistent with reciprocal regulation of autophagy and PPARGC1A expression. Further support for this hypothesis comes from ProfileChaser analysis of A549↑ versus A549↓ datasets, which revealed reduced TGF-β signalling in A549↑ (Section 3.16, Table 3.16.2). This is because TGF-β activates autophagy in human hepatocellular carcinoma cells[140], suggesting that reduced TGF-β signalling in A549↑ cells in turn reduces autophagic flux, compared to autophagic flux in A549↓ cells.

4.3 Modelling the effects of autophagic flux on mtDNA segregation

Reduced autophagic flux could promote a decrease in the proportion of mutant mtDNA by permitting the time-dependent expression of a dysfunctional phenotype in mitochondria homoplasmic or heteroplasmic for mutant mtDNA (Figure 4.3.1). This hypothesis rests upon three assumptions.

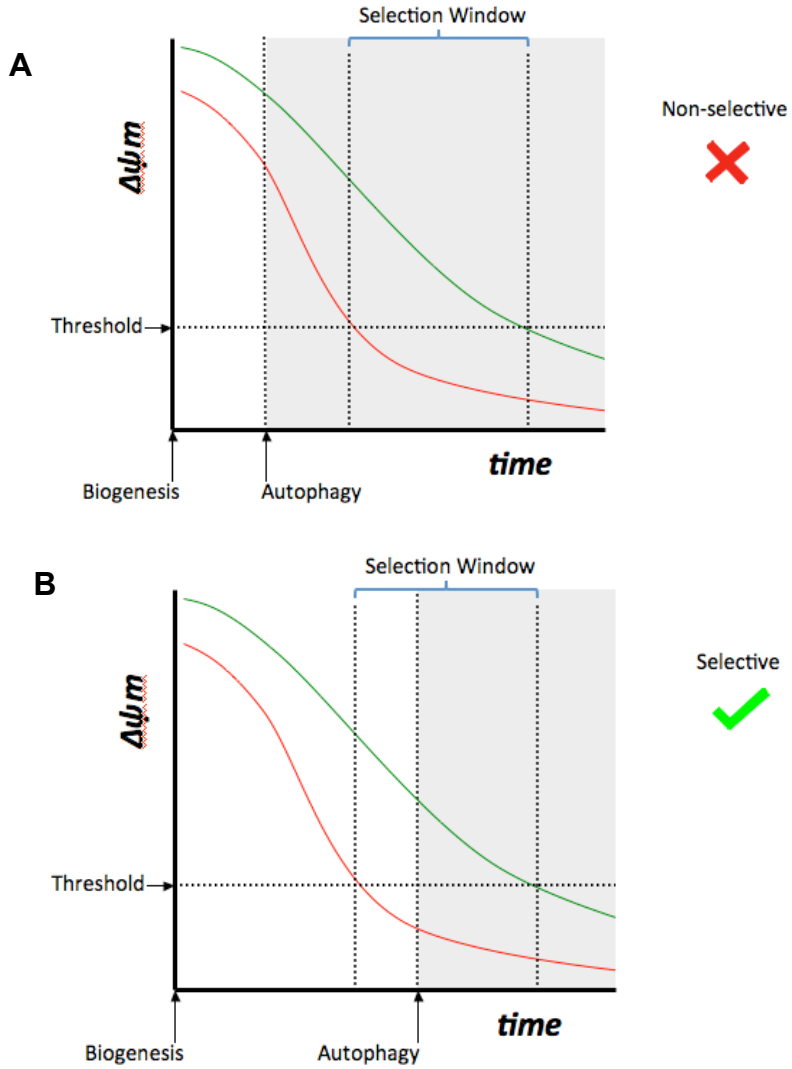


Figure 4.3.1. A model of mutant mtDNA selection by reduced autophagic flux. Following biogenesis, functional mitochondria that are homoplasmic for WT mtDNA (green line) exhibit a loss of Δψm with time, which is accelerated in dysfunctional mitochondria that are homoplasmic or heteroplasmic for mutant mtDNA (red line). (**A**) At basal autophagic flux, functional and dysfunctional mitochondria both maintain a Δψm above threshold at early time points. The autophagic machinery does not discriminate between functional and dysfunctional mitochondria, so mitochondria are indiscriminately removed by autophagy (**B**) Lower autophagic flux prevents indiscrimate autophagy of mitochondria at early time points and permits a 'selection window' where dysfunctional mitochondria exhibiting a Δψm below threshold and are targeted for autophagy, whilst functional mitochondria exhibit a Δψm above threshold and escape autophagy. Targeting of dysfunctional mitochondria for removal by autophagy promotes a decrease in the proportion of mutant mtDNA. The grey zone represents Δψm values inaccessible to mitochondria due to their removal by autophagy.

The first is that mitochondria are only selected for autophagy below a threshold Δψm. This is supported by recent elucidation of a mitochondrial quality control pathway, in which Pink1 is rapidly degraded in functional mitochondria, but accumulates on the outer membrane of Δψm deficient mitochondria, where it recruits Parkin to mediate removal of mitochondria by autophagy[63,64].

The second assumption is that following biogenesis mitochondrial function follows a protein activity decay curve. Since mitochondrial function is dependent on protein activity, this assumption seems warranted. However, mitochondria are not static organelles, rather they experience cycles of fission and fusion with the mitochondrial network. For the inner mitochondrial membrane, this might serve to maintain Δψm by functional complementation despite decreases in local protein activity[141]. Nevertheless, isolation of mitochondria from the network by fission would be expected to prevent complementation and reveal local protein activity. Therefore, the genotype, via the functional phenotype of mtDNA, would be unmasked intermittently.

The third assumption is that mitochondria heteroplasmic for mutant mtDNA exhibit an accelerated decay of mitochondrial function. Since mitochondrial function is dependent on protein activity and integrity, which can both be disrupted by mitochondrial mutations, this assumption seems warranted.

In the 'Delayed-Autophagy' model, basal autophagic flux would permit the indiscrimate autophagy of both functional and dysfunctional mitochondria at early time points, where both maintain a Δψm above threshold (Figure 4.3.1 A). A lower autophagic flux would spare mitochondria from indiscriminate autophagy at early time points and permit, at later time points, a 'selection window' where dysfunctional mitochondria unable to maintain Δψm above threshold are targeted for autophagy, whilst functional mitochondria able to maintain a Δψm above threshold and are spared (Figure 4.3.1 B). If autophagic flux is lower still, Δψm for both functional and dysfunctional mitochondria falls below the threshold, and autophagy once again becomes indiscriminate. However, autophagic flux in a cell would have to be high enough to prevent partial or total loss of Δψm, which can trigger apoptosis[142].

4.4 ROS response and glutathione metabolism

GSEA analysis of A549 datasets versus a group of datasets from cell lines with a known segregation bias towards mutant mtDNA revealed the significant up-regulation of genesets involved in response to ROS, and glutathione metabolism (section 3.5, Figure 3.5.1). Moreover, these genesets were also upregulated when A549 datasets were compared to a group of datasets taken from other cell lines in the NCI60 study, suggesting this is a feature specific to the A549 cell line.

Studies in mice also suggest a link between the response to ROS and biased segregation. Mice heteroplasmic for BALB and NZB mtDNA haplotypes select BALB in hematopoietic tissue and NZB in kidney and liver[81]. Both haplotypes are indistinguishable with the exception that mtDNA copy number and ROS production are higher in cells with NZB mtDNA[89]. Differential response to ROS in hematopoietic tissue and kidney/liver could influence biased segregation according to the following model.

In hematopoietic tissue, elevated expression of genes involved in the defence and detoxification of ROS would prevent premature stimulation of mitochondrial biogenesis by elevated ROS from NZB mtDNA. This would promote preferential removal of NZB mtDNA by the Parkin/Pink1 pathway (or an alternative pathway), since NZB mtDNA has a lower respiratory capacity per mtDNA and will therefore maintain a lower $\Delta\psi m$. This would increase the proportion of the BALB mtDNA haplotype.

In kidney and liver tissue, reduced expression of genes involved in ROS defence and detoxification would permit the premature stimulation of mitochondrial biogenesis, due to elevated ROS from NZB mtDNA. Cells with an increased percentage of NZB mtDNA could arise through mitotic drift coupled to regenerative proliferation. Due to a reduction in respiratory capacity and an increase in ROS associated with the NZB haplotype, these cells would increase their mtDNA copy number by ROS induced mitochondrial biogenesis and increase the percentage of the NZB mtDNA haplotype at the cell population level.

A caveat of the 'Delayed-Autophagy' model is that it rests upon the speculation that A549↑ represents a cellular state synonymous with a spontaneous decrease in the proportion of mutant mtDNA. A caveat of the ROS-based model is that it rests upon the idea that hematopoietic and kidney/liver tissue exhibit differential expression of genes involved in ROS defence/detoxification and that regenerative proliferation in kidney/liver is sufficient to permit mitotic drift of mtDNA haplotypes. Results from Enrichment Map analysis of 2DG treated versus non-treated HT1080 cells, a treatment associated with segregation to WT mtDNA in a sub-population of clones, suggest a third model that does not rely on such speculation.

4.5 MtDNA turnover

2DG treatment of HT1080 cells, which is associated with segregation to WT mtDNA in A549 sub-clones, induces down-regulation of mitochondrial gene expression based on Enrichment Map analysis. This is synonymous to reduced autophagy in respect to establishing a cellular environment where Δψm of dysfunctional mitochondria is permitted to decay to sub-threshold levels. With continuous biogenesis, Δψm of individual functional and dysfunctional mitochondria tracks the green and red lines respectively with time (Figure 4.5.1 A).

Biogenesis of daughter mitochondria and removal of parent mitochondria by autophagy resets the Δψm 'clock', such that Δψm is restored to maximum in new mitochondria before the Δψm of its parent has the opportunity to decay below the threshold. Hence, high 'turnover' may allow a dysfunctional mitochondrial lineage to maintain a Δψm above the threshold and persist.

Conversely, inhibition of mitochondrial biogenesis would reduce turnover. Owing to the aforementioned accelerated protein activity decay in mutant mtDNA containing mitochondria, such organelles would manifest a markedly lower Δψm at later time points compared to mitochondria with WT mtDNA.

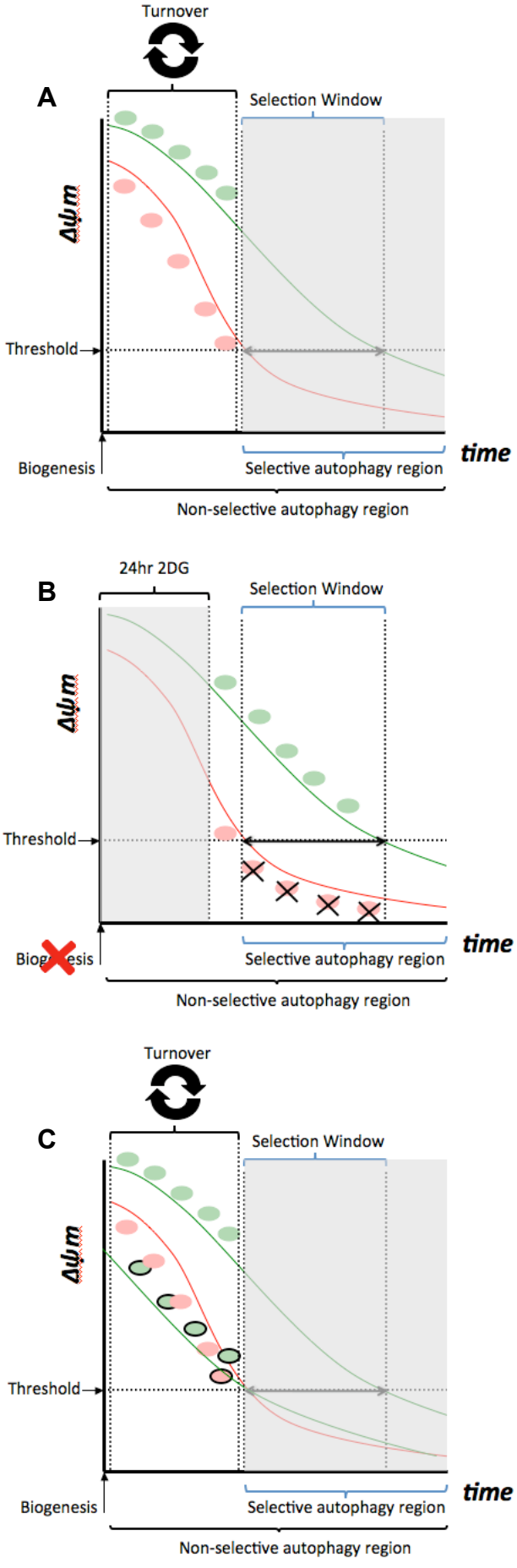

Turnover
A
Selection Window
Δψm
Threshold
Biogenesis
time
Selective autophagy region
Non-selective autophagy region
24hr 2DG
B
Selection Window
Δψm
Threshold
time
Selective autophagy region
Non-selective autophagy region
Turnover
C
Selection Window
Δψm
Threshold
Biogenesis
time
Selective autophagy region
Non-selective autophagy region

Figure 4.5.1. A model of mutant mtDNA selection by inhibition of mitochondrial biogenesis. (**A**) The Δψm of functional (green) and dysfunctional (red) mitochondria track the green and red lines respectively with time, but do not reach the 'selection window' due to the biogenesis of daughter mitochondria and the non-selective autophagy of parental mitochondria, or 'turnover'. (**B**) Inhibition of mitochondrial biogenesis by the ER stressor 2DG allows mitochondria to reach the selection window, where dysfunctional mitochondria are targeted for removal by selective autophagy due to a Δψm value below threshold. (**C**) Upon removal of 2DG and resumption of mitochondrial biogenesis, the proportion of functional mitochondria is increased. Mitochondria that have survived previous 2DG treatment are circled black. The grey zone represents Δψm values that are inaccessible to mitochondria.

Not only would this allow an increase in selective autophagy of dysfunctional mitochondria, *in extremis*, it would reduce the ability of dysfunctional mitochondria to contribute to renewed biogenesis, since biogenesis requires Δψm-dependent mitochondrial protein import[143] (Figure 4.5.1 B).

A caveat of the 'turnover' model is the assumption that during 2DG treatment mitochondria are spared from non-selective autophagy. A finding that supports this assumption is that following stimulation of autophagy by nutrient starvation in HeLa cells, mitochondria elongate and are spared from degradation by autophagy[9]. 2DG also stimulates autophagy in human pancreatic tumor cells[144] and endothelial cells[145], suggesting mitochondria may also elongate to escape autophagy in these contexts.

4.6 ER stress response and biased mtDNA segregation

In this study, A549 m.3243A>G cybrid sub-clones exhibited a ~2-fold higher rate of spontaneous segregation to WT mtDNA and at least a ~4 fold higher rate of segregation to WT mtDNA with intermittent ER stress stimulated by 2DG, compared to HEK293T m.3243A>G cybrid sub-clones expressing Parkin (Table 4.6.1).

Cell line	Spontaneous		2DG	
	Sub-clone	Pool	Sub-clone (+P)	Pool (+P)
HEK293T m.3243A>G	1.82% (1/55)	0% (0/7)	2.33% (2/86) 2.33% (2/86)	13.3% (2/15)
A549 m.3243A>G	3.45% (1/29)	0% (0/2)	10.3% (3/29)	100% (2/2)

Table 4.6.1. The frequency of spontaneous and 2DG induced mtDNA segregation in HEK293T m.3243A>G and A549 m.3243A>G cybrid cell lines. Green text indicates segregation to WT mtDNA whereas red text indicates segregation to mutant mtDNA. (+P) indicates Parkin expression.

These results confirm a single previous report that A549 cells have an innate tendency to segregate to WT mtDNA[77], whose findings were critical to the identification of 2DG as a chemical agent for stimulation of this phenomena. 2DG stimulates an ER stress response by inhibiting glycolysis and N-glycosylation of proteins[146,147]. This causes protein misfolding, detected by the ER localised transmembrane protein kinase PERK, which phosphorylates serine 51 on the α-subunit of eukaryotic translation initiation factor 2 (eIF2α)[128].

Phosphorylation of eIF2α reduces global translation yet induces expression of the transcription factor ATF4, which in turn stimulates expression of genes involved in amino acid import, glutathione biosynthesis and resistance to oxidative stress[128]. 2DG has multiple activities besides activating a response to ER stress including stimulation of autophagy[144,145], inhibition of glycolysis[147] and inhibition of viral infection[132].

The contribution of each activity to a reduction in the proportion of mutant mtDNA remains unknown. Supporting a role for ER stress in the segregation to WT mtDNA, the impetus for 2DG treatment of cells was the bioinformatic identification of a gene expression profile synonymous with the response to ER stress in A549 cells, which exhibit a high frequency of segregation to WT mtDNA. Significant enrichment for genesets involved in glutathione metabolism and ROS response with ER stress links this process to a model for biased segregation (section 4.4). Furthermore, reduction of global translation with ER stress could promote the inhibition of mitochondrial gene expression revealed by Enrichment Map analysis of 2DG treated HT1080 cells. This would link ER stress to the 'turnover' model of biased mtDNA segregation (section 4.5). However, another form of ER stress, amino acid starvation, failed to stimulate segregation to WT mtDNA in 13 A549 m.3243A>G sub-clones and a pooled population (data not shown).

Marked down-regulation of the glycolytic enzyme HK2 in A549 cells compared to cells with a known segregation bias towards mutant mtDNA (section 3.4, table 3.4.1) suggest that inhibition of glycolysis by 2DG[147] may be responsible for stimulation of segregation to WT mtDNA. This hypothesis is supported by the finding that glucose limitation in standard culture conditions promotes segregation to WT mtDNA in HEK m.3243A>G cybrids in a reproducible manner (Catherine Nezich, personal communication).

Alternatively, the inhibition of viral infection by 2DG[132] suggests it might stimulate biased mtDNA segregation via up-regulation of genes involved in viral interference or the 'Interferon-response' (section 3.11, Figure 3.11.1), many of which participate in Parkin-mediated mitophagy[133]. Stimulation of this pathway could increase the turnover of sub-optimal mitochondria and thereby promote a decrease in the proportion of mutant mtDNA.

This is supported by UBP analysis of 2DG treated HT1080 cells, which revealed significant enrichment for UBPs associated with an interferon-response. Moreover, GIMAP3, which co-segregates with the selection of one of two mitochondrial genotypes in hematopoietic cells from mice[84], is from a family of nucleotide binding proteins associated with immune functions and conserved in plants, mice rats and

humans. This suggests involvement in conserved defence mechanisms[85]. However, IFNα/γ treatment did not increase segregation to WT mtDNA in HEK cybrids and inhibited spontaneous segregation to WT mtDNA in an A549 cybrid sub-clone. Therefore, stimulation of the 'Resistome' may be necessary but not sufficient to reduce the proportion of mutant mtDNA. Alternatively, features of the A549 cell line response to interferon may inhibit segregation to WT mtDNA. The 'Autophagic flux' model for biased segregation (section 4.3) requires reduced autophagy to promote segregation to WT mtDNA. Since 2DG has been found to stimulate autophagy[144,145], it is unlikely that induction of biased segregation occurs via 2DG's effect on this process.

2DG did not stimulate a significant increase in the segregation to WT mtDNA in HEK cybrid sub-clones expressing Parkin relative to spontaneous segregation. However, 2DG stimulated an increase in WT mtDNA in 13.3% (2/15) of pooled populations expressing Parkin, whereas heteroplasmy was stable in all 12 pooled populations maintained without Parkin expression and seven pooled populations maintained without 2DG or Parkin expression (Table 4.6.1). More strikingly, whilst segregation to WT mtDNA was observed in 10.3% (3/29) of A549 cybrid sub-clones, it was observed in both pooled populations tested. Although a larger sample size is required to confirm this behaviour, these results suggest that biased segregation to WT mtDNA is enriched in pooled populations.

One explanation for this behaviour is that biased mtDNA segregation may confer a growth advantage to a particular sub-clone, permitting both intracellular and intercellular selection of WT mtDNA in pooled populations. Although differences in growth rate have not been observed between A549 sub-clones with and without the capacity for segregation to WT mtDNA[77], co-culture of different sub-clones may be required to permit a growth advantage. Alternatively, the capacity to segregate may be a dominant transferable phenomenon, which could be communicated by a diffusible factor.

4.7 ER stress and the interferon-response

Interferon is an annotation of some of the most enriched UBPs for both 2DG treatment of HT1080 cells and amino acid starvation of HepG2 cells. An outstanding question is why does ER stress induce features of the interferon-response? One explanation is that ER stress induces the expression of endogenous retroviruses (ERVs). ERVs are inherited cellular genes that encode retroviral gene products or replication-competent retroviruses[148]. ERVs originate from retroviral infection of the germline in a human ancestor, an event that has occurred multiple times[149].

Infectious ERVs have not been identified in human DNA, suggesting that accumulated mutations have rendered them inactive, and ERV products are not normally expressed due to defences in the cell that repress proviral expression[150]. However, multiple stresses including UV light[151], γ radiation[152], serum starvation or infection by exogenous viruses[153] can activate ERV elements. It is tempting to speculate that ER stress may activate ERV elements, whose gene products could trigger the interferon-response observed in both the UBP profile of 2DG-treated HT1080 cells and amino acid starved HepG2 cells.

Alternatively, ER stress could directly induce features of the interferon-response by up-regulation of TNFα, which stimulates the 'Resistome'[131] and interferon pathway[154]. This is because under conditions of ER stress, translational suppression of inhibitory kappa B activates nuclear factor kappa B, stimulating expression of target genes including TNFα[155].

Activation of the interferon-response by ER stress may also provide an explanation for why selection of mtDNA is only observed in a sub-population of clones, and why it is enrichment in pooled populations. Measurement of interferon-ß expression in single Newcastle disease virus (NDV) infected human monocyte-derived dendritic cells (MSCCs) revealed that only a small subset of cells produce and secrete interferon-ß (IFN- ß) in response to NDV infection, whilst the majority of cells are unresponsive[156].

Secretion of IFN-ß in early responders activates the interferon pathway in the whole population by binding to interferon receptors on cells unresponsive to NDV. A model consistent with IFN-ß measurement in a population of cells over time requires that ~5% are early responders. This is similar to the fraction of HEK cybrid sub-clones that exhibit biased mtDNA segregation with ER stress (~2-3%). Therefore, ER stress may activate the expression of ERV products in all sub-clones, but only a small subset of early responders would mount an interferon-response to promote a reduction in the proportion of mutant mtDNA.

The presence of early responders in a pooled population and their isolation upon sub-cloning also provides an explanation for the elevated frequency of heteroplasmy change in pooled populations; early responders can communicate an interferon-response to other cells in a pooled population but not as an isolated sub-clone. However, a minimum percentage or density of early responders may be required for heteroplasmy change since this phenomenon is not observed in all pooled populations.

The hypothesis that activation of the interferon pathway is responsible for inducing heteroplasmy change cannot explain stable heteroplasmy in the majority of HEK cybrid sub-clones with IFNα/γ treatment, or the inhibition of a spontaneous decrease in the proportion of mutant mtDNA in an A549 cybrid sub-clone with IFNα/γ treatment, observed in the absence of treatment. Therefore, activation of the interferon pathway may be necessary but not sufficient to induce segregation of mutant and WT mtDNA in HEK293T m.3243A>G cells, while the elements of the interferon-response needed for mtDNA segregation may be constitutively active in the A549 cells, with further stimulation merely repressing other factors needed for the effect, with PPARGC1A being an obvious possible example.

Intriguingly, ProfileChaser analysis of A549↑ versus A549↓ datasets revealed a low level enrichment for a 24-hour versus six-hour type I and II interferon-response, suggesting 24 hour IFNα/γ treatment induces a cellular state similar to A549↑, which may be associated with decrease in the proportion of mutant mtDNA, whilst six hour IFNα/γ treatment induces a cellular state similar to A549↓.

Out of 15 sub-clones and a pooled population, the transition from cycles of six hour IFNα/γ treatment to cycles of 24 hour combined 2DG/IFNα/γ treatment followed by recovery in standard culture conditions may have permitted the spontaneous decrease in the proportion of mutant mtDNA in a single sub-clone, also observed in the absence of treatment (section 3.14, Figure 3.14.2). Although the presence of 2DG in the 24-hour treatment complicates the interpretation of the results, 2DG treatment is dispensable for the selection of WT mtDNA in this sub-clone, suggesting that duration of IFNα/γ treatment may be significant.

The limited resources available for this study meant that many assumptions were made in experimental design and interpretation. Firstly, it was assumed that all treatments had the expected effect, for example, that 2DG stimulated ER stress and that IFNα/γ stimulated an interferon-response. The possibility remains that failure to stimulate segregation to WT mtDNA in a particular cell line or sub-clone reflects failure to stimulate the desired pathway.

It was also assumed that the global expression profile of 2DG treated HT1080 cells is recapitulated in HEK293 and A549 cybrids. Therefore, features of the global expression profile of 2DG treated HT1080 cells were assumed to have caused successful stimulation of segregation to WT mtDNA in A549 cybrid sub-clones and pooled populations, which led to the decision to implement IFNα/γ treatment in HEK293 and A549 cybrids. In turn, the ability of IFNα/γ to inhibit the spontaneous segregation to WT mtDNA in an A549 cybrid sub-clone was assumed to be due to down-regulation of PPARGC1A expression, a feature observed in the global expression profile of IFNα treated HepG2 cells but not experimentally verified in A549 cybrids.

The single-cell study of 143B m.3243A>G cybrid clonal cell lines revealed that increased variation in heteroplasmy between progeny of a single clone preceded a shift from ~100% mutant mtDNA to near ~100% WT mtDNA[157]. To explain the segregation to WT mtDNA, the authors propose a metastable mtDNA segregation unit, which through infrequent disruption and reorganisation creates segregation units with modified heteroplasmy in a ‘re-arranged cell’.

Segregation units with modified heteroplasmy then become fixed in descendent cells by random mitotic drift whilst growth advantage in a particular descendent causes a shift towards the heteroplasmy of its segregation unit.

According to this model, the decreases in the proportion of mutant mtDNA effected by 2DG treatment would occur by stimulating the disruption and reorganisation of segregation units to generate a 'rearranged cell', followed by reduction in the proportion of mutant mtDNA due to growth advantage of a descendent containing a segregation unit with reduced mutant mtDNA. Therefore, treatment of cells with a single pulse of 2DG would be expected to recapitulate results obtained with intermittent 2DG treatment followed by recovery in standard culture conditions. These meta-stable segregation units could well be physical entities, if not nucleoids, then 'membrane-compartments' bounded by the inner mitochondrial membrane[157].

Electron tomography (ET) of nerve terminal mitochondria revealed lamellar cristae, which are not interconnected and fill the interior of the mitochondria (Figure 4.7.1)[158]. The spaces between these lamellar cristae are a plausible candidate for the 'membrane compartment' posited as a segregation unit. High-resolution microscopy or ET could be used to assess the effect of treatments, such as 2DG and IFNα/γ, on these mitochondrial ultra-structures.

Comparison of A549↑ versus A549↓ datasets by GSEA analysis revealed a functional profile consistent with the response to ER stress, but a profile distinct from the ER stress response induced by 2DG treatment of HT1080 cells or amino acid starvation of HepG2 cells. Since hypersensitivity of PPARGC1A expression to amino acid starvation in A549 cells is not observed in HEK293T or 143B cell lines, it is likely that HepG2 cells lack this hypersensitivity. Therefore, differences between A549 and HepG2 cells undergoing amino acid starvation may simply reflect an absence of PPARGC1A stimulation in HepG2 cells.

This hypothesis is supported by leading-edge analysis of A549↑ versus A549↓ datasets, which indicates that PPARGC1A expression is responsible for many of the functional features enriched in A549↑, including reduced autophagy.

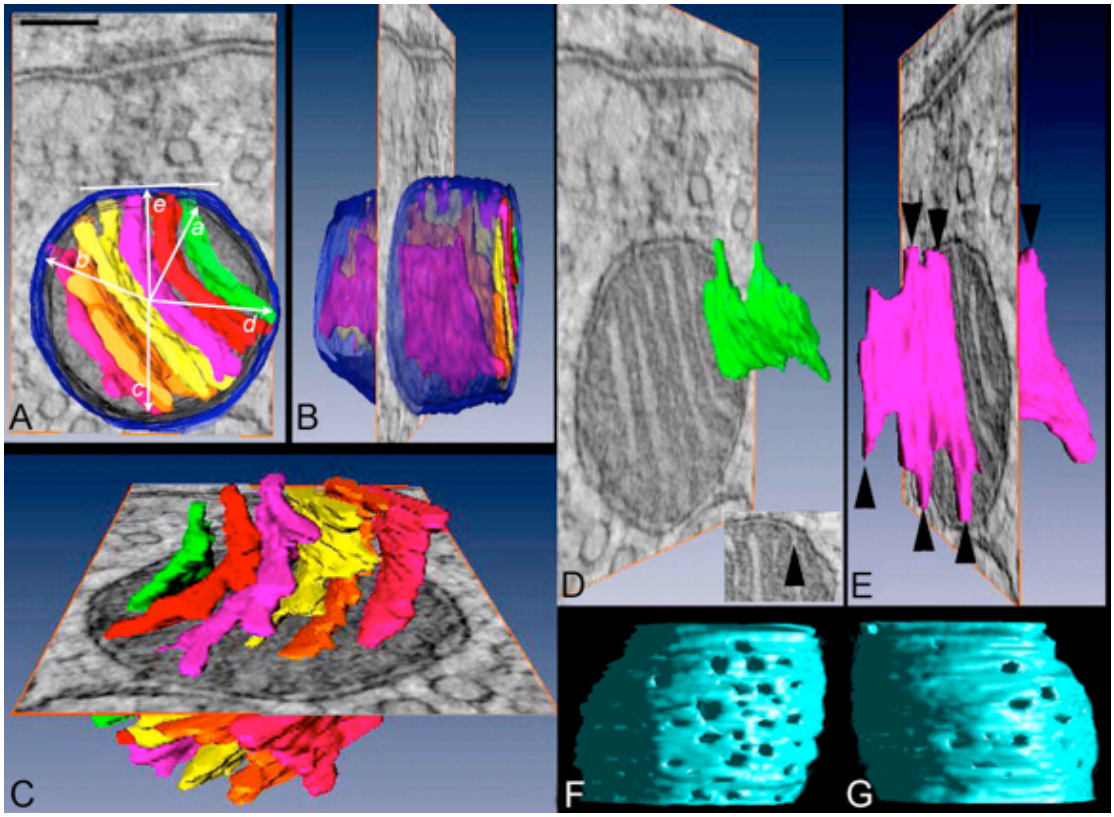


Figure 4.7.1. Electron tomography of nerve terminal mitochondria. (**A**) A top-view of the surface rendered volume of mitochondria superimposed on to an electron micrograph representing a roughly central slice of the reconstruction. The outer mitochondrial membrane (OMM) is shown in translucent blue and each crista in a different colour. (**B/C**) A side-view of the surface rendered volume of mitochondria, with and without the outer mitochondrial membrane. (**D/E**) Examples of two crista with 'fingers' that connect to the periphery via circular openings (arrowheads) termed crista junctions. (**F/G**) View of the inner mitochondrial membrane with the OMM removed. Scale bar (**A**), 100nm (Perkins et al., 2010).

Therefore, amino acid starvation of A549 m.3243A>G cybrids could be used to induce both PPARGC1A expression and a functional profile synonymous with A549↑, which could induce a decrease in the proportion of mutant mtDNA. However, amino acid starvation failed to stimulate segregation to WT mtDNA in 13 A549 m.3243A>G sub-clones and a pooled population (data not shown).

2DG has multiple known activities including stimulation of ER stress[144], stimulation of autophagy[144,145] inhibition of glycolysis[147] and inhibition of viral infection[132]. The contribution of each activity to a reduction in the proportion of mutant mtDNA in A549 m.3243A>G cybrid sub-clones and pooled populations remains to be determined, but could be uncovered by stimulation of cells with alternative agents possessing each, or a combination of 2DG activities.

The failure to induce mtDNA segregation in any HEK cybrids with a variety of mitochondrial respiratory chain inhibitors and poisons (Catherine Nezich, personal communication) suggests that 2DG does not influence biased segregation by directly effecting mitochondrial respiration. The ability to induce segregation in A549 cybrids at a frequency of ~10% of 2DG treated sub-clones will enable A549 cells with and without the capacity for segregation to WT mtDNA to be characterised in detail. Study of biased segregation within a single nuclear background will help to ensure that differences observed in the presence and absence of the phenomenon are mechanistically relevant, and differences unrelated to the phenomenon are minimised.

This study contributes to the ultimate goal of providing a detailed mechanistic understanding of biased segregation to enable the development of new therapies to reduce the proportion of mutant mtDNA in cells and tissues of patients with mitochondrial disease, PD or the aged. Because a relatively small increase in WT mtDNA can boost the respiratory function of cells considerably[53,159], harnessing biased segregation could provide an effective approach to alleviate symptoms or even cure these diseases.

This study also showcases the power of global expression databases, in combination with bioinformatic tools, to extract insight from biological systems. NCBI GEO was utilised directly to acquire and compare global expression datasets from different cells lines with a characterised segregation bias, but was also utilised indirectly to generate an unbiased architecture of the cell[109]. This architecture was subsequently harnessed to generate increased insight, and in future could be used to guide combinatorial intervention.

Bioinformatic tools such as GSEA[95], Connectivity Map[106,107], ProfileChaser[108] and Enrichment Map[110] were critical to the extraction of meaning from lists of differentially expressed genes, without which biological understanding would be seriously limited. These tools revealed a number of genes, genesets, processes and pathways that potentially impinge on mtDNA segregation bias, many of which are susceptible to pharmacological intervention. These include ER stress and interferon-responses, PPARGC1A and autophagy. The ER stress response was identified as the most promising intervention, the ER stressor 2DG was trialed for stimulation of biased mtDNA segregation, and successful stimulation was demonstrated in A549 m.3243A>G cybrids.

The capability to stimulate an increase in the frequency of biased mtDNA segregation will allow a more extensive characterisation of the phenomena, as it markedly reduces the number of cell lines that have to be followed. This will permit the transition from high throughput characterisation techniques, which at present can answer only a limited number of scientific questions, to the large repertoire of low throughput characterisation techniques that allow most if not all questions to be asked. The finding that active segregation is observable and potentially transmissible in pools of clones will further facilitate this transition, and aid the identification of new interventions that are efficacious.

The recent isolation of iPSCs derived from MELAS patients[160] presents the opportunity to study the effects of interventions in a non-aneuoploid system. Moreover, differentiation of iPSC to multiple cell types could be used to explore whether interventions are efficacious in somatic tissues.

5 References

1. Westermann, B. Mitochondrial fusion and fission in cell life and death. *Nat Rev Mol Cell Biol* **11**, 872-884 (2010).
2. Smirnova, E., Shurland, D. L., Ryazantsev, S. N. & van der Bliek, A. M. A human dynamin-related protein controls the distribution of mitochondria. *J Cell Biol* **143**, 351-358 (1998).
3. James, D. I., Parone, P. A., Mattenberger, Y. & Martinou, J. C. hFis1, a novel component of the mammalian mitochondrial fission machinery. *J Biol Chem* **278**, 36373-36379 (2003).
4. Friedman, J. R., Lackner, L. L., West, M., DiBenedetto, J. R., Nunnari, J. & Voeltz, G. K. ER tubules mark sites of mitochondrial division. *Science* **334**, 358-362 (2011).
5. Cipolat, S., Martins de Brito, O., Dal Zilio, B. & Scorrano, L. OPA1 requires mitofusin 1 to promote mitochondrial fusion. *Proc Natl Acad Sci U S A* **101**, 15927-15932 (2004).
6. Santel, A. & Fuller, M. T. Control of mitochondrial morphology by a human mitofusin. *J Cell Sci* **114**, 867-874 (2001).
7. Chen, H., Detmer, S. A., Ewald, A. J., Griffin, E. E., Fraser, S. E. & Chan, D. C. Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *J Cell Biol* **160**, 189-200 (2003).
8. Mitra, K., Wunder, C., Roysam, B., Lin, G. & Lippincott-Schwartz, J. A hyperfused mitochondrial state achieved at G1-S regulates cyclin E buildup and entry into S phase. *Proc Natl Acad Sci U S A* **106**, 11960-11965 (2009).
9. Gomes, L. C., Di Benedetto, G. & Scorrano, L. During autophagy mitochondria elongate, are spared from degradation and sustain cell viability. *Nat Cell Biol* **13**, 589-598 (2011).
10. Skulachev, V. P. Mitochondrial filaments and clusters as intracellular power-transmitting cables. *Trends Biochem Sci* **26**, 23-29 (2001).
11. Frank, S., Gaume, B., Bergmann-Leitner, E. S., Leitner, W. W., Robert, E. G., Catez, F., Smith, C. L. & Youle, R. J. The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Dev Cell* **1**, 515-525 (2001).
12. Sagan, L. On the origin of mitosing cells. *J Theor Biol* **14**, 255-274 (1967).

13. Sirrenberg, C., Bauer, M. F., Guiard, B., Neupert, W. & Brunner, M. Import of carrier proteins into the mitochondrial inner membrane mediated by Tim22. *Nature* **384**, 582-585 (1996).
14. Yano, M., Hoogenraad, N., Terada, K. & Mori, M. Identification and functional analysis of human Tom22 for protein import into mitochondria. *Mol Cell Biol* **20**, 7205-7213 (2000).
15. Shmookler-Reis, R. J., Goldstein, S. Mitochondrial DNA in mortal and immortal human cells. Genome number, integrity, and methylation. *J Biol Chem* **258**, 9078-9085 (1983).
16. Giles, R. E., Blanc, H., Cann, H. M. & Wallace, D. C. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* **77**, 6715-6719 (1980).
17. Sato, M. & Sato, K. Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochim Biophys Acta* **1833**, 1979-84 (2013).
18. Lane, N. Mitonuclear match: optimizing fitness and fertility over generations drives ageing within generations. *Bioessays* **33**, 860-869 (2011).
19. Bokori-Brown, M. & Holt, I. J. Expression of algal nuclear ATP synthase subunit 6 in human cells results in protein targeting to mitochondria but no assembly into ATP synthase. *Rejuvenation Res* **9**, 455-469 (2006).
20. Claros, M. G., Perea, J., Shu, Y., Samatey, F. A., Popot, J. L. & Jacq, C. Limitations to in vivo import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome b. *Eur J Biochem* **228**, 762-771 (1995).
21. Wright, A. F., Murphy, M. P. & Turnbull, D. M. Do organellar genomes function as long-term redox damage sensors? *Trends Genet* **25**, 253-261 (2009).
22. Kasamatsu, H. & Vinograd, J. Unidirectionality of replication in mouse mitochondrial DNA. *Nat New Biol* **241**, 103-105 (1973).
23. Clayton, D. A. Replication of animal mitochondrial DNA. *Cell* **28**, 693-705, (1982).
24. Pohjoismaki, J. L., Holmes, J. B., Wood, S. R., Yang, M. Y., Yasukawa, T., Reyes, A., Bailey, L. J., Cluett, T. J., Goffart, S., Willcox, S., Rigby, R. E., Jackson, A. P., Spelbrink, J. N., Griffith, J. D., Crouch, R. J., Jacobs, H. T. & Holt, I. J. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *J Mol Biol*

397, 1144-1155 (2010).

25. Yang, M. Y., Bowmaker, M., Reyes, A., Vergani, L., Angeli, P., Gringeri, E., Jacobs, H. T. & Holt, I. J. Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* **111**, 495-505 (2002).
26. Yasukawa, T., Reyes, A., Cluett, T. J., Yang, M. Y., Bowmaker, M., Jacobs, H. T. & Holt, I. J. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J* **25**, 5358-5371(2006).
27. Reyes, A., Kazak, L., Wood, S. R., Yasukawa, T., Jacobs, H. T. & Holt, I. J. Mitochondrial DNA replication proceeds via a 'bootlace' mechanism involving the incorporation of processed transcripts. *Nucleic Acids Res* **41**, 5837-5850 (2013).
28. Reyes, A., Yang, M. Y., Bowmaker, M. & Holt, I. J. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *J Biol Chem* **280**, 3242-3250 (2005).
29. Yasukawa, T., Yang, M. Y., Jacobs, H. T. & Holt, I. J. A bidirectional origin of replication maps to the major noncoding region of human mitochondrial DNA. *Mol Cell* **18**, 651-662 (2005).
30. Mita, S., Rizzuto, R., Moraes, C. T., Shanske, S., Arnaudo, E., Fabrizi, G. M., Koga, Y., DiMauro, S. & Schon, E. A. Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. *Nucleic Acids Res* **18**, 561-567 (1990).
31. Holt, I. J. & Reyes, A. Human mitochondrial DNA replication. *Cold Spring Harb Perspect Biol* **82**, 351-355 (2012).
32. Bron, S., Holsappel, S., Venema, G. & Peeters, B. P. Plasmid deletion formation between short direct repeats in Bacillus subtilis is stimulated by single-stranded rolling-circle replication intermediates. *Mol Gen Genet* **226**, 88-96 (1991).
33. Pomerantz, R. T. & O'Donnell, M. What happens when replication and transcription complexes collide? *Cell Cycle* **9**, 2537-2543 (2010).
34. Iborra, F. J., Kimura, H. & Cook, P. R. The functional organization of mitochondrial genomes in human cells. *BMC Biol* **2**, 9, (2004).
35. Legros, F., Malka, F., Frachon, P., Lombes, A. & Rojo, M. Organization and dynamics of human mitochondrial DNA. *J Cell Sci* **117**, 2653-2662 (2004).

36. Gilkerson, R. W., Schon, E. A., Hernandez, E. & Davidson, M. M. Mitochondrial nucleoids maintain genetic autonomy but allow for functional complementation. *J Cell Biol* **181**, 1117-1128 (2008).
37. Kukat, C., Wurm, C. A., Spahr, H., Falkenberg, M., Larsson, N. G., Jakobs, S. Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proc Natl Acad Sci U S A* **108**, 13534-13539.
38. Chen, X. J. & Butow, R. A. The organization and inheritance of the mitochondrial genome. *Nat Rev Genet* **6**, 815-825 (2005).
39. Alam, T. I., Kanki, T., Muta, T., Ukaji, K., Abe, Y., Nakayama, H., Takio, K., Hamasaki, N. & Kang, D. Human mitochondrial DNA is packaged with TFAM. *Nucleic Acids Res* **31**, 1640-1645 (2003).
40. Ekstrand, M. I., Falkenberg, M., Rantanen, A., Park, C. B., Gaspari, M., Hultenby, K., Rustin, P., Gustafsson, C. M. & Larsson, N. G. Mitochondrial transcription factor A regulates mtDNA copy number in mammals. *Hum Mol Genet* **13**, 935-944 (2004).
41. Falkenberg, M., Larsson, N. G. & Gustafsson, C. M. DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem* **76**, 679-699 (2007).
42. He, J. Mao, C. C., Reyes, A., Sembongi, H., Die Re, M., Granycome, C., Clippingdale, A. B., Fearnley, I. M., Harbour, M., Robinson, A. J., Reichelt, S., Spelbrink, J. N., Walker, J. E. & Holt, I. J. The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization. *J Cell Biol* **176**, 141-146 (2007).
43. Holt, I. J., He, J., Mao, C. C., Boyd-Kirkup, J. D., Martinsson, P., Sembongi, H., Reyes, A. & Spelbrink, J. N. Mammalian mitochondrial nucleoids: organizing an independently minded genome. *Mitochondrion* **7**, 311-321 (2007).
44. Holt, I. J. Zen and the art of mitochondrial DNA maintenance. *Trends Genet* **26**, 103-109 (2010).
45. Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M., Elsas, L. J. & Nikoskelainen, E. K. Mitochondrial DNA mutation associated with Leber´s hereditary optic neuropathy. *Science* **242**, 1427-1430 (1988).
46. Holt, I. J., Harding, A. E. & Morgan-Hughes, J. A. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* **331**, 717-719 (1988).

47. Poulton, J., Deadman, M. E. & Gardiner, R. M. Duplications of mitochondrial DNA in mitochondrial myopathy. *Lancet* **1**, 236-240 (1989).
48. Goto, Y., Nonaka, I. & Horai, S. A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* **348**, 651-653 (1990).
49. Chomyn, A., Enriquez, J. A., Micol, V., Fernandez-Silva, P. & Attardi, G. The mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episode syndrome-associated human mitochondrial tRNALeu(UUR) mutation causes aminoacylation deficiency and concomitant reduced association of mRNA with ribosomes. *J Biol Chem* **275**, 19198-19209 (2000).
50. Dubeau, F., De Stefano, N., Zifkin, B. G., Arnold, D. L. & Shoubridge, E. A. Oxidative phosphorylation defect in the brains of carriers of the tRNAleu(UUR) A3243G mutation in a MELAS pedigree. *Ann Neurol* **47**, 179-185 (2000).
51. Krishnan, K. J., Reeve, A. K., Samuels, D. C., Chinnery, P. F., Blackwood, J. K., Taylor, R. W., Wanrooij, S., Spelbrink, J. N., Lightowlers, R. N. & Turnbull, D. M. What causes mitochondrial DNA deletions in human cells? *Nat Genet* **40**, 275-279 (2008).
52. Angelini, C., Bello, L., Spinazzi, M. & Ferrati, C. Mitochondrial disorders of the nuclear genome. *Acta Myol* **28**, 16-23 (2009).
53. Boulet, L., Karpati, G. & Shoubridge, E. A. Distribution and threshold expression of the tRNA(Lys) mutation in skeletal muscle of patients with myoclonic epilepsy and ragged-red fibers (MERRF). *Am J Hum Genet* **51**, 1187-1200 (1992).
54. Chomyn, A., Martinuzzi, A., Yoneda, M., Daga, A., Hurko, O., Johns, D., Lai, S. T., Nonaka, I., Angelini, C. & Attardi, G. MELAS mutation in mtDNA binding site for transcription termination factor causes defects in protein synthesis and in respiration but no change in levels of upstream and downstream mature transcripts. *Proc Natl Acad Sci U S A* **89**, 4221-4225 (1992).
55. King, M. P., Koga, Y., Davidson, M. & Schon, E. A. Defects in mitochondrial protein synthesis and respiratory chain activity segregate with the tRNA(Leu(UUR)) mutation associated with mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes. *Mol Cell Biol* **12**, 480-490 (1992).

56. Obeso, J. A., Rodriguez-Oroz, M. C., Benitez-Temino, B., Blesa, F. J., Guridi, J., Marin, C. & Rodriguez, M. Functional organization of the basal ganglia: therapeutic implications for Parkinson's disease. *Mov Disord* **23**, 548-559 (2008).
57. Bender, A., Krishnan, K. J., Morris, C. M., Taylor, G. A., Reeve, A. K., Perry, R. H., Jaros, E., Hersheson, J. S., Betts, J., Klopstock, T., Taylor, R. W. & Turnbull, D. M. High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat Genet* **38**, 515-517 (2006).
58. Kraytsberg, Y., Kudryavtseva, E., McKee, A. C., Geula, C., Kowall, N. W. & Khrapko, K. Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nat Genet* **38**, 518-520, (2006).
59. Parker, W. D., Jr., Boyson, S. J. & Parks, J. K. Abnormalities of the electron transport chain in idiopathic Parkinson's disease. *Ann Neurol* **26**, 719-723 (1989).
60. Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y. & Shimizu, N. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605-608 (1998).
61. Lucking, C. B., Durr, A., Bonifati, V., Vaughan, J., De Michele, G., Gasser, T., Harhangi, B. S., Meco, G., Denefle, P., Wood, N. W., Agid, Y. & Brice, A. Association between early-onset Parkinson's disease and mutations in the parkin gene. *N Engl J Med* **342**, 1560-1567 (2000).
62. Valente, E. M., Abou-Sleiman, P. M., Caputo, V., Muqit, M. M., Harvey, K., Gispert, S., Ali, Z., Del Turco, D., Bentivoglio, A. R., Healy, D. G., Albanese, A., Nussbaum, R., Gonzalez-Maldonado, R., Deller, T., Salvi, S., Cortelli, P., Gilks, W. P., Latchman, D. S., Harvey, R. J., Dallapiccola, B., Auburger, G. & Wood, N. W. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science* **304**, 1158-1160 (2004).
63. Okatsu, K., Oka, T., Iguchi, M., Imamura, K., Kosako, H., Tani, N., Kimura, M., Go, E., Koyano, F., Funayama, M., Shiba-Fukushima, K., Sato, S., Shimizu, H., Fukunaga, Y., Taniguchi, H., Komatsu, M., Hattori, N., Mihara, K., Tanaka, K. & Matsuda, N. PINK1 autophosphorylation upon membrane potential dissipation is essential for Parkin recruitment to damaged mitochondria. *Nat Commun* **3**, 1016 (2012).

64. Narendra, D., Tanaka, A., Suen, D. F. & Youle, R. J. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J Cell Biol* **183**, 795-803 (2008).
65. Corral-Debrinski, M., Horton, T., Lott, M. T., Shoffner, J. M., Beal, M. F. & Wallace, D. C. Mitochondrial DNA deletions in human brain: regional variability and increase with advanced age. *Nat Genet* **2**, 324-329 (1992).
66. Schwarze, S. R., Lee, C. M., Chung, S. S., Roecker, E. B., Weindruch, R. & Aiken, J. M. High levels of mitochondrial DNA deletions in skeletal muscle of old rhesus monkeys. *Mech Ageing Dev* **83**, 91-101 (1995).
67. Khaidakov, M., Heflich, R. H., Manjanatha, M. G., Myers, M. B. & Aidoo, A. Accumulation of point mutations in mitochondrial DNA of aging mice. *Mutat Res* **526**, 1-7 (2003).
68. Trifunovic, A., Wredenberg, A., Falkenberg, M., Spelbrink, J. N., Rovio, A. T., Bruder, C. E., Bohlooly, Y. M., Gidlof, S., Oldfors, A., Wibom, R., Tornell, J., Jacobs, H. T. & Larsson, N. G. Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417-423 (2004).
69. Ameur, A., Stewart, J. B., Freyer, C., Hagstrom, E., Ingman, M., Larsson, N. G. & Gyllensten, U. Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. *PLoS Genet* **7**, e1002028 (2011).
70. Bailey, L. J., Cluett, T. J., Reyes, A., Prolla, T. A., Poulton, J., Leeuwenburgh, C. & Holt, I. J. Mice expressing an error-prone DNA polymerase in mitochondria display elevated replication pausing and chromosomal breakage at fragile sites of mitochondrial DNA. *Nucleic Acids Res* **37**, 2327-2335 (2009).
71. Mito, T., Kikkawa, Y., Shimizu, A., Hashizume, O., Katada, S., Imanishi, H., Ota, A., Kato, Y., Nakada, K. & Hayashi, J. Mitochondrial DNA mutations in mutator mice confer respiration defects and B-cell lymphoma development. *PLoS One* **8**, e55789 (2013).
72. Vermulst, M., Bielas, J. H., Kujoth, G. C., Ladiges, W. C., Rabinovitch, P. S., Prolla, T. A. & Loeb, L. A. Mitochondrial point mutations do not limit the natural lifespan of mice. *Nat Genet* **39**, 540-543 (2007).
73. Tyynismaa, H., Mjosund, K. P., Wanrooij, S., Lappalainen, I., Ylikallio, E., Jalanko, A., Spelbrink, J. N., Paetau, A. & Suomalainen, A. Mutant mitochondrial helicase Twinkle causes multiple mtDNA deletions and a late-onset mitochondrial disease in mice. *Proc Natl Acad Sci U S A* **102**, 17687-17692, (2005).

74. King, M. P. & Attardi, G. Mitochondria-mediated transformation of human rho(0) cells. *Methods Enzymol* **264**, 313-334 (1996).
75. Hayashi, J., Ohta, S., Kikuchi, A., Takemitsu, M., Goto, Y. & Nonaka, I. Introduction of disease-related mitochondrial DNA deletions into HeLa cells lacking mitochondrial DNA results in mitochondrial dysfunction. *Proc Natl Acad Sci U S A* **88**, 10614-10618 (1991).
76. Yoneda, M., Chomyn, A., Martinuzzi, A., Hurko, O. & Attardi, G. Marked replicative advantage of human mtDNA carrying a point mutation that causes the MELAS encephalomyopathy. *Proc Natl Acad Sci U S A* **89**, 11164-11168 (1992).
77. Dunbar, D. R., Moonie, P. A., Jacobs, H. T. & Holt, I. J. Different cellular backgrounds confer a marked advantage to either mutant or wild-type mitochondrial genomes. *Proc Natl Acad Sci U S A* **92**, 6562-6566 (1995).
78. Holt, I. J., Dunbar, D. R. & Jacobs, H. T. Behaviour of a population of partially duplicated mitochondrial DNA molecules in cell culture: segregation, maintenance and recombination dependent upon nuclear background. *Hum Mol Genet* **6**, 1251-1260 (1997).
79. Jenuth, J. P., Peterson, A. C., Fu, K. & Shoubridge, E. A. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nat Genet* **14**, 146-151 (1996).
80. Cree, L. M., Samuels, D. C., de Sousa Lopes, S. C., Rajasimha, H. K., Wonnapinij, P., Mann, J. R., Dahl, H. H. & Chinnery, P. F. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nat Genet* **40**, 249-254 (2008).
81. Jenuth, J. P., Peterson, A. C. & Shoubridge, E. A. Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. *Nat Genet* **16**, 93-95 (1997).
82. Malena, A., Loro, E., Di Re, M., Holt, I. J. & Vergani, L. Inhibition of mitochondrial fission favours mutant over wild-type mitochondrial DNA. *Hum Mol Genet* **18**, 3407-3416 (2009).
83. Suen, D. F., Narendra, D. P., Tanaka, A., Manfredi, G. & Youle, R. J. Parkin expression selects against a deleterious mtDNA mutation in heteroplasmic cybrid cells. *Proc Natl Acad Sci U S A* **107**, 11835-11840 (2010).
84. Jokinen, R., Marttinen, P., Sandell, H. K., Manninen, T., Teerenhovi, H., Wai, T., Teoli, D., Loredo-Osti, J. C., Shoubridge, E. A. & Battersby, B. J. Gimap3 regulates tissue-specific mitochondrial DNA segregation. *PLoS Genet* **6**,

e1001161 (2010).

85. Zenz, T., Roessner, A., Thomas, A., Frohling, S., Dohner, H., Calabretta, B. & Daheron, L. hIan5: the human ortholog to the rat Ian4/Iddm1/lyp is a new member of the Ian family that is overexpressed in B-cell lymphoid malignancies. *Genes Immun* **5**, 109-116 (2004).
86. Blanc, H. & Dujon, B. Replicator regions of the yeast mitochondrial DNA responsible for suppressiveness. *Proc Natl Acad Sci U S A* **77**, 3942-3946 (1980).
87. Lockshon, D., Zweifel, S. G., Freeman-Cook, L. L., Lorimer, H. E., Brewer, B. J. & Fangman, W. L. A role for recombination junctions in the segregation of mitochondrial DNA in yeast. *Cell* **81**, 947-955 (1995).
88. Hyvarinen, A. K., Pohjoismaki, J. L., Reyes, A., Wanrooij, S., Yasukawa, T., Karhunen, P. J., Spelbrink, J. N., Holt, I. J. & Jacobs, H. T. The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA. *Nucleic Acids Res* **35**, 6458-6474 (2007).
89. Moreno-Loshuertos, R., Acin-Perez, R., Fernandez-Silva, P., Movilla, N., Perez-Martos, A., Rodriguez de Cordoba, S., Gallardo, M. E. & Enriquez, J. A. Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants. *Nat Genet* **38**, 1261-1268 (2006).
90. Hori, A., Yoshida, M., Shibata, T. & Ling, F. Reactive oxygen species regulate DNA copy number in isolated yeast mitochondria by triggering recombination-mediated replication. *Nucleic Acids Res* **37**, 749-761 (2009).
91. Li, L. C., Okino, S. T., Zhao, H., Pookot, D., Place, R. F., Urakami, S., Enokida, H. & Dahiya, R. Small dsRNAs induce transcriptional activation in human cells. *Proc Natl Acad Sci U S A* **103,** 17337-17342 (2006).
92. Janowski, B. A., Younger, S. T., Hardy, D. B., Ram, R., Huffman, K. E. & Corey, D. R. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat Chem Biol* **3**, 166-173 (2007).
93. Fuste, J. M., Wanrooij, S., Jemt, E., Granycome, C. E., Cluett, T. J., Shi, Y., Atanassova, N., Holt, I. J., Gustafsson, C. M. & Falkenberg, M. Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol Cell* **37**, 67-78 (2010).
94. O'Gorman, S., Fox, D. T. & Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-1355 (1991).

95. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S. & Mesirov, J. P. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
96. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type five. *J Gen Virol* **36**, 59-74 (1977).
97. King, M. P. & Attardi, G. Isolation of human cell lines lacking mitochondrial DNA. *Methods Enzymol* **264**, 304-313 (1996).
98. Attardi, G., King. M. P., Chomyn, A., & Loguercio-Polosei. P. Novel genetic and molecular approaches to the study of mitochondrial biogenesis and mitochondrial diseases in human cells. *Progress in Neuropathology* (eds Sato, T., & DiMauro, S.) 75-92 (Raven, New York, 1991).
99. Bodnar, A. G., Cooper, J. M., Holt, I. J., Leonard, J. V. & Schapira, A. H. Nuclear complementation restores mtDNA levels in cultured cells from a patient with mtDNA depletion. *Am J Hum Genet* **53**, 663-669 (1993).
100. Martinus, R. D., Linnane, A. W. & Nagley, P. Growth of rho 0 human Namalwa cells lacking oxidative phosphorylation can be sustained by redox compounds potassium ferricyanide or coenzyme Q10 putatively acting through the plasma membrane oxidase. *Biochem Mol Biol Int* **31**, 997-1005 (1993).
101. Morais, R., Desjardins, P., Turmel, C. & Zinkewich-Peotti, K. Development and characterization of continuous avian cell lines depleted of mitochondrial DNA. *In Vitro Cell Dev Biol*. **24**, 649-58.
102. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166**, 557-580 (1983).
103. Medina, I., Carbonell, J., Pulido, L., Madeira, S. C., Goetz, S., Conesa, A., Tarraga, J., Pascual-Montano, A., Nogales-Cadenas, R., Santoyo, J., Garcia, F., Marba, M., Montaner, D. & Dopazo, J. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res* **38**, W210-213 (2010).
104. Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S. E., Walford, G. A., Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R., Carr, S. A. & Mootha, V. K. A mitochondrial

protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123 (2008).

105. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P. & Mesirov, J. P. GenePattern 2.0. *Nat Genet* **38**, 500-501 (2006).

106. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J. P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S. & Golub, T. R. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929-1935 (2006).

107. Lamb, J. The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer* **7**, 54-60 (2007).

108. Engreitz, J. M., Chen, R., Morgan, A. A., Dudley, J. T., Mallelwar, R. & Butte, A. J. ProfileChaser: searching microarray repositories based on genome-wide patterns of differential expression. *Bioinformatics* **27**, 3317-3318 (2011).

109. Chang, J. T. Deriving transcriptional programs and functional processes from gene expression databases. *Bioinformatics* **28**, 1122-1129 (2012).

110. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984 (2010).

111. Hofhaus, G., Johns, D. R., Hurko, O., Attardi, G. & Chomyn, A. Respiration and growth defects in transmitochondrial cell lines carrying the 11778 mutation associated with Leber´s hereditary optic neuropathy. *J Biol Chem* **271**, 13155-13161 (1996).

112. Bennett, C. M., Baird, A. A., Miller, M. B., Wolford, G. L. Neural Correlates of Interspecies Perspective Taking in the Post-Mortem Atlantic Salmon: An Argument For Multiple Comparisons Correction. Journal of Serendipitous and Unexpected Results. **1**, 1-5 (2010).

113. Smiraglia, D. J., Kulawiec, M., Bistulfi, G. L., Gupta, S. G. & Singh, K. K. A novel role for mitochondria in regulating epigenetic modification in the nucleus. *Cancer Biol Ther* **7**, 1182-1190 (2008).

114. de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. & Vogel, C. Global signatures of protein and mRNA expression levels. *Mol Biosyst* **5**, 1512-1526 (2009).

115. Komili, S. & Silver, P. A. Coupling and coordination in gene expression processes: a systems biology view. *Nat Rev Genet* **9**, 38-48 (2008).

116. Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R. & Hood, L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929-934 (2001).

117. Kondrashov, N., Pusic, A., Stumpf, C. R., Shimizu, K., Hsieh, A. C., Xue, S., Ishijima, J., Shiroishi, T. & Barna, M. Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* **145**, 383-397 (2011).

118. Holcik, M. & Sonenberg, N. Translational control in stress and apoptosis. *Nat Rev Mol Cell Biol* **6**, 318-327 (2005).

119. Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. & Selbach, M. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).

120. Wu, Z., Puigserver, P., Andersson, U., Zhang, C., Adelmant, G., Mootha, V., Troy, A., Cinti, S., Lowell, B., Scarpulla, R. C. & Spiegelman, B. M. Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-one. *Cell* **98**, 115-124 (1999).

121. Shi, L., Reid, L. H., Jones, W. D., Shippy, R., Warrington, J. A., Baker, S. C., Collins, P. J., de Longueville, F., Kawasaki, E. S., Lee, K. Y., Luo, Y., Sun, Y. A., Willey, J. C., Setterquist, R. A., Fischer, G. M., Tong, W., Dragan, Y. P., Dix, D. J., Frueh, F. W., Goodsaid, F. M., Herman, D., Jensen, R. V., Johnson, C. D., Lobenhofer, E. K., Puri, R. K., Schrf, U., Thierry-Mieg, J., Wang, C., Wilson, M., Wolber, P. K., Zhang, L., Amur, S., Bao, W., Barbacioru, C. C., Lucas, A. B., Bertholet, V., Boysen, C., Bromley, B., Brown, D., Brunner, A., Canales, R., Cao, X. M., Cebula, T. A., Chen, J. J., Cheng, J., Chu, T. M., Chudin, E., Corson, J., Corton, J. C., Croner, L. J., Davies, C., Davison, T. S., Delenstarr, G., Deng, X., Dorris, D., Eklund, A. C., Fan, X. H., Fang, H., Fulmer-Smentek, S., Fuscoe, J. C., Gallagher, K., Ge, W., Guo, L., Guo, X., Hager, J., Haje, P. K., Han, J., Han, T., Harbottle, H. C., Harris, S. C., Hatchwell, E., Hauser, C. A., Hester, S., Hong, H., Hurban, P., Jackson, S. A., Ji, H., Knight, C. R., Kuo, W. P., LeClerc, J. E., Levy, S., Li, Q. Z., Liu, C., Liu, Y., Lombardi, M. J., Ma, Y., Magnuson, S. R., Maqsodi, B., McDaniel, T., Mei, N., Myklebost, O., Ning, B., Novoradovskaya, N., Orr, M. S., Osborn, T. W., Papallo, A., Patterson, T. A., Perkins, R. G., Peters,

E. H., Peterson, R., Philips, K. L., Pine, P. S., Pusztai, L., Qian, F., Ren, H., Rosen, M., Rosenzweig, B. A., Samaha, R. R., Schena, M., Schroth, G. P., Shchegrova, S., Smith, D. D., Staedtler, F., Su, Z., Sun, H., Szallasi, Z., Tezak, Z., Thierry-Mieg, D., Thompson, K. L., Tikhonova, I., Turpaz, Y., Vallanat, B., Van, C., Walker, S. J., Wang, S. J., Wang, Y., Wolfinger, R., Wong, A., Wu, J., Xiao, C., Xie, Q., Xu, J., Yang, W., Zhong, S., Zong, Y. & Slikker, W. Jr. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* **24**, 1151-1161 (2006).

122. Guo, L., Lobenhofer, E. K., Wang, C., Shippy, R., Harris, S. C., Zhang, L., Mei, N., Chen, T., Herman, D., Goodsaid, F. M., Hurban, P., Phillips, K. L., Xu, J., Deng, X., Sun, Y. A., Tong, W., Dragan, Y. P. & Shi, L. Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol* **24**, 1162-1169 (2006).

123. Juel, C. & Halestrap, A. P. Lactate transport in skeletal muscle - role and regulation of the monocarboxylate transporter. *J Physiol* **517**, 633-642 (1999).

124. Narendra, D., Kane, L. A., Hauser, D. N., Fearnley, I. M. & Youle, R. J. p62/SQSTM1 is required for Parkin-induced mitochondrial clustering but not mitophagy; VDAC1 is dispensable for both. *Autophagy* **6**, 1090-1106.

125. Kamp, F., Exner, N., Lutz, A. K., Wender, N., Hegermann, J., Brunner, B., Nuscher, B., Bartels, T., Giese, A., Beyer, K., Eimer, S., Winklhofer, K. F. & Haass, C. Inhibition of mitochondrial fusion by α-synuclein is rescued by PINK1, Parkin and DJ-one. *EMBO J* **29,** 3571-3589.

126. Wang, H., Huang, S., Shou, J., Su, E. W., Onyia, J. E., Liao, B. & Li, S. Comparative analysis and integrative classification of NCI60 cell lines and primary tumors using gene expression profiling data. *BMC Genomics* **7**, 166 (2006).

127. Wek, R. C., Jiang, H. Y. & Anthony, T. G. Coping with stress: eIF2 kinases and translational control. *Biochem Soc Trans* **34**, 7-11 (2006).

128. Harding, H. P., Zhang, Y., Zeng, H., Novoa, I., Lu, P. D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., Stojdl, D. F., Bell, J. C., Hettmann, T., Leiden, J. M. & Ron, D. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell* **11**, 619-633 (2003).

129. Pelkmans, L. Using cell-to-cell variability--a new era in molecular biology. *Science* **336**, 425-426 (2012).

130. Gonzalez-Navajas, J. M., Lee, J., David, M. & Raz, E. Immunomodulatory functions of type I interferons. *Nat Rev Immunol* **12**, 125-135 (2012).

131. MacMicking, J. D. Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nat Rev Immunol* **12**, 367-382 (2012).

132. Kang, H. T. & Hwang, E. S. 2-Deoxyglucose: an anticancer and antiviral therapeutic, but not any more a low glucose mimetic. *Life Sci* **78**, 1392-1399 (2006).

133. Orvedahl, A., Sumpter, R., Jr., Xiao, G., Ng, A., Zou, Z., Tang, Y., Narimatsu, M., Gilpin, C., Sun, Q., Roth, M., Forst, C. V., Wrana, J. L., Zhang, Y. E., Luby-Phelps, K., Xavier, R. J., Xie, Y. & Levine, B. Image-based genome-wide siRNA screen identifies selective autophagy factors. *Nature* **480**, 113-117 (2011).

134. Nishikawa, M., Nishiguchi, S., Kioka, K., Tamori, A. & Inoue, M. Interferon reduces somatic mutation of mitochondrial DNA in liver tissues from chronic viral hepatitis patients. *J Viral Hepat* **12**, 494-498 (2005).

135. De Preter, K., De Brouwer, S., Van Maerken, T., Pattyn, F., Schramm, A., Eggert, A., Vandesompele, J. & Speleman, F. Meta-mining of neuroblastoma and neuroblast gene expression profiles reveals candidate therapeutic compounds. *Clin Cancer Res* **15**, 3690-3696 (2009).

136. Braconi, C., Meng, F., Swenson, E., Khrapenko, L., Huang, N. & Patel, T. Candidate therapeutic agents for hepatocellular cancer can be identified from phenotype-associated gene expression signatures. *Cancer* **115**, 3738-3748 (2009).

137. Hammarstrom, S. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* **9**, 67-81(1999).

138. Reeves, M. B., Davies, A. A., McSharry, B. P., Wilkinson, G. W. & Sinclair, J. H. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* **316**, 1345-1348 (2007).

139. Cunningham, J. T., Rodgers, J. T., Arlow, D. H., Vazquez, F., Mootha, V. K. & Puigserver, P. mTOR controls mitochondrial oxidative function through a YY1-PGC-1alpha transcriptional complex. *Nature* **450**, 736-740 (2007).

140. Kiyono, K., Suzuki, H. I., Matsuyama, H., Morishita, Y., Komuro, A., Kano, M. R., Sugimoto, K. & Miyazono, K. Autophagy is activated by TGF-beta and potentiates TGF-beta-mediated growth inhibition in human hepatocellular carcinoma cells. *Cancer Res* **69**, 8844-8852 (2009).

141. Schon, E. A. & Gilkerson, R. W. Functional complementation of mitochondrial DNAs: mobilizing mitochondrial genetics against dysfunction. *Biochim Biophys Acta* **1800**, 245-249 (2010).

142. Wang, C. & Youle, R. J. The role of mitochondria in apoptosis*. *Annu Rev Genet* **43**, 95-118 (2009).

143. Geissler, A., Krimmer, T., Bomer, U., Guiard, B., Rassow, J. & Pfanner, N. Membrane potential-driven protein import into mitochondria. The sorting sequence of cytochrome b(2) modulates the deltapsi-dependence of translocation of the matrix-targeting sequence. *Mol Biol Cell* **11**, 3977-3991 (2000).

144. Xi, H., Barredo, J. C., Merchan, J. R. & Lampidis, T. J. Endoplasmic reticulum stress induced by 2-deoxyglucose but not glucose starvation activates AMPK through CaMKKbeta leading to autophagy. *Biochem Pharmacol* **85**, 1463-1477 (2013).

145. Wang, Q., Liang, B., Shirwany, N. A. & Zou, M. H. 2-Deoxy-D-glucose treatment of endothelial cells induces autophagy by reactive oxygen species-mediated activation of the AMP-activated protein kinase. *PLoS One* **6**, e17234 (2011).

146. Schroder, M. & Kaufman, R. J. ER stress and the unfolded protein response. *Mutat Res* **569**, 29-63 (2005).

147. Brown, J. Effects of 2-deoxyglucose on carbohydrate metabolism: review of the literature and studies in the rat. *Metabolism* **11**, 1098-1112 (1962).

148. Weiss, R. A. The discovery of endogenous retroviruses. *Retrovirology* **3**, 67 (2006).

149. Stoye, J. P. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat Rev Microbiol* **10**, 395-406 (2012).

150. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273-287 (2011).

151. Reiche, J., Pauli, G. & Ellerbrok, H. Differential expression of human endogenous retrovirus K transcripts in primary human melanocytes and melanoma cell lines after UV irradiation. *Melanoma Res* **20**, 435-440 (2010).

152. Lee, J. R., Ahn, K., Kim, Y. J., Jung, Y. D. & Kim, H. S. Radiation-induced human endogenous retrovirus (HERV)-R env gene expression by epigenetic control. *Radiat Res* **178**, 379-384 (2012).

153. Nellaker, C., Yao, Y., Jones-Brando, L., Mallet, F., Yolken, R. H. & Karlsson, H. Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology* **3**, 44 (2006).

154. Yarilina, A., Park-Min, K. H., Antoniv, T., Hu, X. & Ivashkiv, L. B. TNF activates an IRF1-dependent autocrine loop leading to sustained expression of chemokines and STAT1-dependent type I interferon-response genes. *Nat Immunol* **9**, 378-387 (2008).

155. Qiu, J., Hu, X., Nesic, O., Grafe, M. R., Rassin, D. K., Wood, T. G. & Perez-Polo, J. R. Effects of NF-kappaB oligonucleotide "decoys" on gene expression in P7 rat hippocampus after hypoxia/ischemia. *J Neurosci Res* **77**, 108-18 (2004).

156. Hu, J., Nudelman, G., Shimoni, Y., Kumar, M., Ding, Y., Lopez, C., Hayot, F., Wetmur, J. G. & Sealfon, S. C. Role of cell-to-cell variability in activating a positive feedback antiviral response in human dendritic cells. *PLoS One* **6**, e16614 (2011).

157. Raap, A. K., Jahangir Tafrechi, R. S., van de Rijke, F. M., Pyle, A., Wahlby, C., Szuhai, K., Ravelli, R. B., de Coo, R. F., Rajasimha, H. K., Nilsson, M., Chinnery, P. F., Samuels, D. C. & Janssen, G. M. Non-random mtDNA segregation patterns indicate a metastable heteroplasmic segregation unit in m.3243A>G cybrid cells. *PLoS One* **7**, e52080 (2012).

158. Perkins, G. A., Tjong, J., Brown, J. M., Poquiz, P. H., Scott, R. T., Kolson, D. R., Ellisman, M. H. & Spirou, G. A. The micro-architecture of mitochondria at active zones: electron tomography reveals novel anchoring scaffolds and cristae structured for high-rate metabolism. *J Neurosci* **30**, 1015-1026 (2010).

159. Dunbar, D. R., Moonie, P. A., Zeviani, M. & Holt, I. J. Complex I deficiency is associated with 3243G:C mitochondrial DNA in osteosarcoma cell cybrids. *Hum Mol Genet* **5**, 123-129 (1996).

160. Hamalainen, R., H. Manninen, T., Koivumaki, H., Kislin, M., Otonkoski, T., Suomalainen, A. Tissue- and cell-type-specific manifestations of heteroplasmic mtDNA 3243A>G mutation in human induced pluripotent stem cell-derived disease model. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 3622-3630 (2013).

6 Appendix

MITOPHAGY	BIOGENESIS	DYNAMICS	THRESHOLD	NUCLEOID	HEAT-SHOCK	PARKINSONS
AMBRA1	CREB1	AFG3L2	ATPIF1	ATAD3A	HSPA9	AMCSD
ATF4	GABPA	DNM1L	PARK2	ATAD3B	HSPD1	APOE
ATG12	LRPPRC	FIS1	PARL	NIPSNAP1	HSPE1	ATP13A2
ATG4A	MEF2C	MFF	PINK1	GBAS		BDNF
ATG4B	MPV17	MFN1	SLC25A14	NF2		BMP4
ATG4C	MPV17L	MFN2	SLC25A27	POLG2		BST1
ATG4D	MPV17L2	MTFP1	SQSTM1			C20orf82
ATG5	MTOR	MUL1	UCP1			CALB1
ATG7	MYC	OPA1	UCP2			CCDC62
BAD	NFE2L1	SMCR7	UCP3			CYP17A1
BBC3	NFE2L2	SMCR7L				DBC1
BCL2	POLG1	SNCA				DRD2
BCL2L1	POLRMT	SPG7				DRD3
BCL2L11	PPARGC1A	YME1L1				FARP1
BCL2L2	PPARGC1B					FGF20
BECN1	PRC					GAK
BIK	PRKAA1					GBA
BNIP3L	PRKAA2					GLIS1
CISD1	SFXN					HIP1R
DAPK1	SIRT1					HLA-DRB5
DRAM1	SIRT2					ITGA8
DRAM2	SIRT3					LOC100505836
GABARAP	SIRT4					LRP8
GDAP1	SIRT5					LRRK2
HDAC6	SIRT6					MAOB
MAP1B	SIRT7					MAPT
MAP1LC3A	TFAM					MCCC1
MAP1LC3B	TFB2M					MED13
MCL1	DDX58					MTHFR
MOSC1						NSF
MOSC2						NUCKS1
NBR1						PARK16
PARK2						PARK7
PARL						PDXK
PGAM5						PM20D1
PINK1						RAB25
PMAIP1						RAI1
PTEN						RIT2
PTEN1P						SCARB2
RHOT1						SLC25A48
RHOT2						SLC41A1
SH3GLB1						SLCO3A1
SQSTM1						SNCA
STK11						SNCB
TOMM70A						SNCG
TRAK1						SOD2
TRAK2						SREBF1
TRAP1						STK39
TSC1						SYT11
TSC2						SYT4
ULK1						TAS1R2
ULK2						UCHL1
UVRAG						UNC13B
VCP						USP24
WDFY3						USP25
						USP40
						WNT3

Table 6.1 Genes with a link to mitochondria based on the cumulative domain expertise of group members.

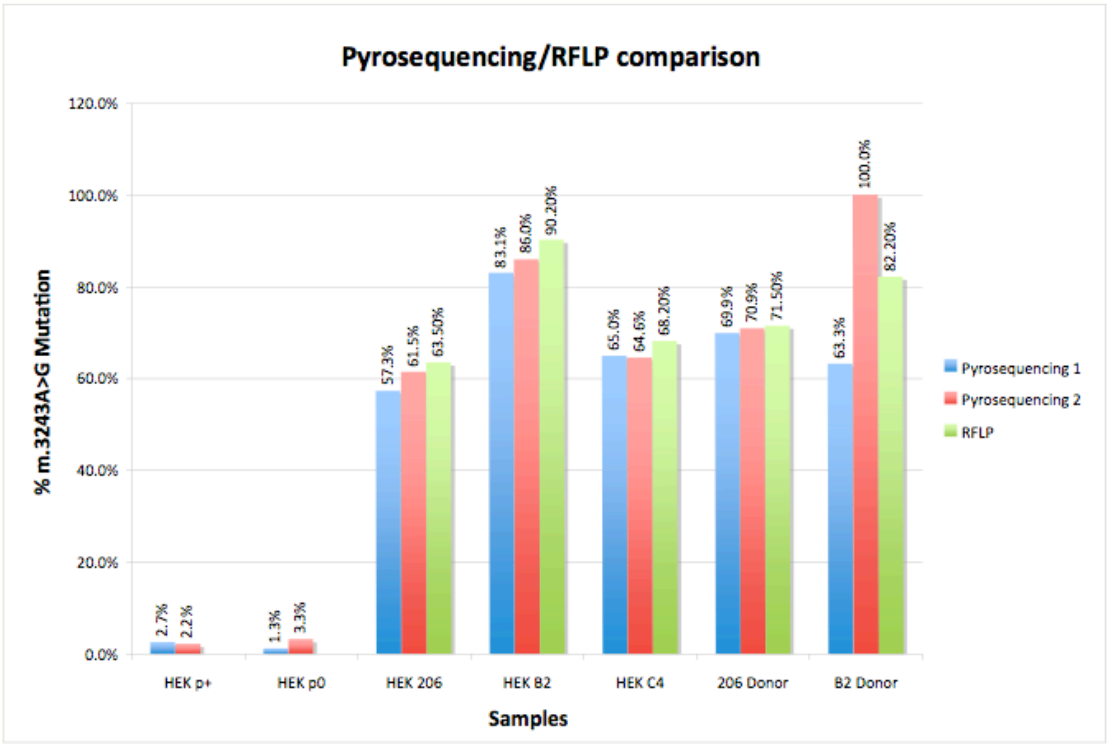


Figure 6.1. m.3243A>G heteroplasmy of seven samples measured by pyrosequencing and simulataneous measurement by RFLP in five samples.