Is there selection against somatic mtDNA mutations?

Colby Luke Samstag

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Leo J. Pallanck, Chair

Philip P. Green

Matt R. Kaeberlein

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Is there selection against somatic mtDNA mutations?

Colby Luke Samstag

Chair of the Supervisory Committee:
Professor Leo J. Pallanck
Department of Genome Sciences

Mitochondrial DNA (mtDNA) mutations are implicated in a number of human diseases, ranging from rare maternally inherited neuromuscular disorders of infancy to late onset neurodegenerative disease and cancer. While much is known of the mechanisms by which damaged mitochondria are selectively degraded in the lysosome through a mitochondrial-specific form of autophagy termed mitophagy, it remains unclear whether mitophagy limits the accumulation of harmful mtDNA mutations. To address this matter, we created a *Drosophila* strain that expresses a mitochondrial DNA polymerase lacking proofreading ability. This strain has up to 55-fold elevation in mtDNA point mutation frequency and a number of accompanying phenotypes, including reduced lifespan, a locomotor deficit, and mitochondrial dysfunction. I hypothesized that mitophagy would selectively remove harmful mutations, and that less

deleterious mutations would persist in mtDNA mutator flies. Surprisingly, my analysis revealed that deleterious mutations were overrepresented in mutator flies. I propose two models to explain this finding: 1) Mitochondria that bear deleterious mtDNA mutations may undergo less oxidative damage, thereby evading mitophagy. 2) Cells that stochastically acquire a deleterious mtDNA mutation induce compensatory mitochondrial biogenesis, leading to the amplification of mutant genomes. The overrepresentation of deleterious mtDNA mutations present in mutator flies suggests that mitophagy is either incapable of selecting against harmful mtDNA mutations or is insufficient to overcome a positive selective force acting in favor of harmful mtDNA mutations. To test whether mitophagy combats the accumulation of deleterious mutations, I employed a novel proteomic approach developed in our lab to identify the turnover rate of cellular proteins. I hypothesized that mutator flies would exhibit faster mitochondrial protein turnover to clear mitochondria damaged by mutant mtDNA, but instead discovered that the rates of mitochondrial protein turnover were unchanged between wild-type and mutator flies. I also tested whether genetic manipulations including knockout and ubiquitous overexpression of the mitophagy-promoting factor *parkin* influence selection against mtDNA mutations, but I found that genetic perturbations to *parkin* did not alter the frequency, spectrum, or pathogenicity of mtDNA mutations. Taken together, my findings suggest that mitophagy does not select against deleterious mtDNA point mutations in somatic tissues.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I have been very fortunate to have many people who have supported my growth and exploration. I first want to thank my mentor, Leo Pallanck, for taking a chance on me and sticking with me and my project through its many twists and turns. Leo possesses an extraordinarily sharp scientific mind, making our conversations about science consistently exhilarating. Leo and I have shared countless thoughtful conversations not only about science, but also politics, history, and philosophy, and I will miss stopping by to hear his perspective. Thank you for always treating me as your colleague.

Thank you to the members of the Pallanck lab, who patiently answered my endless barrage of questions, and to the MCB and Genome Sciences programs for their support and resources. Thank you to Peter Rabinovitch and everyone who makes the Genetic Approaches to Aging Training Grant possible; you have provided a wonderful training environment. Thank you to members of my committee—Phil Green, Celeste Berg, Matt Kaeberlein, and Lawrence Loeb—for your perceptiveness and guidance. Thanks also to Scott Kennedy for your expertise in sequencing. We had many heated discussions, but I am certain that our work came out better as a result. My sincerest gratitude to my colleague and collaborator Simonetta Andreazza—I have been incredibly fortunate to find someone as mystified by mitochondria as I am. Our collaboration helped propel me through the stagnations of grad school, and I am proud of the work we accomplished together. I also want to recognize my good friend Mark Chaisson for introducing me to the computational and statistical techniques that transformed my entire project.

I am extraordinarily grateful to have benefited from some outstanding teachers and mentors who cultivated my curiosity. I am eternally indebted to Steve Scoville and Kristen

Kurzawski in high school, and Robert Rescorla and Ted Abel at the University of Pennsylvania for being instrumental to my academic growth. Thank you to my late friend and mentor, Grant Krow—I miss you dearly. I also want to recognize my high school art instructors at the Manchester Craftsmen's Guild, namely Jamie Matthews, Justin Mazzei, and Jeff Guerrero, who helped teach me how to harness my creative energy.

Thank you to my friends from graduate school and elsewhere. You have provided me with an abundance of rich memories and stories. It's been a crazy ride.

While my three brothers are each very unique, I want to recognize Tyler for not only being intellectually insightful, but also for lending words of wisdom in tough times. Moreover, I owe my successes to my parents, Mom and Dad, who convinced me early on that I could achieve anything I set my mind to. Life has taken me to many unexpected places, but my parents have always stood by me through every decision. Thanks for having faith in me and letting me get back on the bike, even when it wasn't a good idea. You are remarkable parents and grandparents.

I am forever grateful for my two wonderful kids, Renee and Luke. Thank you for making me significantly less productive as I finished graduate school, and in the process, convincing me to manage my time better. As proud as I am of my scientific achievements, you two are undoubtedly my finest work.

Finally, I could never express my gratitude and appreciation for my incredible wife, Marie Y. Davis MD PhD, for her unwavering love and support. No matter how much time I waste listening to the same off-key songs, devouring books and articles, or fixating on my obsession of the week, Marie continues to stay by my side. Marie is not only an unbelievable wife and mother, but she is an exceptionally talented scientist and physician, unparalleled in her work ethic, and the most caring person I have ever met. She is my inspiration.

Chapter 1. **Introduction**

1.1 ***Background***

Mitochondria are the evolutionary descendants of a bacterial endosymbiont that arose ~1.5 years ago (*1, 2*). This ancestry is evident in the many features mitochondria share with bacteria, including a relatively porous outer membrane, a highly impermeable inner membrane, a circular genome, and a similar mechanism of protein synthesis (*1, 2*). Although the benefit that drove this symbiosis event remains a matter of debate, mitochondria now perform a variety of indispensable eukaryotic functions, including the production of up to 90% of a cell's ATP through oxidative phosphorylation, as well as critical roles in Ca2+ buffering and apoptosis (*3-5*). Mitochondria are not discrete, static organelles as they are often depicted. Rather, in most cell types mitochondria are highly motile and form extensively interconnected networks. Mitochondria are also highly dynamic, undergoing continual cycles of fission and fusion. The dynamic nature of the mitochondrial network is mediated by a host of evolutionarily conserved factors, and promotes the mixing of mitochondrial proteins and mRNAs (*6*).

Most bacterial genomes consist of thousands of genes. By contrast, the mitochondrial genome (mtDNA) of most animals consists of merely 37 genes, encoding 13 essential subunits of the electron transport chain complexes as well as rRNAs and tRNAs required to translate these proteins (*7*). The small mitochondrial genome is a consequence of substantial gene flow from the original endosymbiont genome to the nucleus (*2*). Most of the approximately 1000-2000 proteins of the mitochondrial proteome are now encoded by genes that reside in the nucleus (*2*). Nuclear-encoded mitochondrial proteins are translated in the cytosol and contain targeting information that directs them to the mitochondrion. While it is unclear why a small subset of mitochondrial proteins

remains encoded within mtDNA, a leading hypothesis is that these proteins are critical for mitochondrial function yet are too hydrophobic for efficient import into mitochondria (*8*). The complexes of the electron transport chain are comprised of subunits of both nuclear and mitochondrial origin, and their stoichiometry and assembly must be coordinated through regulated import by mitochondrial translocases as well as nuclear-encoded assembly factors (*9*).

The mitochondrial respiratory chain is comprised of five complexes composed of between four subunits (Complex II) and 44 subunits (Complex I). All respiratory chain complexes except Complex II are comprised of at least one mtDNA-encoded subunit (*9*). The electron transport chain functions through a process known as chemiosmosis (*10*). Complexes I, II, and IV maintain a high concentration of protons in the intermembrane space by pumping protons from the mitochondrial matrix across the inner mitochondrial membrane and into the intermembrane space. These complexes harness the energy from a series of electron reducing reactions to push protons against their electrochemical gradient into the intermembrane space. Protons flow back into the inner matrix by passing through the turbine-like ATP synthase (Complex V), which uses proton flow to phosphorylate ADP and produce ATP (*10*). While this process is very efficient at generating energy, electrons can "slip" during reactions at Complexes I and III, forming damaging superoxide radicals (*11*). The reactive oxygen species (ROS) produced as a byproduct of oxidative phosphorylation damage mitochondrial proteins, lipids and DNA.

The assembly and function of mitochondria poses an array of challenges to the cell, and accumulated damage to mitochondria is believed to be a major cause of aging and age-related diseases including Parkinson's disease, Alzheimer's Disease, and cancer (*12-17*). There are a number of reasons mitochondria are prone to accumulating such damage. First, respiration forms damaging ROS as a byproduct of electron transfer. Second, while the mitochondrial genome

encodes proteins essential for the functioning of mitochondria, it is also prone to accumulating mutations. Finally, the stoichiometry of nuclear- and mitochondrially- encoded proteins must be tightly coordinated. In the next section, I will provide an overview of the defenses the cell has evolved to maintain mitochondrial health in the face of these challenges.

1.2 *Mechanisms of mitochondrial quality control*

Due to the critical importance of mitochondria to the cell as well as their susceptibility to molecular damage, mitochondria and their eukaryotic host have evolved a variety of quality control adaptations. There are broadly three general strategies that cells employ to oppose mitochondrial damage (Figure 1.1).

Damage Prevention

One strategy that cells employ to maintain mitochondrial health is to prevent damage from occurring in the first place (Figure 1.1). This strategy is perhaps best exemplified by the array of antioxidant stress response systems that are designed to scavenge ROS before they can do damage. Two antioxidant stress response factors that are believed to play a major role in mitochondria are the mitochondria superoxide dismutase and the glutathione peroxidases. Superoxide anions produced from electron slippage during mitochondrial respiration are catalytically converted by the mitochondrial superoxide dismutase (*SOD2*) into hydrogen peroxide and oxygen. Hydrogen peroxide is deactivated in the cytosol by the enzyme catalase, and in the mitochondria by glutathione peroxidase-1 (*Gpx1*) (*18, 19*). However, superoxide anions that are not detoxified by *SOD2*, as well as excess hydrogen peroxide can form a number of other radical species, including peroxynitrite and the highly reactive hydroxyl radicals. Hydroxyl radicals cannot be enzymatically

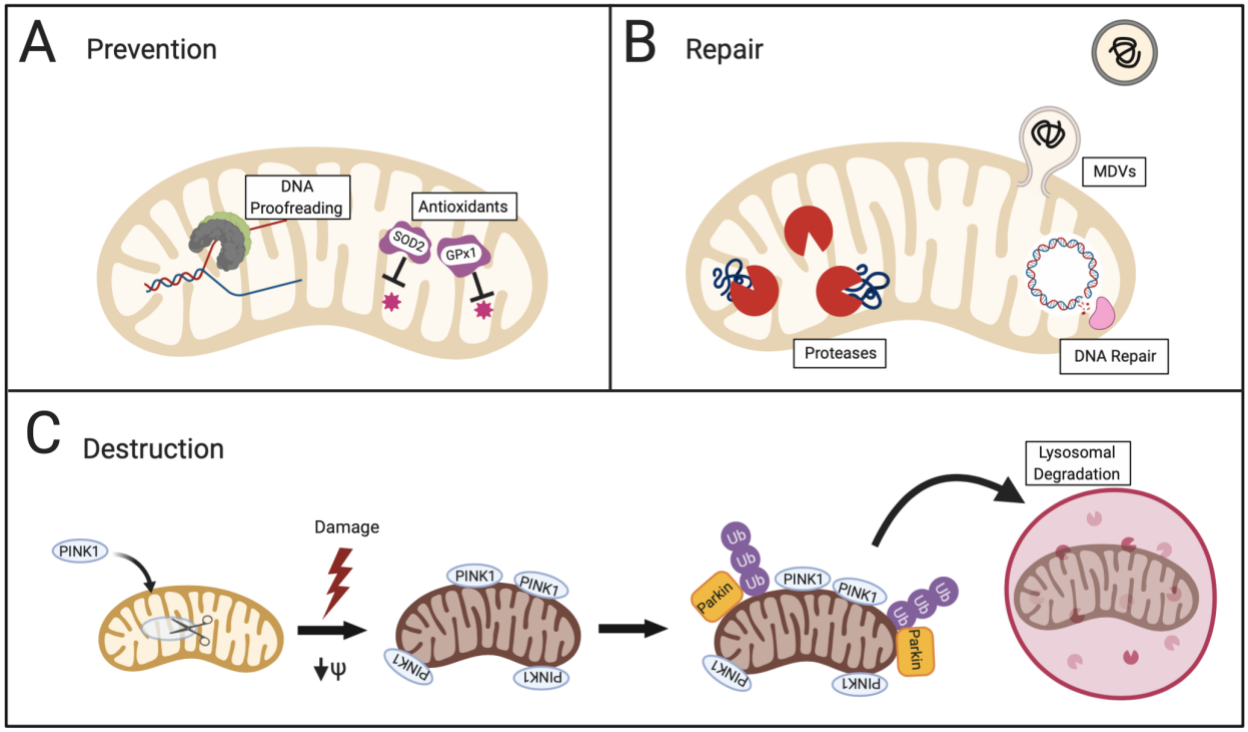


Figure 1.1. Mitochondrial maintenance and quality control pathways. (A) Damage prevention mechanisms include polymerase proofreading that prevents the incorporation of new mutations during mtDNA replication, and antioxidant enzymes that eliminate damaging free radicals generated during oxidative respiration before they can react with mitochondrial lipids, proteins, and DNA. **(B)** Mitochondrial repair pathways include a host of proteases that degrade dysfunctional mitochondrial proteins, multiple DNA repair pathways maintain mtDNA integrity, and mitochondrial derived vesicles (MDV) that package damaged proteins for lysosomal degradation. **(C)** When a mitochondrion becomes damaged beyond repair, it can be targeted for wholesale destruction through mitophagy. In canonical mitophagy, PINK1 selectively accumulates on the outer membrane of damaged mitochondria that can no longer maintain a membrane potential. PINK1 recruits and phosphorylates the E3-Ubiquitin ligase Parkin, which poly-ubiquitinates the mitochondrion to target it for lysosomal degradation.

reduced and are the major source of ROS-mediated damage to the cell (*11, 20, 21*). Knockout of either *SOD2* or *Gpx1* genes is highly detrimental, leading to increased oxidative stress, decreased respiration, and early lethality (*22*). Additionally, targeting the cytosolic hydrogen peroxide scavenger catalase to mitochondria results in extended lifespan and reduced mtDNA mutations (*23*). Curiously overexpression of a key mitochondrially-localized free radical scavenger, superoxide dismutase, does not appear to increase lifespan (*24*), and there have been broadly mixed results following genetic perturbations of other antioxidant enzymes (*25-27*).

Another damage prevention system is the prevention of new mutations arising during mitochondria replication. Because mtDNA replication is uncoupled from the cell cycle, mtDNA is replicated throughout the lifespan and so is subject to many rounds of replication (*12*). Polymerase γ (PolG), the sole mitochondrial DNA polymerase, contains an exonuclease domain that proofreads throughout replication of mtDNA, conferring a 20-fold increase in replication fidelity (*28*). This exonuclease domain stalls replication upon nucleotide mis-incorporation, allowing for the excision of the incorrect nucleotide. In this way, the exonuclease domain prevents misincorporation events that arise during replication from becoming fixed as mutations in mtDNA. Inactivation of the exonuclease domain of PolG has become a prominent model to study mtDNA mutations that I will discuss in greater detail later in this introduction.

Damage correction

Another strategy that cells use to prevent the accumulation of damaged mitochondria is damage correction—to repair or eliminate the damaged mitochondrial component (Figure 1.1B). Damage correction actually constitutes a broad array of individual quality control systems aimed at damaged proteins, lipids and DNA, although my focus in this introduction is the quality control

systems aimed at proteins and DNA. One important mitochondrial protein quality control system consists of mitochondrial proteases. In particular, there are five major degradative proteases of the AAA+ family that localize to different mitochondrial subcompartments where they degrade oxidatively damaged and misfolded protein (*29*). There are also multiple mitochondrial chaperones, the best characterized of which are the Hsp60 and Hsp70 family of chaperones that promote mitochondrial protein folding, as well as unfolding of protein aggregates for degradation (*30*). In addition to mitochondrial proteases, the proteasome has also been shown to degrade mitochondrial proteins. Although there is little evidence that the proteasome resides within the mitochondrion, it is well established that the proteasome can degrade proteins on the outer surface of the mitochondrion under conditions of stress (*31*). There is also evidence that at least some proteins that reside within the mitochondrial matrix can be exported to the cytoplasm and degraded by the proteasome in a manner that crudely resembles how misfolded proteins in the ER lumen are exported and degraded in the cytoplasm (*32*).

The excess accumulation of misfolded mitochondrial proteins can also trigger induction of the mitochondrial unfolded stress response (UPR^{mt}), a pathway whereby dysfunctional mitochondria coordinate with the nucleus to upregulate transcription of mitochondrial proteases and chaperones. The UPR^{mt} was first discovered in mammalian cell culture, when it was observed that mitochondria upregulated these protein quality control pathways in response to administration of ethidium bromide, an agent that depletes mtDNA and caused imbalance between mitochondrially- and nuclear- encoded electron transport chain complex subunits (*33*). The molecular mechanisms of the UPR^{mt} are best understood in *C. elegans*, where the pathway is primarily mediated by the dually localized transcription factor ATFS-1 (*34*). ATFS-1 is a transcription factor that alters the expression of ~500 genes, including reducing transcription of

nuclear-encoded respiratory chain subunits and upregulating mitochondrial chaperones and proteases (*34, 35*). ATFS-1 has an N-terminal targeting sequence directing it to mitochondria. However, import into mitochondria depends on mitochondrial membrane potential. Protein misfolding can impair the membrane potential and block import, in which case a nuclear localization signal in ATFS-1 directs it to the nucleus and upregulates transcription of mitochondrial chaperones and proteases. The UPR^{mt} has been shown to act in *C. elegans* in response to mitochondrial unfolded protein stress, caused by either the accumulation of misfolded mitochondrial proteins or imbalances to the stoichiometry of nuclear- and mitochondrially-encoded electron transport chain subunits. A number of putative mammalian homologs of ATFS-1 have recently been identified, suggesting mitochondrial-nuclear communication is an essential stress response conserved in vertebrates (*36, 37*).

Accumulating evidence also indicates that mitochondria possess a vesicular pathway—Mitochondrial Derived Vesicles (MDVs)—by which damaged mitochondrial proteins can be selectively packaged into vesicles that bud from the surface of mitochondria and are subsequently trafficked to the lysosome for degradation (*38*). It appears that vesicles have many different cargo and targets, some of which carry oxidized proteins to the lysosome for destruction. The formation of these vesicles is dependent upon PINK1 and Parkin, essential components of the mitochondrial autophagy pathway (see *Mitochondrial Destruction* below), but it is unclear how PINK1 and Parkin regulate vesicle formation (*39*).

Mitochondria are also host to an assortment of DNA repair mechanisms necessary to maintain the integrity of the mitochondrial genome. Due to the complexity of the pathways involved, I will only highlight the existence of the major pathways in this introduction (see *Alexeyev et. al, 2013* (*40*) for a comprehensive review). Following the observation that

mitochondria were unable to repair mutations induced by exposure to UV-radiation, it was initially believed that mitochondria lack DNA repair systems analogous to those that operate in the nucleus (*41*). However, it has since been revealed that mitochondria express some form of nearly every DNA lesion repair pathway, including repair mechanisms for single and double strand break repair, short- and long-patch base excision repair, and DNA mismatch repair (*40*). It now appears that nucleotide excision repair, the pathway responsible for correcting UV-introduced mutations, is one of the few known DNA repair mechanisms lacking in mitochondria. In addition to mtDNA repair, there is also evidence that mitochondria can selectively degrade mtDNA that has sustained extensive damage, such as through oxidative lesions or double strand breaks (*42*), through an unknown, autophagy-independent mechanism.

Mitochondrial destruction

A third strategy that cells use to prevent the accumulation of damaged mitochondria is to selectively target the damaged mitochondria for degradation (Figure 1.1C). Ultrastructural studies performed in the 1950s provided the first evidence of mitochondrial autophagy by capturing images of entire mitochondria contained within lysosomes (*43, 44*). While subsequent work led to the identification of evolutionarily conserved factors that promote autophagy and demonstrated that mitochondria are degraded in response to starvation and other cellular signals, it was only discovered relatively recently that cells could selectively detect and degrade damaged mitochondria. Studies of two genes associated with autosomal recessive parkinsonism, *PINK1* and *parkin*, have revealed that these genes encode components of a mitochondrial quality control system that promotes the selective degradation of damaged mitochondria through a mitochondria-selective form of autophagy termed mitophagy (*45, 46*).

The *PINK1* gene encodes a mitochondrially targeted serine/threonine kinase, and *parkin* encodes a cytosolic E3 ubiquitin-protein ligase (*47*). Genetic studies of *PINK1* and *parkin* in the fruit fly *Drosophila* provided the first evidence that these proteins act in a common pathway to promote mitochondrial integrity (*48-50*). Mutations in the *Drosophila PINK1* and *parkin* genes result in the accumulation of enlarged mitochondria, followed by the degeneration of flight muscle, sperm, and dopaminergic neurons. The finding that *Drosophila PINK1* and *parkin* mutants accumulated enlarged mitochondria led our lab and others to test the hypothesis that the PINK1-Parkin pathway regulates mitochondrial morphology (*51-54*).

Mitochondrial fusion is mediated primarily through evolutionarily conserved factors known as mitofusins, a family of GTPases that joins two mitochondria by tethering their outer membranes (*55*). We and others found that Parkin targets Mitofusin for ubiquitin-mediated turnover (*51-54*), leading us to propose that PINK1 and Parkin also degrade Mitofusin to promote the isolation and degradation of damaged segments of the mitochondrial network. Subsequent experiments primarily in vertebrate cell culture validated this hypothesis and clarified the mechanism of action of this pathway. Together, these studies showed that loss of mitochondrial membrane potential (depolarization) leads to accumulation of PINK1 on the mitochondrial outer membrane, triggering recruitment of Parkin to the mitochondria. Parkin then ubiquitinates proteins on the outer mitochondrial membrane, leading to the isolation and eventual degradation of the depolarized mitochondrion in the lysosome.

While studies performed *in vitro* suggest quality control pathways including mitophagy can reduce levels of mutant mtDNA (*56*), it remains contentious whether such mechanisms eliminate mutant mtDNA within somatic tissues. Before discussing mitophagy's role in mtDNA selection further, I will first examine the reasons mtDNA mutations accumulate in somatic tissues.

1.3 *mtDNA acquires mutations throughout the lifespan*

mtDNA has a high rate of mtDNA mutations, exceeding that of the nuclear genome by about tenfold (*57*). Unlike the nuclear genome, mtDNA is not compacted by histones. Instead, mtDNA is packed into an assembly of proteins and DNA known as a nucleoid (*58*). Each molecule of mtDNA is coated with many molecules of the protein Transcription factor A, mitochondrial (TFAM), which constricts the mtDNA and regulates its accessibility for transcription and translation (*59*). It was previously believed that the high rate of mtDNA mutations was a consequence of its lack of histone protection and its proximity to metabolically derived reactive ROS (*26*). However, recent experiments demonstrate that mtDNA shows robust resistance *in vivo* to damage from exogenously-administered oxidizing agents (*60*), and analyses of the spectrum of mutations indicates that ROS-mediated damage plays a minor role in the accumulation of mtDNA mutations (*61, 62*). It is now believed that errors that arise during replication are the predominant source of mtDNA mutations (*61, 62*).

Due to inherited and acquired mutations, an organism's pool of mtDNA can be highly heterogenous (*63, 64*). Within every somatic cell resides hundreds to thousands of copies of mtDNA. Mutations that are present in every copy of mtDNA within a cell are referred to as *homoplasmic*. However, mutations typically share residence with wild-type copies of the genome within the same cell, resulting in a mixed mtDNA population referred to as *heteroplasmy*. In maternally inherited mitochondrial diseases, single mutations may be inherited at high abundance, resulting in mitochondrial dysfunction and cellular pathologies. Studies of cultured cells suggest that deleterious mtDNA mutations do not present pathogenic effects until mutant heteroplasmy rises above a threshold (Figure 1.2), but this threshold can vary from 35-90% depending on the nature of the mutation and the cell type in which the mutation resides (*65-68*).

Cells may be resistant to phenotypic consequences of mutations at low frequency because they benefit from *functional complementation*—the mixing of gene products through the mitochondrial reticulum. Complementation through fusion and fission provides mixing of mtDNA-encoded tRNAs, mRNAs, and proteins (Figure 1.3), and is believed to protect the cell from the effects of damaging mtDNA mutations until they reach high frequency (Figure 1.2)(*69*). It had previously been thought that the constriction of mtDNA into nucleoids provided another mechanism for complementation. Nucleoids were believed to contain multiple copies of mtDNA and exchange mtDNA with one another during mitochondrial fission-fusion, but this view has recently been challenged (*70*-*72*). Superresolution microscopy reveals that most nucleoids contain just one mtDNA molecule (*73*), are distributed nonrandomly throughout mitochondria (*74*, *75*), and do not exchange mtDNA with one another, so it now appears that complementation is result of product mixing or exchange of entire nucleoids.

By contrast, somatically acquired mutations arise sporadically throughout the body, and so individual mutations accumulate at relatively low heteroplasmy (*64*). The physiological consequences of acquiring diverse low frequency heteroplasmies are still poorly understood. Despite the fact that individual mutations generally remain at heteroplasmy levels below the thresholds observed in maternally inherited mitochondrial diseases, the accumulation of somatic mtDNA mutations over the lifetime has been hypothesized to contribute to cellular aging (*76*). Furthermore, increased abundance of mtDNA mutations has been observed in age-related pathologies and disease, including decreased metabolic efficiency, decreased cardiac function, and neurodegeneration. However, recent work suggests that mtDNA mutations accumulate over the lifetime even in healthy individuals, calling into question their physiologic relevance (*77*). Additionally, mouse models with elevated mtDNA deletions do not show progeroid phenotypes

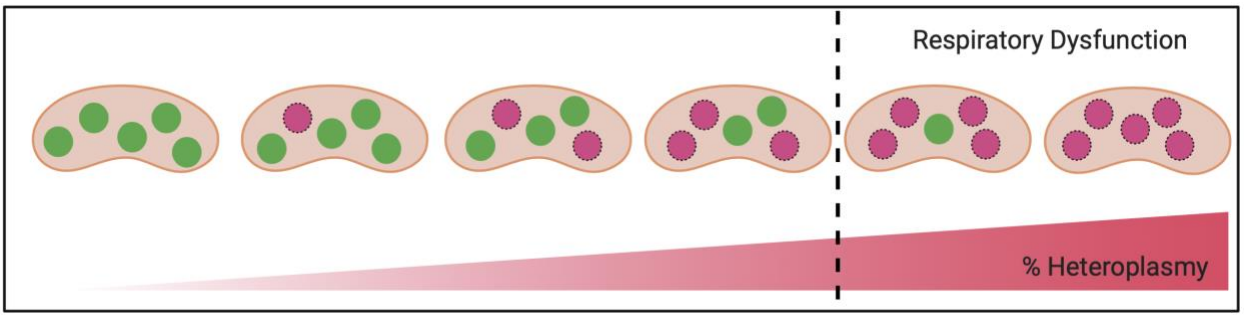


Figure 1.2. Levels of heteroplasmy correlate with the severity of disease. Due to functional complementation, cells function without deficits in oxidative phosphorylation or clinical presentation until reaching high heteroplasmy. Wild-type and mutant mtDNA are depicted in green and red, respectively. Adapted from (*78*)

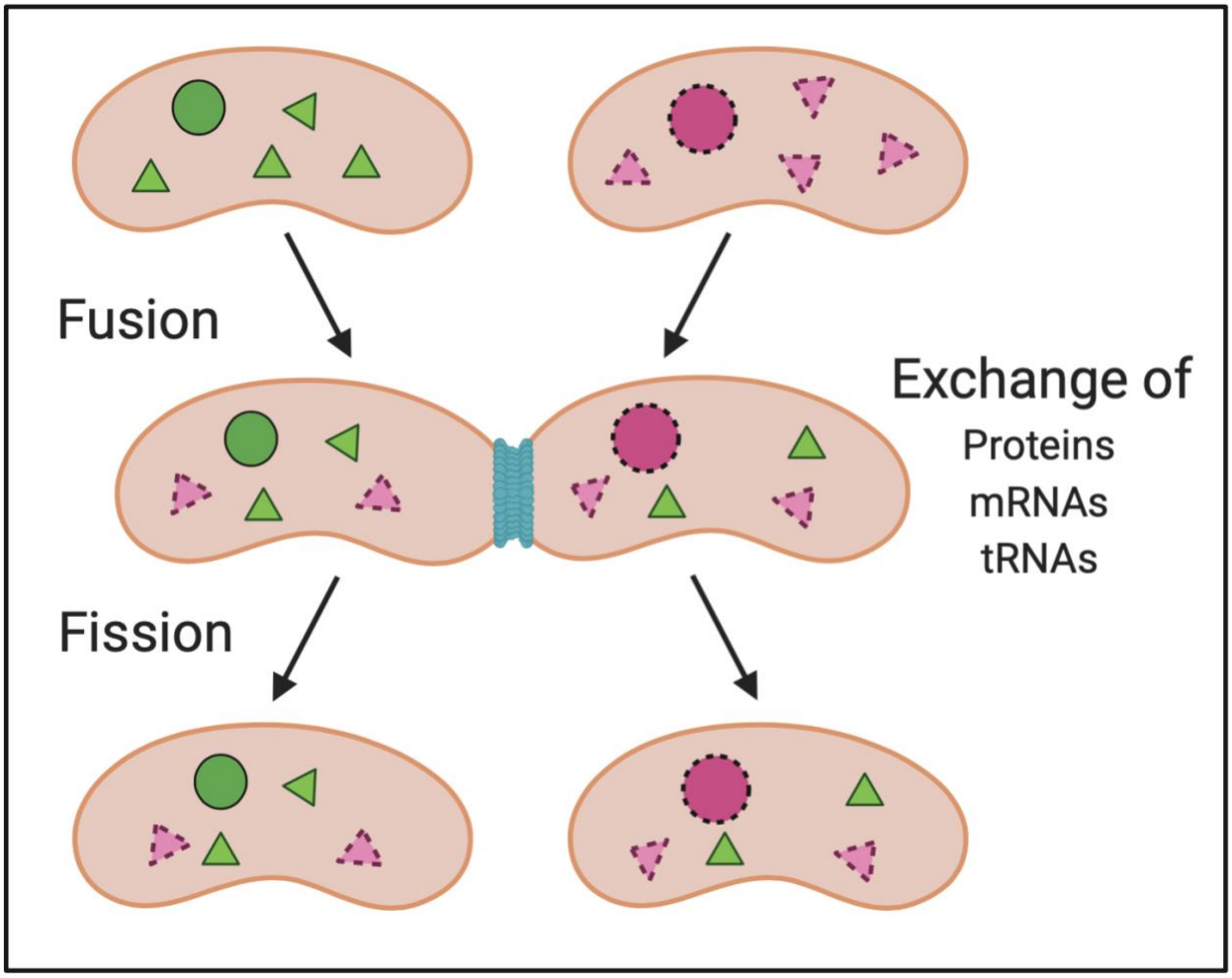


Figure 1.3. Mitochondrial dynamics promotes the mixing of mitochondrial products. Dynamic cycles of fusion and fission results in the frequent mixing of mitochondrial proteins, mRNAs, and tRNAs. Green and red circles depict wild-type and mutated mtDNA, respectively, and triangles represent their gene products. Mitochondrial fusion allows for the sharing of wild-type gene products with mitochondria bearing mutated mtDNA, and this complementation is believed to mask phenotypic consequences of mtDNA mutations.

(*79*), and other mouse models do not show progeroid phenotypes from point mutations until well above physiologic levels (*80*). It remains possible that only a subset of cells in a tissue acquire pathogenic mutations, and these drive pathologies of aging.

To better evaluate whether cellular factors influence mtDNA selection in somatic tissues, I will first review the evidence supporting the existence of mtDNA selection in the germline. After a brief overview of this literature, I will discuss several key distinctions between germline and somatic cells that might influence their abilities to select against mtDNA mutations.

1.4 *Germline selection reduces the transmission of mtDNA mutations*

Studies of mtDNA evolution and of family pedigrees have long suggested that selective mechanisms operate to reduce transmission of mtDNA mutations in the germline (*81*). Studies of human mother-daughter pairs as well as mouse models suggest germline selection against deleterious nonsynonymous mutations (*82, 83*). It has recently been shown that the human oocytes harbor fewer mutations than the maternal peripheral blood (*84*), and it is thought that the reduction of deleterious mutations coincides with a genetic bottleneck facilitated by a reduction in the number of mtDNA molecules (*85*). It has been suggested that mice can purge deleterious mtDNA through germline transmission in as few as four generations (*86*). Although there is accumulating evidence that germline selection can reduce the transmission of mutations at high clonality, selection may be less effective at eliminating the low-frequency heteroplasmies that more typical of human populations, and technical limitations have impeded the analysis of such mutations.

Recent work in *Drosophila* has begun to uncover molecular mechanisms governing germline selection. A *Drosophila* model bearing a heteroplasmic temperature-sensitive mutation in cytochrome oxidase displays reduced cytochrome oxidase stability and activity when shifted to

high temperatures (*87, 88*). Flies reared for multiple generations at high temperature show a reduction of heteroplasmy with each subsequent generation. Surprisingly, the reduction in germline heteroplasmy was found to be Parkin-independent, suggesting PINK1-Parkin mitophagy was not responsible for selection. Two models have since emerged to explain selection in these flies—the first model proposes that PINK1 accumulates on the outer mitochondrial membrane of defective mitochondria and phosphorylates the protein translational stimulator Larp. Phosphorylation of Larp inhibits translation selectively within dysfunctional mitochondria, blocking their expansion during early development (*89*). A second model suggests that hyper-fragmentation of mitochondria in early oogenesis promotes the isolation of individual mtDNA molecules, which are subject to selection by a non-canonical BNIP-3-dependent mitophagy pathway (*90*). While these models do not conflict with one another, future work will need to distinguish the relative contributions of these mechanisms in *Drosophila* germline selection, and to investigate whether similar mechanisms operate in vertebrates.

Germline cells maintain several advantages over somatic cells that might facilitate efficient mtDNA selection (*91, 92*). First, germline mitochondria become highly fragmented, promoting the formation of small mitochondria harboring fewer copies of mtDNA and reducing their ability to benefit from functional complementation. In contrast, mitochondria in most somatic tissues remain highly dynamic, presenting a challenge for efficient identification of mutant mtDNA. Second, germline mitochondria undergo one or more bottleneck events whereby mtDNA copy number is dramatically reduced, potentially further reducing the ability of complementation to mask deficiencies. By contrast, somatic mtDNA are typically present in thousands of copies per cell, making the defects of any individual molecule less consequential to the cell and more difficult to isolate for destruction. Third, rapid mitochondrial biogenesis supplies the developing embryo. If

germline selection is accomplished through the selective replication of healthy mitochondria, this mechanism might not effectively remove mutant mtDNA in somatic tissues where mitochondria are turned over far less frequently.

1.5 *Selection is inefficient at combating somatic mtDNA mutations, but may be enhanced through genetic or pharmacologic manipulation*

Somatic selection appears to be either absent or inefficient in the clearance of mutant mtDNA mutations that accumulate over the lifespan. This is evident as somatic mtDNA point mutations and deletions rise in abundance in many eukaryotes, including humans, mice, rats, and *Drosophila*. Several studies have failed to find evidence of selection in either germline or somatic tissues, and computational modeling suggests that random drift is sufficient to explain the apparent rise of clonal expansion with age (*93*-*95*). An analysis of the mutation spectra in human colon shows no evidence of selection (*96*). In the *Drosophila* temperature-sensitive mutant model described above, robust selection was found to influence mtDNA mutations in the germline, but no changes were observed in heteroplasmy in somatic tissues (*97*).

It has been suggested that mitochondrially-derived proteins are integrated into respiratory chain complexes in close proximity to the nucleoid from which they originated (*6*). This would at least transiently pair the phenotypic effects of a mutation to the mtDNA molecule from which it was derived, facilitating the destruction of both concurrently. However, nucleoids also move rapidly throughout the mitochondrial matrix and undergo frequent transient attachment to one another, making it unclear how tightly mtDNA phenotypes are linked to their associated genotypes (*98*). While functional complementation may provide resilience to the damaging effects of mtDNA mutations, it too comes at a cost: rescue of mitochondrial function likely hinders efficient

recognition of mitochondria bearing mtDNA mutations in somatic tissues. If mitochondrial dysfunction is not spatially constrained to mutant mtDNA, then efficient selection cannot occur.

Many studies suggest that the dynamic nature of somatic mitochondria as a key reason that somatic mtDNA mutations are not efficiently selected against. It remains an open question, however, whether pharmacologic or genetic induction of mitochondrial quality control pathways can enhance selection against somatic mtDNA mutations. Two promising approaches have emerged to selectively target mutation-bearing mitochondria. The first strategy involves encouraging mitochondrial fission to promote the isolation and destruction of mutant mtDNA. The second strategy involves directly promoting mitophagy, primarily through increased expression of PINK1 and Parkin.

Mitochondrial dynamics have long been suspected to hinder somatic mtDNA selection, and mathematical modeling suggests that the rate of mitochondrial fusion-fission is a critical determinant to the effectiveness of mtDNA selective mechanisms. Increasing mitochondrial fission extends lifespan in *Drosophila* in a mitophagy-dependent manner, suggesting that mitochondrial fission and mixing of mitochondrial products promotes mitochondrial health (*99*). Conversely, inhibition of mitochondrial fission results in an increase in mtDNA mutations, suggesting that the segregation of mutation-bearing mitochondria from the mitochondrial network may facilitate selection (*100*).

Administration of rapamycin, a pharmacological inducer of mitophagy, improves mitochondrial health *in vivo,* and reduces the frequency of heteroplasmic mutations in cell culture (*101-104*). Work in heteroplasmic cell lines also suggests that Parkin overexpression can reduce the abundance of mutant mtDNA, and additional stimulation of mitochondrial fission promoted

Parkin localization on mutant-bearing mitochondria (*105*). Together, these findings suggest that Parkin-mediated mitophagy has the capacity to target mutant-bearing mitochondria *in vitro*.

The predominant model that has emerged to study mtDNA mutations is the PolG mutator mouse, a mouse model bearing elevated mtDNA point mutations and deletions (*106*). These mice express an exonuclease-deficient copy of the (nuclear-encoded) mtDNA polymerase PolG. Exonuclease-deficient PolG loses its capacity to proofread, resulting in mis-incorporation of nucleotides during mtDNA replication. Recent reports using mitochondrially-targeted pH sensitive reporter suggests increased mitochondrial turnover in mutator mice, suggesting mitophagy is induced in response to high levels of mutations (*107*). However, this work does not show whether mitophagy is selective to mutant-bearing mitochondria. Knocking out *parkin* in PolG mutator mice causes dopaminergic neuron death, suggesting that Parkin promotes neuronal survival in these animals (*108*). And while sequencing *parkin* knockout mutator mice revealed a shift in their mutation spectrum, the difference is very slight. Exercise has been shown to rescue some of the pathologies of mutator mice, but it is unclear whether exercise leads to the reduction of mtDNA mutations (*109, 110*). Overall, the evidence that cellular selection eliminates somatic mtDNA mutations *in vivo*, even in mice with dramatically elevated mutations, remains inconclusive.

Recent work in a *Drosophila* model bearing a heteroplasmic large deletion in mtDNA provided support for the hypothesis that Parkin induction can select against mtDNA mutations *in vivo* (*111*). This mtDNA deletion was generated specifically within flight muscle using a mitochondrially-targeted restriction enzyme and ligase. Flies accrued high levels of the mtDNA deletion (~70%) during development but display no other cellular or organismal phenotypes. In the absence of additional genetic perturbations, this heteroplasmic deletion remained stable, again implying that selection is inefficient at removing somatic mtDNA mutations. However, co-

overexpression of Parkin in this model completely eliminated mutant mtDNA, concomitant with very little change in the total mtDNA copy number. There are many crucial questions that remain about these findings. It is unclear how or why mitophagy would efficiently clear mtDNA deletions in the absence of cellular phenotypes. Further, while a time course is provided showing the rise in deletion abundance, such time course is not provided for its destruction and replacement. This is problematic because somatic mitochondrial proteins can have half lives of many days to weeks in *Drosophila (112)*, yet the complete replacement of mutant mtDNA in this model occurs within days. Such dramatic replacement of mitochondria has not been shown previously. Interestingly, total mtDNA copy number is hardly reduced in these flies, implying not only significant mitochondrial turnover but also selective replication of wild-type mtDNA. While compelling, many questions remain about the elimination of mutant mtDNA in this model.

It remains highly controversial whether Parkin can be induced to select against mtDNA mutations. Parkin is finally being overexpressed within mutator mouse models, to disappointing results. Parkin overexpression has no impact on mitochondrial turnover in the hearts of mutator mice despite extensive pathology (*113*). Future work will be necessary to understand whether Parkin targets mitochondria bearing mutant mtDNA, and if so, whether overexpression of Parkin is sufficient to appreciably reduce cellular heteroplasmy.

1.6 *Positive selection (sometimes) promotes the expansion of mtDNA mutations*

There are also several reports of positive selection favoring the expansion of mtDNA mutations. A simple hypothesis for why mtDNA mutations might rise to high abundance in certain tissues is simply that they are better suited for the metabolic demands of a particular somatic tissue type. We generally inherit homoplasmic mtDNA, which populates all the somatic tissues of the

body. However, somatic tissues can have considerably different metabolic needs and mitochondrial morphologies. In cell culture, the same mutation leads to appreciably different metabolic manifestations depending on the host cell type (*67, 68*). Furthermore, the level of mitochondrial fragmentation and the rates of mitophagy in a cell affect its preference for wild-type or mutant mtDNA (*67, 68*). Heteroplasmic induced pluripotent stem cells (iPSCs) will rapidly shift to homoplasmy towards either wild-type or mutant mtDNA once differentiated, demonstrating that the cellular environment can affect positive selection for mtDNA. In mice and humans, mtDNA haplotypes segregate nonrandomly between tissues, implying either that mtDNA segregates based on haplotype during cellular differentiation or that the different cellular environments shape mtDNA evolution to meet different metabolic needs of these tissues (*114*). Furthermore, both humans and PolG mutator mice have been found to accumulate deleterious mtDNA mutations within liver tissues (*115, 116*).

In addition to mtDNA selection being mediated by the cellular environment, it is thought that deleterious mutations can also rise in abundance through clonal expansion. Mitochondrial genomes bearing a large deletion have been found to act as a selfish genetic element in germline mtDNA transmission, expanding with each generation (*117*). iPSCs derived from a young patient with Pearson's disease bearing a large 6kb deletion reproducibly rose to ~70% abundance across iPSC clones. Furthermore, cells bearing this deletion at low frequency were injected into mice, and formed teratomas with a high abundance of the deletion (*118*). In these cases, deleted mtDNA is hypothesized to clonally expand by selfish replication—smaller mtDNA molecules replicate faster than wild-type molecules (Figure 1.4A) (*119*). However, deletions in mtDNA are not the only mutations that may confer a replicative advantage—high mutation rates in the control region have been thought to favor the replication of mutant mtDNA (*120, 121*). Together, these studies

lead to the model that particular mtDNA mutations confer a replicative advantage over WT molecules, facilitating their clonal expansion. However, mathematical modeling challenges this hypothesis, largely due to the fact that the intervals between mtDNA replication are far longer than the time it takes to replicate (*122, 123*), Additionally, there are many reports of deleterious mutations that are not expected to alter the kinetics of replication, yet still rise in abundance. It is unlikely that the replicative advantage hypothesis adequately explains the rise of such mutations.

Another model that may explain the increased abundance of damaging mtDNA mutations is the "survival of the slowest" hypothesis (Figure 1.4B)(*124*) This model proposes that mitochondria bearing deleterious mutations become less metabolically active, and as a consequence, generate fewer ROS and sustain less ROS-mediated damage. According to this model, mitochondria bearing deleterious mutations are turned over less frequently than healthy mitochondria because they evade regular turnover through mitochondrial quality control. This model may explain why liver mitochondria accumulate damaging mtDNA mutations (*115, 116*). This may be because the liver is rich with heavy metals, and so liver mitochondria are particularly susceptible to forming damaging free radicals through the Fenton reaction. In accordance with the survival of the slowest model, mutations in liver may be abundant because they decrease the turnover of mutant mitochondria. Furthermore, some human cancers harbor an increased abundance of highly pathogenic mutations, suggesting positive selection for these variants (*125*).

A final model—the hitchhiker model—posits that mitochondria bearing deleterious mutations produce insufficient ATP (Figure 1.4C). In response, those cells harboring mutant mtDNA attempt to rescue the defect through increased mitochondrial replication. While increased biogenesis might generate enough WT mtDNA to fulfill the energetic needs of the cell, mutant mtDNA can be amplified in the process. There are two main lines of support for this model— first,

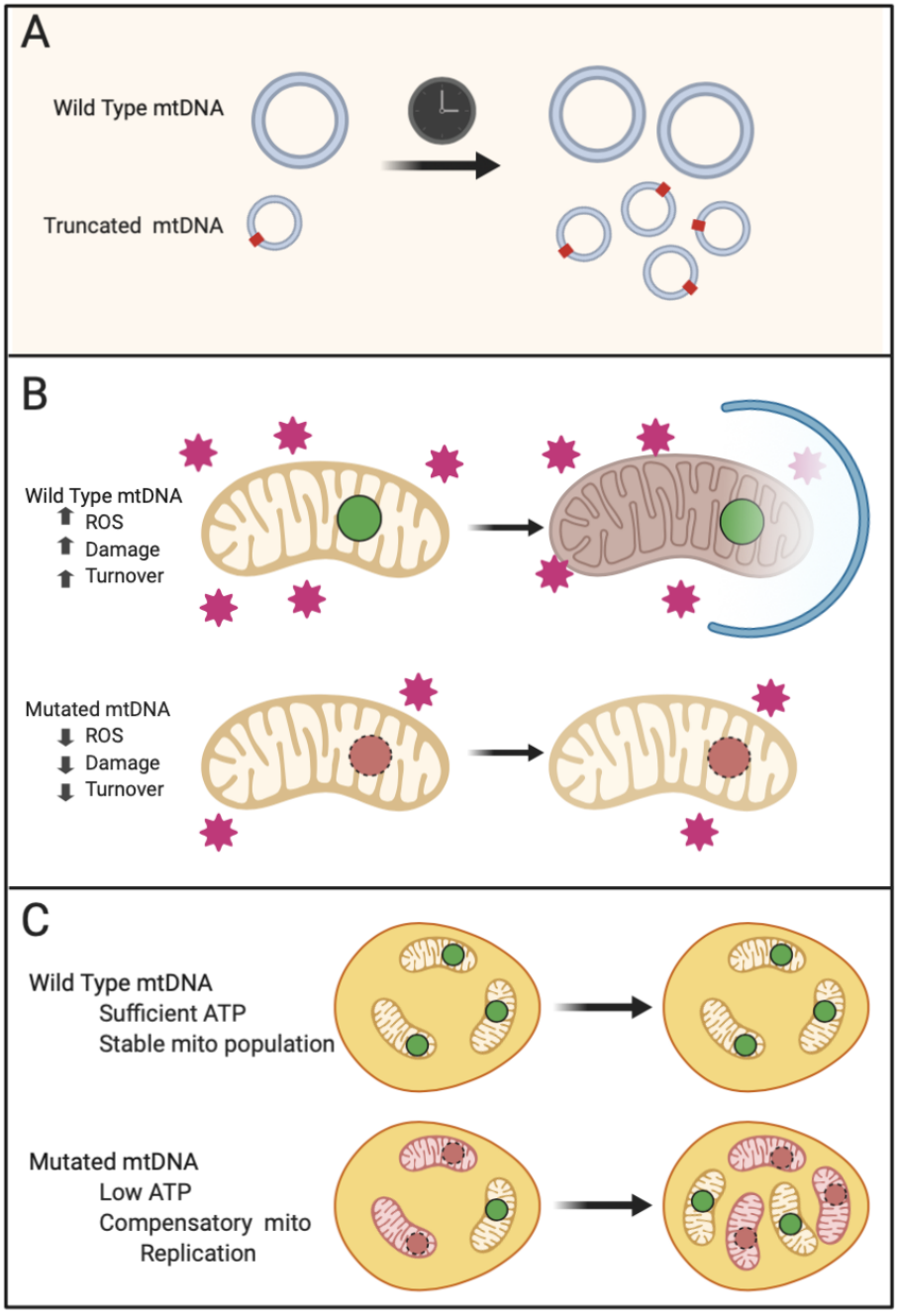


Figure 1.4. Putative mechanisms promoting positive selection of deleterious mtDNA mutations. (A) Deletion-bearing mtDNA replicate faster than full-length wild-type mtDNA, conferring a replicative advantage. (**B)** Survival of the slowest hypothesis: Wild-type mitochondria undergo oxidative phosphorylation, which produces high levels of ROS (represented by red stars) and promotes their frequent turnover. Mitochondria bearing mutant mtDNA undergo less respiration, produce less ROS, sustain less ROS-mediated damage, and turn over less frequently. (**C)** Hitchhiker hypothesis: Cells harboring mutant mtDNA produce insufficient ATP, triggering compensatory mitochondrial biogenesis.

mutations hitchhiking “futile” compensatory replication have been observed in multiple human diseases, including the mitochondrial disease Myoclonic epilepsy with ragged red fibers, (*126, 127*), Chronic Obstructive Pulmonary Disease (*128*), and the age-associated accumulation of mtDNA deletions (*129, 130*).

Further support for the hitchhiker model emerged from recent studies in a *C. elegans* model bearing a stable heteroplasmic large deletion that encompasses 11 mitochondrial genes. Worms from this strain never become homoplasmic for either wild-type or truncated mtDNA (*131*), even with many generations of selective propagation of high- and low-heteroplasmy worms (*132*). Intriguingly, the copy number of deletion-bearing mtDNA varies dramatically between individuals, whereas the copy number of wild-type mtDNA remains remarkably stable. This observation implies that cells maintain levels of wild-type gene products, but that mtDNA replication does not readily distinguish between wild-type and mutant mtDNA. According to this model, cells that acquire high levels of mutant mtDNA stimulate mitochondrial biogenesis to maintain wild-type copy numbers, but replication is unable to distinguish between wild-type and mutant genomes. The mutant genomes occasionally become replicated, too, ‘hitchhiking’ the cell’s homeostatic response. Remarkably, it appears that the UPR_{mt} is permissive for high levels of mutant mtDNA. Paradoxically, mitochondrial chaperones and proteases enable the cell to tolerate the detrimental effects of mutations, shielding mutations from mitophagic degradation. Knockdown of central UPR_{mt} genes eliminates this protection from mutations, and leads to Parkin-mediated degradation of mtDNA deletions (*133, 134*). Taken together, these studies suggest that some mtDNA mutations hitchhike the cell’s replication machinery to rise in abundance. Importantly, they also lend support to the presence of both positive and negative selective processes operating simultaneously to influence mtDNA mutation accumulation.

1.7 *Challenges to studying somatic mtDNA selection*

To conclude this introduction, I would like to briefly summarize some of the major technical and statistical obstacles that have historically hindered the study of somatic selection. While a number of animal models have recently been developed to study somatic mutations, technical limitations have stymied the accurate detection and interpretation of heteroplasmic mutations. Furthermore, due to the unique composition of the mitochondrial genome, statistical oversights have plagued the study of selection. Understanding the causes and consequences of these errors more fully will be critical to future interpretation of mtDNA sequence data.

1.7.1 *Exploring cell-to-cell mtDNA heterogeneity*

The acquisition of mtDNA mutations occurs stochastically within cells throughout the body, resulting in a highly heterogeneous pool of mtDNA whereby the frequency of any individual mutation is at very low abundance. In studies of mtDNA mutations, mtDNA is typically extracted from complex mixtures containing multiple cell types, resulting in a mixed population of mtDNA. The overall mutation frequency from this population is therefore an average of a very heterogeneous population and does not reflect the cell-to-cell variation of mutations (Figure 1.5). This has important biological ramifications for the interpretation of selection.

Hundreds to thousands of copies of mtDNA reside in each cell, distributed between highly dynamic organelles. Mutations arise independently in each somatic cell and are independently subject to selective forces that dictate their fate; most mutations likely persist at relatively low heteroplasmy. There have been attempts to disentangle these levels of variability, such as looking at COX-negative cells within heterogeneous cell populations (*67*). In all the work in this thesis, I limit cell type heterogeneity by performing all of my sequencing on individual fly heads. It is

imperative in future work to limit cellular heterogeneity through focused dissection of individual cell types, or eventually, sequencing of single cells.

1.7.2 *Improving sequencing of mtDNA*

Due to the heterogeneity of somatic mutations, it remains a technical challenge to accurately detect low frequency heteroplasmies. Calculation of the mtDNA mutation frequency of wild-type animals varies by as much as three orders of magnitude among published studies. While some of this variability likely reflects true biological differences between organisms, tissue type, and ages studied, much of this variability likely reflects differences in sequencing methodologies. Indeed, many technical challenges hinder the accurate detection of mtDNA mutations. Some common molecular biology tools cannot be used effectively to study mtDNA mutations, as they introduce excessive errors. Because mtDNA mutations typically occur at low frequency within a mixed population of cells, most sequencing technologies lack the sensitivity to detect these mutations (*135*). The high error rates of next-generation sequencing has relatively little impact in determining the consensus sequence of nuclear DNA, but completely obscures the signal of low frequency heteroplasmies (*136-138*).

One challenge of identifying mtDNA mutations is amplifying DNA template prior to sequencing. PCR amplification has been used to study mtDNA mutations for many years, but the polymerases traditionally used for PCR have high error rates that obscure the *in vivo* mtDNA mutation frequency (*139*). Failing to distinguish between signal and errors makes meaningful interpretation of this data impossible. High-fidelity polymerases decreases polymerase errors, but oxidative damage to the mtDNA template still become fixed as false mutations (*139*). Strategies

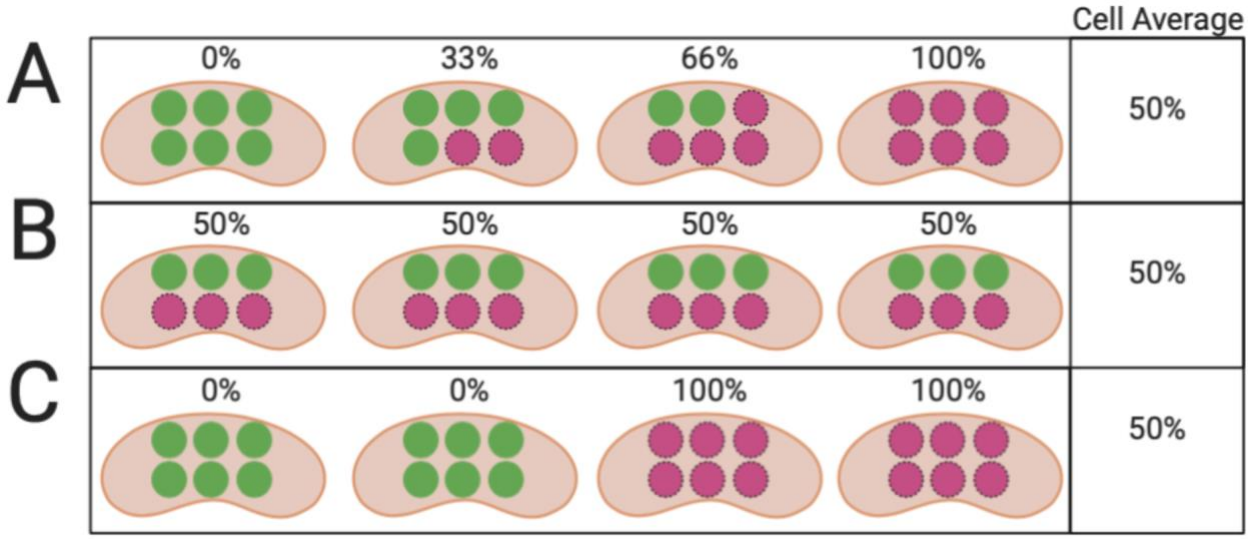


Figure 1.5. Measures of mtDNA heteroplasmy from pooled samples do not reflect the sub-cellular distribution of mutations. Each cell in a population harbors a distribution of mtDNA mutations that is obscured from pooled sample averages, but likely manifests in different cellular phenotypes. Each of the three cells shown bears an average heteroplasmy of 50%, despite different distributions of mutations among mitochondria. **(A)** The cell has a mixed population of mitochondria with variable levels of heteroplasmy. **(B)** The cell's mitochondria are all heteroplasmic at levels equal to the cell's overall heteroplasmy. **(C)** The cell bears only mitochondria homoplasmic for wild-type or homoplasmic for mutant mtDNA.

to amplify mtDNA but reduce error rate through downstream analyses include rolling-circle PCR, cloning and sequencing, and duplex sequencing.

Sequencing mutations presents an additional set of challenges. Traditional Sanger sequencing has a low error rate, but it is also laborious and provides low coverage of the genome. Next-generation sequencing is now widely available, and provides high depth and coverage of the mtDNA. However, next generation sequencing also produces a high error rate that results in many false mutation calls (*139*). Placing stringent quality control cutoffs reduces the number of false positive mutations, but also severely limits the detection of most low frequency mutations. Furthermore, this strategy does not eliminate early-PCR errors.

Three notable high-fidelity methods have recently been developed for the study of mtDNA mutations. The first is the Random Mutation Capture (RMC) assay, which quantifies the frequency by which a given TaqI restriction site in the mitochondrial genome is lost as a result of mutation (*135, 140*). TaqI restriction enzyme is added to mtDNA, where it cleaves all mtDNA in the sample except rare molecules bearing mutations in TaqI sites, which remain resistant to cleavage. qPCR primers are chosen to flank the TaqI site, and in an unrelated site in the genome. The ratio of TaqI-resistant molecules to total mtDNA yields the mutation frequency. RMC is exquisitely accurate because TaqI restriction is performed prior to PCR amplification, so it is not affected by PCR-introduced errors. But RMC is expensive, labor intensive, and limited throughput (*140*). Another technique, the Digital Deletion Detection (3D) assay, was created to address these weaknesses. The 3D assay essentially performs RMC on mtDNA molecules that have been suspended individually within water-in-oil droplets, and quantified using droplet digital PCR (*141*). This refinement greatly increases the sensitivity and throughput of RMC. While the RMC and 3D assays works very well at quantifying mtDNA mutations, they only provide mutational estimates at the

few TaqI sites in the mitochondrial genome that are amenable to this analysis, severely limiting the study of selection on random genome-wide mutations.

I use another high-fidelity technique, Duplex Sequencing, throughout all of my work. In Duplex Sequencing, barcodes are ligated to both strands of the mtDNA molecule, allowing us to trace sequence reads to the original starting molecule (*139*). Duplex Sequencing eliminates PCR-introduced errors and sequencing artifacts, and has the advantages of high sensitivity, sequencing depth, and genomic coverage, coupled with a substantial reduction in the error rate of next-generation sequencing (*61, 139, 142*). Unlike restriction-enzyme based methods which only screen for mutations within small restriction sites, Duplex Sequencing provides us with accurate sequence information for the entire mtDNA coding sequence. It is imperative that future studies of mtDNA selection forego the use of next-generation sequencing that do not employ techniques to improve the fidelity of detection. Estimates of mutation frequencies of the same sample differ dramatically between Sanger Sequencing and next-generation sequencing (*143*), and there is no correlation between the mutation frequencies obtained by next-generation sequencing and RMC (*144*). The widespread use of low sensitivity, error-prone sequencing suggests that much of the published literature on mtDNA mutations may ultimately need to be reevaluated. Traditional sequencing techniques simply do not yield accurate estimates of the mutation frequency, and it is implausible that one can draw robust inferences about mutational spectra and selection with these methods.

1.7.3 *Improving statistical analysis of mtDNA selection*

Perhaps due to the challenges of accurately identifying mtDNA mutations, the statistical framework for analyzing mtDNA mutations remains in its infancy. As a result, simple summary statistics on the frequency and spectrum of mtDNA mutations are typically reported, without deeper analysis or appropriate normalization of these numbers. It is still common practice to report

raw numbers of mutations, or the number of mutated sites, without consideration to the nucleotide composition of mtDNA or variations in sequencing depth. Multiple studies, for example, make claims about particular contexts being *overrepresented* simply because they are mutated more highly mutated than in wild-type mice, but this claim is unsurprising given the ~10-fold increase in mutation frequency of mutator mice (*115, 145*). Similarly, the mutability of a trinucleotide context is a function of not only the number of mutations within a given context, but also that context's prevalence in the genome. Any claims about the over- or underrepresentation of particular mutation contexts can only be inferred if the number of mutations is normalized by how many times that context appears in the genome.

Additionally, the composition of the genome is critical to both the mathematical and biological interpretations of selection in mtDNA. For example, simple comparisons of the mutation frequencies by codon position is not informative as to the mutability of these sites if the nucleotide composition is not taken into account (*146*). mtDNA has evolved such that many animals' mtDNA has low G/C-content (<50% G/C) (Figure 1.6). Moreover, third codon positions display a pronounced depletion of G/C base pairs in commonly used model organisms (Figure 1.6). However, both wild-type and PolG mutator animals display predominantly G/C->A/T mutations. Therefore, we would expect a higher mutation frequency in certain sites, such as in first and second codon positions, based solely on the mutational biases of the polymerase (irrespective of any selective processes). In **Chapter 2**, I address the dual effects of mutational biases and genomic content by randomly permuting observed mutations using Monte Carlo simulations to generate neutral mutational models. It is imperative that future work these genomic biases are considered for both the statistical analyses and biological interpretations of mtDNA mutations.

1.8 *Summary and Outline*

It is now widely believed that selection limits the transmission of pathogenic mtDNA mutations in the female germline, but it is not clear whether such mechanisms operate in the soma. Conflicting reports suggest that mitochondrial quality control pathways may be induced to select against deleterious somatic mtDNA mutations, but we know little about the cell's response to deleterious mtDNA mutations. In this thesis, my central focus is addressing whether or not somatic cells select against mtDNA mutations, but I will also explore the factors that influence the pathogenicity of heteroplasmic mutations. In **Chapter 2**, I describe the creation and characterization of a PolG mutator fly model, which revealed that positive selection favors the accumulation of deleterious mutations. In **Chapter 3**, I compare the mutation spectra of two mtDNA mutator fly models to understand differences in the severity of their phenotypes. In **Chapter 4**, I test whether Parkin overexpression reduces the frequency of mtDNA mutations. In **Chapter 5**, I will suggest future directions that build upon my findings. In addition to expanding upon our understanding of the molecular mechanisms influencing mtDNA selection, it is my goal that this body of work will contribute to future statistical analyses of mtDNA mutations.

1.9 *Notes and Acknowledgements*

Figures for this chapter were generated using www.biorender.com. Special thanks to Marie Y Davis for critical reading of this chapter.

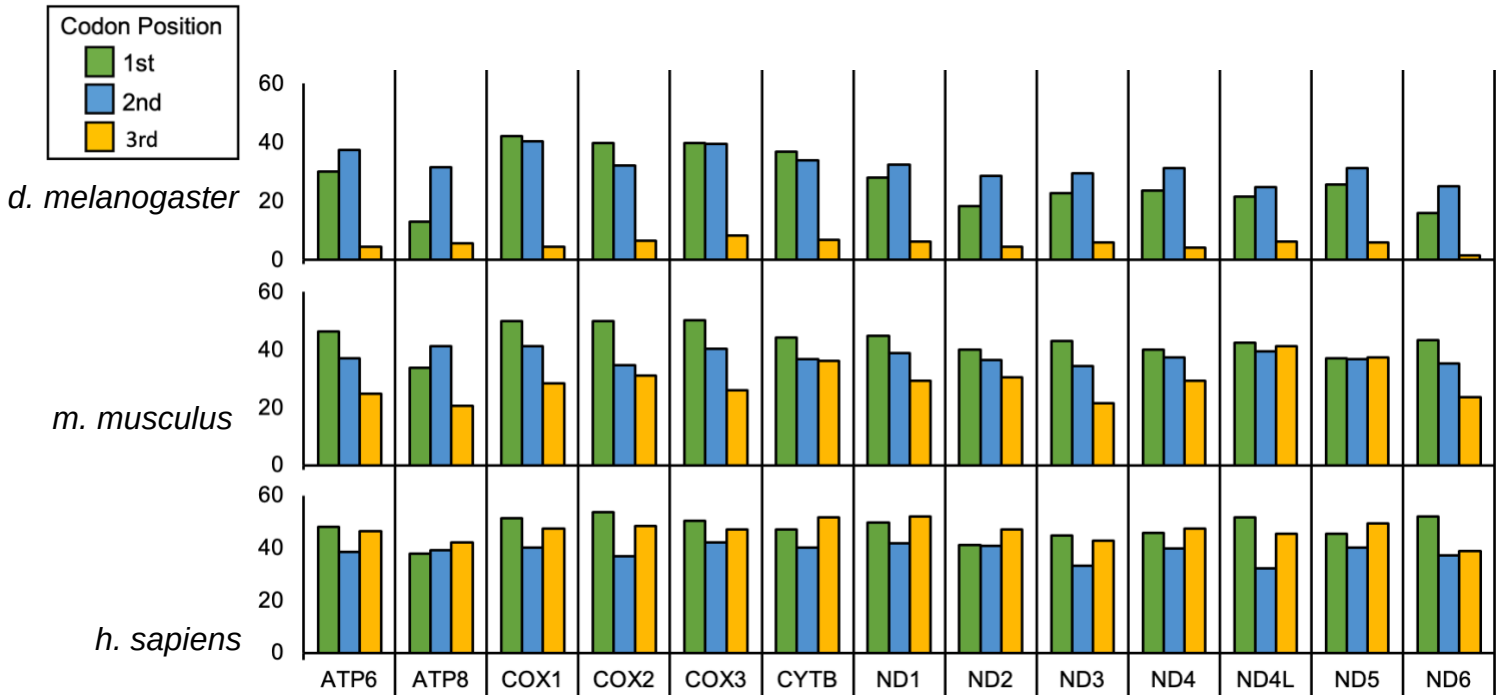


Figure 1.6. Model organisms display marked depletion of G/C base pairs in mtDNA third codon positions. Many eukaryotes show depletion in G/C base pairs (<50%) across the protein-coding sequence in mtDNA. Common model organisms including *Drosophila* and mice show a striking depletion in third codon position G/C sites, but this trend is not recapitulated in human mtDNA. G/C->A/T mutations are the predominant mutation type in wild-type animals as well as commonly used mutator mod

Chapter 2. **Deleterious mitochondrial DNA Point mutations are Overrepresented in *Drosophila* Expressing a Proofreading-Defective Polymerase γ**

Over the past 15 years, the Pallanck Lab has used the fruit fly *Drosophila* as a model system to study mitochondrial quality control. Our previous work on fly homologs of the Parkinson's disease related genes, *PINK1* and *parkin* played an important role in revealing that PINK1 and Parkin are required for the selective fragmentation and mitophagic degradation of depolarized mitochondria in mitophagy (*48, 51, 147, 148*). Dr. Leslie Itsara, a former graduate student in our lab, characterized the frequency and spectrum of mutations in wild-type *Drosophila* (*62*). In her work, Dr. Itsara found that many features associated with somatic mtDNA mutations in mammals are conserved in *Drosophila,* including a similar mtDNA mutation frequency, spectra, and accumulation throughout the lifespan (*62*). My work in this chapter has extended the utility of this work by creating and characterizing a new *Drosophila* model of elevated mtDNA mutation frequency.

This chapter has been published, in slightly modified form, as:

> **Samstag CL**, Hoekstra JG, Huang C-H, Chaisson MJ, Youle RJ, Kennedy SR, et al. Deleterious mitochondrial DNA point mutations are overrepresented in Drosophila expressing a proofreading-defective DNA polymerase γ. PLOS Genetics. 2018;14: e1007805. doi:10.1371/journal.pgen.1007805 (*149*)

2.1 *Abstract*

Mitochondrial DNA (mtDNA) mutations cause severe maternally inherited syndromes and the accumulation of somatic mtDNA mutations is implicated in aging and common diseases. However, the mechanisms that influence the frequency and pathogenicity of mtDNA mutations are poorly understood. To address this matter, we created a *Drosophila* mtDNA mutator strain expressing a proofreading-deficient form of the mitochondrial DNA polymerase. Mutator flies have a dramatically increased somatic mtDNA mutation frequency that correlates with the dosage of the proofreading-deficient polymerase. Mutator flies also exhibit mitochondrial dysfunction, shortened lifespan, a progressive locomotor deficit, and loss of dopaminergic neurons. Surprisingly, the frequency of nonsynonymous, pathogenic, and conserved-site mutations in mutator flies exceeded predictions of a neutral mutational model, indicating the existence of a positive selection mechanism that favors deleterious mtDNA variants. We propose from these findings that deleterious mtDNA mutations are overrepresented because they selectively evade quality control surveillance or because they are amplified through compensatory mitochondrial biogenesis.

2.2 *Introduction*

Mitochondria contain the electron transport chain complexes responsible for generation of most of a cell's energy and also play crucial roles in Ca^{2+} buffering, metabolite synthesis, and apoptosis (*3*-*5*). In addition to the ~1,000-2,000 nuclear genes that encode mitochondrial proteins, mitochondria contain a separate small circular genome, densely packed with 37 genes, that is essential for mitochondrial function. Mitochondrial DNA (mtDNA) mutations transmitted through the female germline are responsible for a host of incurable mitochondrial syndromes (*150*). In

addition, accumulation of mtDNA mutations in somatic tissues is implicated in aging (*151, 152*) and common diseases of the elderly including cancer (*153*) and neurodegenerative diseases (*13*). There are typically thousands of copies of the mitochondrial genome in a single cell, such that when mtDNA mutations occur, they frequently share residence with wild-type mtDNA, a condition known as heteroplasmy. High levels of heteroplasmic mutations correlate with the severity of mitochondrial diseases (*12*), yet we know little about the factors that influence the frequency of mtDNA mutations or the emergence of their associated phenotypes (*77, 154*).

To explore the cellular mechanisms that influence the frequency and pathogenicity of mtDNA mutations, we are using the fruit fly *Drosophila* as a model system. In previous work, we found that many fundamental features associated with somatic mtDNA mutations in mammals are conserved in *Drosophila*, including a similar mtDNA mutation frequency, a preponderance of transition mutations, and an increased frequency of mtDNA mutations with age (*62*). Bratic et al. further extended the utility of using *Drosophila* to study mtDNA mutations by knocking in exonuclease- and polymerase-deficient forms of DNA polymerase γ (PolG), the polymerase responsible for replicating the mitochondrial genome (*155*). While these mutant strains are developmentally lethal in the larval stage as homozygotes, heterozygotes for the exonuclease-deficient PolG exhibited increased mtDNA mutation frequency. However, the physiological effects of this high mutation burden and the possibility of negative selection acting against the resulting mtDNA mutations were not explored.

In our current work, we created a transgenic proofreading-deficient version of the *Drosophila* mtDNA polymerase (designated $PolG_{mut}$) that confers a 10- to 55-fold increase in the mtDNA mutation frequency, depending on transgene dosage without exhibiting embryonic lethality. $PolG_{mut}$ expressing flies exhibited dosage-dependent phenotypes analogous to those of

human mitochondrial diseases, including shortened lifespan, a progressive locomotor defect, and loss of dopaminergic neurons. Analysis of the frequency and distribution of mtDNA mutations in these mutator flies revealed an unexpectedly high ratio of nonsynonymous to synonymous mtDNA mutations. The mutations detected in mutator flies also tended to occur preferentially at conserved mtDNA sequences and resulted in pathogenic alterations. Together, these findings suggest that positive selection acts in favor of deleterious mitochondrial variants, either through the selective evasion of mutant-bearing mitochondria from negative selection or because cells that stochastically acquire deleterious mtDNA mutations activate compensatory mitochondrial biogenesis. Future work with this mtDNA mutator model will facilitate the study of cellular mechanisms that influence the frequency and pathogenesis of mtDNA mutations, as well as the identification of molecular factors that influence these processes.

2.3 ***Results***

2.3.1 *Generation of a Drosophila mtDNA mutator strain*

Previous work has shown that altering a conserved aspartate residue in the second exonuclease domain of PolG to alanine impairs proofreading ability and results in a dramatically elevated mtDNA mutation frequency in multiple species (*106, 155-158*). Thus, we generated a *Drosophila PolG* transgenic construct with an alanine substitution at the equivalent site (designated $PolG_{mut}$; Figure 2.1A). Because overexpression of *PolG* in *Drosophila* results in mtDNA depletion (*159*), we created our transgene using a genomic DNA fragment containing both the endogenous *PolG* gene and its associated *cis*-regulatory transcriptional elements to avoid artifacts associated with overexpression (Figure 2.1A). This construct was then used to create transgenic flies using standard methodologies (see Materials and Methods).

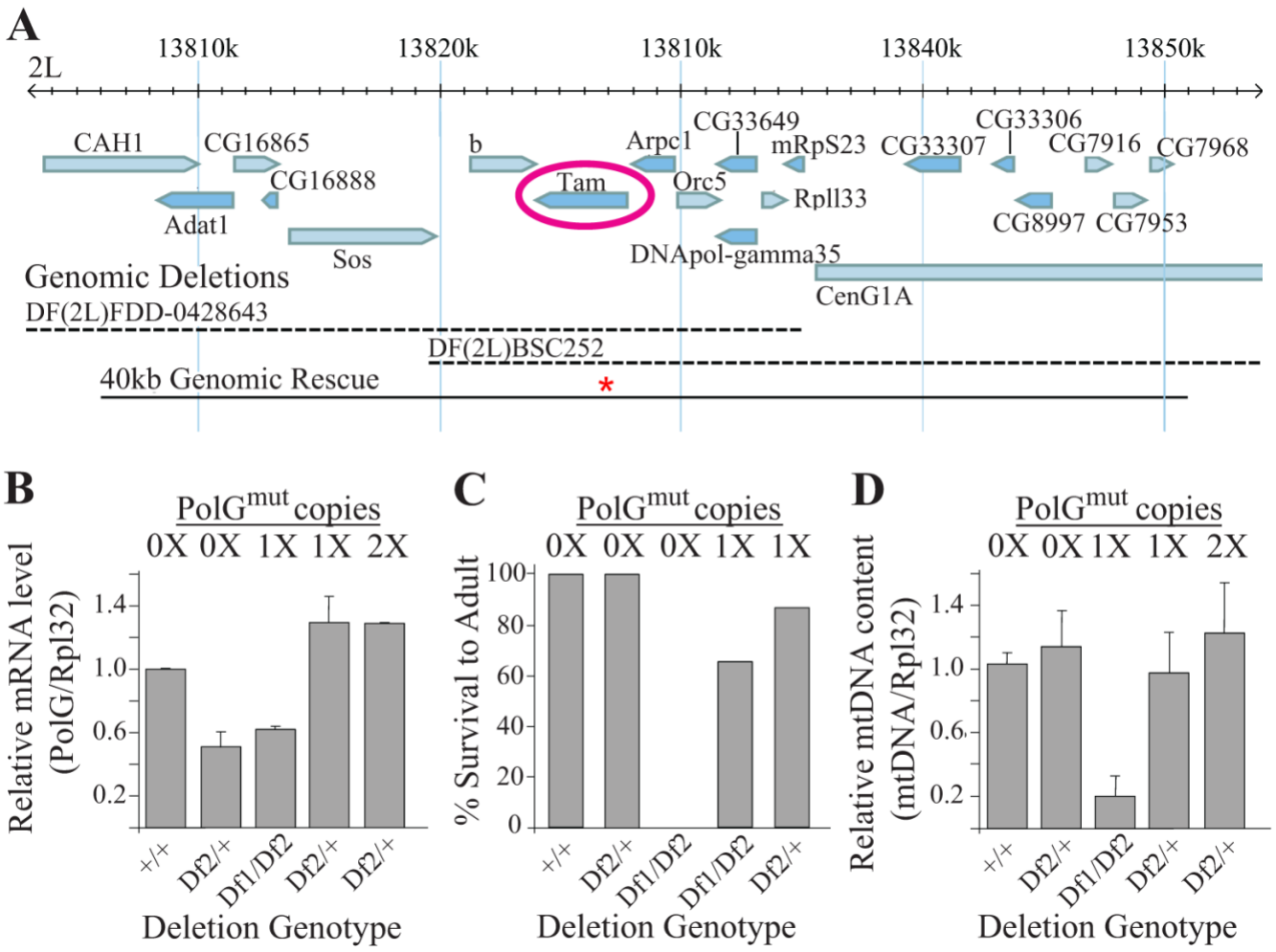


Figure 2.1. Transgenic expression of an exonuclease-deficient PolG rescues the larval lethality caused by overlapping deletions that remove *Drosophila PolG*. (A) The genomic region containing the *Drosophila* PolG gene *tamas* (*tam*). Blue bars represent known or predicted transcripts. Dashed lines correspond to sequences eliminated by the deletion alleles *Df(2L)FDD-0428643* (DF1) and *Df(2L)BSC252* (DF2). The solid black bar corresponds to the 40kb transgenic construct containing the D263->A263 mutation in *Drosophila* PolG (designated by the red asterisk), which spans the overlap between the Df(2L)FDD-0428643 and Df(2L)BSC252 deletions. **(B)** Total RNA was extracted and qPCR was used to measure PolG and Rpl32 transcript abundance from 7-day-old flies of the given genotype (n = 4 per genotype, three independent replicates) and their ratios are indicated. **(C)** The percentage of larvae of the given genotype that survived until the adult stage of development is indicated ($n \geq 236$ flies per genotype). **(D)** DNA was extracted from 7-day-old flies and qPCR was used to measure COX1 and Rpl32 DNA abundance from flies of the given genotype (n = 2 7-day-old male flies per genotype, three independent replicates) and their ratios are indicated. Error bars in panels B and D represent standard error.

Flies bearing one or two copies of the *PolG^{mut}* transgene expressed similar levels of *PolG* mRNA as the endogenous *PolG* gene, thus confirming that this transgene does not cause PolG overexpression (Figure 2.1B). Moreover, a single copy of the *PolG^{mut}* transgene was capable of rescuing the recessive lethal phenotype caused by an overlapping set of deletions that remove the endogenous *PolG* gene, thus confirming that the *PolG^{mut}* transgene encodes a functional mtDNA polymerase (Figure 2.1A,C). However, the rescued flies displayed a marked reduction in mtDNA copy number (Figure 2.1D), consistent with previously published work suggesting that *Drosophila* PolG bearing this proofreading alteration is not fully functional (*155*). In contrast, flies expressing one or two copies of the *PolG^{mut}* transgene in a strain hemizygous for the endogenous *PolG* gene did not display mtDNA depletion (Figure 2.1D). Given that mtDNA depletion could potentially confer phenotypes that are unrelated to mtDNA mutations, all of our remaining work involved the use of flies hemizygous for the endogenous *PolG* gene (Df2) and bearing zero, one, or two copies of the *PolG^{mut}* transgene, which we refer to as 0xPolG^{mut}, 1xPolG^{mut}, and 2xPolG^{mut}, respectively.

To test whether the *PolG^{mut}* transgene conferred an increased mtDNA mutation frequency, we performed Duplex Sequencing (DS) on mtDNA isolated from individual heads of 1-day-old transgenic flies. DS is a high-accuracy next-generation sequencing approach capable of detecting a single mutation in $>10^7$ wild-type bases (*139*). We addressed the possible influence of genetic background by comparing *PolG^{mut}* transgenic flies to control non-transgenic flies hemizygous for the endogenous *PolG*. Furthermore, fly strains were outcrossed to the same WT strain prior to sequence analysis to ensure that all of the genotypes being compared inherit their mtDNA from the same parental strain, as well as to replace mitochondrial genomes that had potentially accumulated mtDNA mutations over multiple generations of replication by the mutator polymerase (Figure 2.2). We first tested whether the *Polg^{mut}* transgene introduces greater

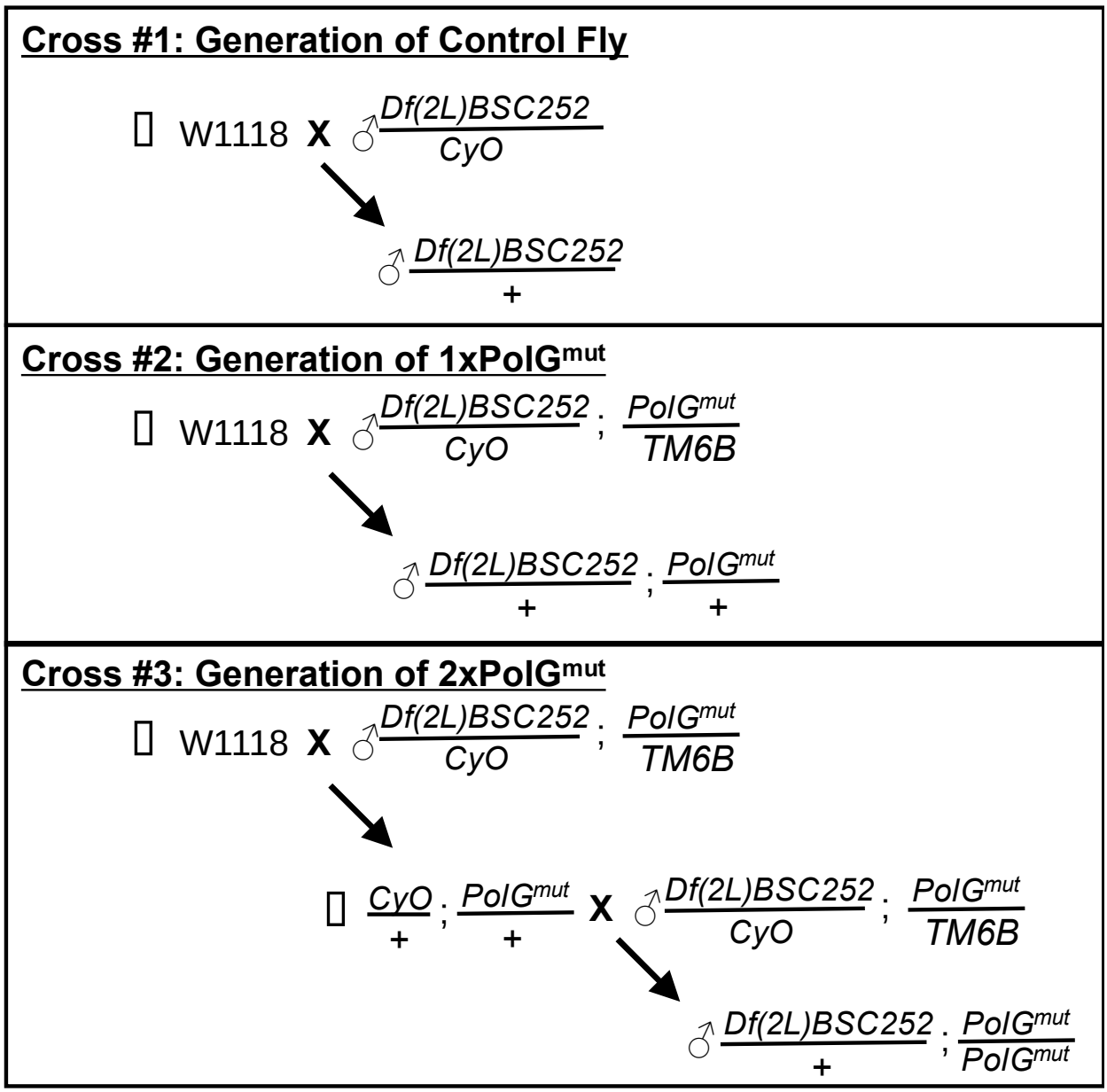


Figure 2.2. The crossing schemes used in our work. The crossing schemes used to generate control 0xPolG$_{mut}$ (Cross #1), 1xPolG$_{mut}$ (Cross #2), and 2xPolG$_{mut}$ (Cross #3) flies. Flies were outcrossed to females from the same isogenic w1118 strain prior to sequencing to eliminate accumulated mutations and control for genetic background effects.

replication errors by measuring the frequency of unique mutations (see materials and methods for description of unique mutations) in 1-day-old flies. 0xPolG$_{mut}$ (control) flies had a mutation frequency comparable to the frequency previously reported for WT flies ($3.4x10^{-6} \pm 8.9x10^{-7}$). By contrast 1xPolG$_{mut}$ flies exhibited significant increases in the point mutation, insertion and deletion frequency relative to controls (Figure 2.3A, Figure 2.4), and the frequency of these mutations increased further in 2xPolG$_{mut}$ flies (Figure 2.3A, Figure 2.4, Figure 2.5). Because 2xPolg$_{mut}$ flies inherit one copy of the *PolG$_{mut}$* transgene maternally (Figure 2.2), the increased mtDNA mutation frequency in 2x relative to 1xPolG$_{mut}$ animals may derive from an increased somatic mtDNA mutation frequency, as well as mutations that arise in the female germline. Consistent with previous work involving exonuclease-deficient mtDNA polymerases, mutator flies exhibited an increased prevalence of predominantly G:C to A:T transition mutations (Figure 2.3D) (*156, 160, 161*).

We next explored the influence of age on mtDNA mutation frequency in mutator flies by sequencing mtDNA from flies aged 25 and 50 days. The highest mtDNA mutation frequencies were observed in 50-day-old 2xPolG$_{mut}$ flies, in which the point mutation frequency of unique mutations was elevated ~55-fold relative to age-matched controls (Figure 2.3A, Table 2.1). Although there was a trend towards increased mutation frequency with age for all mutation types detected, only 2xPolG$_{mut}$ flies exhibited a significant age-associated increase in point mutations relative to young flies of the same genotype (Figure 2.3A). Like young mutator flies, aged mutator flies also exhibited a prevalence of G:C to A:T transition mutations (Figure 2.6). Additionally, when we combined data from control 0xPolG$_{mut}$ animals of all ages to increase the total number of mutations detected, G:C to A:T transition mutations were also the most frequent mutation type detected, consistent with our previously published work (*62*).

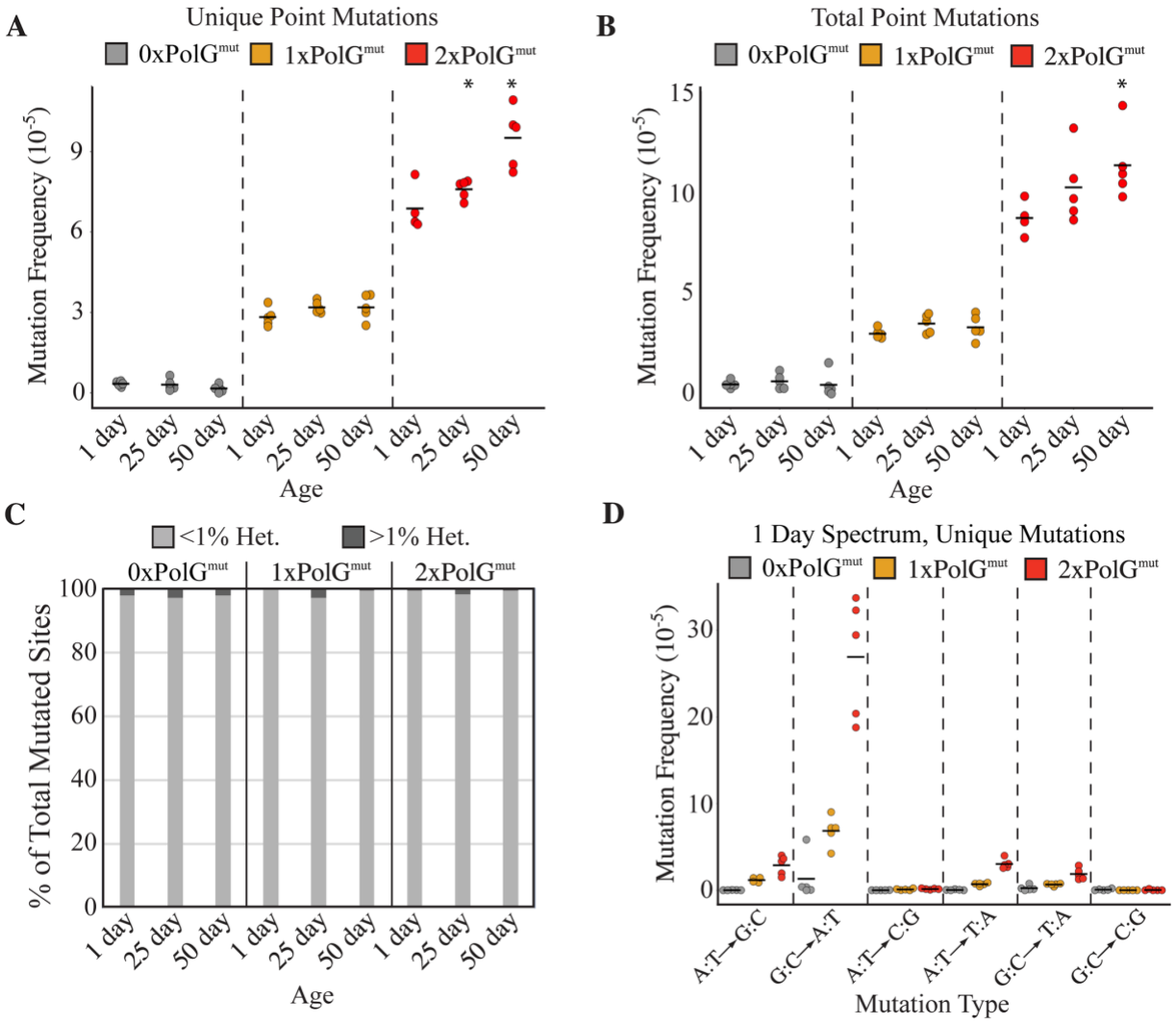


Figure 2.3. Transgenic expression of an exonuclease-deficient PolG results in a dose-dependent increase in mutation frequency. The frequencies of base substitution mutations were quantified in 1-day-old, 25-day-old, and 50-day-old flies of the given genotype using DS. **(A)** The mutation load of unique mutations, representing each unique mutation counted once, thus reflecting the de novo somatic mutation frequency. **(B)** The frequency of total mutations. **(C)** The percentage of mutated sites at sub-clonal levels (<1% heteroplasmy) and clonally expanded mutations (≥1% heteroplasmy) in flies of the indicated age and genotype. **(D)** The frequency of unique mutations of each type of base substitution mutation is indicated in 1-day-old flies of the indicated genotypes. $n = 5$ per genotype per time point. Horizontal bars in panels A, B, and D represent the mean frequency of the indicated group. *$p < 0.05$ compared to 1-day-old flies of the same genotype by Wilcoxon rank-sum test. 1x and 2xPolG^{mut} flies displayed significantly elevated mutation frequencies compared to age-matched control 0xPolG^{mut} flies at all time points ($p < 0.05$ by Wilcoxon rank-sum test).

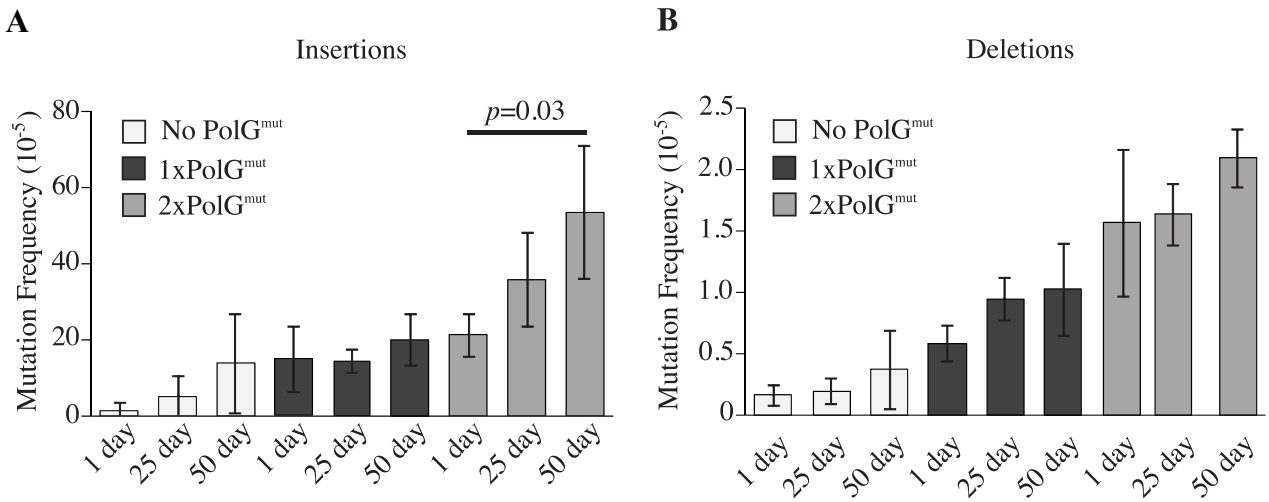


Figure 2.4. Mutator flies show a dose-dependent increase in small insertion and deletion frequency. (A) The frequencies of insertion mutations were quantified in flies of the given ages and genotype using DS. **(B)** The frequencies of deletion mutations (≤5bp) were quantified in flies of the indicated ages and genotypes using DS. *p*-values determined by Student's t-test.

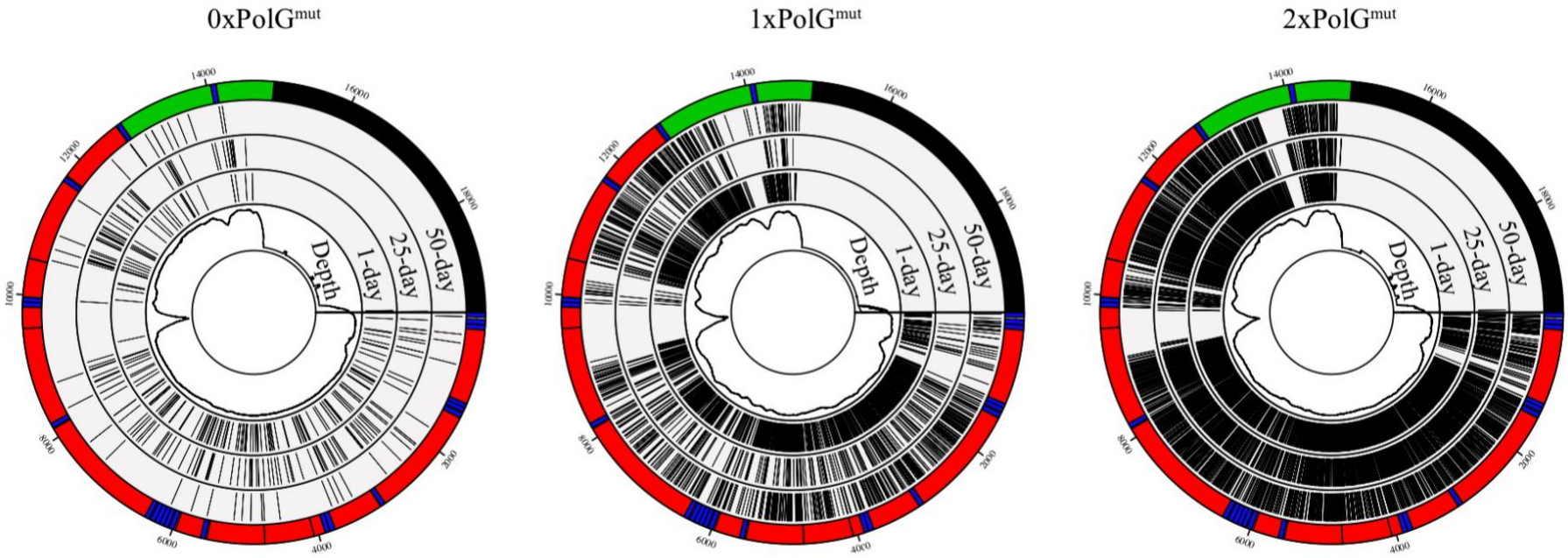


Figure 2.5. The distribution of mutations in mutator flies. Plots of mtDNA mutations identified in 0xPolG$_{mut}$, 1xPolG$_{mut}$ and 2xPolG$_{mut}$ flies. The outermost track of each plot designates functional elements within the mitochondrial genome. Red = Protein coding; Blue = tRNA; Green = rRNA; Black = Control Region. The middle track of each plot depicts the sites where mutations were observed within animals of the age indicated. The innermost track depicts the log-transformed average sequencing depth for flies of the indicated genotype. Note the absence of sequence coverage in the AT-rich control region, as well as the region between ~(ChrM ~9100-9850) not efficiently captured in our sequencing.

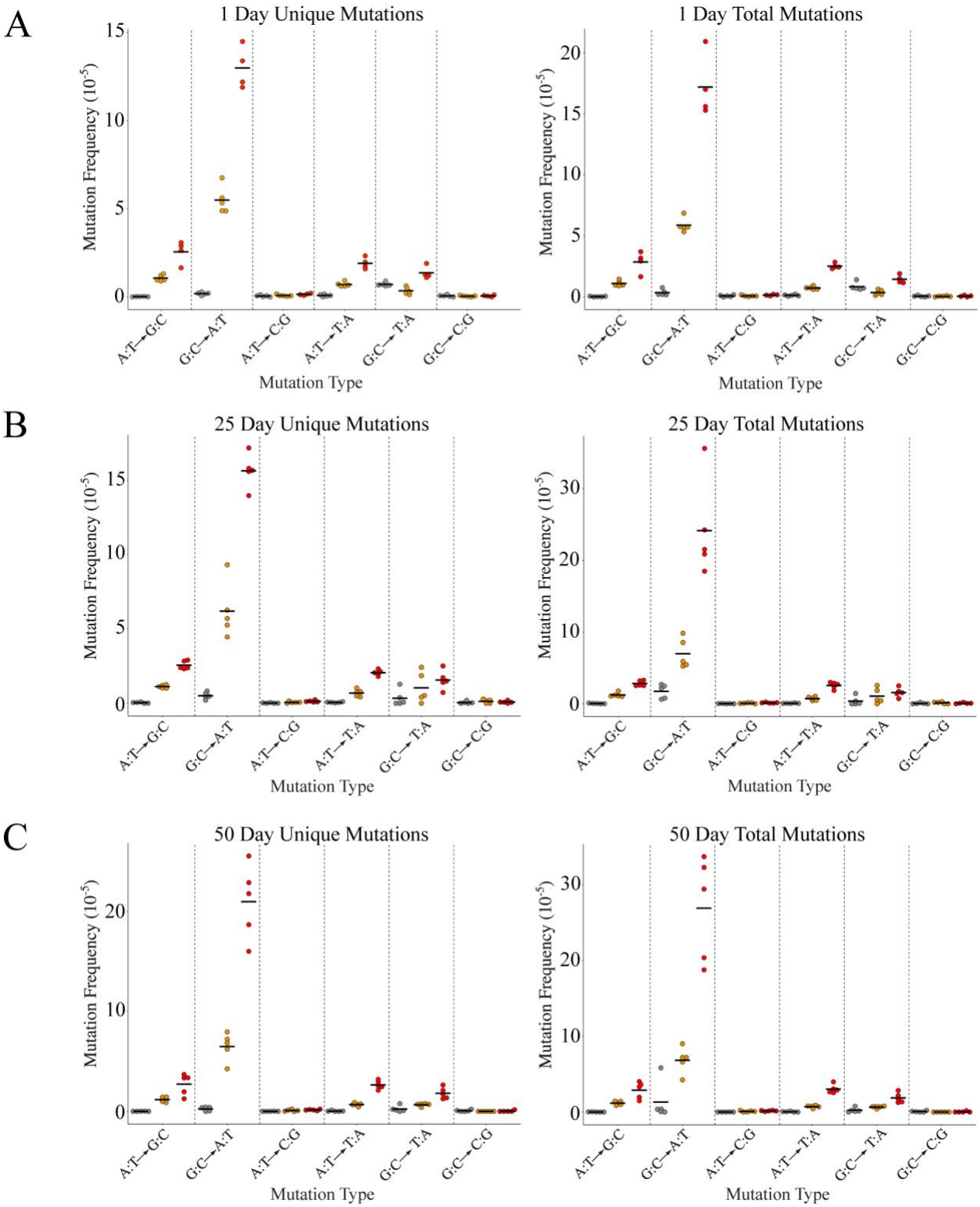


Figure 2.6. The spectra of point mutations in 1-, 25-, and 50-day-old flies. The frequency of each type of base substitution mutation for Unique and Total mutations observed in in **(A)** 1-day-old, **(B)** 25-day-old, and **(C)** 50-day-old flies of the indicated genotype, *n* = 5 flies per genotype per time point.

Table 2.1. Summary statistics of unique mutation frequencies obtained from duplex sequencing of control, 1xPolG$_{mut}$, and 2xPolG$_{mut}$ flies. The number of nucleotides sequenced, mutations identified, and genome-wide mutation frequency of uniquely mutated sites are shown for each fly. These sequencing data represent nucleotides for which sequence reads could be mapped to the *Drosophila* mtDNA reference sequence and duplex consensus strands could be constructed.

1-day-old Flies				25-day-old Flies				50-day-old Flies			
Fly genotype	Nucleotides Sequenced	Number of Mutations	Mutation Frequency	Fly Genotype	Nucleotides Sequenced	Number of Mutations	Mutation Frequency	Fly Genotype	Nucleotides Sequenced	Number of Mutations	Mutation Frequency
0xPolG^{mut} 1	42518691	90	$2.12x10^{-6}$	0xPolG^{mut} 1	5406269	10	$1.85x10^{-6}$	0xPolG^{mut} 1	5679645	13	$2.29x10^{-6}$
0xPolG^{mut} 2	8199467	33	$4.02x10^{-6}$	0xPolG^{mut} 2	19092032	123	$6.44x10^{-6}$	0xPolG^{mut} 2	2215855	8	$3.61x10^{-6}$
0xPolG^{mut} 3	9037683	27	$2.99x10^{-6}$	0xPolG^{mut} 3	2661304	6	$2.25x10^{-6}$	0xPolG^{mut} 3	9677840	7	$7.23x10^{-7}$
0xPolG^{mut} 4	6147590	27	$4.39x10^{-6}$	0xPolG^{mut} 4	11853341	12	$1.01x10^{-6}$	0xPolG^{mut} 4	9506293	16	$1.68x10^{-6}$
0xPolG^{mut} 5	6741158	23	$3.41x10^{-6}$	0xPolG^{mut} 5	4407413	16	$3.63x10^{-6}$	0xPolG^{mut} 5	3364513	0	$<2.97x10^{-7}$
1xPolG^{mut} 1	8070242	226	$2.80x10^{-5}$	1xPolG^{mut} 1	2450795	86	$3.51x10^{-5}$	1xPolG^{mut} 1	8110386	243	$3.00x10^{-5}$
1xPolG^{mut} 2	6288317	181	$2.88x10^{-5}$	1xPolG^{mut} 2	1542728	46	$2.98x10^{-5}$	1xPolG^{mut} 2	9385321	343	$3.65x10^{-5}$
1xPolG^{mut} 3	4066001	137	$3.37x10^{-5}$	1xPolG^{mut} 3	2015717	61	$3.03x10^{-5}$	1xPolG^{mut} 3	1788143	45	$2.52x10^{-5}$
1xPolG^{mut} 4	14242560	372	$2.61x10^{-5}$	1xPolG^{mut} 4	3447601	107	$3.10x10^{-5}$	1xPolG^{mut} 4	1204779	38	$3.15x10^{-5}$
1xPolG^{mut} 5	24762695	613	$2.48x10^{-5}$	1xPolG^{mut} 5	2938793	98	$3.33x10^{-5}$	1xPolG^{mut} 5	3081790	112	$3.63x10^{-5}$
1xPolG$^{mut-Chr2}$ 1	52868988	1186	$2.24x10^{-5}$	2xPolG^{mut} 1	5462486	404	$7.40x10^{-5}$	2xPolG^{mut} 1	2422131	242	$9.99x10^{-5}$
1xPolG$^{mut-Chr2}$ 2	65209719	1526	$2.34x10^{-5}$	2xPolG^{mut} 2	6547279	517	$7.90x10^{-5}$	2xPolG^{mut} 2	1896758	188	$9.91x10^{-5}$
1xPolG$^{mut-Chr2}$ 3	70944805	1480	$2.09x10^{-5}$	2xPolG^{mut} 3	4274349	333	$7.79x10^{-5}$	2xPolG^{mut} 3	3432156	375	$1.09x10^{-4}$
1xPolG$^{mut-Chr2}$ 4	74409046	1688	$2.27x10^{-5}$	2xPolG^{mut} 4	7056676	553	$7.84x10^{-5}$	2xPolG^{mut} 4	6967151	594	$8.53x10^{-5}$
2xPolG^{mut} 1	14409626	967	$6.71x10^{-5}$	2xPolG^{mut} 5	7133239	505	$7.08x10^{-5}$	2xPolG^{mut} 5	5222775	430	$8.23x10^{-5}$
2xPolG^{mut} 2	1951359	159	$8.15x10^{-5}$								
2xPolG^{mut} 3	14911301	951	$6.38x10^{-5}$								
2xPolG^{mut} 4	10731036	675	$6.29x10^{-5}$								

To search for evidence of clonal expansion, we also quantified the total mutation frequency, which in contrast to the unique mutation frequency, includes multiple occurrences of the same mutation. Control flies harbor a low total mutation burden (1-day-old flies had just 4.7 $x10^{-6}$ ± 1.8 $x10^{-6}$ mutations/nucleotide, or ~1 point mutation per 14 mtDNA molecules; Figure 2.3B, Table 2.2) that is similar to that of the unique mutation frequency, suggesting little clonal expansion occurs in WT flies (*62*). The total mutation frequency increased in a dose-dependent manner in mutator flies: 1-day-old $1xPolG^{mut}$ flies had a mutation frequency of $3.0x10^{-5}$ ± $0.4x10^{-5}$ (~one point mutation per 2.2 molecules of mtDNA Figure 2.3B, Table 2.2) and 1-day-old $2xPolG^{mut}$ flies had a mutation frequency of 8.8 $x10^{-5}$ ± 0.9 $x10^{-5}$ (~1.3 point mutations/mtDNA molecule; Figure 2.3B, Table 2.2). Although the total mutation frequency in mutator flies exceeded that of the unique mutation frequency, the increase was small and we discovered very few clonally-expanded mutation sites in flies of any genotype (Figure 2.3C). Furthermore, very few mutations rose above 1% heteroplasmy even in old $2xPolG^{mut}$ flies (Figure 2.7A). These findings suggest that clonal expansion of mtDNA mutations is either restricted to individual cells, or that the short lifespan of *Drosophila* is incompatible with extensive clonal expansion of mtDNA mutations.

Our previous work to measure the mtDNA mutation frequency in *Drosophila* involved the Random Mutation Capture method, which only allowed us to analyze three small parts of the mitochondrial genome, thus precluding detailed analysis of the distribution of mtDNA mutations (*62*). By contrast, DS enabled us to characterize the frequency of mutations across the entire mitochondrial coding sequence. Only the non-coding control region [ChrM:14917-19524] was refractory to DS owing to its high A:T content (~95%), which prevents efficient sequence capture and accurate reassembly. We found that mutations were distributed relatively uniformly between tRNA, rRNA, and protein-coding genes with no apparent mutational hotspots or mutational deserts

Table 2.2. Summary statistics of the overall (total) mutation frequencies obtained from duplex sequencing of control, 1xPolG$_{mut}$, and 2xPolG$_{mut}$ flies. The number of nucleotides sequenced, mutations identified, and genome-wide mutation frequency are shown for each fly for all mutations. These sequencing data represent nucleotides for which sequence reads could be mapped to the *Drosophila* mtDNA reference sequence and duplex consensus strands could be constructed.

1-day-old Flies				25-day-old Flies				50-day-old Flies			
Fly genotype	Nucleotides Sequenced	Number of Mutations	Mutation Frequency	Fly Genotype	Nucleotides Sequenced	Number of Mutations	Mutation Frequency	Fly Genotype	Nucleotides Sequenced	Number of Mutations	Mutation Frequency
0xPolG^{mut} 1	42518691	111	$2.61x10^{-6}$	0xPolG^{mut} 1	5406269	33	$6.10x10^{-6}$	0xPolG^{mut} 1	5679645	13	$2.29x10^{-6}$
0xPolG^{mut} 2	8199467	35	$4.27x10^{-6}$	0xPolG^{mut} 2	19092032	222	$1.16x10^{-5}$	0xPolG^{mut} 2	2215855	8	$3.61x10^{-6}$
0xPolG^{mut} 3	9037683	43	$4.76x10^{-6}$	0xPolG^{mut} 3	2661304	7	$2.63x10^{-6}$	0xPolG^{mut} 3	9677840	11	$1.14x10^{-6}$
0xPolG^{mut} 4	6147590	27	$4.39x10^{-6}$	0xPolG^{mut} 4	11853341	32	$2.70x10^{-6}$	0xPolG^{mut} 4	9506293	147	$1.55x10^{-5}$
0xPolG^{mut} 5	6741158	51	$7.57x10^{-6}$	0xPolG^{mut} 5	4407413	35	$7.94x10^{-6}$	0xPolG^{mut} 5	3364513	0	$<2.97x10^{-7}$
1xPolG^{mut} 1	8070242	240	$2.97x10^{-5}$	1xPolG^{mut} 1	2450795	89	$3.63x10^{-5}$	1xPolG^{mut} 1	8110386	255	$3.14x10^{-5}$
1xPolG^{mut} 2	6288317	192	$3.05x10^{-5}$	1xPolG^{mut} 2	1542728	46	$2.98x10^{-5}$	1xPolG^{mut} 2	9385321	383	$4.08x10^{-5}$
1xPolG^{mut} 3	4066001	138	$3.39x10^{-5}$	1xPolG^{mut} 3	2015717	62	$3.08x10^{-5}$	1xPolG^{mut} 3	1788143	45	$2.52x10^{-5}$
1xPolG^{mut} 4	14242560	398	$2.79x10^{-5}$	1xPolG^{mut} 4	3447601	134	$3.89x10^{-5}$	1xPolG^{mut} 4	1204779	38	$3.15x10^{-5}$
1xPolG^{mut} 5	24762695	709	$2.86x10^{-5}$	1xPolG^{mut} 5	2938793	118	$4.02x10^{-5}$	1xPolG^{mut} 5	3081790	116	$3.76x10^{-5}$
1xPolG$^{mut-Chr2}$ 1	52868988	1686	$3.19x10^{-5}$	2xPolG^{mut} 1	5462486	476	$8.71x10^{-5}$	2xPolG^{mut} 1	2422131	267	$1.10x10^{-4}$
1xPolG$^{mut-Chr2}$ 2	65209719	3052	$4.68x10^{-5}$	2xPolG^{mut} 2	6547279	872	$1.33x10^{-4}$	2xPolG^{mut} 2	1896758	200	$1.05x10^{-4}$
1xPolG$^{mut-Chr2}$ 3	70944805	1742	$2.46x10^{-5}$	2xPolG^{mut} 3	4274349	461	$1.08x10^{-4}$	2xPolG^{mut} 3	3432156	496	$1.45x10^{-4}$
1xPolG$^{mut-Chr2}$ 4	74409046	2094	$2.81x10^{-5}$	2xPolG^{mut} 4	7056676	690	$9.78x10^{-5}$	2xPolG^{mut} 4	6967151	688	$9.87x10^{-5}$
2xPolG^{mut} 1	14409626	1427	$9.90x10^{-5}$	2xPolG^{mut} 5	7133239	654	$9.17x10^{-5}$	2xPolG^{mut} 5	5222775	595	$1.14x10^{-4}$
2xPolG^{mut} 2	1951359	174	$8.92x10^{-5}$								
2xPolG^{mut} 3	14911301	1286	$8.62x10^{-5}$								
2xPolG^{mut} 4	10731036	839	$7.82x10^{-5}$								

(Figure 2.5, Figure 2.7A,B). The mild variation in mutation frequency detected between genes is likely explained by differences in GC content and the G:C to A:T mutation bias of the polymerase (Figure 2.7C).

2.3.2 *Mutator flies exhibit reduced longevity, progressively worsening locomotor ability, neurodegeneration and mitochondrial dysfunction*

Mutator mice display premature aging phenotypes, including a reduced lifespan, kyphosis, and hair loss (*106*). Therefore, we tested whether mutator flies also exhibit signs of premature aging by examining lifespan and locomotor activity. Control flies had a median lifespan of 75 days,

whereas the 1xPolG$_{mut}$ flies had a modest reduction in lifespan, displaying a median lifespan of 64 days (Figure 2.8A). 2xPolG$_{mut}$ flies showed a further reduction in lifespan, with a median lifespan of 53 days. Mutator flies also displayed a defect in locomotor performance using a simple test of climbing ability. Normal flies exhibit negative geotaxis, climbing to the top of a vial after being tapped to the bottom, and this behavior declines naturally as flies age. The presence of the *PolG$_{mut}$* transgene did not influence climbing ability in young flies, but conferred a dose-dependent decline in climbing behavior in older flies relative to controls (Figure 2.8B). The decline in climbing ability preceded the onset of decreased viability in 2xPolG$_{mut}$ flies and ultimately culminated in a complete failure in climbing at ages >25 days. Approximately 10% of 2xPolG$_{mut}$ flies also exhibited a rhythmic seizure phenotype beginning approximately 24 hours prior to death. This phenotype was not observed in any other genotype.

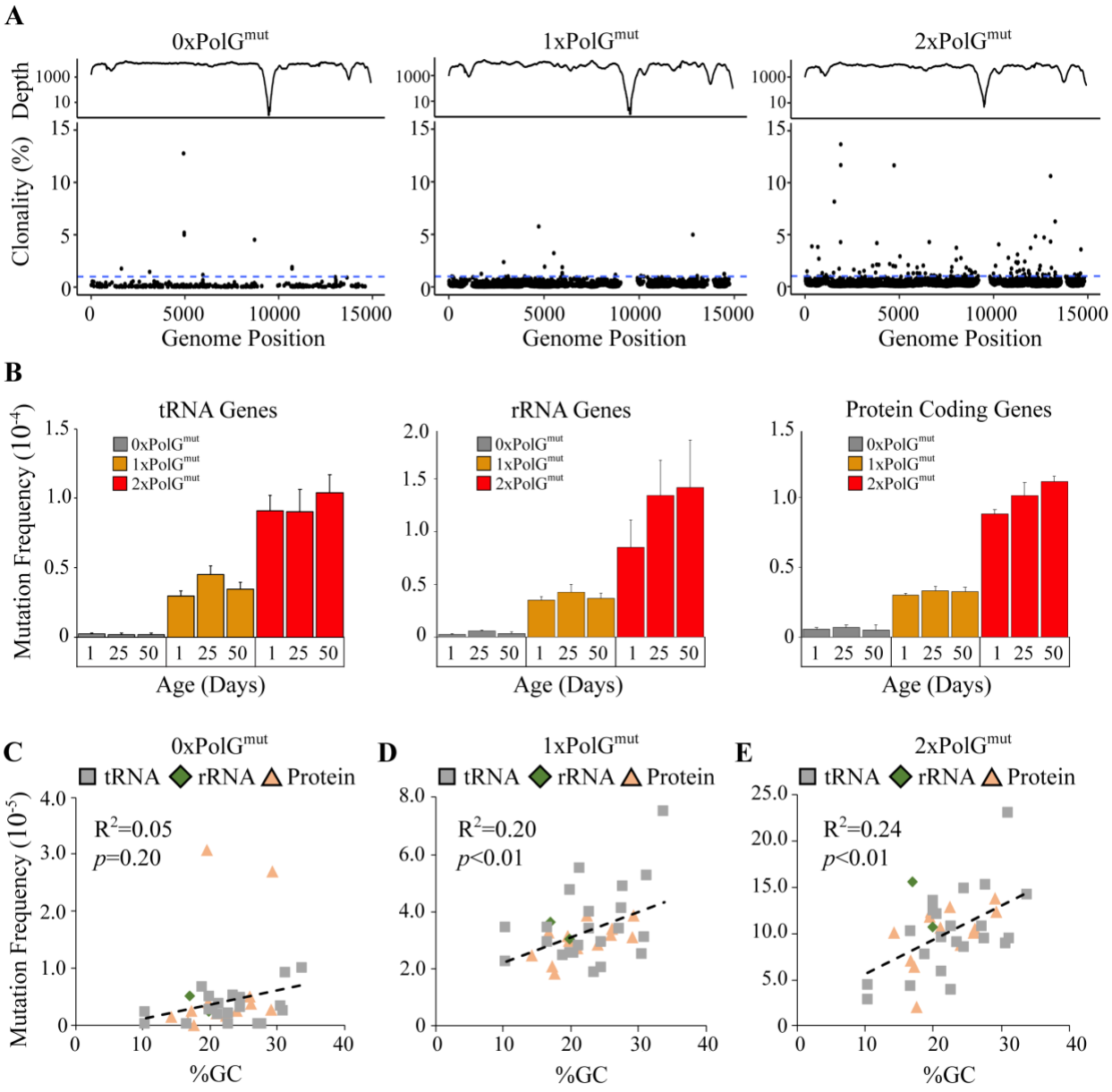


Figure 2.7. Mutation frequency is relatively uniform across the mitochondrial genome. (A) Sequencing depth (log_{10}-transformed y-axis) and percent heteroplasmy of mutations across the mitochondrial genome in flies of all ages for the indicated genotype. 1% heteroplasmy is indicated by the dashed blue line. **(B)** Mutation frequency of mitochondrial tRNAs, rRNAs and protein coding genes for flies of the given genotype and age are shown. The correlation between GC content and mutation frequency in **(C)** tRNA, **(D)** rRNA, and **(E)** protein coding sequences for flies of all ages in the genotypes indicated. *p*-value was determined by Pearson correlation.

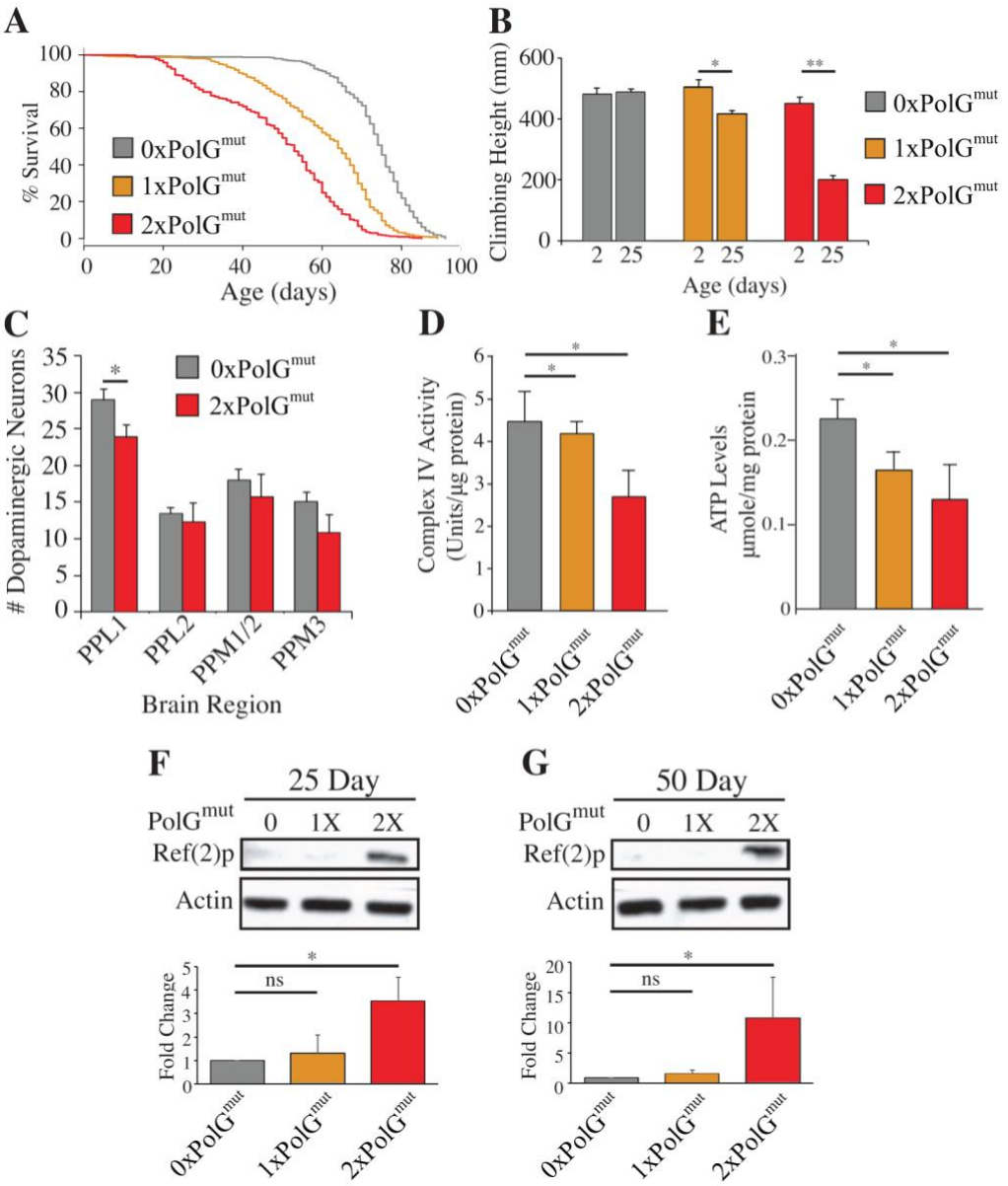


Figure 2.8. mtDNA mutator flies exhibit shortened lifespan, a locomotion defect, neuron loss, and mitochondrial dysfunction. (A) The lifespan of male flies of the indicated genotypes are shown (n≥364 flies per genotype). **p* < .001 determined by log-rank test. **(B)** The mean distance climbed by male flies of the given ages and genotypes in 3 seconds (n≥54 flies per genotype per time point). **(C)** Quantification of dopaminergic neurons within the indicated clusters of in 50-day-old control and 2xPolGmut flies. PPL1, protocerebral posterior lateral; PPM, protocerebral posterior medial **(D)** Complex IV activity is inversely correlated with PolGmut dosage in 14-day-old flies (*n* = 4 male flies per genotype). **(E)** ATP abundance is similarly decreased in 14-day-old 1xPolGmut and 2xPolGmut flies. **(F) and (G)** Ref(2)P/p62 abundance is selectively increased in 25- and 50-day-old 2xPolGmut flies, respectively. Error bars in panels B and C represent standard deviation, error bars in panels D-G represent standard error. **p* < 0.05; ***p* < 0.01 by Student's t-test.

A number of previous observations indicate that dopaminergic neurons are particularly sensitive to mitochondrial stress (*115, 162-164*). In particular, mitochondrial toxins and mutations affecting mitochondrial quality control pathway components result in the selective death of dopaminergic neurons in humans and animal models (*165*). Moreover, a recent study has shown that mice expressing a proofreading-deficient form of DNA polymerase γ exhibit selective dopaminergic neuron death when crossed to *parkin* mutant mice (*108*). These observations prompted us to test whether an increased load of mtDNA mutations leads to degeneration of dopaminergic neurons in *Drosophila*. To perform this analysis, we dissected whole brains and immunostained with antiserum to tyrosine hydroxylase to quantify the number of dopaminergic (DA) neurons in 50-day-old flies. The number of DA neurons in the protocerebral posterior lateral 1 (PPL1) cluster was significantly reduced in 2xPolG$_{mut}$ flies (Figure 2.8C, Figure 2.9), suggesting that a high mtDNA mutational load triggers the loss of a subset of dopaminergic neurons consistent with these prior reports (*147*).

To test whether the premature aging phenotypes of mutator flies are caused by mitochondrial dysfunction, we monitored several mitochondrial functional parameters. Since PolG mutator mice show reduced Complex IV activity and assembly (*108*) , we assayed Complex IV activity in mutator flies and found it to be significantly reduced in a dose-dependent fashion relative to age-matched controls (Figure 2.8D). Consistent with this deficit, mutator flies also had a reduced abundance of ATP (Figure 2.8E). 2xPolG$_{mut}$ flies also had elevated levels of the autophagy marker Ref(2)p (the *Drosophila* homolog of p62), possibly suggesting that there is a buildup of autophagic intermediates in response to an upregulation of autophagy (Figure 2.8F,G). However, the abundance of the mitochondrial unfolded protein stress markers HSP60 and mitochondrial HSP70 were unchanged in mutator flies (Figure 2.10), indicating that the increased

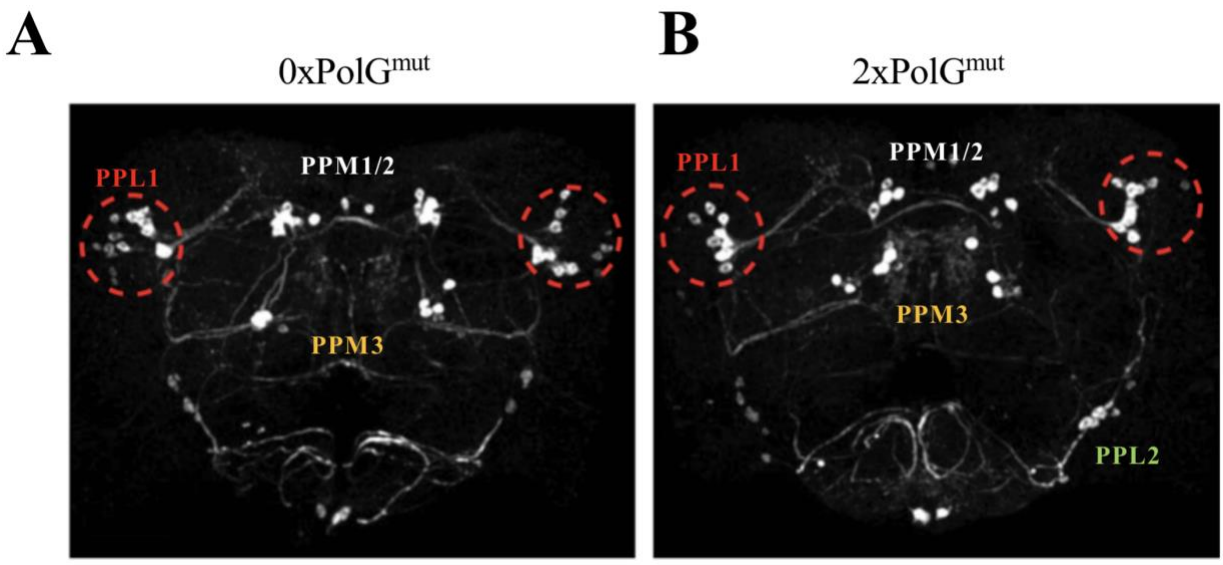


Figure 2.9. Dopaminergic neuron loss in mutator flies. Representative confocal images of immunostained brains of 50-day-old **(A)** 0xPolG$_{mut}$ and **(B)** 2xPolG$_{mut}$ flies. Dopaminergic neurons were immunostained using tyrosine hydroxylase antiserum. The PPL1-2 and PPM1-3 clusters of dopaminergic neurons are indicated.

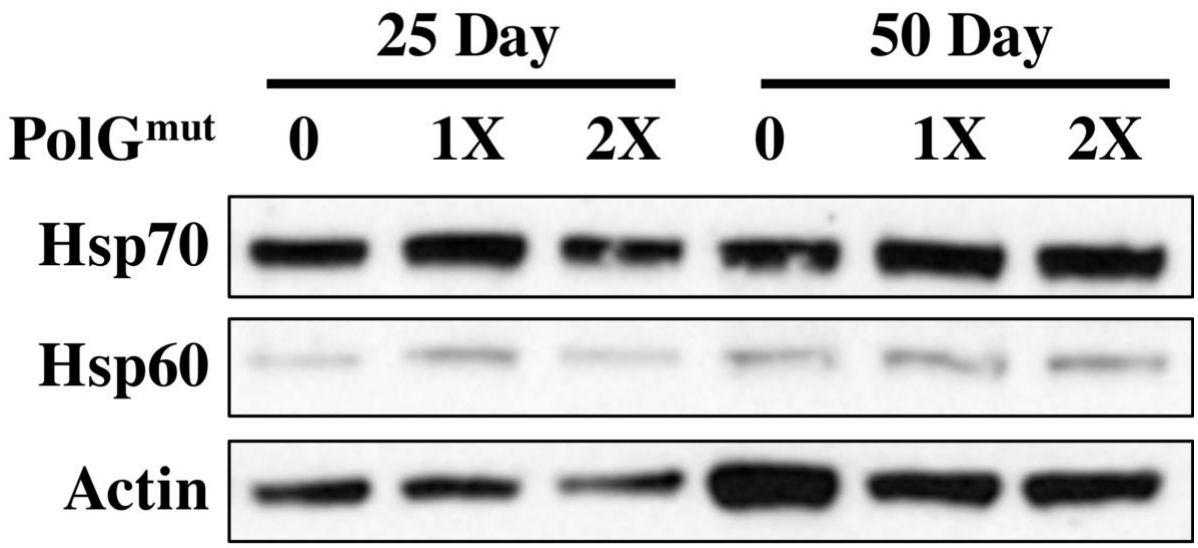


Figure 2.10. The abundance of the mitochondrial unfolded protein stress markers Hsp60 and mitochondrial Hsp70 are unchanged in 1xPolG$_{mut}$ or 2xPolG$_{mut}$ flies. Western blots of 25- and 50-day-old 1xPolG$_{mut}$ (1X) and 2xPolG$_{mut}$ (2X) flies. Actin was used as a loading control. Images are representative of four biological replicates.

load of mtDNA mutations does not result in sufficient protein misfolding to trigger activation of the mitochondrial unfolded protein stress response. We also observed an abnormal downturned wing posture in 2xPolG$_{mut}$ mutator flies at 35 days of age, similar to that seen in *Drosophila PINK1* and *parkin* mutants (*48*). However, in contrast to *PINK1* and *parkin* mutants, both of which exhibit apoptotic muscle degeneration, there was no gross evidence of muscle degeneration or apoptosis in mutator flies. Ultrastructural examination of flight muscle tissue also failed to detect alterations in mitochondrial morphology or integrity (Figure 2.11). Together, these results suggest that flies harboring high mutation loads suffer from non-structural muscle abnormalities, thus potentially making these flies a suitable model for the study of mitochondrial myopathies.

2.3.3 *Deleterious mtDNA mutations are overrepresented in mutator flies*

Studies in cultured cells have indicated the existence of pathways that can be manipulated to decrease the frequency of a deleterious heteroplasmic mutation (*102, 105, 166, 167*). However, there is little evidence that these pathways are normally operative in the somatic tissues of an intact animal

model. To address this matter, we subjected mutator flies to a variety of analyses aimed at the detection of selective forces acting against harmful mutations. To account for the clonality of mutations, we used the frequency of total mutations in all of our remaining analyses. To diminish the influence of sampling bias and increase the number of mutations detected per animal, we re-sequenced 1-day-old 1xPolG$_{mut}$ flies using a technical advance in reagent preparation for the Duplex Sequencing protocol that was developed during the course of our study, thus providing us with a high-quality dataset with very high sequencing depth. The use of 1xPolG$_{mut}$ flies for this

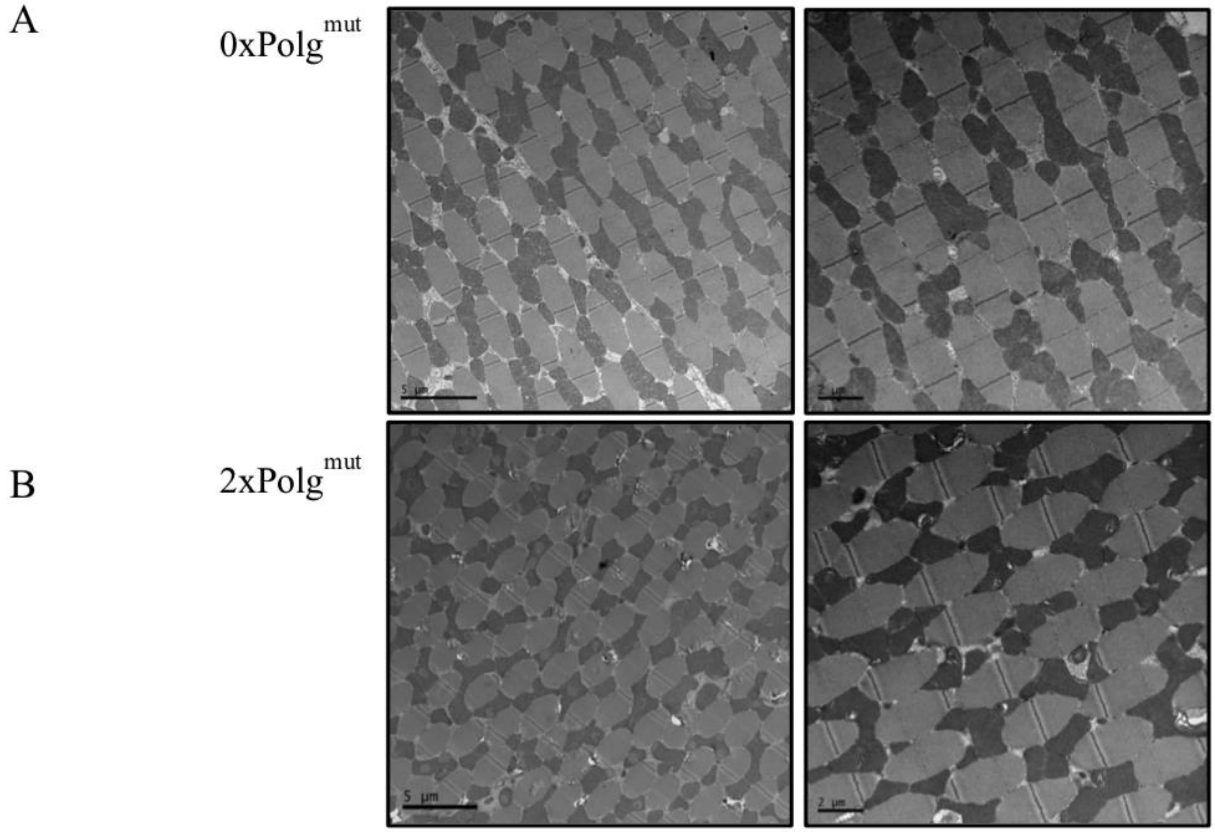


Figure 2.11. Mutator flies do not have mitochondrial morphological alterations. Electron microscopy of indirect flight muscle of 50-day-old **(A)** 0xPolG$_{mut}$ and **(B)** 2xPolG$_{mut}$ flies does not reveal additional mitochondrial ultrastructural defects in mutator flies.

analysis also ensures that the mutations detected from sequencing are acquired in somatic tissues (Figure 2.2).

Because the third codon position is often degenerate, we hypothesized that negative selection acting against deleterious variants would result in a lower frequency of mutations at the first two codon positions relative to the third codon position. Our results were in complete reverse to our expectations: we detected higher mutation frequencies at the first and second codon positions (Figure 2.12A). However, a confound in this analysis concerns the high frequency of mutations at G:C base pairs, and the relative deficiency of G:C base pairs in the third codon position (Figure 2.13). Notably, the AT-rich *Drosophila* mitochondrial genome, like many other insect species, primarily consists of A:T bases at four-fold degenerate (i.e., synonymous) sites, where the frequency of A:T base pairs is 94% (*168*, *169*). By contrast, G:C base pairs are predominantly located at nonsynonymous (NS) sites. To circumvent this confound, we compared the mutation frequency at NS sites and at four-fold synonymous (S) sites separately for A:T and G:C base pairs. Because mutations arising at four-fold degenerate sites do not alter the encoded amino acid, such mutations should be present at higher frequency relative to those at NS sites in the context of negative selection acting against deleterious variants. We detected no significant difference in the mutation frequency between NS sites and S sites at A:T positions. However, at G:C positions the mutation frequency was higher at NS sites relative to S sites, in complete opposition to the expectations of a negative selection model (Figure 2.12B).

The unexpected finding that deleterious mtDNA variants were overrepresented in *Drosophila* prompted us to examine this matter further. Specifically, we performed Monte Carlo simulations of random mutagenesis, such that we could compare our findings from sequencing mutator flies to a neutral mutational model derived from simulations. Because the mutation

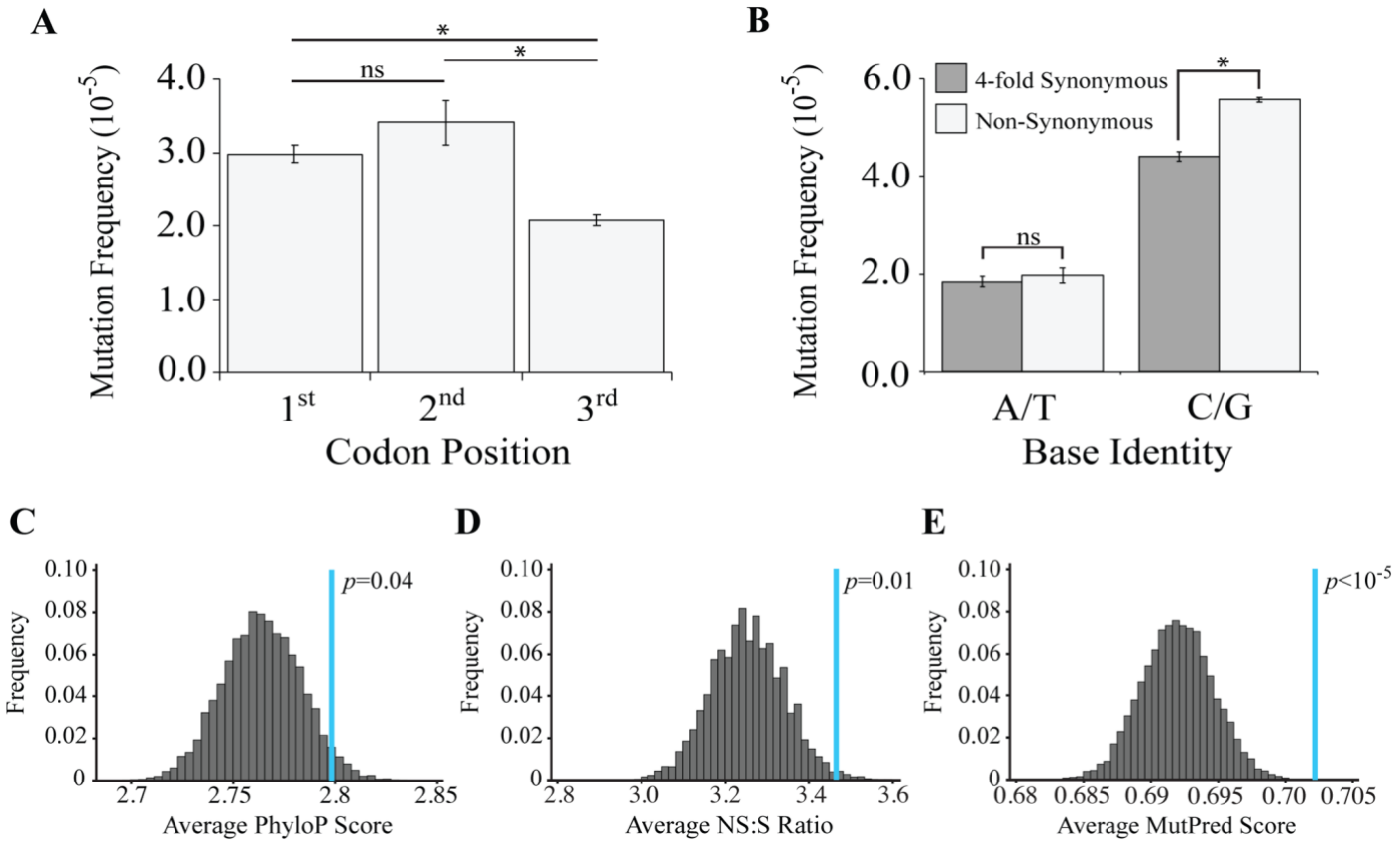


Figure 2.12. Mutator flies accumulate an excess of deleterious mtDNA mutations. (A) The mutation frequency of 1-day-old 1xPolGmut flies per codon position. **(B)** The mutation frequency at four-fold degenerate (synonymous; S) and non-synonymous (NS) sites. **(C)** The distribution of PhyloP scores from 10,000 trials of simulated mutagenesis of the *Drosophila* mitochondrial genome under conditions of neutrality. The blue line indicates the observed average PhyloP score from 1-day-old 1xPolGmut flies. **(D)** The distribution of NS/S ratios observed from 10,000 trials of simulated mutagenesis of the *Drosophila* mitochondrial genome under conditions of neutrality. The blue line indicates the observed average NS/S ratio in 1-day-old 1xPolGmut flies. **(E)** The distribution of MutPred scores from 10,000 trials of simulated mutagenesis of the *Drosophila* mitochondrial genome under conditions of neutrality. The blue line indicates the observed average MutPred score in 1-day-old 1xPolGmut flies. *p*-values for A and B determined through pooled sample Z-test. *p*-values in panels C-E determined empirically

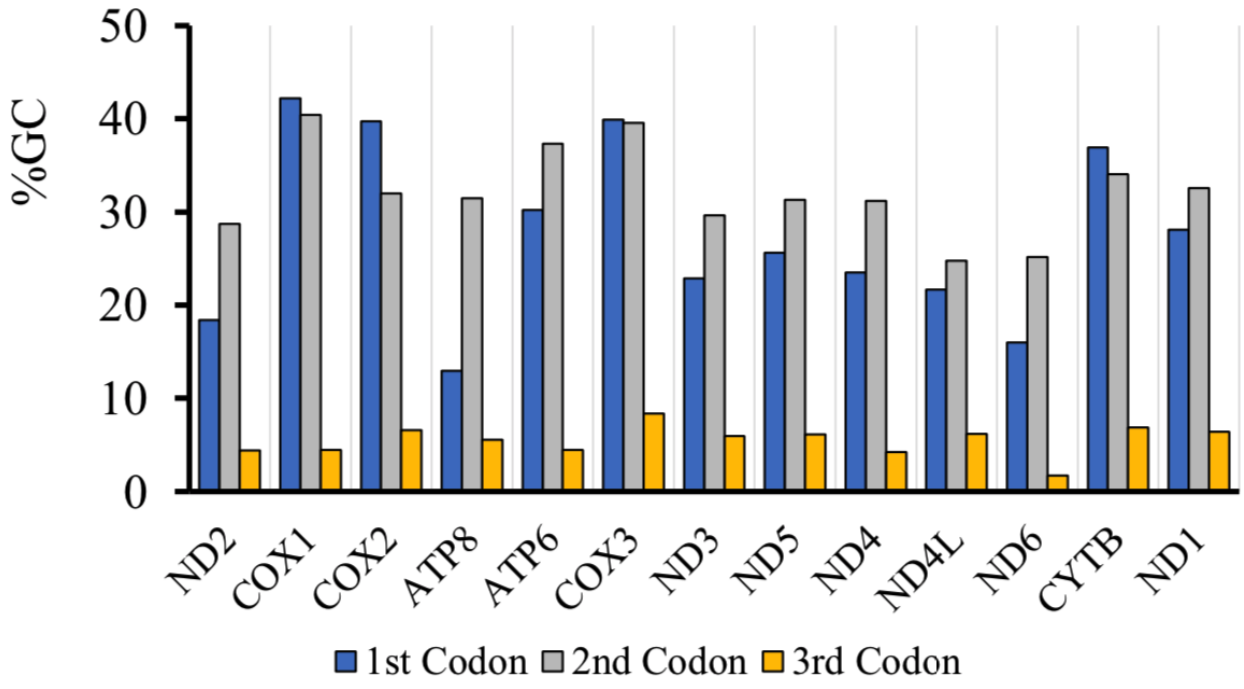


Figure 2.13. Third codon sites in protein-coding genes have reduced GC content. All protein-coding genes show depleted GC content in third codon positions, as calculated from the *Drosophila* reference genome.

frequency at G:C sites greatly exceeds that at A:T sites, we performed simulations to precisely mirror the mutational biases detected in mutator flies. Each round of simulation selected the proportion of each type of mutation detected within protein-coding regions of 1xPolG$_{mut}$ flies and computationally redistributed these mutations randomly throughout the coding portions of the mtDNA. Moreover, because the probability of detecting a mutation at any given site in the mitochondrial genome is directly proportional to the sequencing depth at that site, we also weighted the probability of detecting a nucleotide alteration at each site according to sequencing depth at that site in our simulations (for further details, see the Materials and Methods section). Simulations were repeated 10,000 times to create a distribution of neutral outcomes, thus providing a framework to compare findings from 1xPolG$_{mut}$ flies.

We performed three analyses designed to detect whether negative selection reduces the frequency of pathogenic mtDNA mutations. First, we tested the hypothesis that mutations at evolutionarily conserved mtDNA sites are more susceptible to negative selection than those at sites of low conservation. Second, we tested the hypothesis that mutations resulting in nonsynonymous (NS) alterations are more prone to negative selection than synonymous (S) mutations. Third, we tested the hypothesis that mutations resulting in deleterious amino acid alterations are more susceptible to negative selection than those that result in conservative amino acid alterations. Our hypotheses predict that mtDNA mutations arising at highly conserved sites, particularly those mutations that result in deleterious NS amino acid alterations, are eliminated through negative selection. Consequently, these deleterious mutations should be underrepresented in mutator flies relative to a distribution of randomly generated mutations.

To test the prediction that mutations at conserved sites are underrepresented in mutator flies, we used the PhyloP algorithm, which calculates mtDNA positional conservation scores from

pairwise comparisons between 27 insect species (*170*). The PhyloP algorithm assigns a logarithmic score to each nucleotide position indicating the degree of evolutionary conservation at that site; a score of zero indicates neutral evolution, whereas negative and positive scores suggest accelerated evolution and increased conservation, respectively. We then compared the average PhyloP scores from the mutations identified in mutator flies to a distribution of PhyloP scores created from Monte Carlo simulations of random mutagenesis. Consistent with our comparison of NS and four-fold S sites, we found that mutations at sites with high PhyloP scores are overrepresented, again suggesting that deleterious variants are overrepresented in mutator flies (Figure 2.12C).

We next tested whether NS mutations were underrepresented in mutator flies by calculating the NS/S ratio and comparing this ratio to a distribution of NS/S ratios obtained from Monte Carlo simulation of random mutagenesis (neutrality). Our hypothesis that negative selection preferentially eliminates nonsynonymous mutations predicts a reduction in the occurrence of nonsynonymous mutations and thus the observed NS/S ratio should be lower than expected from neutrality. In contrast to our hypothesis, the NS/S ratio detected in mutator flies is significantly elevated relative to neutrality, indicating that nonsynonymous mutations are overrepresented in mutator flies (Figure 2.12D). These findings are inconsistent with the hypothesis that negative selection acts against nonsynonymous mutations.

As a final test of the model that negative selection acts against deleterious mtDNA variants, we used the MutPred algorithm (*171*) to compare the pathogenicity of nonsynonymous mutations found in mutator flies to a distribution of MutPred scores created from Monte Carlo simulations of random mutagenesis. The MutPred algorithm uses the structural and functional properties of a protein to predict the functional consequence of a nonsynonymous amino acid substitution, and previous work has established the validity of the MutPred algorithm to predict the consequences

of mtDNA mutations (*96, 172, 173*). MutPred assigns scores ranging from 0 to 1 to quantify the pathogenicity of a particular variant, with higher scores indicating a greater likelihood of pathogenicity. If negative selection acts to preferentially remove the most deleterious mutations, mutator flies should accumulate variants with low MutPred scores. In contrast to this prediction, the detected mutations have an average MutPred score that is significantly higher than expected from a neutral mutational model (Figure 2.12E). To confirm findings from simulations, we added the combined data from our previously acquired low-depth sequencing of 1x and 2x$PolG_{mut}$ flies in an effort to increase the total number of mutations. We then reran our simulations using this data and applied the same three metrics to ask if there is selection against harmful mutations. Results of this analysis again revealed that deleterious mutations are overrepresented in mutator flies (Figure 2.14).

Our previous work revealed a strand asymmetry in the occurrence of C:G to T:A mutations (*62*), and this phenomenon could potentially influence our distributions of simulated mutations. To eliminate the potential confound of a mutational strand asymmetry, we repeated our analysis using the coding sequence of the *Mitochondrial Cytochrome c oxidase subunit I* (*COX1*) gene, which does not exhibit strand asymmetry in the mutation spectrum (Figure 2.15A). While we did not detect a significant difference in the evolutionary conservation of mutations arising in *COX1* (Figure 2.15B) relative to neutrality, we discovered that these mutations displayed an elevated NS/S ratio (Figure 2.15C) and an overrepresentation of mutations with high MutPred Scores (Figure 2.15D), indicating that our findings are not a consequence of mutational strand asymmetry. Taken together, our analyses do not support the model that selection acts against deleterious mtDNA mutations; instead, our findings indicate that pathogenic mtDNA variants are overrepresented in mutator flies.

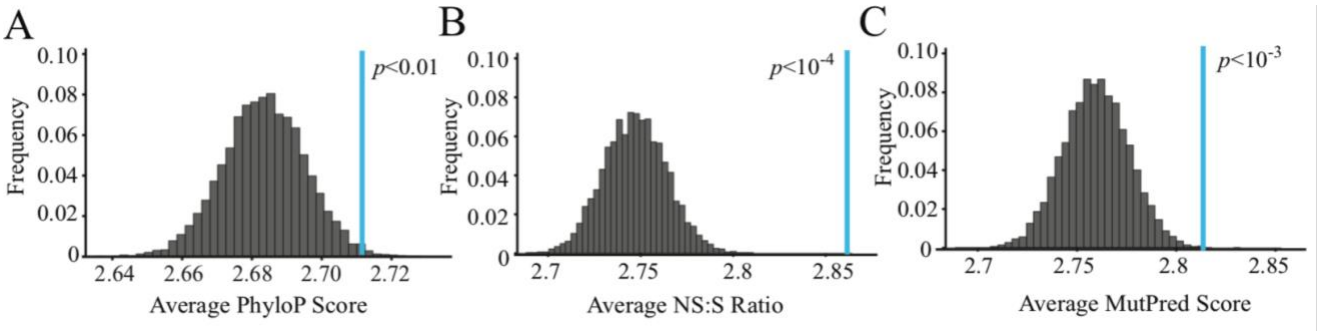


Figure 2.14. Deleterious nonsynonymous mutations are overrepresented in a pooled sample of all mutator flies. (A) The distribution of average PhyloP scores of mutations in the protein-coding sequence expected under neutrality, calculated from 10,000 trials of simulated mutagenesis under conditions of neutrality. Sequence data from 1x and 2xPolG$_{mut}$ flies of all ages were aggregated and Monte Carlo resampling simulations were performed as above. The blue line indicates the average PhyloP score of all mutations observed in the 1xPolG$_{mut}$ and 2xPolG$_{mut}$ flies. **(B)** The distribution of average NS/S ratios of mutations expected under neutrality, calculated from 10,000 simulations of random mutagenesis from the combined sequence data of 1xPolG$_{mut}$ and 2xPol*G$_{mut}$* flies. The blue line indicates the average NS/S ratio of mutations observed across 1xPolG$_{mut}$ and 2xPolG$_{mut}$ flies. **(C)** The distribution predicted average MutPred scores of NS variants calculated from 10,000 simulations of random mutagenesis as described above. The blue line indicates the observed average MutPred score in NS variants found in 1xPolG$_{mut}$ and 2xPolG$_{mut}$ flies. *p*-values were determined empirically.

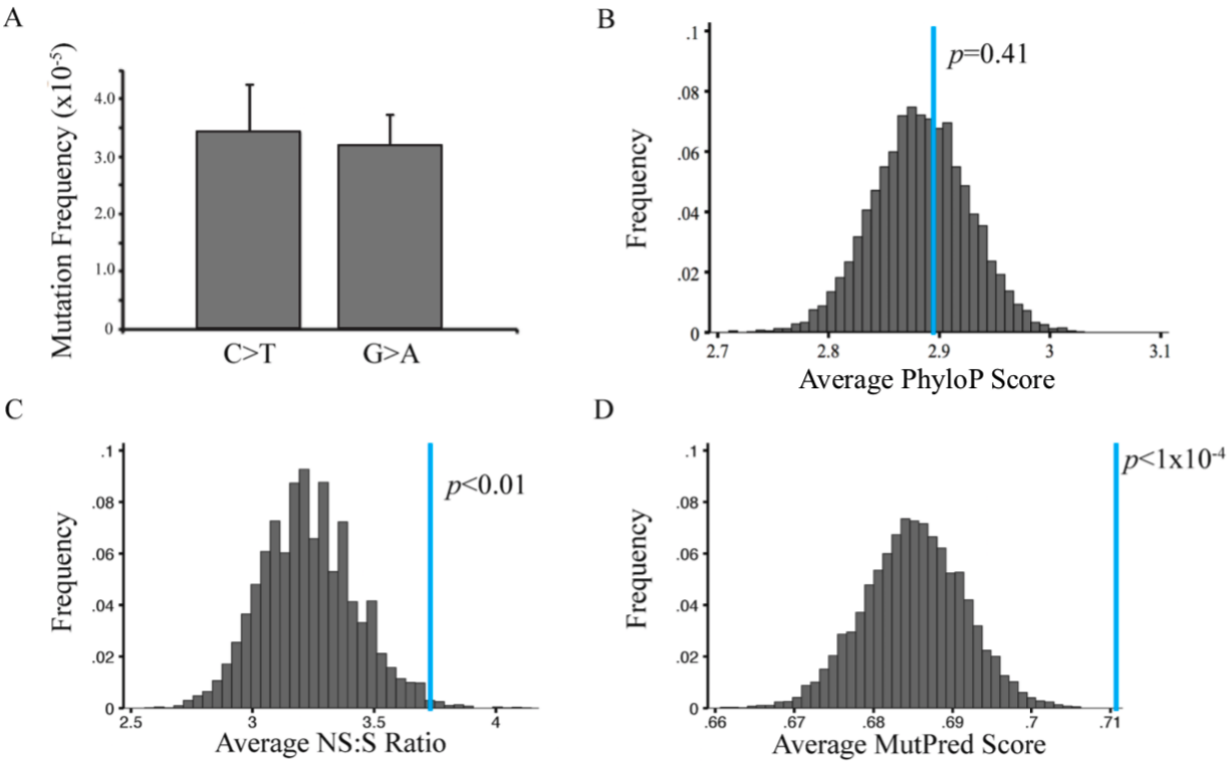


Figure 2.15 Deleterious nonsynonymous mutations are overrepresented within the coding sequence of the *COX1* gene. (A) Mutations within the *COX1* coding sequence display no strand bias for the predominant mutation type, G:C to A:T transitions. **(B)** The distribution of average PhyloP scores of mutations in the COX1 coding sequence expected under neutrality, calculated from 10,000 trials of simulated mutagenesis under conditions of neutrality. Only mutations observed within the COX1 gene [ChrM 1474-3009] were used for Monte Carlo resampling in these simulations. The blue line indicates the average PhyloP score of mutations observed in the COX1 coding sequence of 1-day-old 1xPolG$_{mut}$ flies. **(C)** The distribution of average NS/S ratios of mutations in the *COX1* coding sequence expected under neutrality, calculated from 10,000 simulations of random mutagenesis as described above. The blue line indicates the average NS/S ratio of mutations observed in the *COX1* coding sequence of 1-day-old 1xPolG$_{mut}$ flies. **(D)** The distribution predicted average MutPred scores of NS variants observed in the Drosophila *COX1* coding region under conditions of neutrality, calculated from 10,000 simulations of random mutagenesis as described above. The blue line indicates the observed average MutPred score in NS variants found in the *COXI* sequence in 1-day-old 1xPolG$_{mut}$ flies. *p*-values in panels B-D determined empirically.

2.3.4 *The trinucleotide context of mutation sites does not explain the overrepresentation of deleterious mutations*

One potential explanation for the overabundance of deleterious mutations in mutator flies is that nucleotide context influences the mutation frequency and, by chance, results in a higher mutation frequency at functionally important sites. Such an occurrence was observed in mutator mice where the high abundance of C>T transitions within the 'TCA' context was suggested to explain the high pathogenicity of mtDNA mutations in this organism (*115*). To explore this model, we examined the influence of trinucleotide context of the nucleotides directly 5' and 3' to the mutation bearing sites on mutation frequency. We first began by computing the number of mutations observed in all 96 trinucleotide contexts. Like mutator mice, we found that C>T mutations within the TCA context represented one of the most abundant trinucleotide mutation categories. However, unlike mutator mice, the TCA context was not the most abundant mutation category and the abundance of mutations within the TCA context was similar to the frequency of other C>T mutation contexts (Figure 2.16A). Also, despite the high frequency of G:C>A:T mutations, trinucleotide contexts associated with T>C transitions, such as the ATT trinucleotide context, were similarly abundant to those associated with G:C>A:T mutations (Figure 2.16A).

A limitation of comparing the total abundance of mutations within a particular trinucleotide context is that this value will be influenced by the prevalence of that context in the genome. In addition, our ability to detect mutations at any given site in the genome will depend on sequencing depth at that site. Thus, to test whether the trinucleotide context influences mutation frequency, we calculated the trinucleotide mutation frequency by normalizing the raw number of mutations observed in each trinucleotide context to their prevalence in the genome and to the sequencing depth at these sites (see Materials and Methods). Analysis of the trinucleotide mutation frequency substantially decreased the heterogeneity across the entire C>T trinucleotide context

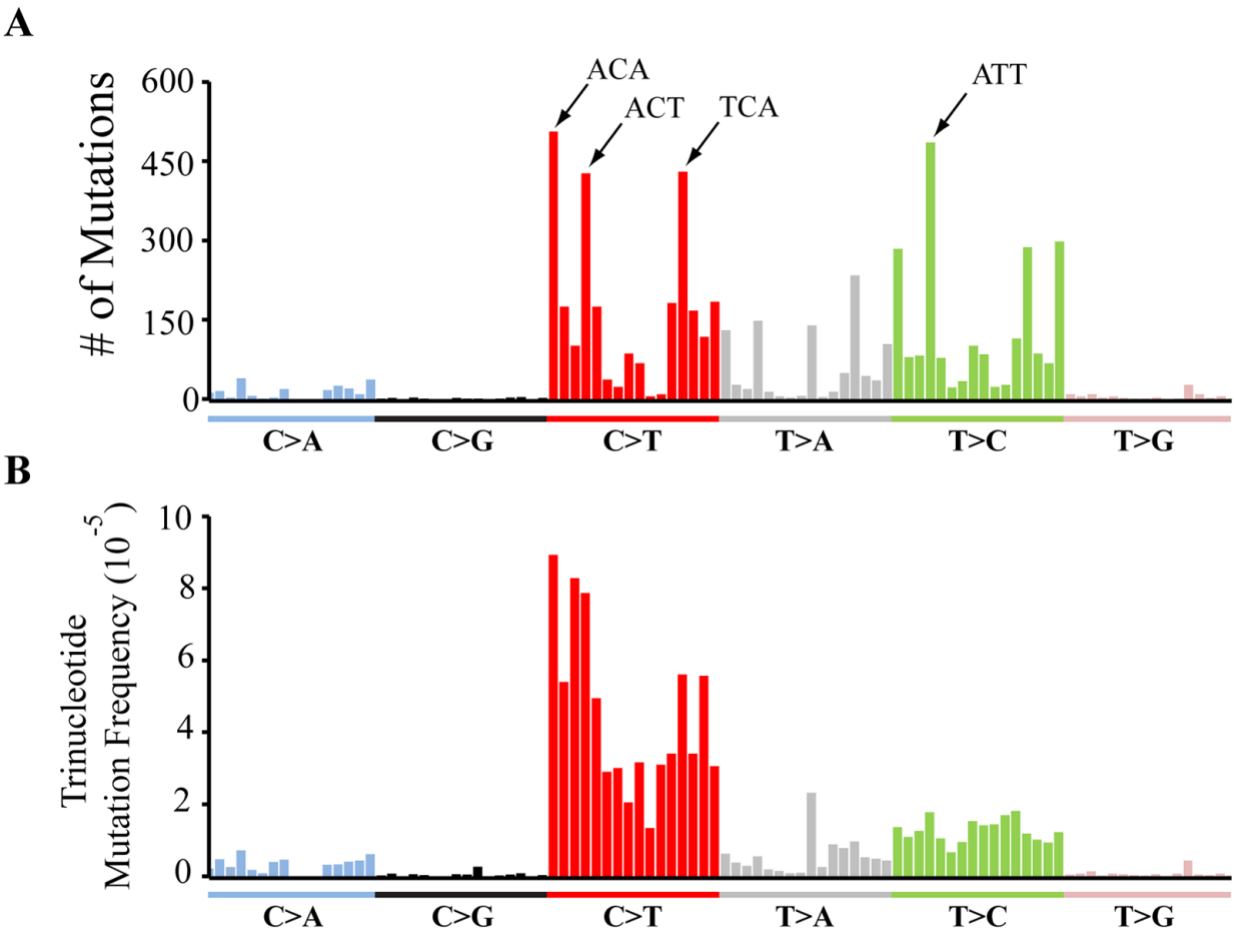


Figure 2.16. The trinucleotide context of mutations does not explain the overrepresentation of deleterious mutations. (A) The raw number of mutations detected in 1xPolG$_{mut}$ flies within all 96 possible trinucleotide contexts. **(B)** The trinucleotide mutation frequency at all 96 possible trinucleotide contexts. Trinucleotide contexts are presented in the order previously characterized (*174*), and the four most highly mutated contexts annotated in panel A.

(Figure 2.16B). Normalizing our data in this fashion also revealed that the high abundance of particular T>C mutation contexts reflects the composition of the *Drosophila* mitochondrial genome rather than an influence of sequence context on mutation frequency (Figure 2.16B). Given the similarity of the mutation frequency within any given trinucleotide context, we conclude that the trinucleotide context does not play a major role in the overrepresentation of deleterious mutations.

2.4 ***Discussion***

The progressive accumulation of deleterious mtDNA mutations in somatic tissues is implicated in aging and common diseases of the elderly. Although the frequency of these mutations correlates with the severity of the symptoms that they cause (*12*), little is known about the pathways that influence the frequency and pathogenicity of mtDNA mutations. In previous work, we showed that many of the features associated with mtDNA mutations are similar in flies and vertebrates, including a low frequency of mutations and a preponderance of transition mutations (*62, 155*). Our current work offers further evidence in support of these previous findings by using a high accuracy next-generation sequencing approach that allowed us to sensitively detect mtDNA mutations over a broader region of the mitochondrial genome (*139*). We have also extended the utility of *Drosophila* as a model system for studying mtDNA mutations by creating a *Drosophila* mtDNA mutator strain that exhibits a dramatically increased mtDNA mutation frequency and displays features associated with mitochondrial diseases and premature aging including mitochondrial dysfunction, reduced lifespan, a progressive locomotor deficit, and loss of dopaminergic neurons. Surprisingly, deleterious mtDNA mutations were overrepresented in mutator flies, suggesting the existence of a novel selective mechanism underlying this phenomenon. Our current work provides a foundation to explore the factors responsible for the

overabundance of harmful mtDNA mutations and the pathways that influence the pathogenicity of these mutations.

Our findings warrant comparison to another recently described *Drosophila* mtDNA mutator strain that was created by introducing the same exonuclease mutation used in our current study into the endogenous *PolG* locus (*155*). While heterozygotes for this *PolG* knock-in allele displayed a similar increase in the mtDNA mutation frequency to flies that bear a single copy of our mutator transgene, homozygotes for the knock-in *PolG* allele did not survive beyond the larval stage of development. By contrast, a single copy of our $PolG^{mut}$ transgene rescued the recessive lethal phenotype of an overlapping set of deletions that completely remove the endogenous *PolG* gene. Given that our mutator transgene appears to express endogenous levels of *PolG* transcript, we are at a loss to explain this discrepancy. Moreover, heterozygotes for the *PolG* knock-in allele displayed no obvious effect on lifespan (*175*), whereas our mutator lines displayed a dose dependent decrease in lifespan. The explanation for this difference is also unknown but may reflect differences in the mtDNA genetic background in these two studies, which we have found in unpublished work can influence lifespan. Further study of the knock-in *PolG* mutant revealed that the mtDNA mutation frequency progressively increased in subsequent generations, indicating that purifying selection is unable to fully keep pace with an increased mtDNA mutation rate (*155*). Although we have not directly explored the influence of our mutator transgene in the female germline, we find that mutator stocks lose viability within several generations without periodic outcrossing, consistent with the model that the accumulation of mtDNA mutations is responsible for this loss of viability.

The finding that somatic mtDNA mutations accumulate with age at an accelerated rate relative to mutations in the nuclear genome has led to the suggestion that aging may be a

consequence of accumulated mtDNA mutations (*12*). While the phenotypes of homozygous PolG mutator mice were initially offered as support for this hypothesis, it was later found that heterozygous mutator mice live a normal lifespan despite having a mtDNA point mutation frequency that greatly exceeds that of elderly WT mice (*80, 176*). Our findings also indicate that a high mtDNA point mutation load can negatively affect longevity. However, our data do not support the model that the shortened lifespan of mutator flies results from the progressive accumulation of mtDNA mutations. Although there is a trend towards increased mtDNA point mutation frequency with age in mutator flies, this age-dependent increase in mtDNA point mutations is small in comparison to the mutation frequency in young mutator flies. For example, the mutation frequency in 50-day-old 1xPolG$_{mut}$ is approximately 19-fold higher than age-matched controls, but fully 89% of the mutations detected in 50-day-old 1xPolG$_{mut}$ are acquired by the time these flies reach 1 day of age. These findings suggest that a high mtDNA mutational load throughout life simply increases the probability of death later in life. The delayed phenotypic effects of mtDNA mutations may reflect the long half-lives of mitochondrial proteins (*102, 148, 177, 178*), such that the consequences of a mtDNA mutation would require substantial time to develop. Additionally, our finding that most of the somatic mtDNA mutations in mutator flies are acquired during development, coupled with the fact that at least most of these mutations presumably result from replication errors, indicates that there is likely relatively little mtDNA replication in adult flies. This conclusion is further bolstered by the finding that key components of the mtDNA replication apparatus, including PolG, the Twinkle helicase, and the mitochondrial biogenesis factor Spargel, are primarily expressed in the female ovary, and early in development when most tissue growth occurs (*179, 180*). The importance of early arising mtDNA mutations to aging phenotypes may be conserved in vertebrates. Young mutator mice also display a dramatic

increase in mtDNA mutation frequency relative to controls, but the mtDNA mutation frequency increases little with age, consistent with the model that these mutations occur primarily during development (*106*).

Previous work testing whether negative selection acts against deleterious mtDNA mutations has led to conflicting findings. Studies in cultured cells have demonstrated that pharmacological and genetic perturbations that activate mitophagic pathways can select against certain severe heteroplasmic mtDNA mutations (*102, 105, 166*). Work in the nematode *C. elegans* has also shown that autophagy is required for the elimination of radiation-induced mtDNA damage (*181*), and that inactivation of the mitophagy-promoting factor Parkin results in increased abundance of point mutations in a mitochondrial mutator background and increased abundance of a deleterious mtDNA deletion when the mitochondrial unfolded protein stress pathway is activated (*133, 134, 182*). Overexpression of autophagy-promoting factors in *Drosophila* also reduced the frequency of a heteroplasmic deletion created by expression of a mitochondrially-targeted restriction endonuclease (*111*). Studies in mice expressing a proofreading-defective mtDNA polymerase also indicate that mitochondrial turnover is increased relative to controls (*107*). However, reducing the activity of PINK1 and Parkin in an otherwise WT *C. elegans* genetic background did not significantly influence the frequency of point mutations or a large mtDNA deletion, suggesting that this pathway does not ordinarily select against deleterious mtDNA mutations (*133, 134, 182*). Moreover, mutator mice do not exhibit an altered NS/S mutation ratio relative to WT mice (*108, 146, 160*), and mutator mice lacking the mitophagy-promoting factor Parkin do not exhibit an increased mtDNA mutation frequency or an altered NS/S mutation ratio relative to mutator mice (*108*). Previous work in *Drosophila* also suggests that negative selection does not act against a heteroplasmic mtDNA mutation in somatic tissues in the absence of extreme

measures to induce autophagy (*97*, *111*), and our current findings are fully consistent with this observation.

In vitro studies indicate that PINK1 and Parkin selectively target depolarized mitochondria for lysosomal degradation (*108*). Thus, one possible explanation for the apparent absence of negative selection opposing the accumulation of deleterious mtDNA mutations is that the mitochondria that bear these mutations are not sufficiently depolarized to trigger activation of the PINK1-Parkin pathway *in vivo*. Even in the event of a mtDNA mutation sufficient to trigger severe depolarization, ATP synthase (Complex V) is capable of coupling the hydrolysis of ATP to maintain partial membrane potential (*105, 183, 184*). Alternatively, the fusion of mitochondria bearing mutations with healthy mitochondria containing WT genomes may allow deleterious mtDNA mutations to evade negative selection through genetic complementation. Indeed, mitochondrial stress frequently elicits mitochondrial fusion as a compensatory response (*185*). While such compensatory mechanisms may prove useful when mtDNA mutations are present at low abundance, the phenotypes associated with high mutational loads in worms, flies, mice and humans indicate that these potential compensation pathways are incapable of fully preventing the deleterious consequences of a high mtDNA mutational load.

While we detect no evidence of negative selection acting against deleterious mtDNA mutations, a simple absence of negative selection does not fully explain our findings. If no selective forces were acting on mtDNA, the frequency of deleterious mtDNA mutations in mutator flies should match predictions from our simulations of neutrality. However, we found that deleterious mtDNA mutations were consistently overrepresented in mutator flies. One possible explanation for this finding is that there is positive selection for deleterious mtDNA mutations, which has previously been reported in vertebrates (*116*, *186*). Because many of the mtDNA

mutations subjected to positive selection in vertebrates reside in or near the sequences that control mtDNA replication, it has been proposed that these mutations confer a replicative advantage, thus accounting for their overrepresentation (*186*). However, a recent report has found strong evidence of positive selection acting on mutations that reside in protein coding sequences of human mtDNA with many of these variants appearing to result in deleterious effects on protein function (*116*). One potential explanation for the excess accumulation of deleterious mutations in mtDNA coding sequences is offered by the "survival of the slowest model" (*187*). In brief, this model posits that the mitochondrial quality control apparatus selectively targets oxidatively damaged mitochondria for degradation. According to this model, mitochondria that bear defective genomes are less metabolically active than fully functional WT mitochondria, and therefore less prone to damage from reactive oxygen species. This in turn makes these mutant-bearing mitochondria less prone to targeted degradation by quality control surveillance relative to fully functional mitochondria.

There are also at least two alternative models to explain the overabundance of deleterious mtDNA mutations in mutator flies. One possible explanation is that the exonuclease-deficient polymerase induces mutations in a sequence context-dependent fashion and that these contexts are enriched at critical residues. Although our analyses do not offer support for this model, we only examined the nucleotides immediately neighboring mutation sites, leaving open the possibility that other features of sequence context explain the overabundance of deleterious mutations in mutator flies. A second alternative model to explain the overabundance of deleterious mtDNA mutations in mutator flies is that the subset of cells that acquire deleterious mtDNA mutations may compensate for the presence of these defective genomes by inducing mitochondrial biogenesis. If the mitochondrial biogenesis machinery is incapable of distinguishing between mitochondria with defective and WT genomes, this phenomenon would inadvertently result in a higher replication

frequency of deleterious mutations relative to benign mutations because replication would be selectively induced in cells that bear deleterious mutations. Although we do not detect overt evidence of increased mitochondrial biogenesis in mutator flies (Figure 2.1D), our model posits that mitochondrial biogenesis would only be induced in a subset of cells, and this modest level of induction may not be readily detectable on a macroscopic level. Two recent studies in *C. elegans* offer potential support for this model by showing that worm strains bearing a heteroplasmic mtDNA deletion maintain tight regulation of WT mtDNA abundance, but that the abundance of the deletion can vary dramatically between individuals in the population (*133, 134*). These findings suggest that mutant genomes can "hitchhike" to high frequency as a consequence of a compensatory mitochondrial biogenesis response to decreased mitochondrial activity. However, a difference between these studies and our current work is that the mitochondrial unfolded protein stress pathway appears be an active participant, if not the driver, of the high mtDNA deletion frequency in *C. elegans* (*133, 134*), whereas we detected no evidence of mitochondrial unfolded protein stress pathway activation in mutator flies (Figure 2.10). Future experiments will be required to more fully investigate the potential role of the mitochondrial unfolded protein stress pathway and other candidate pathways that may influence the frequency of deleterious mtDNA mutations in mutator flies.

2.5 *Materials and Methods*

Fly strains and animal husbandry: All experiments were performed with flies raised and maintained at 25°C on standard cornmeal-molasses food unless otherwise stated. The w^{1118} isogenic fly strain, *Df(2L)BSC252* strain, and *Df(2L)FDD-0428643* strain were obtained from the Bloomington Drosophila Stock Center.

Generation of the PolG$_{mut}$ transgenic fly: The *Drosophila* POLG genomic region was PCR amplified from genomic DNA with primers (5'- TAAATCAATGTGACCGCCGC and 5'- TGTCCTTGCCTTGGGAACTG) and cloned into TOPO vector (Invitrogen). The D263A mutation was introduced by PCR using primers (5'- CAATGTCTCCTACGCAAGGGCGCGACTGAAG and 5'- CTTCAGTCGCGCCCTTGCGTAGGAGACATTG). Kanamycin resistance selection cassette (loxp-kanamycin-loxp) was PCR amplified and cloned into C-terminus of dPOLG-D263A using the PmeI site on the TOPO vector.

PCR primers containing 70 bp homology arms were used to generate linearized fragment for recombineering (5'- TACGGAGGAGTGTGTGGTCGCAGGTTGGACTTCAGTTGCCTTAAAGGATGTTTCCTT TATTAAAACGAGGATGCAGTTCCACCTGATCAG and 5'- TGCCTTGGGAACTGGGAAACGTATCGGCAACAGGATGCTTTAAATGCAAGGTTATTT AAAAACATAGTGATGTACAAGAAAGCTGGGTCG). The recombineering and Kanamycin excision process were performed using a published procedure (*188*, *189*).

To increase transgenesis efficiency, a 40 kb P[acman] clone with dPOLGD263A was generated from modified 100 kb P[acman] constructs. In brief, two 500 bp homology arms flanking the POLG 40 kb fragment were PCR amplified using primers (left arm: 5'- AGGCGCGCCTGTATTGCCTCAGCCGGTTG and 5'- CGCGGATCCTCGCTGTGTCGATAAGGAAC; right arm: 5'- CGCGGATCCGTTCGATTTGGTCAACCTGC and 5'- AACTTAATTAAGAGTCCAATGGGATTCCACA) and cloned into attB-P[acman]-ApR vector

(*188*). The 40 kb fragment with dPOLGD263A was then introduced into attB-P[acman]-ApR by recombineering. A list of the genes situated on this 40 kb fragment, along with their presumed functions is shown in Table 2.3.

To increase the copy number and allow for transgenesis, DNA from confirmed colonies was extracted and transformed into EPI300 cells. To create the $PolG_{mut}$ transgenic fly, modified 40 kb P[acman] constructs were integrated specifically into the 92F1 and 28E7 regions of the *Drosophila* genome using a φ31-mediated transformation protocol (BestGene., Rainbow Transgenic Flies).

Lifespan analysis: One- to two-day-old flies were collected into vials in groups of up to 20 flies. Flies were transferred every two to three days onto fresh food and the number of deaths was recorded. Lifespans were repeated at least three times with a minimum of 350 flies per genotype. The seizure phenotype of moribund mutator flies was noted during the lifespan analysis. All survivorship data were calculated using R software. The R package "survival" was used to generate Kaplan-Meier survival curves. Genotypes were compared using the log-rank test to determine significance.

Analysis of climbing: Climbing was assayed using a modified protocol for the previously published rapid iterative negative geotaxis (RING) assay (*190*, *191*). Briefly, clean vials containing up to 20 files were placed into the RING apparatus. Flies were manually tapped to the bottom of the vials and their climbing behavior recorded using a digital video camera. This procedure was repeated for a total of three trials per vial and a minimum of 60 flies per genotype at each time point. Still

Table 2.3. Genes contained within the PolG$_{mut}$ transgene along with their presumed functions.

Gene name/identifier	Biological function
CAH1 (carbonic anhydrase 1)	pH regulation
Adat1 (Adenosine deaminase-tRNA specific 1)	tRNA modifying enzyme
CG16865	Presumed splicesome component
CG16888	unknown
Sos (son of sevenless)	guanine nucleotide exchange factor
B (black)	aspartate decarboxylase
tamas (PolG)	mitochondrial DNA replication
Arpc1 (Actin-related protein 2/3 complex, subunit 1)	regulator of F-actin polymerization
Orc5 (Oigin recognition complex subunit 5)	DNA replication
GatC (Glutamyl-tRNA amidotransferase, subunit C)	mitochondrial translation
DNApol-gamma35 (DNA polymerase gamma 35kD)	mitochondrial DNA replication
Rpl133 (RNA polymerase II 33kD subunit)	transcription
mRpS23 (mitochondrial ribosomal protein S23)	translation
CenG1A (Centaurin gamma 1A)	unknown
CG33307	unknown
CG33306	unknown
CG8997	unknown
CG7916	unknown
CG7953	unknown
CG7968	unknown

images were captured using iMovie software (iMovie '11 v9.0.8. Apple Inc.) three seconds after the flies were tapped down. The vertical height attained by each fly was scored using Fiji software (an open-source image processing software) (*192*). Genotypes were compared to age-matched controls using the Wilcoxon rank-sum test.

ATP measurements: Four adult flies were homogenized in 200 µl of lysis buffer (100 mM Tris, 4 mM EDTA) followed by flash freezing in liquid nitrogen. Samples were boiled for 3 min before centrifugation at 8000xg for 5 min. Supernatant was then diluted 50-fold with lysis buffer for ATP quantification. ATP levels were measured using a commercial ATP Determination Kit (Invitrogen) as previously described (*193*).

Complex IV assay: Cytochrome *c* oxidase was measured as previously described (*194*). Homogenates were prepared from 4 male flies homogenized with a hand-held rotor (VWR) in PBS with 0.1 % Triton X-100 and protease inhibitor cocktail (Roche). Absorbance was measured with a SpectraMax M2 plate reader (Molecular Devices). Activities were normalized to total protein, and quantified using DC protein assay (Bio-Rad).

Western blot analysis: To generate protein extracts for Western blot analyses, four male flies were collected and frozen in liquid nitrogen. Flies were thawed and manually homogenized using a pestle in 100µL of lysis buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% NP-40, 10% glycerol, 10 mM NaF, 1 mM Na3VO4, 100 µg/ml PMSF, Sigma protease inhibitor cocktail (Sigma #P8340)). Cell debris was removed from the lysate through centrifugation at 13,000rpm for 15 minutes at 4°C. The protein lysate was boiled in SDS-PAGE sample buffer with 2% beta-mercaptoethanol for 10 minutes, and the resulting proteins subject to Western blot analysis.

For Western blot analyses, extracts were separated by SDS-PAGE on 4–20% Bis-Tris gels (GenScript #M42012) and transferred onto PVDF membranes overnight. Membranes were blocked in iBind Flex Solution Kit (SLF2020) and Western blots were performed using the iBind Flex Western Device (SLF2000) according to the manufacturer's protocol. Immunodetections were performed using the following antibodies: 1:8000 mouse anti-Actin (#MAB1501, Chemicon/Bioscience Research Reagents), 1:200 rabbit anti-Ref(2)P (Ab178440, Abcam), 1:500 rabbit anti-HSP60 (#4870S, Cell Signaling Technology), 1:1000 rabbit anti-GRP 75/mt-Hsp70 (sc-13967, Santa Cruz Biotechnology, Inc.). The secondary antibody anti-mouse HRP (BioRad) was used at 1:1000 for actin. The secondary antibody anti-rabbit HRP (BioRad) was used at 1:2000 for Ref(2)P, and at 1:500 for HSP60 and GRP 75/mt-Hsp70. Signal was detected using Thermo Scientific electrochemoluminescence reagents. Quantification of western blot images was performed using Fiji software (*192*). Western blot data were normalized using log-transformation to stabilize variance before means were compared using Student t-test. Each experiment was repeated with at least three biological replicates.

Transmission electron microscopy: TEM was performed as previously described with minor modifications (*195*). Briefly, indirect flight muscles were dissected from 50-day-old control and 2xPolG$_{mut}$ flies and placed in fixative containing 2.5% glutaraldehyde, and 2% paraformaldehyde in 0.1 M sodium cacodylate buffer, pH 7.4, and incubated overnight at 4°C. Fixed tissues were then postfixed in 1% OsO4, dehydrated in an ethanol series, and embedded using Epon. Samples were subjected to ultra-thin sectioning at 70 nm and stained with 6% uranyl acetate and a Reynolds lead citrate solution before TEM examination. Grids were viewed using a JEOL JEM 1400 transmission electron microscope.

Quantification of dopaminergic (DA) neurons: Adult brain dissection, fixation, immunohistochemistry, and imaging were performed as described previously (*49*). DA neurons were labeled with anti-TH antiserum (1:50, Immunostar). Serial optical sections were taken at 1-μm intervals and the confocal image stacks were analyzed using Imaris software (Bitplane Inc). The number of TH-positive neurons within each of the major DA neuron clusters was determined by visual inspection of individual confocal Z-series images.

DNA isolation for sequencing: One- to two-day-old male flies were collected and transferred into fresh vials every 2-3 days. Once flies reached the appropriate age for sequencing, heads were harvested using a razor blade, flash frozen in liquid nitrogen, and stored at -80°C. Total DNA was isolated from individual fly heads using the QIAamp DNA Micro isolation kit following the manufacturer's instructions. DNA yield for a single head typically ranged between 20-30ng of total DNA.

Duplex Sequencing: Total DNA was prepared for DS using a previously described protocol (*142*) with several modifications. Briefly, ~20 ng of total DNA was sonicated in 60 μL of nuclease-free ddH_2O using a Covaris AFA system with a duty cycle of 10%, intensity of 5, cycles/burst 100, time 20 seconds x 5, temperature of 4°C. After sonication, each sample was subjected to end-repair and 3'-dA-tailing using the NEBNext Ultra End-repair/dA-tailing kit (New England Biolabs) according to the vendor's instructions. Each sample was then ligated with 2 μL of 15 μM DS adapters, prepared as described (*142*) using the NEBNext Ultra Ligation kit (New England Biolabs) according to the manufacturer's instructions. Each sample was then cleaned of excess

adapters using AgenCourt AmpureXP magnetic beads and PCR amplified, as previously described (*142*). After library construction, mtDNA was enriched for sequencing by targeted DNA capture using IDT xGen Lockdown probes (Integrated DNA Technologies) specific for non-repetitive and non-low complexity regions of the *Drosophila* mitochondrial genome, as designated by RepeatMasker (http://www.repeatmasker.org), using the manufacturer's instructions. Probe sequences can be found in the published manuscript. Duplex Sequencing adapters used in collecting data for analyzing mutation selection were chemically synthesized as a collaborative effort with Integrated DNA Technologies to develop a prototype synthesis method.

The captured DNA samples were sequenced on an Illumina NextSeq500 using 150bp paired-end sequencing. The resulting reads were aligned against the *Drosophila* genome (BDGP Release 6 + ISO1 MT/dm6) using the Burrows-Wheeler Aligner and Samtools (*196*) coupled with a custom software workflow described previously (*142*). Reads not uniquely mapping to the mitochondrial genome were excluded from further analysis. Reads mapping to a large repetitive region [ChrM:5961..5983] were excluded from our analyses to avoid artifacts caused by misalignment. The breakpoints of the repetitive region were determined using the RepeatMasker Web Server v.4.0.6 (*197*). Sequence data has been uploaded to the Sequence Read Archive (SRA) repository, and can be accessed at SRA accession PRJNA495611. A heteroplasmy cutoff of 70% was applied to filter polymorphisms from the reference genome. After processing, we called unique somatic mutations by counting every mutation only once at each position of the genome. Total mutation frequency counts all mutations detected, including multiple occurrences at the same site.

Analysis of mtDNA mutation spectrum and trinucleotide context: Spectrum data, mutation frequency by codon position, mutation context by GC content, and four-fold degenerate site analyses were performed using scripts developed in Python v.2.7.

Parsing of the mitochondrial genome as well as GC content analyses were performed using Biopython (*198*). All measures of mutation frequency are calculated as a fraction:

$$\text{mutation frequency} = [\text{total \# mutations} / \text{total sequenced bases}]$$

of the indicated mutation type. For calculations of trinucleotide mutation frequency, the denominator of this equation is calculated for each of the 96 trinucleotide contexts by tabulating the sequencing depth at each nucleotide in the protein-coding sequence and then grouping according to its 3' and 5' flanking nucleotides. Scripts to analyze trinucleotide sequence context as well as trinucleotide mutation frequency were developed in Python v.3.4, and modified from a previously published duplex sequencing workflow (*199*). Circular plots of the distribution of mutations were generated using the R package, 'circlize.' Statistics were performed and graphs were generated in RStudio v1.1.383, Microsoft Excel for Mac v.16.10, and StataCorp Stata v.12.1.

Analysis of mtDNA mutation selection: To search for evidence of selection, a bioinformatics workflow was developed in Python v.2.7 (https://github.com/csamstag/mito-mutations) to analyze mtDNA mutation data obtained from duplex sequencing, and to run simulations of mutagenesis mimicking the mutation spectra obtained from sequencing mutator flies. Data for simulations were obtained by aggregating the results from sequencing four 1-day-old 1xPolG$_{mut}$ flies. Point mutations were identified as described above, except that multiple mutations of a given type at the same site were also included in our analyses. Monte Carlo simulations were performed to generate a distribution of random mutations for statistical comparison to experimental findings. To control

for the observed mutational biases in our data, we designed our scripts to match parameters observed from sequencing mutator flies (i.e., the same number of G:C to A:T mutations, G:C to T:A mutations, *etc*.). To account for variation in sequencing depth, random mutations were generated using a probability function that was weighted according to the total read depth at each position. Each round of simulation mirrored the spectrum and number of mutations observed in the protein-coding regions of mutator flies. Simulations were repeated 10,000 times, and each simulation was analyzed using three metrics designed to detect selective forces.

To examine the relationship between mutation frequency and sequence conservation, we obtained the PhyloP scores for each position in the *Drosophila melanogaster* mitochondrial genome from the UCSC Genome Browser. These values were used to calculate the average PhyloP score of the mutations detected from sequencing mutator flies. Similarly, we calculated the average PhyloP values of the mutations identified from simulations performed as described above. Empirical *p*-values represent the fraction of time a simulation displayed an average PhyloP score greater than or equal to the average PhyloP score observed in mutator flies.

To test whether selection influences the frequency of NS variants, we compared the average NS/S ratio obtained from sequencing mutator flies to the NS/S ratios obtained from simulations. We performed simulations as described above and binned those mutations that map to coding sequences according to whether they induce NS or S alterations. We then calculated the average NS/S ratio for each simulation. The empirical *p*-value reflects the fraction of simulations in which the NS/S ratio was greater than or equal to the average NS/S ratio observed in 1xPolG$_{mut}$ flies. A similar approach was used to compare the NS/S ratio in the *COX1* gene from sequencing mutator flies to the distribution of NS/S ratios from simulations of mutagenesis of the *COX1* gene.

Simulated mutagenesis of the *COX1* gene was performed as described above, including adjustments for sequencing bias and sequencing depth.

To test whether selection influences the frequency of pathogenic mutations, we used MutPred (*171*) software to calculate pathogenicity scores for all NS variants detected in mutator flies, as well as all NS mutations generated from simulations. The average MutPred score from mutator flies was then compared to a distribution of MutPred scores from simulations. The empirical *p*-value was determined as the fraction of simulations in which the average MutPred score was greater than or equal to the average MutPred score observed in mutator flies. A similar approach was used to analyze the pathogenicity of NS mutations occurring within the *COX1* gene. Adjustments were made to simulations to account for mutational bias and sequencing depth within the *COX1* gene.

2.6 *Acknowledgements*

We thank Mary Walton, Selina Yu, and Marie Y. Davis (VA Puget Sound Healthcare System; University of Washington) for technical assistance, Alicia Pickrell (Virginia Tech) for help in Complex IV activity assays, and LaDonne Pallanck for help with computational analysis of mtDNA sequence data. We also thank Phil Green (University of Washington) and Jim Thomas (University of Washington) for help with our simulations of mutagenesis, Gautam Pareek and Glen MacDonald (Microscopy and Imaging Facility, Virginia Merrill Bloedel Hearing Research Center, University of Washington) for technical assistance with confocal imaging, Bobbie Schneider (Fred Hutchinson Cancer Research Center) for assistance with electron microcopy, Maulik Patel (Vanderbilt University), Ruth Thomas and Evvie Vincow for critical reading of the manuscript, and all members of the Pallanck Lab for critical review of this work.

Chapter 3. **Mitochondrially-targeted APOBEC1 is a potent mtDNA mutator affecting mitochondrial function and organismal fitness in *Drosophila***

As reviewed in **Chapter 1**, there are many questions that remain about the factors that influence the pathogenicity of mtDNA mutations, and our ability to accurately detect mutations that arise in somatic mtDNA has hindered the statistical analysis of these mutations. I have also demonstrated that the global mutation frequency, a frequently used metric, does not sufficiently capture the pathogenicity of mutations. This chapter explores how the clonality and spectrum influences the pathogenicity of mutations.

While studying the overrepresentation of deleterious mtDNA mutations in $PolG_{mut}$ mutator flies, I was approached by a colleague in the field who had obtained puzzling data. The Whitworth Lab of Cambridge University had developed an additional *Drosophila* mtDNA mutator model through transgenic expression of mitochondrially-targeted APOBEC1, a cytidine deaminase. Mito-APOBEC1 induces a high abundance of C->T mutations in mtDNA. The Whitworth lab discovered that these flies displayed far more severe phenotypes than PolG mutator fly models. Furthermore, they had also used Duplex Sequencing to study mutations in these flies, but were surprised to find that the frequency of point mutations and the spectrum of mutations was broadly comparable between mito-APOBEC1 and PolG flies. They were, therefore, at a loss to explain the discrepancies in phenotypes between these two fly models.

To better understand the pathogenicity of APOBEC1-induced mutations, it was clear that the metrics commonly used to summarize mutations were insufficient. Historically, the rich data obtained through sequencing of mtDNA is compressed and summarized simply using simple

metrics of mutation frequency. I improved upon my computational pipelines from **Chapter 2** to better compare the mutation spectra of these two mutator models. My analysis challenges the conventional wisdom that elevated mtDNA mutation alone is sufficient to cause severe organismal phenotypes in mutator models. Additionally, it highlights why a deeper analysis of the spectrum and clonality of mutations will be critical to our future understanding of these models.

This work was performed in close collaboration with Dr. Simonetta Andreazza of Dr. Whitworth's Lab. Dr. Andreazza created and performed behavioral assays on the mito-APOBEC fly model, performed extensive biochemical characterization, and was the principal author of the resulting publication. I performed mtDNA sequence analysis, visualization of this data, and technical assistance in writing the publication. This chapter has been published, in slightly modified form:

> Andreazza S, **Samstag CL**, Sanchez-Martinez A, Fernandez-Vizarra E, Gomez-Duran A, Lee JJ, et al. Mitochondrially-targeted APOBEC1 is a potent mtDNA mutator affecting mitochondrial function and organismal fitness in Drosophila. Nature Communications. 2019;10: 3280. doi:10.1038/s41467-019-10857-y (*200*)

3.1 *Abstract*

Somatic mutations in the mitochondrial genome (mtDNA) have been linked to multiple disease conditions and to aging itself. In *Drosophila,* knock-in of a proof-reading deficient mtDNA polymerase (*POLG*) generates high levels of somatic point mutations and also small indels, but surprisingly limited impact on organismal longevity or fitness. Here we describe a new mtDNA mutator model based on a mitochondrially-targeted cytidine deaminase, APOBEC1. *mito-APOBEC1* acts as a potent mutagen which exclusively induces C:G>T:A transitions with no indels or mtDNA depletion. In these flies, the presence of multiple non-synonymous substitutions, even at modest heteroplasmy, disrupts mitochondrial function and dramatically impacts organismal fitness. A detailed analysis of the mutation profile in the *POLG* and *mito-APOBEC1* models reveals that mutation type (quality) rather than quantity is a critical factor in impacting organismal fitness. The specificity for transition mutations and the severe phenotypes make *mito-APOBEC1* an excellent mtDNA mutator model for aging research.

3.2 *Introduction*

Mitochondria play essential roles in many cellular metabolic and signaling processes, most notably in the production of molecular energy in the form of ATP and in the regulation of cell death. Mitochondria also contain their own genome – in animals, a circular DNA molecule that contains 37 genes encoding 13 subunits of the respiratory chain and ATP synthase, along with the tRNAs and rRNAs necessary for their production (*7*). As such, the transmission of germline mtDNA mutations, which occurs exclusively via the maternal lineage, can give rise to devastating mitochondrial diseases (*201, 202*). Furthermore, since mtDNA exists in multiple copies per mitochondrion, and thus many copies per cell, this can give rise to a mixed population of wild-type and mutated genomes, a condition known as heteroplasmy.

MtDNA has long been considered to be particularly vulnerable to accumulating spontaneous mutations due to its high replication rate, proximity to the major source of reactive oxygen species (the respiratory chain), the absence of protective histones and limited repair mechanisms (*203*). Consistent with this idea, mtDNA mutations have been seen to accumulate with age in somatic tissues in a wide range of organisms including humans, rodents and invertebrates, and thus have been implicated as a driving force in the aging process (*12, 204*). Moreover, high levels of mtDNA mutations have also been found in various age-related conditions including neurodegenerative disorders such as Parkinson's and Alzheimer's diseases (*13, 14*).

To explore the consequences of increasing the level of mtDNA mutations on cellular and organismal fitness, and the mechanisms that counteract such mutations, a number of model systems have been developed. The best-established model system for inducing high levels of mtDNA point mutations is via the expression of a proofreading (exonuclease) deficient variant of the sole DNA polymerase responsible for replicating mtDNA, POLG. Mouse strains of this model system exhibit many characteristics of a premature aging syndrome, including kyphosis, alopecia, and osteoporosis, all correlating with a shortened lifespan (*106, 158*). However, the mutational heterogeneity that arises in this model has led to considerable debate about the pathogenic entity and whether point mutations are actually driving aging (*15, 80, 176, 205*).

Equivalent 'mutator' models have been established in *Drosophila*, either by homologous recombination targeting the endogenous locus of the *POLG* homologue, *tamas* (*tamas*$^{exo-}$ flies) (*155*), or using a genomic transgene bearing a defective *tamas* (*149*). Recent studies have shown that mutations that arise in these models have very limited impact on adult fly health or lifespan (*149, 175*), questioning whether mtDNA mutations are capable of rising to levels sufficient to impact aging in flies (*175*).

To circumvent the limitations of the POLG model, we sought to develop an alternative mtDNA mutator system by employing the activity of the cytidine deaminase APOBEC1. APOBEC1 (Apolipoprotein B (apoB) mRNA editing catalytic polypeptide 1) is a vertebrate-specific, zinc-dependent deaminase which, in complex with complementing specificity factor (ACF), catalyzes the deamination of cytosine to uracil (C>U) in apoB mRNA (*206-208*). Despite a physiological role in mRNA editing, which is fully dependent upon an RNA-binding accessory subunit, APOBEC1 alone was found to be a potent mutator of DNA both *in vitro* (*209*) and *in vivo* (*210*). APOBEC1's effects as a DNA mutator are becoming more widely recognized and this ability is being exploited as a targeting mutagen in combination with zinc-finger peptides and CRISPR/Cas9 (*211, 212*). Mechanistically, uracil (U) present in DNA is recognized as mutagenic and is usually removed by uracil-DNA glycosylases, generating an abasic site, which can itself be inappropriately filled. However, if uncorrected, in subsequent replication adenine (A) is paired with U to cause a mutagenic transition of a C:G to U(T):A, resulting in a stable C:G>T:A point mutation. Importantly, this phenomenon recapitulates the predominant mutation profile in human aging (*61*). Thus, we sought to exploit this deaminase activity, directing APOBEC1 specifically to mitochondria via a construct we termed *mito-APOBEC1*.

In contrast to heterozygous *tamas*$_{exo-}$ flies, somatic *mito-APOBEC1* expression severely limits mitochondrial function, organismal vitality and lifespan. A detailed analysis of the mutation spectra in these models reveals that the mutation profile (i.e. quality), rather than the overall mutation load (quantity), correlates with a reduced organismal fitness.

3.3 ***Results***

3.3.1 *Analysis of tam$_{exo-}$ mtDNA mutagenesis*

A *Drosophila* model equivalent to the 'mutator' mouse has been generated by knock-in of a proofreading (exonuclease) deficient variant of the *POLG* homologue, *tamas* (*tam*$_{exo-}$)(*155*). Surprisingly, *tam*$_{exo-}$ adults have shown very limited defects on organismal health or lifespan (*175*), prompting us to examine in greater detail their mutation spectra. Since the *tam*$_{exo-}$ strain is homozygous lethal, all our analyses were conducted on heterozygous animals, inherited either paternally (+/*tam*$_{exo-}$) or maternally (*tam*$_{exo-}$/+). Reinsertion of the wild-type coding region, referred to as *tam*$_{rescue}$, served as a control genotype. As an additional control we also compared the *tam* lines to a strain bearing the closest nuclear and mitochondrial genetic background; *white*^{Dahomey} (*w*^{Dah}) for maternal inheritance or *white*^{1118} (*w*^{1118}) for paternal inheritance (see Methods).

Sequence analysis of the *tam*$_{exo-}$ strains and controls has previously been performed using a PCR-based cloning and sequencing strategy (*155, 175*). We sought to extend this analysis by performing Duplex Sequencing of the mitochondrial genome on individual brains from mutator and control animals (see Methods). Duplex Sequencing is a high-accuracy next-generation sequencing approach capable of detecting a single mutation in >10^7 wild-type bases (*139*). We initially examined the mtDNA mutation profile of 10-day-old flies to allow for a modest amount of post-mitotic accumulation of mtDNA mutations. Our sequencing analysis provided a depth of ~500-1000 unique reads per site. Although coverage is not uniform across the genome, patterns of sequencing depth were consistent across all samples tested, suggesting that no samples bore large deletions within mtDNA (Figure 3.1). Sequence analysis identified multiple mutations at varying levels of heteroplasmy and several homoplasmic polymorphisms, depending on the background strain (Table 3.1). Because we believe high heteroplasmy mutations could contribute to organismal

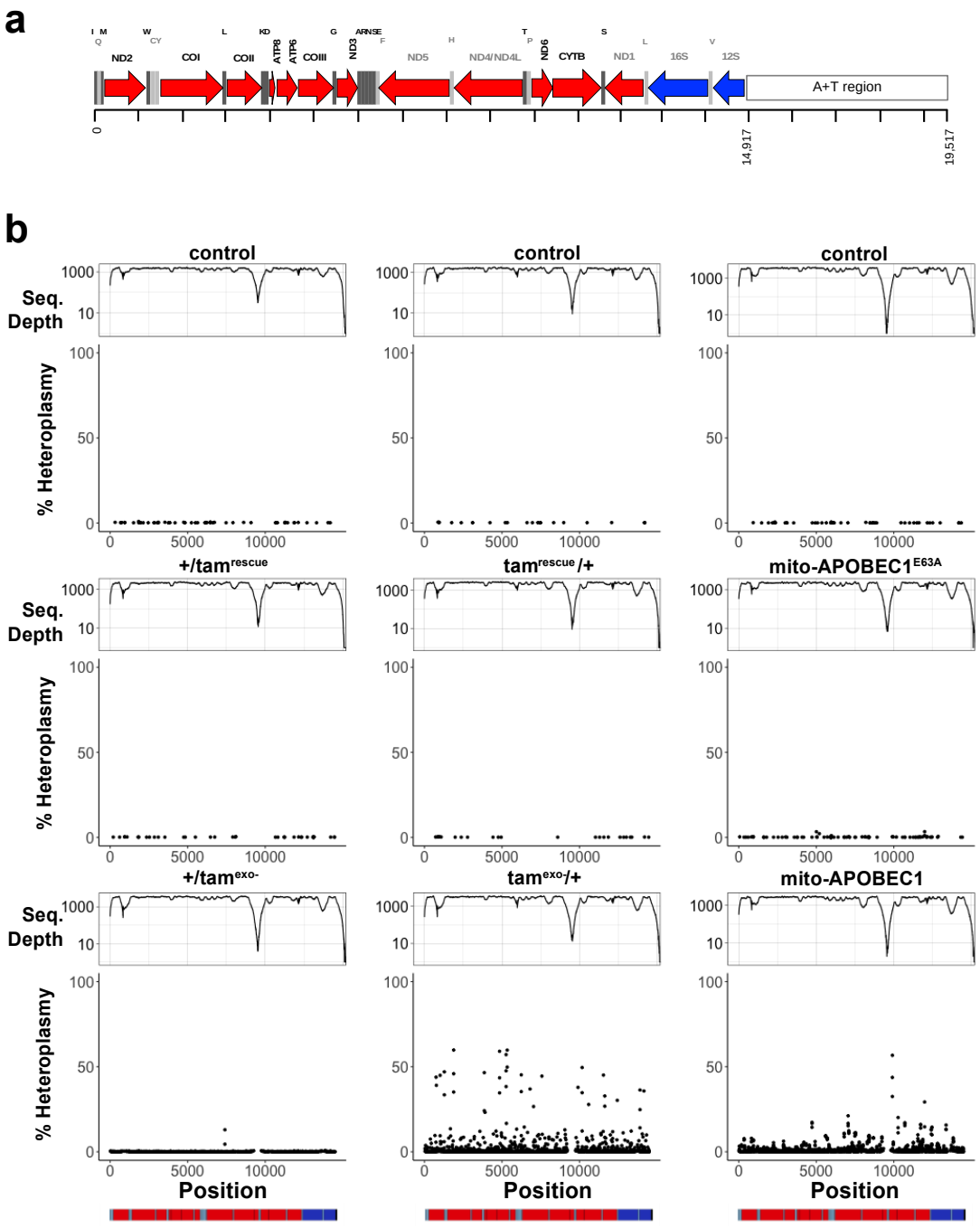


Figure 3.1. Coverage of Duplex Sequencing across the mitochondrial genome and heteroplasmy levels recovered at each position. (**A**) Detailed schematic of the *Drosophila* mtDNA genome map. Red: protein coding genes; grey: tRNAs; dark blue: rRNAs. Gene names and one-letter tRNA symbols are indicated. (**B**) Sequencing depth (log_{10}- transformed *y*-axis) and heteroplasmy level of mutations across the mitochondrial genome in 10-day-old flies of indicated genotypes. Each dot represents an identified mutation shown at the observed heteroplasmy level. Each chart shows the sum of all mutations identified per genotype (n = 3 animals).

Table 3.1. Point mutations identified by Duplex Sequencing in 10-day-old flies.
All mutations observed at ≤70% clonality. DCS denotes the duplex consensus sequence, the number of post-processing nucleotides sequenced.

		# DCS	# mutated sites	# mutations	Mutation load
paternal tamas, 10 days old	w1118_1	3010367	16	20	6.64E-06
	w1118_2	7961128	18	233	2.93E-05
	w1118_3	8110354	47	116	1.43E-05
	w1118_AVG		27		1.67E-05
	+/tam_{rescue}_1	8728147	14	110	1.26E-05
	+/tam_{rescue}_2	9143056	11	112	1.22E-05
	+/tam_{rescue}_3	8100463	35	271	3.35E-05
	+/tam_{rescue}_AVG		20		1.94E-05
	+/tam_{exo-}_1	17481755	733	1025	5.86E-05
	+/tam_{exo-}_2	14005318	577	694	4.96E-05
	+/tam_{exo-}_3	9852168	317	617	6.26E-05
	+/tam_{exo-}_AVG		542		5.69E-05
maternal tamas, 10 days old	wDah_1	5693981	9	9	1.58E-06
	wDah_2	7246279	13	21	2.90E-06
	wDah_3	6577020	28	59	8.97E-06
	wDah_AVG		17		4.48E-06
	tam_{rescue}/+_1	9499339	8	10	1.05E-06
	tam_{rescue}/+_2	10713046	17	28	2.61E-06
	tam_{rescue}/+_3	9734980	35	75	7.70E-06
	tam_{rescue}/+_AVG		20		3.79E-06
	tam_{exo-}/+_1	11067470	917	8081	7.30E-04
	tam_{exo-}/+_2	14561594	973	10129	6.96E-04
	tam_{exo-}/+_3	11483065	905	7559	6.58E-04
	tam_{exo-}/+_AVG		932		6.95E-04
mito-APOBEC1, 10 days old	mito-GFP_1	14922865	18	26	1.74E-06
	mito-GFP_2	12278047	29	36	2.93E-06
	mito-GFP_3	14130656	23	29	2.05E-06
	mito-GFP_AVG		23		2.24E-06
	mito-APOBEC1$_{E63A}$_1	10853172	42	71	6.54E-06
	mito-APOBEC1$_{E63A}$_2	11007973	61	79	7.18E-06
	mito-APOBEC1$_{E63A}$_3	5266167	28	64	1.22E-05
	mito-APOBEC1$_{E63A}$_AVG		44		8.62E-06
	mito-APOBEC1_1	10162080	734	5405	5.32E-04
	mito-APOBEC1_2	11318537	789	6573	5.81E-04
	mito-APOBEC1_3	8666010	690	4538	5.24E-04
	mito-APOBEC1_AVG		738		5.45E-04

phenotypes, we considered a threshold of 70% heteroplasmy (the maximum we observed) so as to exclude homoplasmic sequence polymorphisms present in the background parental strain but capture all other mutations. Hence, for all our analyses we have included the total number of observed mutations, including multiple counts for mutations at a same site (heteroplasmy). Mutation loads were calculated by dividing all mutations by the total number of nucleotides sequenced, considering the whole or a subset of the genome, as relevant to each analysis.

The background mtDNA mutation load varied a little by genotype and individual animal (see Table 3.1), being 1.5-17 x10^{-6} depending on background strain, but comparable with previous measurements (*62*). The *tam*$_{rescue}$ controls exhibited similar levels of mtDNA mutations to the background controls (Table 3.1). In animals with paternally inherited *tam*$_{exo-}$, and hence accumulating only somatic (and not germline-inherited) mutations, the mtDNA mutation load increased to 5.7 x10^{-5} (Figure 3.2A). However, in animals with maternally inherited *tam*$_{exo-}$, mtDNA mutation levels were substantially higher (6.9 x10^{-4}; Figure 3.2A). Importantly, maternally inherited *tam*$_{exo-}$ flies showed no significant alterations in mtDNA copy number (see below). The strong increase in mutation load of maternally inherited *tam*$_{exo-}$ is likely a combination of inherited heteroplasmic mtDNA copies from the *tam*$_{exo-}$-bearing mothers and clonal expansion of replication errors occurring during the peak mtDNA replication that happens part-way through embryogenesis (*213*). This is at least partially supported by the relatively similar number of sites (<2-fold difference) that are mutated in the paternally and maternally transmitted mutator flies (Figure 3.2B). A much larger mutation load in a similar number of sites indicates that the same mutated sites are represented many more times in the maternal *tam*$_{exo-}$/+. These results indicate that zygotically expressed (paternally inherited) heterozygous *tam*$_{exo-}$ is not a potent mutator, but mtDNA mutations can rise to high loads when *tam*$_{exo-}$ is maternally inherited. However, maternally

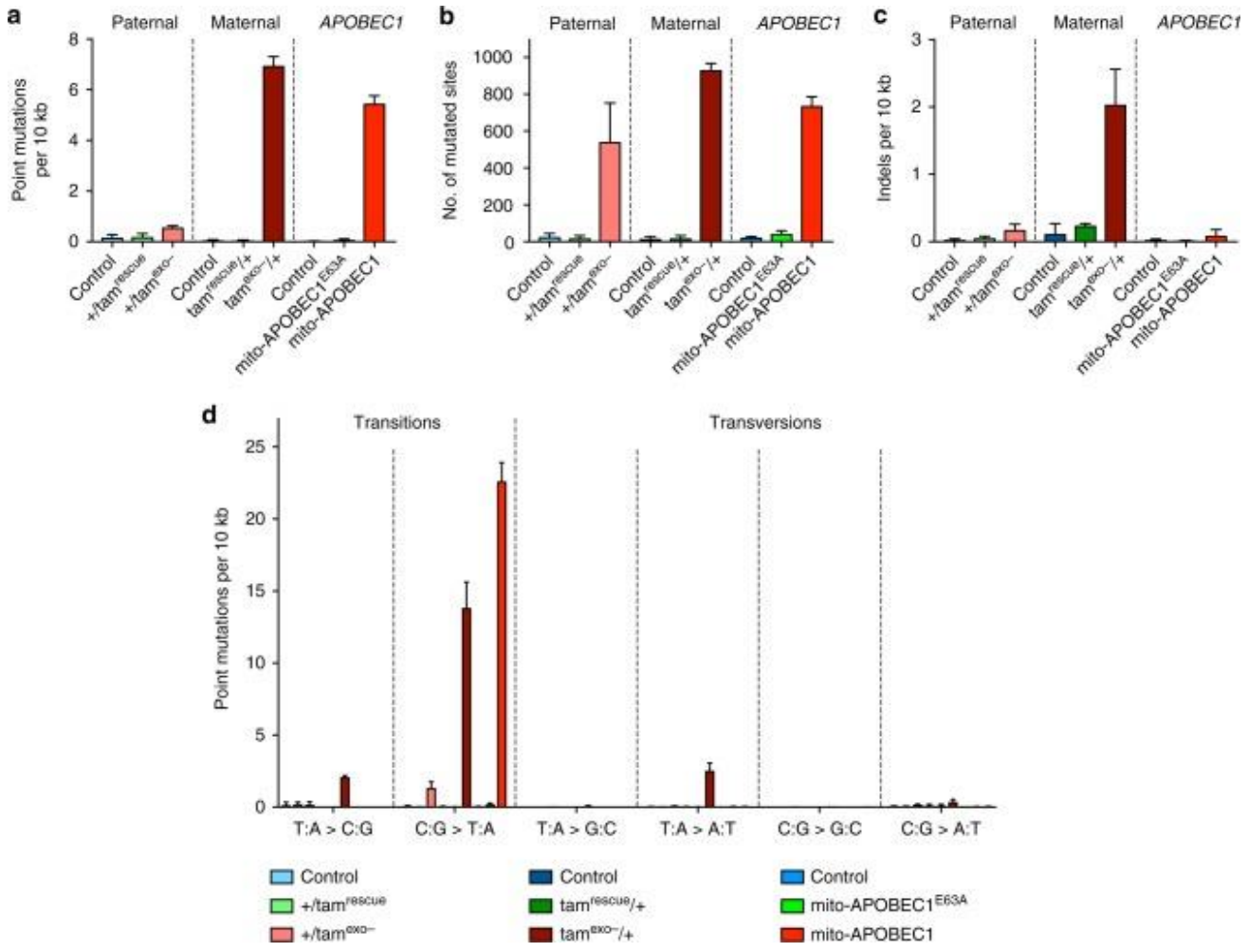


Figure 3.2. Maternally inherited *tam*$_{exo-}$ and *mito-APOBEC1* flies generate high mtDNA mutation level. Quantification of mtDNA mutations in 10-day-old paternally inherited *tam*$_{exo-}$ (+/*tam*$_{exo-}$), maternally inherited *tam*$_{exo-}$ (*tam*$_{exo-}$/+) and *mito-APOBEC1* mutator flies and respective controls. (**A**) mtDNA mutation load, expressed as number of mutations observed, including heteroplasmic mutations, per total number of bases sequenced [Note that coverage at the non-coding A+T region, because of its repetitive high A:T content, is minimal and therefore negligibly contributes to the overall mutation load], (**B**) the number of sites with at least one mutation, (**C**) indel mutation load, and (**D**) the mutation load identified by mutation type as indicated. Charts show mean ± SD, *n* = 3 animals per genotype.

inherited *tam*$_{exo-}$ also introduces a significant number of small insertion/deletion (indel) mutations (Figure 3.2C, Table 3.2), which can confound the physiological interpretation of mtDNA point mutations. Thus, we sought to develop an alternative genetically encoded mtDNA mutator system with a greater specificity for inducing point mutations.

3.3.2 *Generation of a new mtDNA mutator model*

We generated an inducible mutator model by exploiting the nucleic acid editing capability of the cytidine deaminase APOBEC1, which has been shown to cause C:G>T:A transition mutations in DNA (*210*). Rat APOBEC1 was cloned into a 'UAS' transgenic expression vector with an N-terminal mitochondrial targeting sequence (MTS) followed by a hemagglutinin (HA) tag (Figure 3.3A). A nuclear export sequence (NES) was also included to minimize any potential for nuclear localization. We termed this construct '*mito-APOBEC1*'. As a control, we also generated an equivalent construct expressing a catalytically inactive mutant, E63A ('*mito-APOBEC1*^{E63A})(*214*). Independent transgenic lines were established with the constructs in the same genomic locus to ensure comparable expression levels and genetic background (Figure 3.3B). Sub-cellular fractionation analyses determined that *mito-APOBEC1* and *mito-APOBEC1*^{E63A} were successfully targeted to mitochondria (Figure 3.3C); this was further confirmed by immunofluorescence analysis in larval epidermal cells (Figure 3.3D).

Initial tests determined that the expression of *mito-APOBEC1* induced by a range of GAL4 drivers had no gross effects on development or viability. We chose to conduct the current study using the ubiquitous driver, *da-GAL4*. In all instances, transgene expression was only induced in the zygote, thus avoiding maternal transmission of germline mutations. Sequence analysis of mtDNA mutations revealed that zygotic expression of *mito-APOBEC1* is a highly effective mutagen, creating mutation loads of ~5.5 x10^{-4}, similar to maternally inherited *tam*$_{exo-}$, and

Table 3.2. Indels identified by Duplex Sequence in 10-day-old flies. DCS denotes the duplex consensus sequence, the number of post-processing nucleotides sequenced.

		# DCS	# INDELs	# INs	# DELs	INDELs	INs	DELs
paternal tamas, 10 days old	w1118_1	3010367	4	3	1	1.33E-06	9.97E-07	3.32E-07
	w1118_2	7961128	35	1	34	4.40E-06	1.26E-07	4.27E-06
	w1118_3	8110354	26	2	24	3.21E-06	2.47E-07	2.96E-06
	w1118_AVG					2.98E-06	4.56E-07	2.52E-06
	+/tam$_{rescue}$_1	8728147	68	56	12	7.79E-06	6.42E-06	1.37E-06
	+/tam$_{rescue}$_2	9143056	58	49	9	6.34E-06	5.36E-06	9.84E-07
	+/tam$_{rescue}$_3	8100463	23	8	15	2.84E-06	9.88E-07	1.85E-06
	+/tam$_{rescue}$_AVG					5.66E-06	4.25E-06	1.40E-06
	+/tam$_{exo-}$_1	17481755	457	308	149	2.61E-05	1.76E-05	8.52E-06
	+/tam$_{exo-}$_2	14005318	247	147	100	1.76E-05	1.05E-05	7.14E-06
	+/tam$_{exo-}$_3	9852168	95	20	75	9.64E-06	2.03E-06	7.61E-06
	+/tam$_{exo-}$_AVG					1.78E-05	1.00E-05	7.76E-06
maternal tamas, 10 days old	wDah_1	5693981	167	162	5	2.93E-05	2.85E-05	8.78E-07
	wDah_2	7246279	18	0	18	2.48E-06	0.00E+00	2.48E-06
	wDah_3	6577020	23	0	23	3.50E-06	0.00E+00	3.50E-06
	wDah_AVG					1.18E-05	9.48E-06	2.29E-06
	tam$_{rescue}$/+_1	9499339	209	181	28	2.20E-05	1.91E-05	2.95E-06
	tam$_{rescue}$/+_2	10713046	241	229	12	2.25E-05	2.14E-05	1.12E-06
	tam$_{rescue}$/+_3	9734980	268	249	19	2.75E-05	2.56E-05	1.95E-06
	tam$_{rescue}$/+_AVG					2.40E-05	2.20E-05	2.01E-06
	tam$_{exo-}$/+_1	11067470	2051	256	1795	1.85E-04	2.31E-05	1.62E-04
	tam$_{exo-}$/+_2	14561594	2373	451	1922	1.63E-04	3.10E-05	1.32E-04
	tam$_{exo-}$/+_3	11483065	3014	397	2617	2.62E-04	3.46E-05	2.28E-04
	tam$_{exo-}$/+_AVG					2.04E-04	2.96E-05	1.74E-04
mito-APOBEC1, 10 days old	mito-GFP_1	14922865	21	8	13	1.41E-06	5.36E-07	8.71E-07
	mito-GFP_2	12278047	50	3	47	4.07E-06	2.44E-07	3.83E-06
	mito-GFP_3	14130656	19	2	18	1.34E-06	1.42E-07	1.27E-06
	mito-GFP_AVG					2.27E-06	3.07E-07	1.99E-06
	mito-APOBEC1$_{E63A}$_1	10853172	14	0	14	1.29E-06	0.00E+00	1.29E-06
	mito-APOBEC1$_{E63A}$_2	11007973	6	3	3	5.45E-07	2.73E-07	2.73E-07
	mito-APOBEC1$_{E63A}$_3	5266167	12	8	4	2.28E-06	1.52E-06	7.60E-07
	mito-APOBEC1$_{E63A}$_AVG					1.37E-06	5.97E-07	7.74E-07
	mito-APOBEC1_1	10162080	188	0	188	1.85E-05	0.00E+00	1.85E-05
	mito-APOBEC1_2	11318537	93	1	92	8.22E-06	8.84E-08	8.13E-06
	mito-APOBEC1_3	8666010	8	2	6	9.23E-07	2.31E-07	6.92E-07
	mito-APOBEC1_AVG					9.21E-06	1.06E-07	9.11E-06

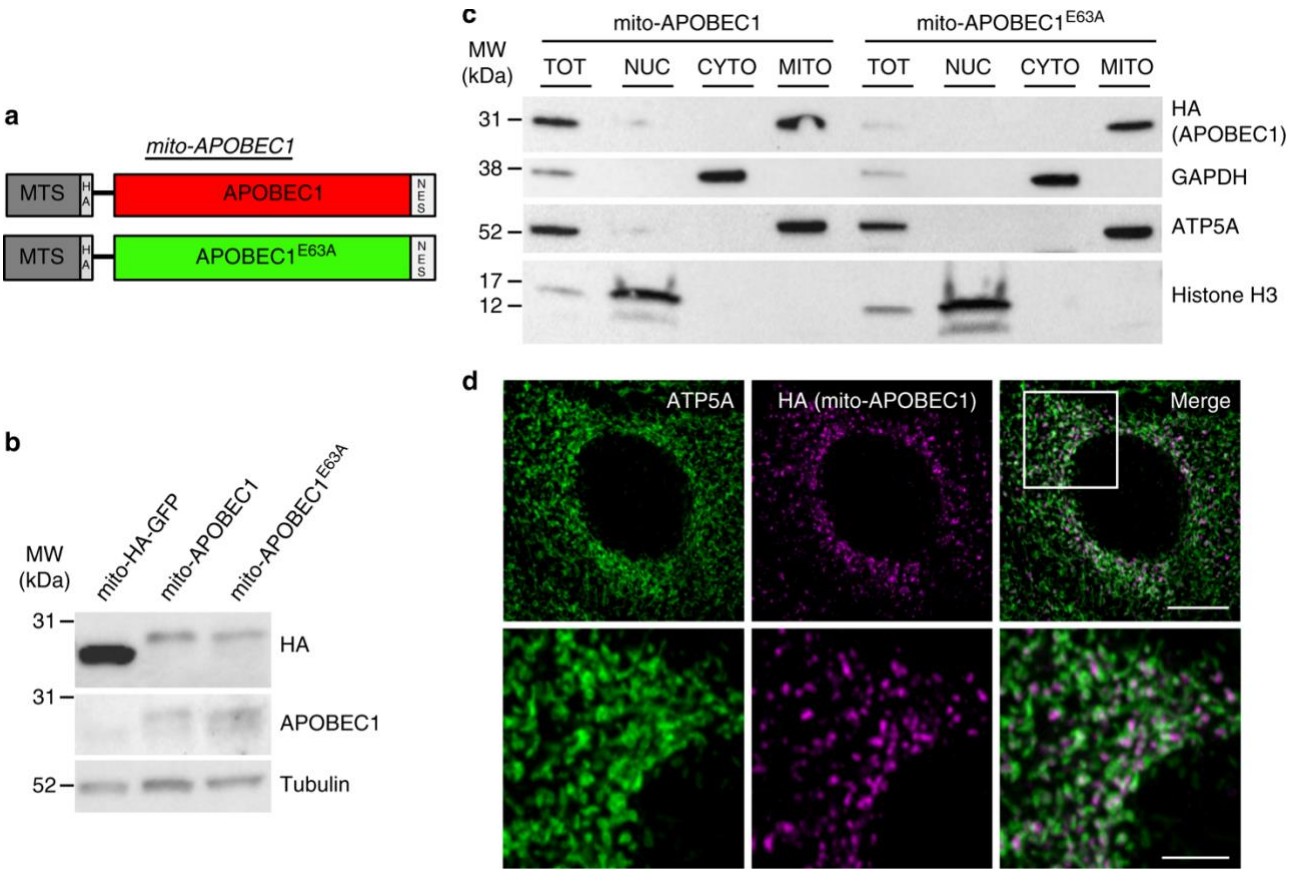


Figure 3.3. Generation and validation of the *mito-APOBEC1* mutator model. (**A**) Schematic of *mito-APOBEC1* construct and inactive mutant (E63A). (**B**) Immunoblot showing *mito-APOBEC1* transgene expression in whole fly extracts, using anti-APOBEC1 and anti-HA antisera. Transgenic *mito-HA-GFP* expression is used as a positive control, and anti-Tubulin immunostaining as loading control. (**C**) Tissue fractionation and immunoblotting shows that mito-APOBEC1 and mito-APOBEC1$_{E63A}$ proteins localize to the mitochondrial fraction (ATP5A-positive). GAPDH and Histone H3 mark the cytoplasmic and nuclear fractions, respectively. (**D**) Immunohistochemistry analysis of *mito-APOBEC1* expression in larval epidermal cells, revealed by anti-HA staining, co-stained with the mitochondrial membrane marker ATP5A. Scale bars = 10 μm (top panels) and 4 μm (bottom panels).

substantially more effective than zygotically expressed, paternally inherited *tam*$_{exo-}$ (Table 3.2A, Table 3.1). Interestingly, *mito-APOBEC1* mutates a similar number of mitochondrial genome sites as *tam*$_{exo-}$ (Figure 3.2B). Also, *mito-APOBEC1* expression has minimal impact on mtDNA levels (Figure 3.4B, D) and does not affect mtDNA integrity (Figure 3.4C, E).

Analyzing the mutation profile we found that *mito-APOBEC1*-induced changes were almost entirely composed of C:G>T:A transitions (Figure 3.2D, and Figure 3.5A), consistent with its mechanism of action. In contrast, although the majority of *tam*$_{exo-}$ mutations were also C:G>T:A, a considerable proportion were T:A>C:G transitions or T:A>A:T transversions (Figure 3.2D, Figure 3.5A), as observed in the mouse (*108*). Interestingly, mapping reciprocal mutations by strand showed a dramatic bias for *mito-APOBEC1*, with most of the cytidine deamination events occurring in the minor strand (G>A transitions in the major strand) (Figure 3.6A). This is in line with the strand pattern and type of mutations naturally occurring in *Drosophila* during aging and across generations (*62, 169*). Importantly, in contrast to maternally inherited *tam*$_{exo-}$, *mito-APOBEC1* does not induce small indels (Figure 3.2, Table 3.2), nor larger deletions (Figure 3.4e). Taken together, these results show that *mito-APOBEC1* is a highly effective, inducible mtDNA mutator that exclusively causes C:G>T:A transitions at high loads without causing indel mutations or substantial mtDNA depletion.

3.3.3 *The two mutator models affect organismal fitness differently*

We then evaluated the impact of the mutator systems on organismal fitness. Foremost, with the long-standing association of mtDNA mutations and aging, we assessed the effects on lifespan. In agreement with recent findings (*175*), we found that both paternally and maternally inherited *tam*$_{exo-}$ showed minimal impact on lifespan (Figure 3.7A, C). In stark contrast, expression of *mito-APOBEC1* caused a dramatic reduction in median and maximal lifespan (Figure 3.7E).

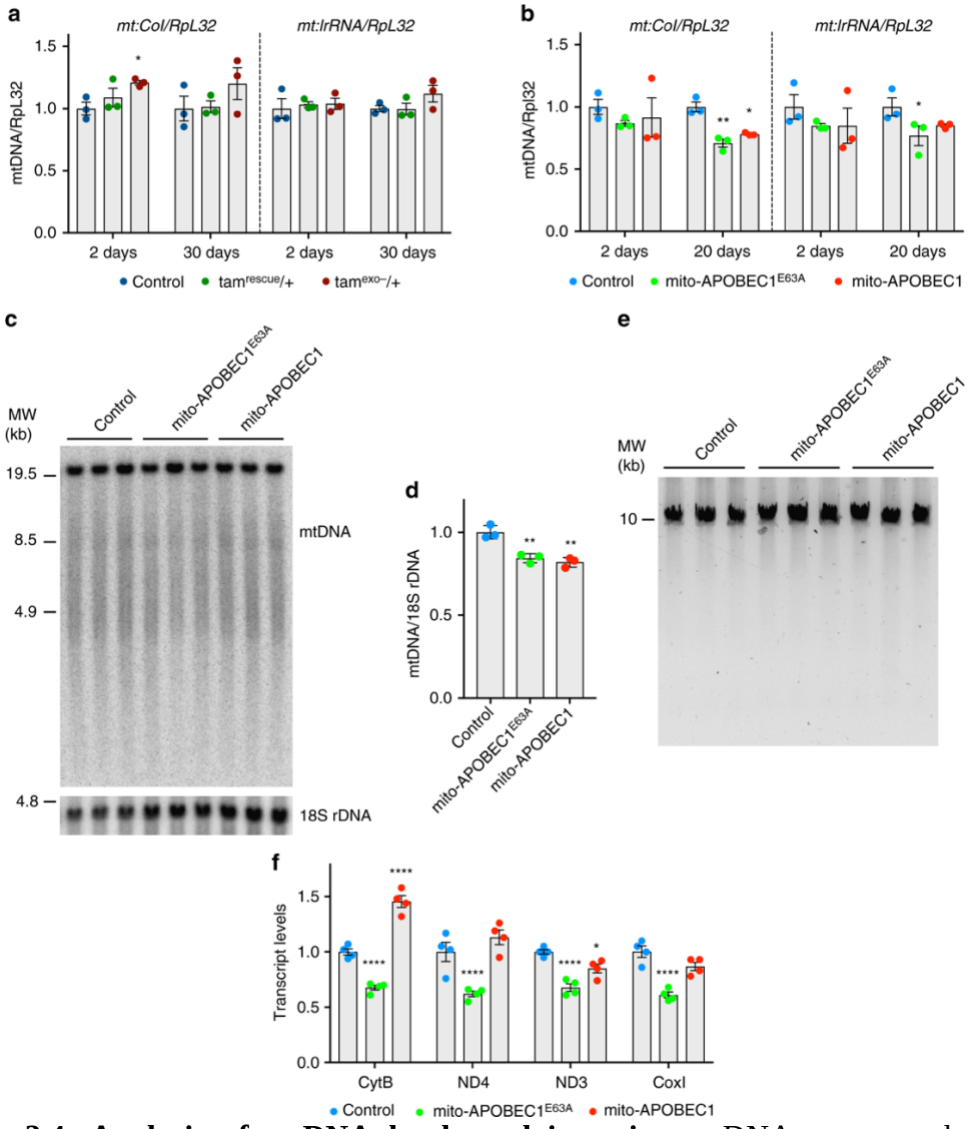


Figure 3.4. Analysis of mtDNA levels and integrity. mtDNA copy number was analyzed in 10-day-old (**A**) maternally inherited *tam*$_{exo-}$ and (**B**) *mito-APOBEC1* mutator flies and respective controls by quantitative-PCR of two regions of the mitochondrial genome (*mt:CoI* and *mt:lrRNA*) against the level of nuclear gene, *RpL32*. Charts show means ± SEM, $n = 3$ biologically independent samples; data points indicate independent experiments. Statistical analysis used two-way ANOVA with Turkey's multiple comparison test. * $p < 0.05$, ** $p < 0.01$. (**C**) Southern blot, quantified in (**D**), and (**E**) long-range PCR analyses further confirm mtDNA integrity and levels in *mito-APOBEC1* flies and relative controls. Charts in (**D**) show means ± SD, $n = 3$ biologically independent samples. Statistical analysis used one-way ANOVA with Sidak's multiple comparison test. * $p < 0.05$. (**F**) Levels of mitochondrial transcripts in one-week-old *mito-APOBEC1* flies and controls. Charts show means ± SEM, $n = 4$ biologically independent samples.

Statistical analysis used two-way ANOVA with Turkey's multiple comparison test. * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$. All other comparisons are non-significant.

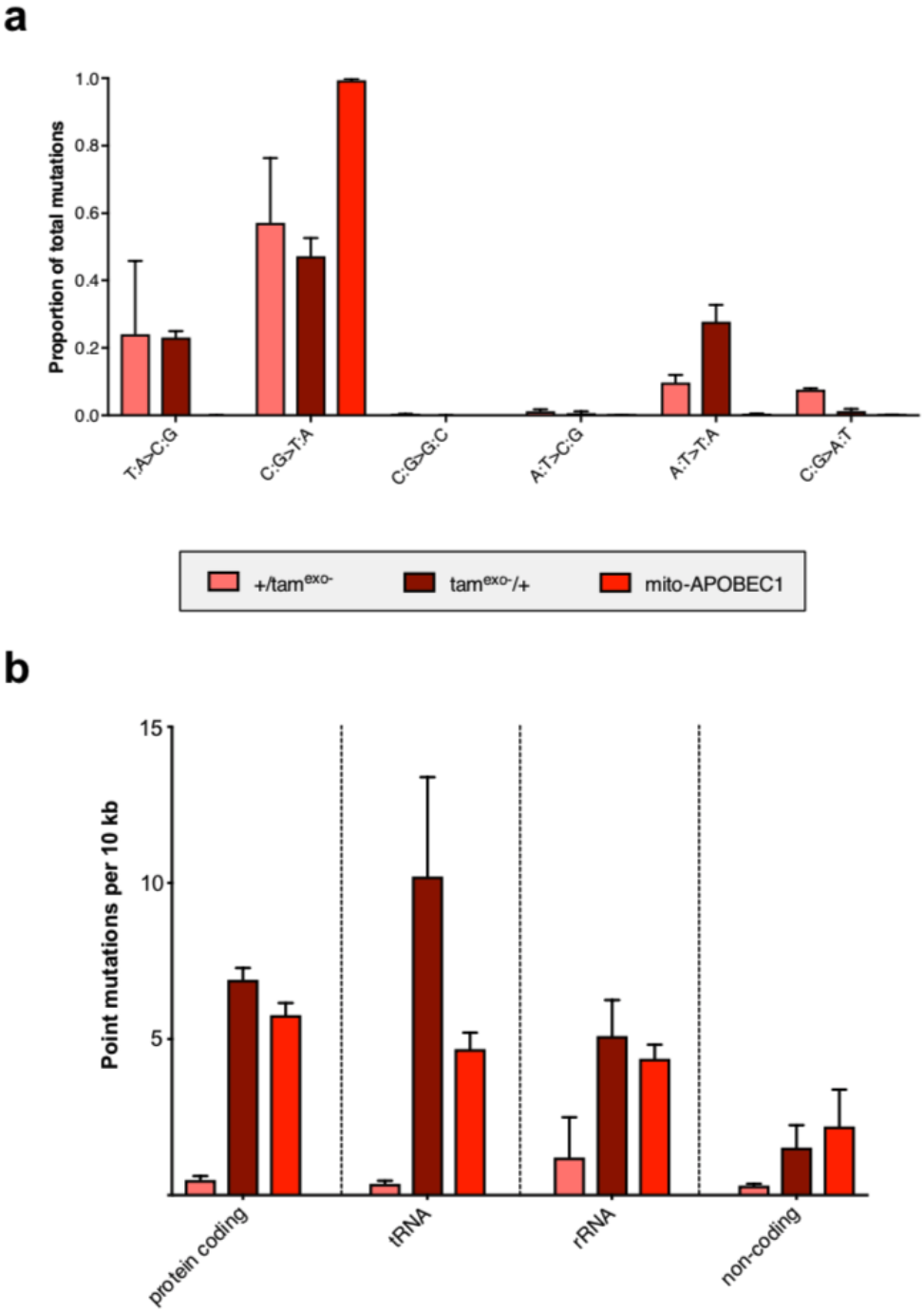


Figure 3.5. *mito-APOBEC1* is highly specific for C:G>T:A transitions across the genome. (**A**) Proportion of total mutations observed expressed by mutation type, as indicated; and (**B**) point mutation load in protein coding genes, tRNAs, rRNAs or non-coding sequence for paternally inherited *tam*$^{exo-}$ (+/*tam*$^{exo-}$), maternally inherited *tam*$^{exo-}$ (*tam*$^{exo-}$/+) and *mito-APOBEC1* mutator flies. Charts show mean ± SD, $n = 3$ animals.

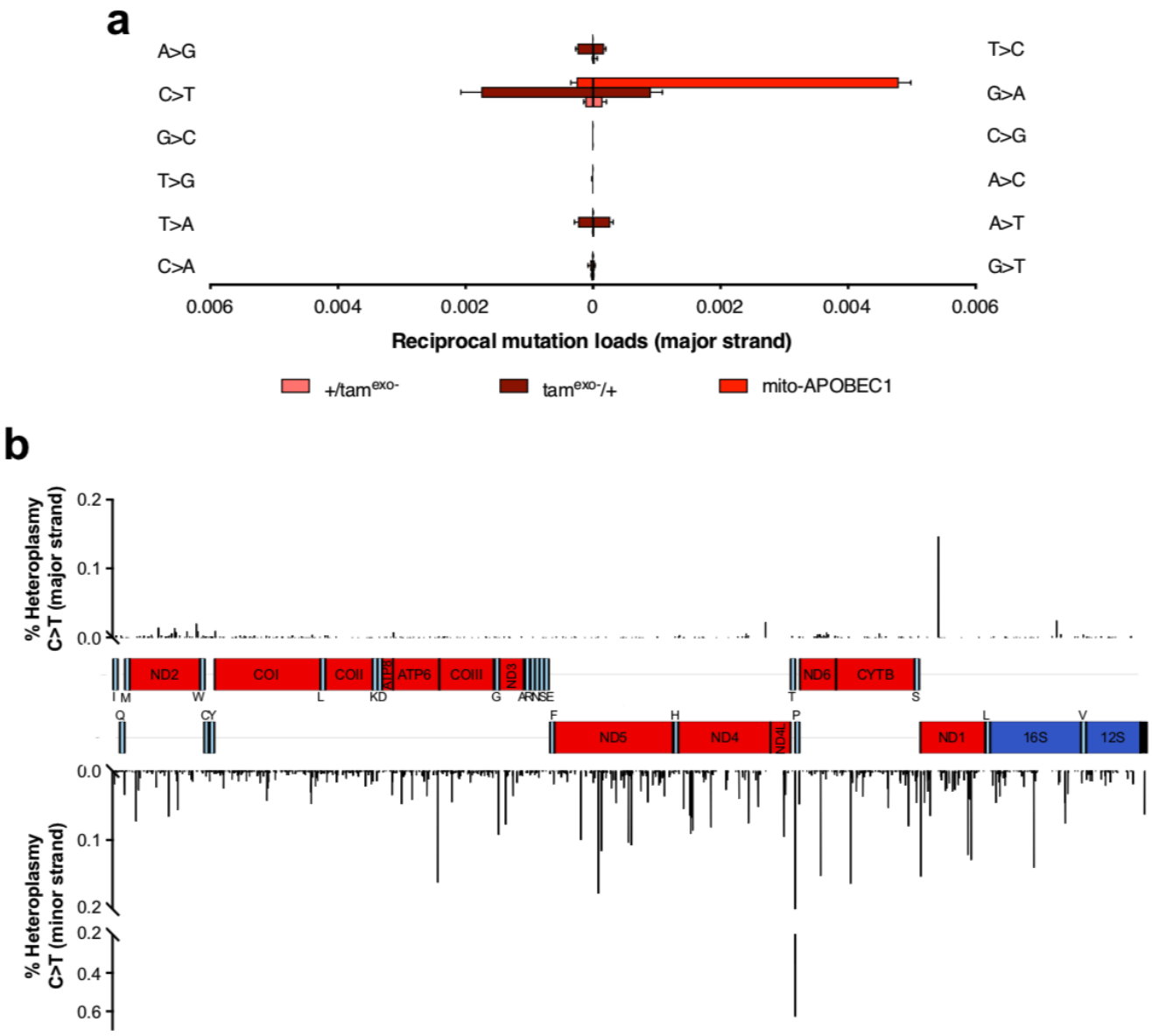


Figure 3.6. *mito-APOBEC1* preferentially targets the minor strand. (**A**) Frequency of reciprocal mutation types on the major strand (reference sequence) in 10-day-old paternally inherited *tam*$^{exo-}$ (+/*tam*$^{exo-}$), maternally inherited *tam*$^{exo-}$ (*tam*$^{exo-}$/+) and *mito-APOBEC1* mutator flies. Chart show mean ± SD, *n* = 3 animals per genotype. (**B**) Distribution of average heteroplasmies for C>T mutations along the major strand (top) and minor strand (i.e. G>A substitutions in the major strand, bottom) in *mito-APOBEC1* flies. Genome map per each strand is depicted as red: protein coding genes; light blue: tRNAs; dark blue: rRNAs. Gene names and one-letter tRNA symbols are indicated.

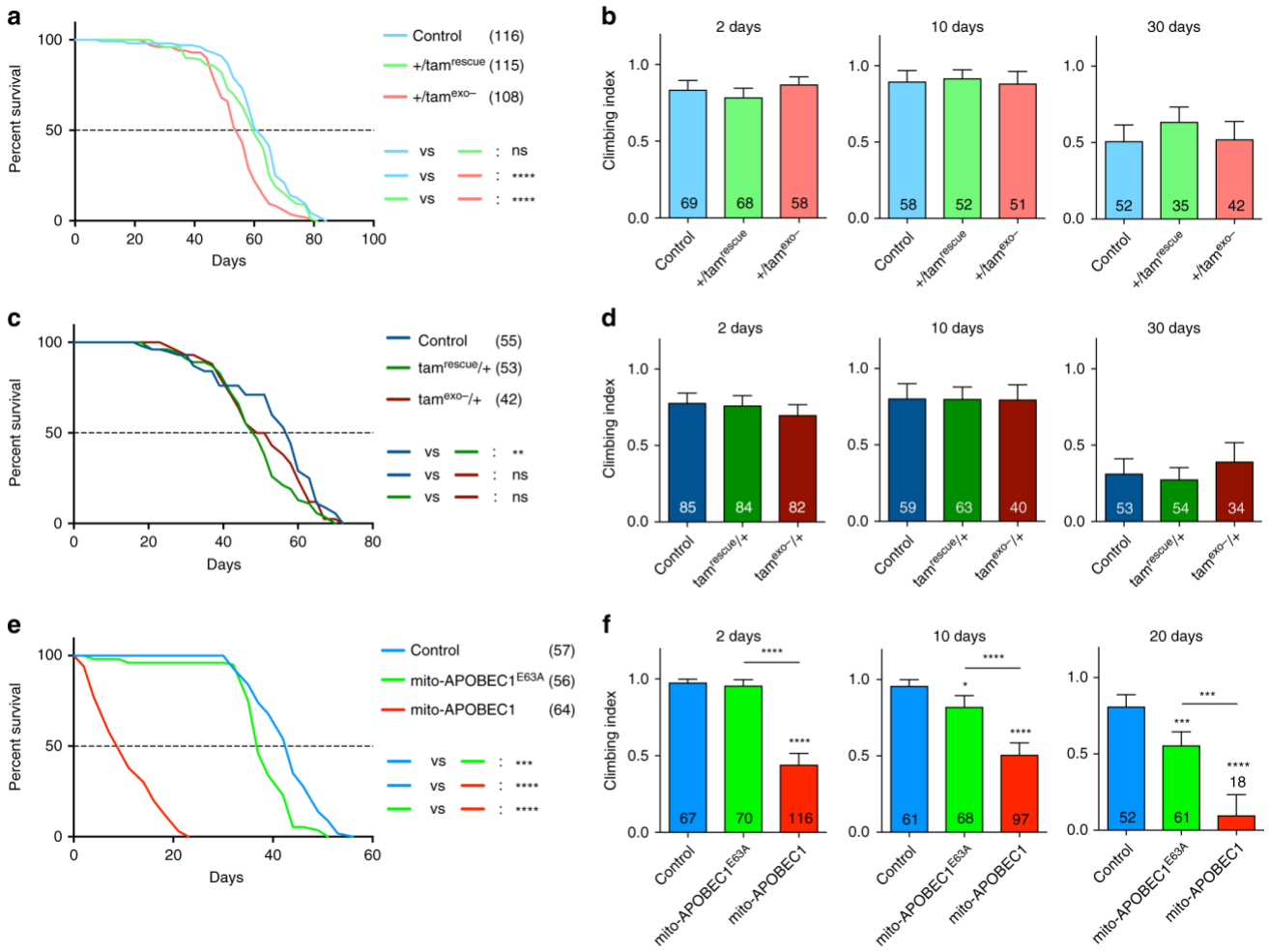


Figure 3.7. *mito-APOBEC1* but not *tam*exo- impacts fly lifespan and locomotor ability. (**A**,**C**, **E**) Lifespan and (**B**, **D**, **F**) locomotor (climbing) ability were assessed for (**A**,**B**) paternally inherited *tam*exo-, (**C**, **D**) maternally inherited *tam*exo- and (**E**, **F**) *mito-APOBEC1* flies and respective controls. Only male flies were tested. For lifespan, statistical significance was determined by Log-rank (Mantel Cox) test; the number of animals tested are indicated in the legend. For locomotor assays, charts show mean ± 95% confidence interval. The number of animals tested are shown per column. Statistical significance was determined by Kruskal-Wallis non-parametric test with Dunn's multiple comparisons correction. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. All other comparisons are non-significant.

This striking difference in lifespan led us to investigate whether there was a similar differential effect on vitality during aging as an indicator of healthspan. Locomotor assays, such as startle-induced negative geotaxis (climbing), offer a sensitive and reliable read-out of the neuromuscular circuit function. Analyzing the climbing ability of the *tamas* mutator lines, either paternally or maternally inherited, we found no difference between any of the genotypes in 2-, 10- and 30-day-old flies (Figure 3.7B, D). However, consistent with the effects on lifespan, expression of *mito-APOBEC1* caused a significant reduction in climbing ability across all ages tested (Figure 3.7F).

Given the strong phenotypic effects of *mito-APOBEC1* but not *tam*$_{exo-}$, we sought to address potentially unrecognized extra-mitochondrial effects of this construct. To this end, we generated transgenic lines expressing HA-tagged wild-type and mutant *APOBEC1* variants lacking the MTS but retaining the NES sequences (Figure 3.8A), which we termed '*cyto-APOBEC1*' and '*cyto-APOBEC1*$_{E63A}$'. The *cyto-APOBEC1* lines appear to express the transgene at higher level than *mito-APOBEC1* lines, in comparison to *mito-HA-GFP* (compare Figure 3.3B with Figure 3.8B); yet encouragingly, we observed no effect of *cyto-APOBEC1* expression upon lifespan or locomotor ability with age (Figure 3.8C, D). Thus, there appears to be no detrimental effects of *cyto-APOBEC1*, substantiating the idea that the impact of *mito-APOBEC1* expression on organismal fitness derives exclusively from mtDNA mutations.

3.3.4 *mito-APOBEC1 but not tam*$_{exo-}$ *affects mitochondrial function*

Having established the relative impact of the mutator models on organismal fitness, we next sought to understand the impact on mitochondrial function. We first assessed the oxygen consumption rate (OCR) using high-resolution respirometry. Analyzing the maternally inherited *tam*$_{exo-}$ heterozygotes we found no significant change to either Complex I- or Complex II-linked

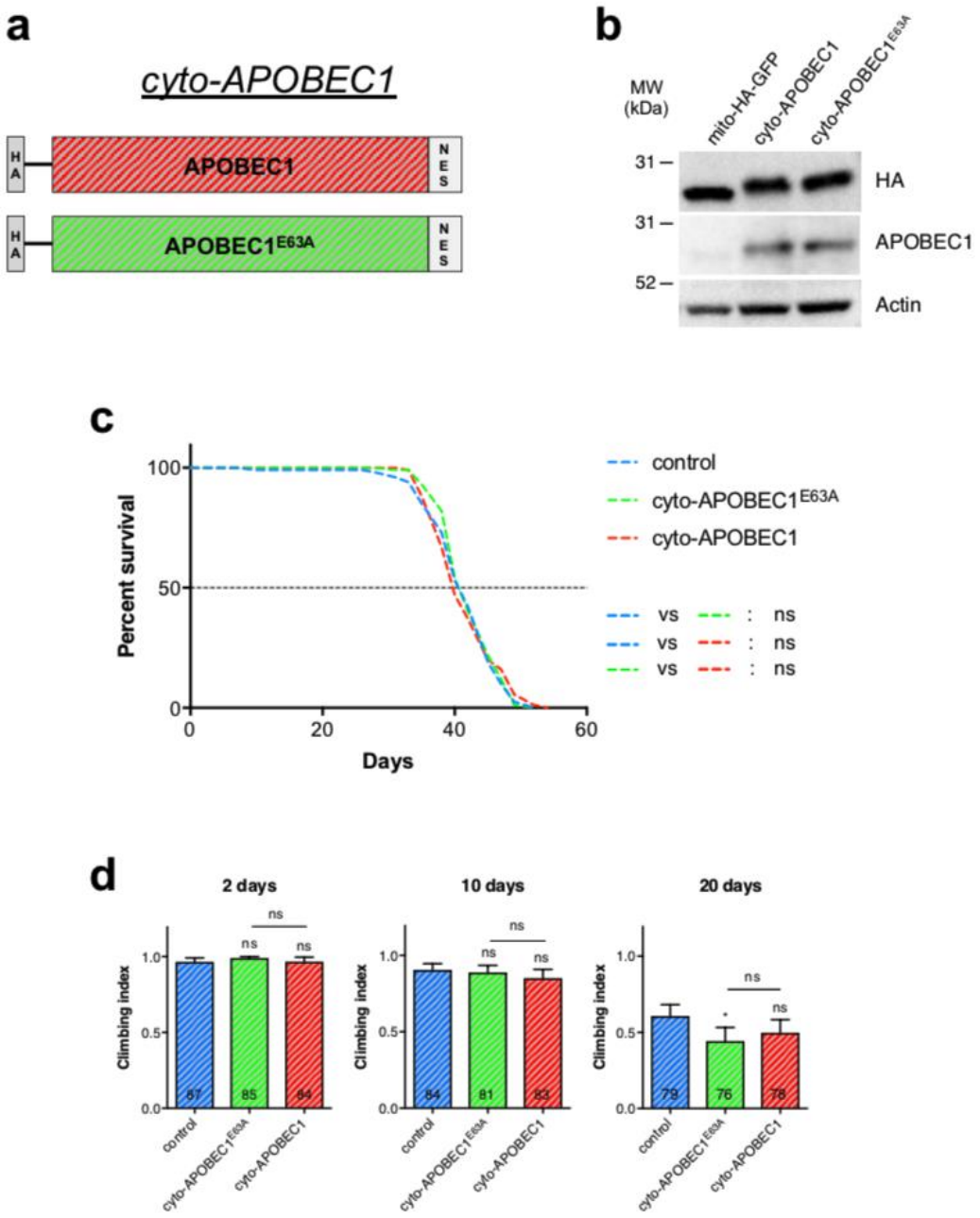


Figure 3.8. Expression of *cyto-APOBEC1* does not cause lifespan or behavioral defects. (**A**) Schematic of *cyto-APOBEC1* construct and inactive (E63A) mutant. (**B**) Immunoblot showing *cyto- APOBEC1* transgene expression in whole fly extracts, using both an anti-APOBEC1 antibody and anti-HA. Transgenic *mito-HA-GFP* expression is used as a positive control, and anti-Actin immunostaining as loading control. (**C**) Lifespan of *cyto-APOBEC1* flies is no different to that of *cyto-APOBEC1*^{E63A} and control flies expressing *mito-HA-GFP*. Log-rank (Mantel-Cox) test. (**D**) Climbing ability of *cyto-APOBEC1* flies is not affected across different ages. Kruskal-Wallis test with

Dunn's multiple comparisons correction. The number of flies tested are shown per column. Charts show mean ± 95% confidence interval. * $p < 0.05$.

respiration, even after substantial aging (Figure 3.9A). In contrast, *mito-APOBEC1* expressing flies showed a significant decrease in Complex I-linked respiration, already in very young flies (Figure 3.9B). Although the downturn in Complex II-linked respiration did not reach significance at 2 days, by 10 days old this too was significantly reduced (Figure 3.9B). Because Complex-II is exclusively composed by nuclear-encoded proteins, we hypothesized the defect was due to dysfunctional Complex-III and/or IV. Consistent with these findings, BN-PAGE and in-gel activity assays showed a decrease in Complex-I and Complex-IV activity, while Complex-II activity was undiminished, even possibly increased (Figure 3.9C). Assessing the steady-state levels of the assembled complexes by native-PAGE immunoblotting we found that *mito-APOBEC1* flies consistently displayed decreased amounts of higher-order assemblies of Complexes I, III, IV and V (Figure 3.9D), with partial accumulation of non-functional intermediates.

MtDNA mutations could potentially also impact transcription, causing a decline in transcript levels. To assess this, we analyzed several transcripts from the mitochondrial genome by quantitative RT-PCR. Expression of the catalytically dead *mito-APOBEC1*^{E63A} resulted in a moderate but significant reduction of all mitochondrial transcripts analyzed (Figure 3.4F), which mirrored the lowered amounts of mtDNA copies in these flies (Figure 3.4B, D). In contrast, transcript levels in *mito-APOBEC1* flies were comparable or slightly higher than controls (Figure 3.4F). Taken together, these data indicate that *mito-APOBEC1*-induced mutations compromise OXPHOS assembly and respiration likely by the cumulative impact on mitochondrial-encoded OXPHOS components.

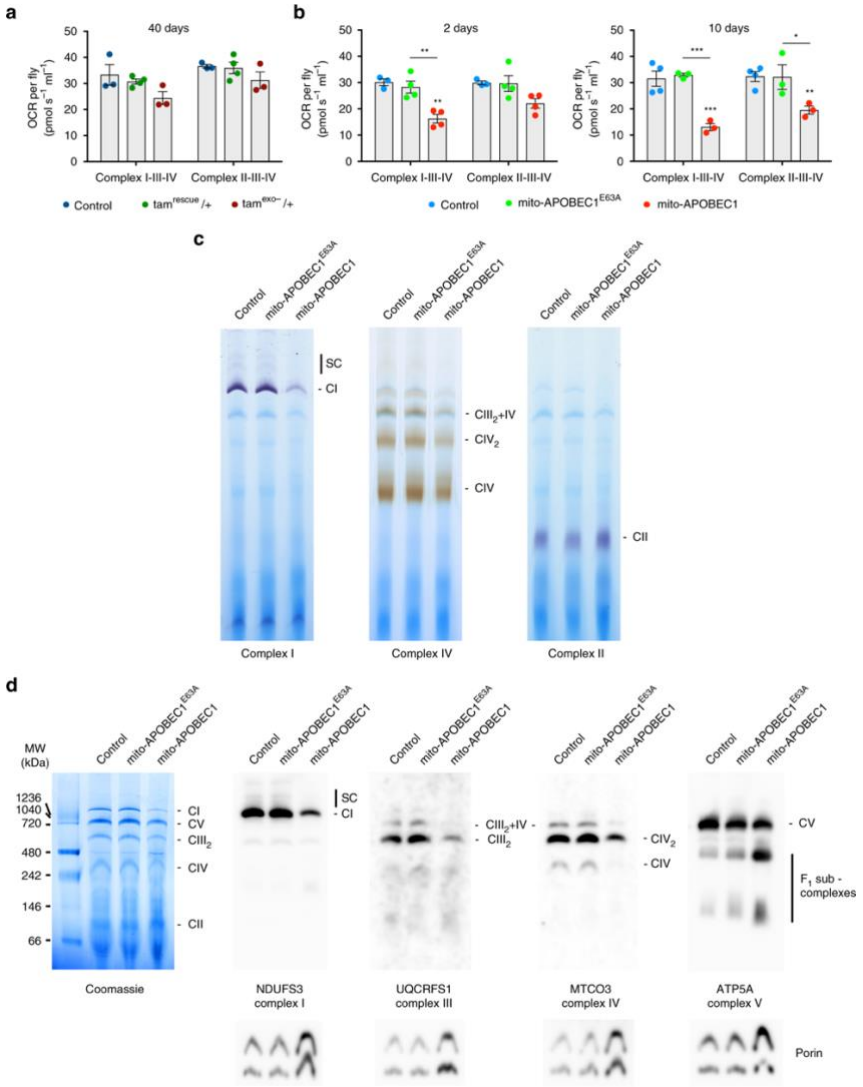


Figure 3.9. Mitochondrial respiration is compromised in *mito-APOBEC1* but not *tam*$_{exo-}$ flies. Respiratory activity was measured in (**A**) maternally inherited *tam*$_{exo-}$ and (**B**) *mito-APOBEC1* flies and respective controls by Complex I- or Complex II-linked oxygen consumption rates (OCR) in isolated mitochondria from flies at the indicated age. Values were normalized per fly. Charts show mean ± SEM, n = 3 biologically independent samples. Statistical analysis used one-way ANOVA with Sidak's multiple comparisons test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All other comparisons are non-significant (NS). Blue Native-PAGE followed by in-gel activity assay (**C**) or immunoblotting (**D**) of mitochondrial-enriched samples from *mito-APOBEC1* flies. (**c**) Gels were incubated for Complex-I or Complex-II activities (violet), or Complex-IV activity (brown). Equal amounts of total protein were loaded per lane. (**D**) Blue-Native gels were Coomassie stained or used for immunoblotting against NDUFS3 (*Dm* ND-30; Complex-I), UQCRFS1/Rieske (*Dm* RFeSP; Complex-III), MTCO3 (*Dm* mt:CoIII; Complex-IV) and ATP5A (*Dm* Blw; Complex-V). Anti-Porin was used as a loading

control. CI, Complex-I; CII, Complex-II; CIII, Complex-III; $CIII_2$, Complex-III dimer; CIV, Complex-IV; CIV_2, Complex-IV dimer; CV, Complex-V; SC, super-complexes.

3.3.5 *Analysis of the mtDNA mutation profiles*

An unexpected finding from our work concerns the similar mutation loads found in maternal *tam*$_{exo-}$ and *mito-APOBEC1* flies, but the strikingly different organismal consequences of those mutations. This suggests that the amount of mtDNA mutations alone is not sufficient to explain the difference in pathogenicity. Therefore, we undertook a deeper analysis of the mtDNA mutation profile observed in the two mutator systems. We first assessed the degree to which the observed mutations may have undergone clonal expansion. Mutations that occur at less than 1% heteroplasmy are generally considered to represent recently acquired or *de novo* mutational events, while mutations present at higher than 1% heteroplasmy are thought to have expanded clonally through replication of an initial event. We compared the overall mutation load, up to 70% heteroplasmy, with that found at less than 1% heteroplasmy (Figure 3.10A-C, Table 3.1, Table 3.3). Paternally inherited *tam*$_{exo-}$ flies showed very little difference in mutation load between <1% and <70% heteroplasmy (Figure 3.10A), consistent with these mutations being recently acquired. Indeed, only a few sites in paternally inherited *tam*$_{exo-}$ flies showed >1% heteroplasmy (Figure 3.10D, Figure 3.1B). In contrast, a substantial proportion of mutations in *mito-APOBEC1* and maternally inherited *tam*$_{exo-}$ occurred at higher than 1% heteroplasmy (Figure 3.10B, C, and Figure 3.1B). Notably, 23% of all mutated sites in *mito-APOBEC1* (Figure 3.10D) displayed a moderate level of heteroplasmy (mainly between 1-10% heteroplasmy, Figure 3.10F), while in *tam*$_{exo-}$/+ a smaller proportion of clonally mutated sites (13% of all mutated sites, Figure 3.10D) distributed as a few moderately heteroplasmic positions together with a handful of high clonally (>30% heteroplasmy) mutated sites (Figure 3.10E).

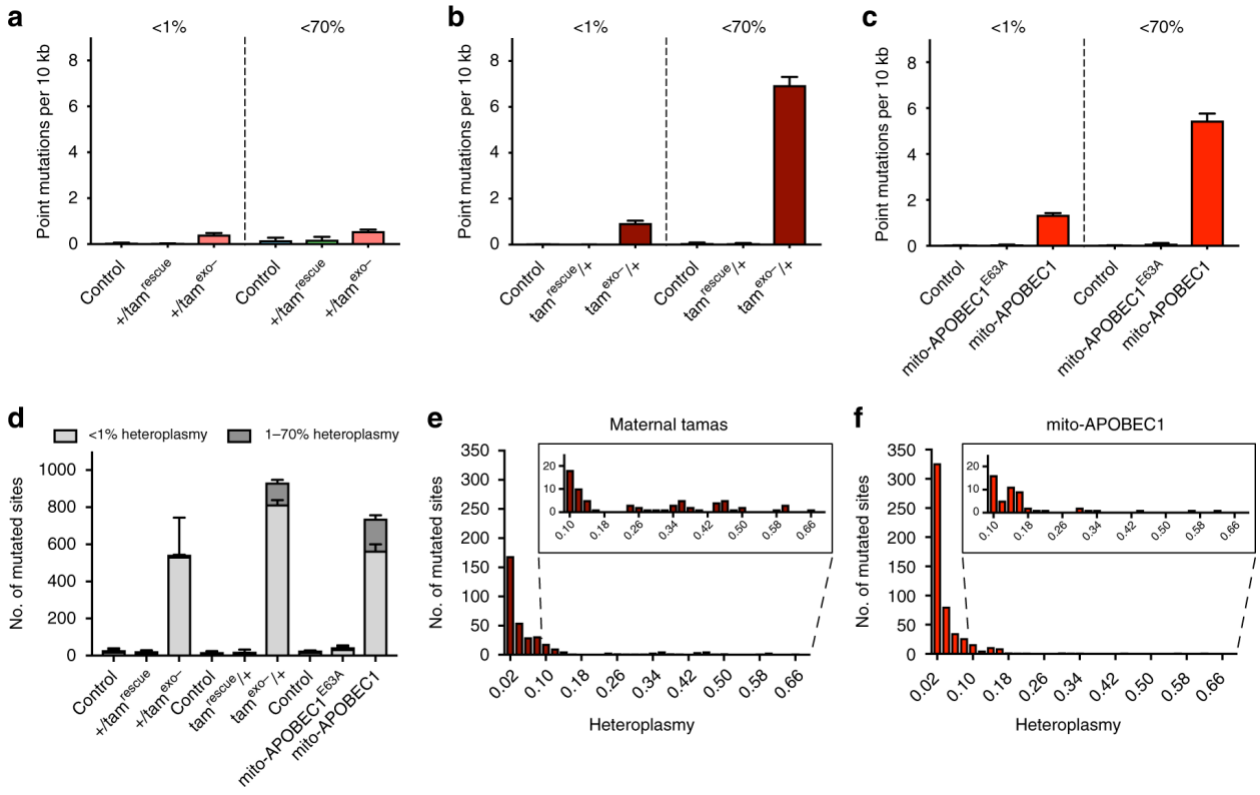


Figure 3.10. High levels of heteroplasmy occurs in maternal *tam*$_{exo-}$ and *mito-APOBEC1*. Point mutation load of low heteroplasmy events (<1%) compared to all observed mutations (<70% heteroplasmy) found in 10-day-old (**A**) paternally inherited *tam*$_{exo-}$, (**B**) maternally inherited *tam*$_{exo-}$ and (**C**) *mito-APOBEC1* mutator flies and respective controls. Mutation loads for total mutations (<70% heteroplasmy) are the same as shown in Fig. 3.2A and reproduced here for comparison. (**D**) Number of sites that show low (<1%) or moderate (1-70%) heteroplasmy. Charts show mean ± SD, n = 3 animals. (**E**, **F**) Frequency distribution of >1% heteroplasmic sites in (**E**) maternally inherited *tam*$_{exo-}$ and (**F**) *mito-APOBEC1* flies. Each chart shows the frequency distribution of the number of mutated sites with heteroplasmy bin width = 0.02, centered as indicated. Sites with heteroplasmy spanning 9% to 69% are zoomed.

Table 3.3. *de novo* point mutations identified by Duplex Sequencing in 10-day-old flies. *de novo* mutations observed at <1% clonality. DCS denotes the duplex consensus sequence, the number of post-processing nucleotides sequenced.

		# DCS	# mutated sites	# mutations	Mutation load
paternal tamas, 10 days old	w1118_1	3010367	5	6	1.99E-06
	w1118_2	7961128	14	14	1.76E-06
	w1118_3	8110354	42	56	6.90E-06
	w1118_AVG		20		3.55E-06
	+/tam_{rescue}_1	8728147	9	9	1.03E-06
	+/tam_{rescue}_2	9143056	9	9	9.84E-07
	+/tam_{rescue}_3	8100463	31	32	3.95E-06
	+/tam_{rescue}_AVG		16		1.99E-06
	+/tam_{exo-}_1	17481755	725	814	4.66E-05
	+/tam_{exo-}_2	14005318	567	623	4.45E-05
	+/tam_{exo-}_3	9852168	310	324	3.29E-05
	+/tam_{exo-}_AVG		534		4.13E-05
maternal tamas, 10 days old	wDah_1	5693981	9	9	1.58E-06
	wDah_2	7246279	7	7	9.66E-07
	wDah_3	6577020	13	13	1.98E-06
	wDah_AVG		10		1.51E-06
	tam_{rescue}/+_1	9499339	7	7	7.37E-07
	tam_{rescue}/+_2	10713046	11	12	1.12E-06
	tam_{rescue}/+_3	9734980	12	14	1.44E-06
	tam_{rescue}/+_AVG		10		1.10E-06
	tam_{exo-}/+_1	11067470	799	1137	1.03E-04
	tam_{exo-}/+_2	14561594	841	1202	8.25E-05
	tam_{exo-}/+_3	11483065	803	1111	9.68E-05
	tam_{exo-}/+_AVG		814		9.40E-05
mito-APOBEC1, 10 days old	mito-GFP_1	14922865	17	24	1.61E-06
	mito-GFP_2	12278047	29	36	2.93E-06
	mito-GFP_3	14130656	14	20	1.42E-06
	mito-GFP_AVG		20		1.99E-06
	mito-APOBEC1$_{E63A}$_1	10853172	36	39	3.59E-06
	mito-APOBEC1$_{E63A}$_2	11007973	55	64	5.81E-06
	mito-APOBEC1$_{E63A}$_3	5266167	9	10	1.90E-06
	mito-APOBEC1$_{E63A}$_AVG		33		3.77E-06
	mito-APOBEC1_1	10162080	571	1456	1.43E-04
	mito-APOBEC1_2	11318537	596	1531	1.35E-04
	mito-APOBEC1_3	8666010	528	1116	1.29E-04
	mito-APOBEC1_AVG		565		1.36E-04

Since the preceding mutation analyses were performed on 10-day-old flies, we wanted to determine the extent to which these mutations were present in early adulthood, in 2-day-old flies (Table 3.4, Table 3.5). Mutation loads were similar between 2- and 10-day-old flies across all genotypes tested (Figure 3.11A). However, while a similar number of sites were mutated in *mito-APOBEC1* flies at 2 and 10 days, the number of mutated sites approximately doubled in both *tam*$_{exo-}$ conditions (Figure 3.11B). Therefore, during this short time frame, *tam*$_{exo-}$ was able to introduce mutations at additional genomic positions, consistent with reports that mtDNA replication indeed occurs in adult fly tissues (*62*). In contrast, the lack of increase in *mito-APOBEC1*-targeted sites suggests the mutable positions may have already been saturated by 2 days. Consistent with this, comparing the heteroplasmy profile of 2- and 10-day-old *tam*$_{exo-}$ flies showed no appreciable increase in the number of clonally mutated sites (1-70% heteroplasmy) while the number of sites with *de novo* mutations (<1% heteroplasmy) specifically increased (Figure 3.11B). It should be noted that the additional mutations at <1% heteroplasmy in 10-day-old flies are numerically very small compared to the large number of mutations present at higher (1-70%) heteroplasmy, accounting for why the overall point mutation load does not significantly increase with age. As above, the heteroplasmy profile of *mito-APOBEC1*-targeted sites did not change between 2 and 10 days (Figure 3.11B).

Being unable to ascribe the divergent organismal outcome of *tam*$_{exo-}$ and *mito-APOBEC1* flies to difference in their heteroplasmy profile, we next considered whether the topography of the mutations may offer some explanation as to the different pathogenicity. In all three mutator models, point mutations were widely distributed across the mtDNA genome (Figure 3.1), with no particular bias towards any particular region or genomic feature; protein coding, tRNA, rRNA or non-coding (Figure 3.5B). Focusing on the protein coding genes, we calculated the ratio of non-

Table 3.4. Point mutations identified by Duplex Sequence in 2-day-old flies. Total mutations observed at ≤70% clonality. DCS denotes the duplex consensus sequence, the number of post-processing nucleotides sequenced.

		# DCS	# mutated sites	# mutations	Mutation load
tamas 2 days old	w1118_1	12585825	18	29	2.30E-06
	w1118_2	11627685	44	163	1.40E-05
	w1118_3	17727788	28	346	1.95E-05
	w1118_AVG		30		1.19E-05
	+/tam$_{exo-}$_1	6933830	299	360	5.19E-05
	+/tam$_{exo-}$_2	2532004	107	115	4.54E-05
	+/tam$_{exo-}$_3	9200101	355	560	6.09E-05
	+/tam$_{exo-}$_AVG		254		5.27E-05
	tam$_{exo-}$/+_1	4928489	441	3595	7.29E-04
	tam$_{exo-}$/+_2	3343360	414	1162	3.48E-04
	tam$_{exo-}$/+_3	5386567	488	4162	7.73E-04
	tam$_{exo-}$/+_AVG		448		6.17E-04
mito-APOBEC1 2 days old	control_1	14243779	23	49	3.44E-06
	control_2	12231441	11	48	3.92E-06
	control_3	10058436	16	64	6.36E-06
	control_AVG		17		4.58E-06
	mito-APOBEC1$_{E63A}$_1	9477753	31	75	7.91E-06
	mito-APOBEC1$_{E63A}$_2	16084817	21	272	1.69E-05
	mito-APOBEC1$_{E63A}$_3	13035706	15	16	1.23E-06
	mito-APOBEC1$_{E63A}$_AVG		22		8.68E-06
	mito-APOBEC1_1	14125154	768	6665	4.72E-04
	mito-APOBEC1_2	10334602	683	4568	4.42E-04
	mito-APOBEC1_3	10431852	725	5435	5.21E-04
	mito-APOBEC1_AVG		725		4.78E-04

Table 3.5. de novo point mutations identified by Duplex Sequencing in 2-day-old flies. *de novo* mutations observed at <1% clonality. DCS denotes the duplex consensus sequence, the number of post-processing nucleotides sequenced.

		# DCS	# mutated sites	# mutations	Mutation load
tamas 2 days old	w1118_1	12585825	18	29	2.30418E-06
	w1118_2	11627685	39	50	4.30008E-06
	w1118_3	17727788	27	36	2.03071E-06
	w1118_AVG		28		2.87832E-06
	+/tam$_{exo-}$_1	6933830	351	380	5.48038E-05
	+/tam$_{exo-}$_2	2532004	107	115	4.54186E-05
	+/tam$_{exo-}$_3	9200101	355	560	6.08689E-05
	+/tam$_{exo-}$_AVG		271		5.36971E-05
	tam$_{exo-}$/+_1	4928489	387	467	9.47552E-05
	tam$_{exo-}$/+_2	3343360	310	346	0.000103489
	tam$_{exo-}$/+_3	5386567	340	386	7.16597E-05
	tam$_{exo-}$/+_AVG		346		8.99679E-05
mito-APOBEC1 2 days old	control_1	14243779	20	32	2.24659E-06
	control_2	12231441	7	7	5.72296E-07
	control_3	10058436	13	13	1.29245E-06
	control_AVG		13		1.37045E-06
	mito-APOBEC1$_{E63A}$_1	9477753	28	28	2.95429E-06
	mito-APOBEC1$_{E63A}$_2	16084817	18	21	1.30558E-06
	mito-APOBEC1$_{E63A}$_3	13035706	13	13	9.97261E-07
	mito-APOBEC1$_{E63A}$_AVG		20		1.75238E-06
	mito-APOBEC1_1	14125154	606	1804	0.000127715
	mito-APOBEC1_2	10334602	536	1325	0.00012821
	mito-APOBEC1_3	10431852	556	1413	0.000135451
	mito-APOBEC1_AVG		566		0.000130459

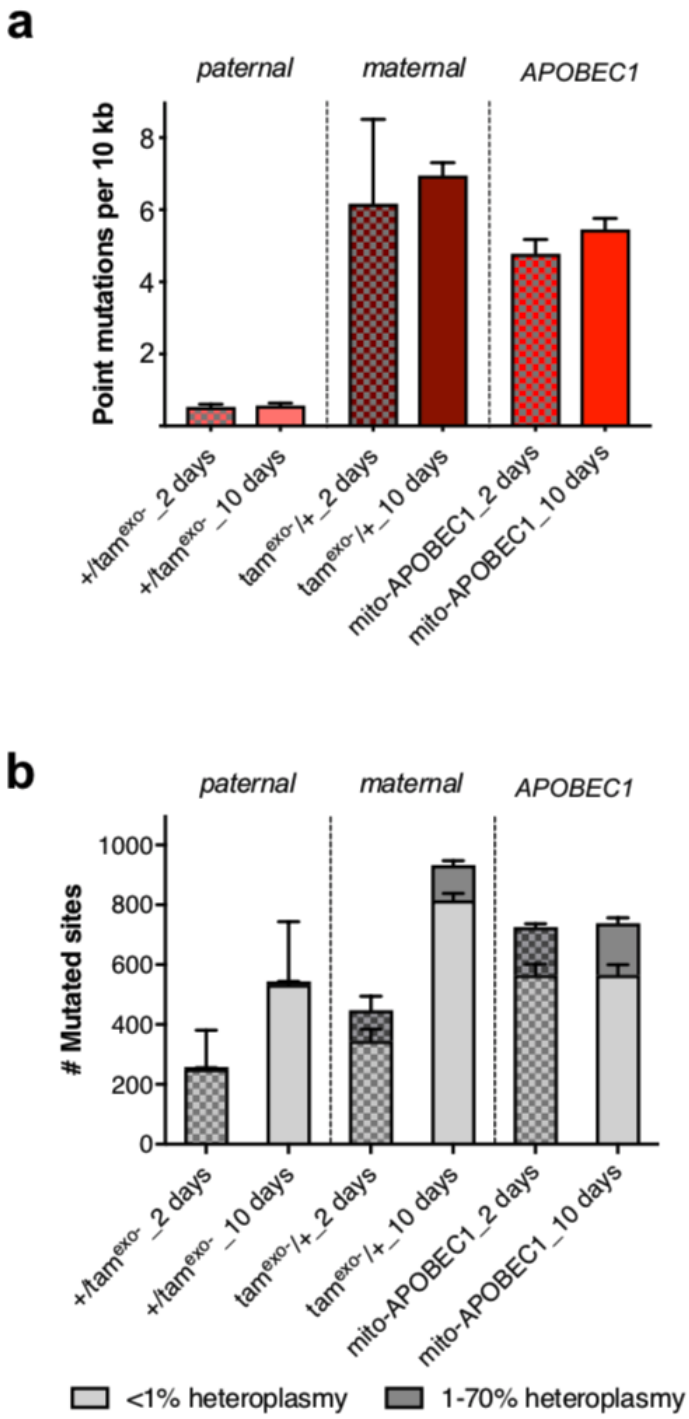


Figure 3.11. The number of mutated sites increases in *tam*$^{exo-}$ but not in *mito-APOBEC1* flies over time. (**A**) Point mutation load and (**B**) number of mutated sites in paternally inherited *tam*$^{exo-}$ (+/*tam*$^{exo-}$), maternally inherited *tam*$^{exo-}$ (*tam*$^{exo-}$/+) and *mito-APOBEC1* mutator flies at 2 and 10 days of age. Charts show mean ± SD, *n* = 3 animals.

synonymous to synonymous (NS:S) mutations found in each genotype (Figure 3.12A). Non-synonymous mutations in either of the tamexo- mutator lines essentially exist as recent or de novo events (<1% heteroplasmy, Figure 3.13A), and the NS:S ratio in these lines is low (Figure 3.13B) and similar to the ratio found in controls (Figure 3.12A). In contrast, mito-APOBEC1 produced a high number of NS mutations, particularly at sites with >1% heteroplasmy (Figure 3.13A), resulting in a dramatic increase in the NS:S ratio (Figure 3.13B). Several characteristics of the mutation profile likely contribute to this difference between the tamexo- and mito-APOBEC1 mutators. First, considering that mito-APOBEC1 exclusively targets C:G nucleotides (Figure 3.5A), it is notable that C:Gs are predominantly found in protein coding regions (Figure 3.12B), particularly in the critical first and second codon-base positions (Figure 3.12C). Indeed, mito-APOBEC1 disproportionately (~98% of all recovered mutations) mutates the first or second codon-base position (Figure 3.12D). Consistent with tamexo- creating a greater diversity of substitutions (Figure 3.5A), a substantial portion (21%) of tamexo--induced mutations affects nucleotides at the third position in the codon (Figure 3.12D), sites that tend to be degenerate (mutations lead to synonymous substitutions). Second, mito-APOBEC1 is highly selective for targeting deamination of cytidine in a TCn context, particularly in the TCT trinucleotide sequence (

Figure 3.14), as previously reported (215). Therefore, mito-APOBEC1-induced mutations are restricted to fewer sites in the genome, which may expand to higher heteroplasmy due to multiple, independent targeting events. In contrast, similar to the codon position analysis, tamexo- mutators show much less selectivity for the surrounding nucleotide context (

Figure 3.14). Third, a significant proportion of *tam*$^{exo-}$-induced mutations are T>C substitutions, which are much more likely to result in a synonymous change (Figure 3.12E). Taken together, these data demonstrate that *mito-APOBEC1* induces predominantly non-synonymous mutations.

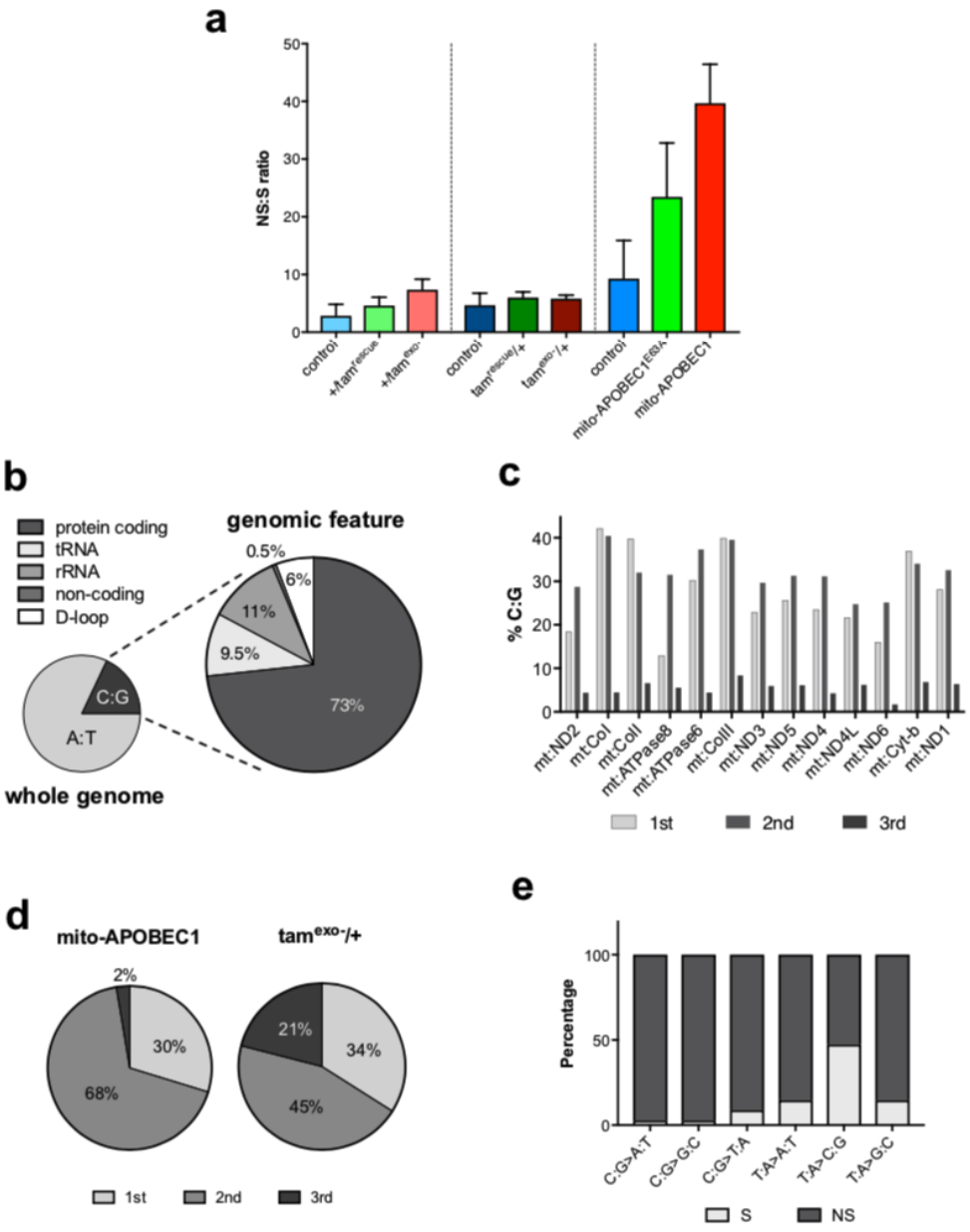


Figure 3.12. mito-APOBEC1 targets 1st and 2nd codon positions, producing non-synonymous substitutions. (**A**) Non-synonymous:synonymous (NS:S) ratio of all substitutions found in protein coding genes for mutator flies and respective controls. (**B**) Distribution of C:G nucleotides across the *Drosophila* mtDNA genome. C:G nucleotides represent 18% of the whole mtDNA genome; and most of them are located in protein coding sequences. (**C**) Percentage of C:G content per codon position for each of the mtDNA protein-coding genes. (**D**) Proportion of all protein-coding mutations from in

mito-APOBEC1 and maternally inherited *tam*$^{exo-}$ (*tam*$^{exo-}$/+) mutator flies in relation to their codon-base position. (**E**) Percentage of C:G and T:A mutations that produce either synonymous (S) or non-synonymous (NS) substitutions based on *Drosophila* mtDNA reference sequence.

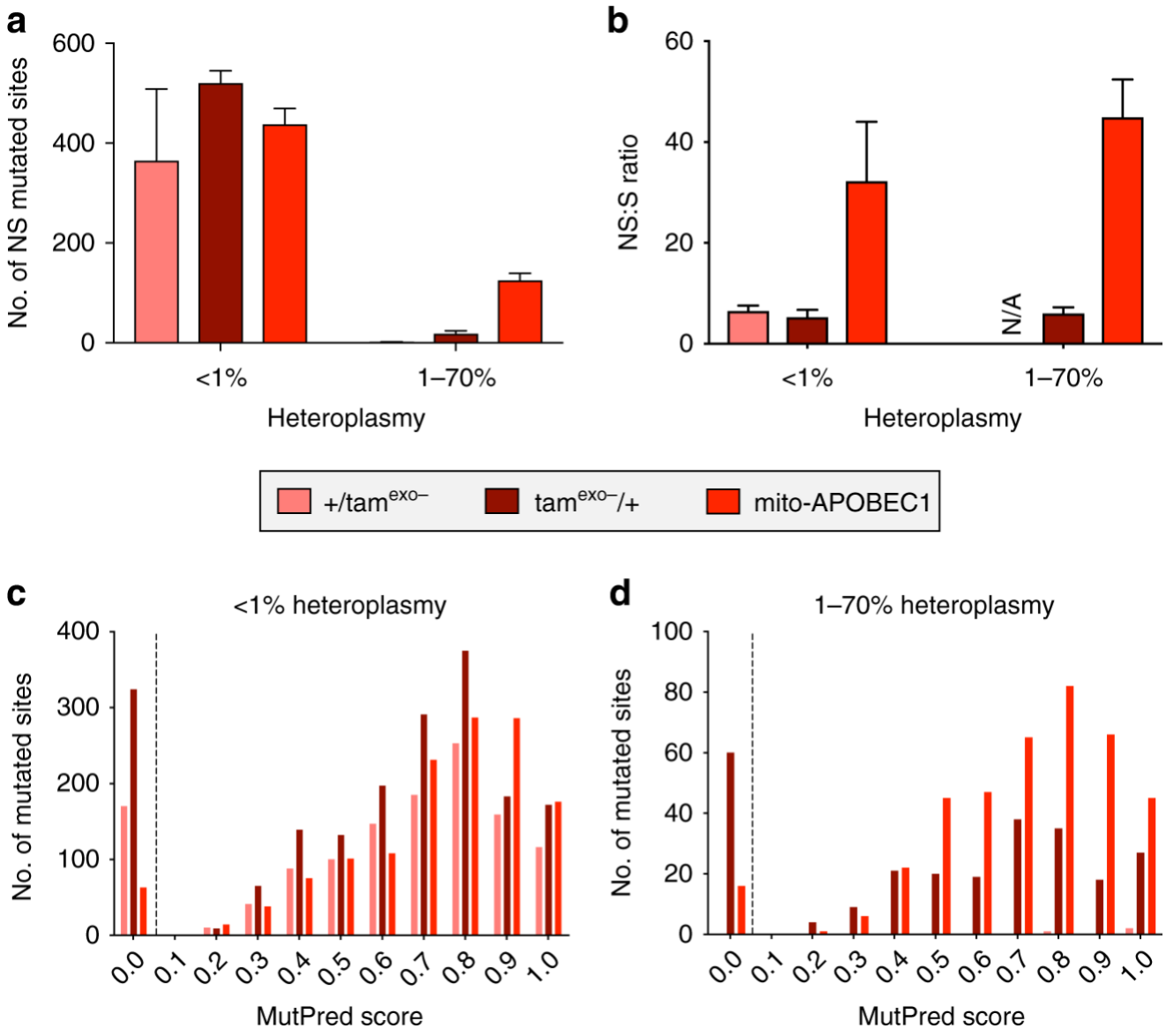


Figure 3.13. mito-APOBEC1 causes high levels of non-synonymous pathogenic mutations. (**A**) Number of genomic sites presenting non-synonymous (NS) substitutions, at <1% or 1-70% heteroplasmy, in the indicated genotypes. (**B**) Non-synonymous:synonymous (NS:S) ratio of all substitutions recovered in protein coding genes, separated by heteroplasmy level as previously indicated [Note: Only one non-synonymous substitution was found at >1% heteroplasmy in paternally inherited *tam*$^{exo-}$ and 0 synonymous mutations; therefore, the ratio is indicated as N/A]. Charts show mean ± SD, *n* = 3 animals. (**C**, **D**) Frequency distribution of MutPred scores, bin width = 0.05, for each mutation found in the indicated genotypes analyzed by (**C**) <1% or (**D**) 1-70% heteroplasmies. (*n* = 3 animals, data from each animal were sum by genotype). A dotted line separates the synonymous (MutPred score = 0) from non-synonymous mutations.

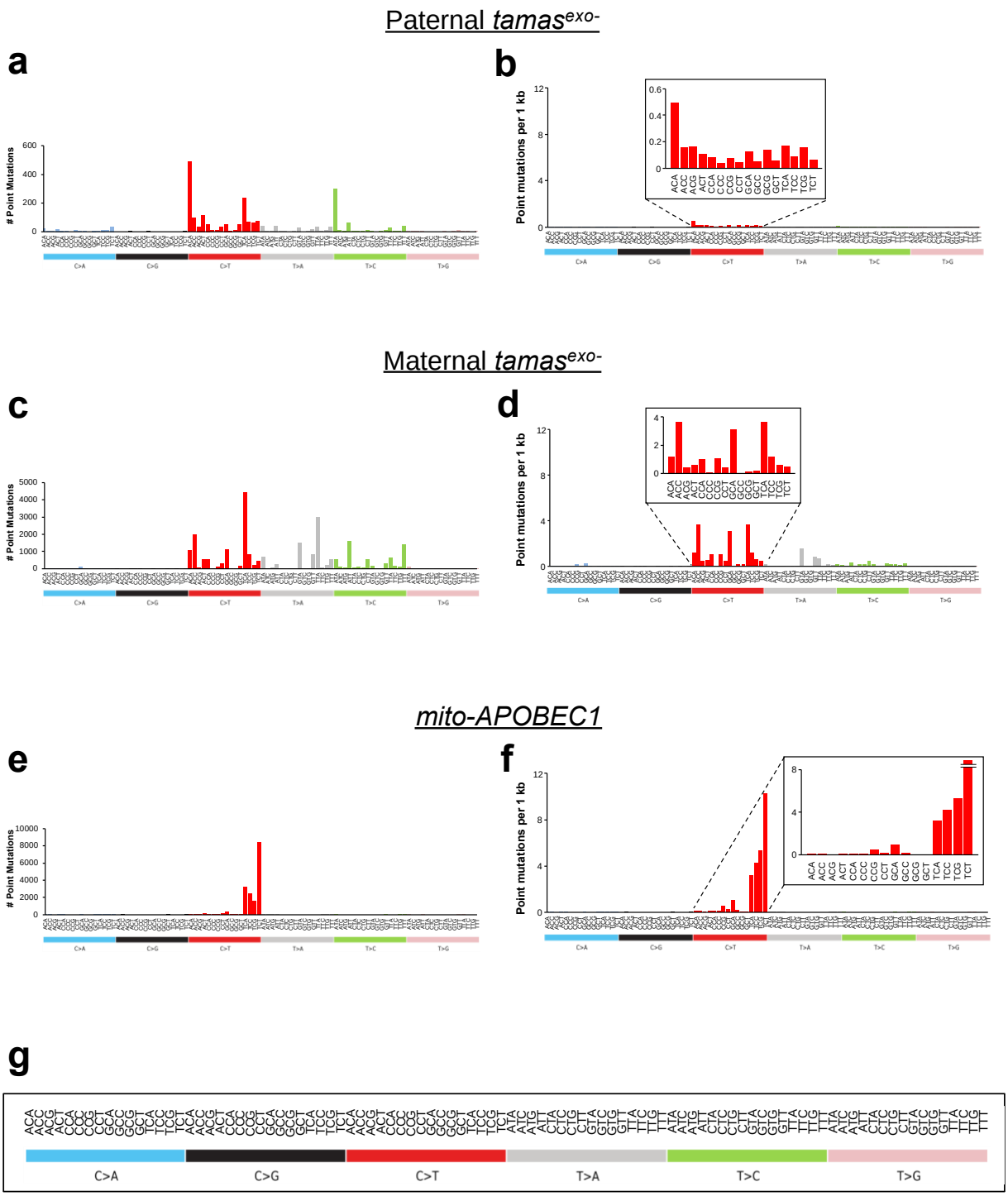


Figure 3.14. *mito-APOBEC1* targets TC dinucleotides with high specificity. (**A**, **C**, **E**) Raw number of C:G and T:A mutations and mutation load (**B**, **D**, **F**) within all 96 possible trinucleotide contexts, detected in 10-day-old flies from (**A**, **B**) paternally inherited *tam*$^{exo-}$ (+/*tam*$^{exo-}$), (**C**, **D**) maternally inherited *tam*$^{exo-}$ (*tam*$^{exo-}$/+) and (**E**, **F**) *mito-*

APOBEC1 mutator flies. (**G**) An expanded view of all 96 possible trinucleotide contexts, centered on the mutated pyrimidine.

As a further estimate of the pathogenicity of the mutations that each model introduces at protein coding genes, we calculated the MutPred pathogenicity score (see Methods) for every mutated site recovered in each of our mutator genotypes at 10 days old. All three mutators show a similar MutPred score distribution when considering sites at <1% heteroplasmy, with either of *tam*$_{exo-}$ mutators having a substantial abundance of synonymous (0-scored) sites (Figure 3.13C). However, higher heteroplasmy sites (1-70%) in *mito-APOBEC1* show a strong preponderance for elevated MutPred scores (Figure 3.13D). As for <1% heteroplasmy sites, a substantial proportion of the 1-70% heteroplasmy positions in *tam*$_{exo-}$/+ are synonymous and therefore non-pathogenic (Figure 3.13D).

Taken together, these analyses indicate that the high selectivity of *mito-APOBEC1* for C:G>T:A transitions generates a remarkably high proportion of non-synonymous and potentially pathogenic mutations which, despite achieving only moderate heteroplasmy, confer a strongly detrimental effect on mitochondria. The different pathogenicity of the mutations introduced by *mito-APOBEC1* or *tamas* mutators offers a plausible explanation for the striking differences on organismal lifespan and vitality between the two models.

3.4 ***Discussion***

The accumulation of mtDNA mutations has been implicated in various age-related pathologies and in the aging process itself (*12, 204*). The *POLG*-based mutator systems, which display a striking progeroid syndrome in mice (*106, 158*), have emerged as the predominant model for obtaining high levels of stochastic mtDNA mutations. However, the mutational heterogeneity that arises in this model, which includes point mutations as well as small indels, large deletions, and mtDNA depletion, has led to considerable debate about the pathogenic entity and whether

point mutations are actually driving aging (*15, 80, 176, 205*). Recently described fly and worm *POLG* mtDNA mutators, which offer an attractive system for their powerful genetic approaches, also show evidence of this mutational heterogeneity (*149, 155, 182*). However, while the first fly model equivalent to the *POLG*-mutator was homozygous lethal (*155*), a similar model recently reported is not lethal (*149*), raising uncertainty about the cause of the lethality. To establish a 'cleaner' mutator system we have developed an alternative mutator model, *mito-APOBEC1*, which we showed is a potent mtDNA mutagen that exclusively induces high levels of C:G>T:A transition mutations, in the absence of indels or mtDNA depletion.

Because we believe the level of heteroplasmy will contribute to the pathogenicity of a mutation profile, the mutation loads calculated here reflect the overall amount of mutations we recovered, including multiple events at a same site. This clonality is generally assumed to be due to clonal propagation of an initial mutational event via replication. However, it is also possible that the same mutational event can independently occur in different molecules, in an iterative mechanism that is independent of mtDNA replication. This is particularly relevant for our *mito-APOBEC1* model which has a very high degree of sequence selectivity and is not directly linked to replication unlike a polymerase-based (*tam*$_{exo-}$) model.

Comparing the POLG/*tamas* and *mito-APOBEC1* model systems, we find that although a similarly high mtDNA mutation load can arise in both *mito-APOBEC1* and maternally inherited *tam*$_{exo-}$ flies, they cause dramatically different effects on organismal fitness. While mutations generated by *mito-APOBEC1* lead to profound shortening of lifespan and loss of vitality (Figure 3.7E, F), associated with a broad spectrum of mitochondrial impairments (Figure 3.9B-D), *tam*$_{exo-}$ does not provoke any major defects either at the organelle or organismal level (Figure 3.7A-D, and Figure 3.9A).

The in-depth mutation analysis method we used allowed us to gain a detailed view of the mutation profiles in these models. First, there is a striking difference (>10-fold) in the overall mutation load when *tam*$_{exo-}$ is inherited paternally versus maternally (Figure 3.2A). This is accounted for by a number of clonally expanded mutations (Figure 3.10A, B, D, E) which likely arise through replication of germline transmitted mutations or very early mutagenesis by maternally deposited Tam$_{exo-}$ protein, which is abundant enough to allow *tam* mutants to development to the third instar stage (*216*). The increase in *de novo* but not clonally mutated sites between 2- and 10-day-old *tam*$_{exo-}$ flies (Figure 3.11B) further supports that mtDNA continues to undergo replication in adults (*62*), but that clonal expansion events likely occur only in development, probably during the peak of mtDNA replication part-way through embryogenesis (*213*). Interestingly, *mito-APOBEC1*, which is only expressed in the zygote and not maternally transmitted, also produces a substantial amount of highly heteroplasmic sites (Figure 3.10C), in stark contrast to paternally-inherited (zygotically expressed) *tam*$_{exo-}$ (Figure 3.10A). *mito-APOBEC1*-induced mutations could expand to high heteroplasmy through replication but also via iterative targeting of cytidines on different mtDNA molecules. Further work will be needed to determine the mechanistic basis for this phenomenon, but we hypothesize that, since APOBEC1 is known to bind single-stranded DNA (*208*), mutations may occur during transcription in addition to during replication, as it has previously been suggested (*215*). Remarkably, the mutation load, number of mutated sites and degree of heteroplasmy did not change in the *mito-APOBEC1* flies between 2 and 10 days (Figure 3.11). We hypothesize that *mito-APOBEC1* is such an efficient mutator that the mutable sites compatible with life are saturated early in development. Further insight into these mechanisms could be gained from analyzing the changes in mutation profile

during developmental stages in these models. Similarly, it will also be informative to explore which, if any, additional changes occur during a more substantial period of aging.

Eliminating differences in heteroplasmy levels as a mechanistic explanation for the differing organismal phenotypes between *tam*$_{exo-}$ and *mito-APOBEC1*, we focused our analyses on the mutation profile. While both models show a predisposition to create C:G>T:A mutations (Figure 3.2D, Figure 3.5A), *mito-APOBEC1* leads to a higher proportion of potentially damaging mutations and non-synonymous substitutions (Figure 3.13). The increased pathogenicity of the *mito-APOBEC1* spectrum is likely caused by its exclusive targeting of C:G nucleotides (Figure 3.5A), particularly in a TCn context (

Figure 3.14), while *tam*$_{exo-}$ has a more variable spectrum of substitutions, including transversions (Figure 3.5A), and a less stringent neighboring sequence requirement (

Figure 3.14). Notably, *Drosophila* mtDNA has evolved to have an unusually low C:G content (Figure 3.12B), likely retaining them predominantly at critical sites (Figure 3.12C); thus, C:G mutations are presumably more likely to be damaging (*169, 217*). Consequently, we surmise that mtDNA mutation type (quality) rather than quantity is a critical factor in impacting organismal fitness. One limitation of our work is that we have restricted our sequence analyses predominantly to protein-coding sequence. Future work will determine whether the context specificity of *mito-APOBEC1* also leads to an enrichment of pathogenic mutations within tRNA and rRNA sequences, and how these may impact organismal fitness and aging.

The fact that *mito-APOBEC1* flies show such detrimental phenotypes but only modest levels of heteroplasmy is also intriguing. Typically, for mitochondrial diseases resulting from mtDNA mutations, heteroplasmy levels must rise above a certain threshold (>60-90%), to cause mitochondrial functional impairment (*17, 92*). This phenomenon is also replicated in a number of *Drosophila* strains with homoplasmic mtDNA mutations that present strong mitochondrial and neuromuscular defects (*218-220*). However, a threshold level can be reached not only by one specific mutation but by the sum of a number of different mutations (*15*). Our results from *mito-*

APOBEC1 flies support this view. Mutations can produce cumulative defects on the same protein or multi-protein complexes that limit the possibility of complementation. Alternatively, mutations can arise locally to very high heteroplasmy and create a mosaic of affected cells within a tissue. Current sequencing technologies, however, cannot distinguish between these scenarios and would appear as apparent lower level heteroplasmies on aggregate. Such a death-by-a-thousand-cuts model could help explaining how the variety of mtDNA mutations frequently observed in aging and in a number of neurodegenerative diseases could have a phenotypic effect without ever reaching the high level of heteroplasmy typically associated with mitochondrial diseases (*61*, *74*, *221*, *222*).

Finally, we posit that the *mito-APOBEC1* model circumvents several limitations of the existing *POLG*-based mutator models and offers several advantages. First, it allows us to specifically address the impact of point mutations, in the absence of indels or mtDNA depletion. Second, the stronger phenotypes induced in this model provide an opportunity to screen for genetic modifiers to identify factors influencing the pathogenicity of mtDNA mutations. Third, as *mito-APOBEC1* is based on the classic GAL4/UAS transgenic approach this opens it up to the full power and versatility of *Drosophila* genetics, providing an unparalleled toolbox for a range of spatial and temporal manipulation, such as tissue-specific or life course-specific expression. Beyond *Drosophila*, this study acts as a proof-of-principle to extend the inducible *mito-APOBEC1* model system to other organisms where it would provide analogous advantages over existing mutator models to further decipher the pathophysiology of mtDNA mutations.

3.5 ***Methods***

Drosophila stocks and husbandry: Flies were raised under standard conditions in a humidified, temperature-controlled incubator with a 12h:12h light:dark cycle at 25°C, on food consisting of

agar, cornmeal, molasses, propionic acid and yeast. Transgene expression was induced using the ubiquitous *da-GAL4* driver. The following strains were obtained from the Bloomington *Drosophila* Stock Center (Bloomington Drosophila Stock Center, RRID:SCR_006457): w^{1118} (RRID:BDSC_6326), *da-GAL4* (RRID:BDSC_55850), *UAS-mito-HA-GFP* (RRID:BDSC_8443). w^{Dah} (*w-* [Dahomey, Wolbachia-free]), tam^{exo-} and tam^{rescue} flies were a gift from N-G. Larsson (Max Planck Institute for Ageing, Cologne). Unless otherwise stated, all experiments were conducted using male flies. The genotypes used in our study are listed in Table 3.6.

Generation of APOBEC1 transgenic lines: *mito-APOBEC1* fusion constructs were made by fusing the mitochondrial targeting sequence (MTS) of human ATP synthase subunit F1β (ATP5F1B), a haemagglutinin (HA) epitope and a flexible linker sequence 5' to the rat APOBEC1 cDNA which lacked the initial ATG and stop codon. A nuclear exporting sequence (NES) was inserted 3' to APOBEC1. The Glu63Ala ($APOBEC1^{E63A}$) catalytically dead mutant was created by changing the codon GAA to GCT at amino acid position 63 of the cDNA by site-directed mutagenesis. Mitochondria-excluded, cytoplasmic versions of APOBEC1 constructs (*cyto-APOBEC1*) were generated by removing the MTS sequence from the *mito-APOBEC1* transgenes. Engineered sequences were cloned into the *pUASTattB* vector and injected in *Drosophila* embryos for phiC31-mediated transgenesis at the attP40 site (BestGene, RRID:SCR_012605).

Table 3.6. Genotypes used in this study

Group	Label	Genotype
paternal tamas	control	*w*1118*; +/+*
	+/*tam*rescue	*w*1118*; +/Tl{Tl}tam*Rescue
	+/*tam*exo-	*w*1118*; +/ Tl{Tl}tam*D263A
maternal tamas	control	*w*Dah*; +/+*
	*tam*rescue /+	*w*Dah*; Tl{Tl}tam*Rescue*/+*
	*tam*exo-/+	*w*Dah*; Tl{Tl}tam*D263A*/+*
mito-APOBEC1	control	*w*1118*; P{UAS-mito-HA-GFP}/+; P{da-GAL4}/+*
	*mito-APOBEC1*E63A	*w*1118*; P{UAS-mito-HA-APOBEC1*E63A*}/+; P{da-GAL4}/+*
	mito-APOBEC1	*w*1118*; P{UAS-mito-HA-APOBEC1}/+; P{da-GAL4}/+*

DNA isolation for sequencing: Brains from individual male flies of the indicated age were dissected in PBS, flash frozen in dry ice, and stored at -80 °C. Total DNA was isolated from individual brains using the QIAamp DNA Micro isolation kit following the manufacturer's instructions (QIAGEN, RRID:SCR_008539). DNA yield for a single brain typically ranged between 20-30 ng.

Duplex Sequencing and mutation calling: DNA was prepared for Duplex Sequencing using a previously described protocol (*142*) with several modifications. Briefly, ~20 ng of total DNA was sonicated in 60 μL of nuclease-free ddH_2O using a Covaris AFA system with a duty cycle of 10%, intensity of 5, cycles/burst 100, time 20 seconds x 5, temperature of 4 °C. After sonication, each sample was subjected to end-repair and 3'-dA-tailing using the NEBNext Ultra End-repair/dA-tailing kit (New England Biolabs, RRID:SCR_013517) according to the manufacturer's instructions. Each sample was then ligated with 2 μL of 15 μM Duplex Sequencing adapters, using the NEBNext Ultra Ligation kit (New England Biolabs, RRID:SCR_013517) according to the manufacturer's instructions. Duplex Sequencing adapters used in collecting data for analysing mutation selection were chemically synthesized as a collaborative effort with Integrated DNA Technologies to develop a prototype synthesis method. Each sample was then cleaned of excess adapters using AgenCourt AmpureXP magnetic beads, and PCR amplified (*142*). The mitochondrial DNA was isolated by targeted DNA capture using IDT xGen Lockdown probes (Integrated DNA Technologies OligoAnalyzer, RRID:SCR_001363) specific for non-repeat regions of the *Drosophila* mitochondrial genome, using the manufacturer's instructions.

The captured DNA samples were sequenced on an Illumina NextSeq500 using 150bp paired-end sequencing. The resulting reads were aligned against the *Drosophila* genome (BDGP

Release 6 + ISO1 MT/dm6) using the Burrows-Wheeler Aligner (BWA, RRID:SCR_010910) and Samtools (*196*) (SAMTOOLS, RRID:SCR_002105) coupled with a custom software workflow (*142*). Reads not uniquely mapping to the mitochondrial genome were excluded from further analysis. The non-coding A+T region [ChrM:14917-19524] has a repetitive high A:T content (~95%) which confounds sequence capture and mapping. The majority of A+T region is therefore not included in the calculation of mutation load. After processing, we called *de novo* mutations by using a cut-off that excluded variants occurring at >1% heteroplasmy, while *global* mutation loads included all mutations with heteroplasmy <70% at each position.

mtDNA mutation analysis: For each animal, point mutation or indel loads were calculated as the total number of mutations, insertions or deletions per total number of nucleotides sequenced, considering the whole or a subset of the genome, as relevant to each analysis. For heteroplasmy distribution analysis, results from individual flies per genotype were aggregated.

The trinucleotide context of mutations was calculated using previously published pipelines on the aggregated sequence data from flies of the same genotype (*149, 199*). Each mutation was characterized according to both the identity of the mutation as well as the nucleotides 3' and 5' of the mutation site, and binned into one of the 96 total possible mutation types. To calculate the load of mutations in each trinucleotide context, we first determined the total genome-wide post-processing duplex depth for each of the 16 total trinucleotide contexts. We then calculated the trinucleotide mutation load for each of the 96 mutation types by dividing the total number of mutations for each mutation type by its post-processing duplex depth.

Calculations for the distribution of mutation positions, trinucleotide context and mutation loads throughout the genome were calculated using Microsoft Excel for Mac v.16 (Microsoft

Excel, RRID:SCR_016137) and scripts developed in Python 2.7 or Python 3.6 (Python Programming Language, RRID:SCR_008394). Graphs were generated using GraphPad Prism software (GraphPad Prism, RRID:SCR_002798) or R software (R Project for Statistical Computing, RRID:SCR_001905).

MutPred software (MutPred, RRID:SCR_010778)(*171*) was used to calculate pathogenicity scores for all non-synonymous (NS) variants detected in mutator flies. In addition, we assigned a score of 0 to all non-synonymous changes, and a score of 1 to nonsense mutations or mutations changing a stop codon to coding. MutPred scores for all detected single nucleotide changes were aggregated by genotype and plotted as a frequency distribution.

mtDNA copy number: Total DNA was extracted from 10-20 male flies using the DNeasy Blood and Tissue kit (Qiagen) and following manufacturer's instructions. Quantification of mtDNA was performed in triplicate by multiplex TaqMan qPCR amplification of the mitochondrial genes mt:CoI and mt:lrRNA, and the nuclear gene RpL32 as normalizing reference. The following primers were used: (mt:CoI) 5'-TTCTACCTCCTGCTCTTTCTTTAC and 5'-CAGCGGATAGAGGTGGATAAAC, probe 5'-FAM-AATGGAGCTGGGACAGGATGAACT-TAMRA; (mt:lrRNA) 5'-AGATAGAAACCAACCTGGCTTAC and 5'-TTGGGTGTAGCCGTTCAAAT, probe 5'-FAM-ACCGGTTTGAACTCAGATCATGTAAGA-TAMRA; (RpL32) 5'-CACCGGAAACTCAATGGATACT and 5'-CACACAAGGTGTCCCACTAAT, probe 5'-FAM-CCAAGAAGCTAGCCCAACCTGGTT-TAMRA. PCR reactions were performed according to standard conditions for TaqMan (Applied Biosystems, RRID:SCR_005039): 50°C for 2 min; 95°C for 10min; 40 cycles at 95°C 15 sec, 60°C 1 min. The expression of mtDNA copy number

relative to nuclear DNA was determined using the $2^{-\Delta\Delta CT}$ method. The relative quantification was corrected for PCR efficiency of each primer pair.

Southern Blot: Total DNA was extracted from 150-180 7-day old flies using phenol:chloroform following the protocol in the VDRC website (Vienna Drosophila Stock Center, RRID:SCR_013805). 5 ug of DNA was digested with *Pst*I (New England Biolabs, RRID:SCR_013517) according to manufacturer's instructions, and resolved on a 0.55% agarose gel. Blotting, labelling of probe DNA and membrane hybridization were performed using standard procedures (*223*). Membranes were exposed to a storage phosphor screen for 24–96 h before imaging with an Amersham Typhoon RGB scanner and processing with ImageQuant software (ImageQuant, RRID:SCR_014246). Probe primer sequences were as follows: (mt:CoII - 3,276-3,840) 5'-AACTATTTTACCAGCAATTATTTTACT and 5'-CAGTCATCTAATGAAGAGTTATTTCTA; (18S rDNA) 5'-CGATGCCAGCTAGCAATTGGGTGTAG and 5'-CTACACCCAATTGCTAGCTGG.

Long-range PCR: Total DNA was extracted from 10-20 male flies using the DNeasy Blood and Tissue kit (Qiagen) and following manufacturer's instructions. PCR amplification was performed using PrimeSTAR GXL DNA Polymerase under manufacturer's conditions (Clontech, RRID:SCR_004423), with 1 μM of primers and the following amplification conditions: 94^{o}C, 1 min; 98^{o}C, 30 sec; 68^{o}C, 13 min (30 cycles); 72^{o}C, 10 min. Primers used (14.2F: 5'-GCCGCTCCTTTCCATTTTTGATTTCC and 14.2R: 5'-TGCCAGCAGTCGCGGTTATACCA) amplify a product encompassing almost the complete mtDNA molecule. The PCR products were then visualized after electrophoresis on 0.8% agarose and 2X Invitrogen SYBR Safe DNA Gel

Stain (Thermo Fisher Scientific, RRID:SCR_008452) and 1 kb DNA Ladder (New England Biolabs, RRID:SCR_013517).

qRT-PCR of mtDNA transcripts: Total RNA was extracted from ten 7-day-old male flies using the Direct-zol MiniPrep kit (Zymo Research, RRID:SCR_008968). Genomic DNA contamination was removed from RNA using Turbo DNAse (Ambion Inc, RRID:SCR_008406), according to manufacturer's protocol. cDNAs were generated from 80 ng of total RNA using the Maxima H Minus cDNA Synthesis Master Mix with dsDNAse (Thermo Fisher Scientific, RRID:SCR_008452), following manufacturer's instructions. To ensure that RNA had no genomic DNA contamination, a control reaction was included in which no reverse transcription was carried out. Reactions were carried out using a QuantStudio 3 Real-Time PCR Systems (Thermo Fisher Scientific, RRID:SCR_008452) with Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific, RRID:SCR_008452). Primers were as follows: (mt:CoI) 5'-CAGGATGAACTGTTTATCCACCTTT and 5'-AATCCCTGCTAAATGTAGAGAAAAAATAG; (mt:ND3) 5'-AAAAAGCTTTAATCGACCGAGA and 5'-CGTAAAGAAAATGGTAATCGAGATG; (mt:CytB) 5'-AAATTTATTGGGAGACCCTGATAAC and 5'-GGAATAGATCGTAAAATAGCATAAGCA; (mt:ND4) 5'-AACCCAGAAGAACATAAACCA and 5'-TGCTTATTCATCTGTTGCTCA; (mt:lrRNA) 5'-ACCTGGCTTACACCGGTTT and 5'-GGGTGTAGCCGTTCAAATTT; (RpL32) 5'-AAACGCGGTTCTGCATGAG and 5'-GCCGCTTCAAGGGACAGTATCTG; (Tub84b) 5'-TGGGCCCGTCTGGACCACAA and 5'- TCGCCGTCACCGGAGTCCAT. Primers' specificity was assessed by melting curve profile and their efficiency ranged from 0.92 to 0.99. Mitochondrial

transcripts levels were normalized to a geometric mean of both Rpl32 and Tub84b reference genes, using the comparative Ct method. The relative quantification was corrected for PCR efficiency of each primer's couple.

Locomotor and lifespan assays: The startle induced negative geotaxis (climbing) assay was performed using a counter-current apparatus. Briefly, 20-23 males were placed into the first chamber, tapped to the bottom, and given 10s to climb a 10cm distance. This procedure was repeated five times (five chambers), and the number of flies that has remained into each chamber counted. The weighted performance of several group of flies for each genotype was normalized to the maximum possible score and expressed as *Climbing index* (*48*).

For lifespan experiments, flies were grown under identical conditions at low-density. Progeny were collected under very light anesthesia and kept in tubes of approximately 25 males each. Flies were transferred every 2-3 days to fresh medium and the number of dead flies recorded. Percent survival was calculated at the end of the experiment after correcting for any accidental loss.

Immunofluorescence experiments: For APOBEC1 immunostaining, larval epidermis was dissected in PBS and fixed in 4% formaldehyde for 30 min at RT, followed by permeabilization in 0.3% Triton X-100 for 30 min and blocking with 0.3% Triton X-100, 1% BSA in PBS for 1 h at RT. Tissues were incubated with anti- HA and anti-ATP5A antibodies diluted in blocking buffer overnight at 4$_{o}$C, and with the appropriate fluorescent secondary antibodies for 2 h at RT. Samples were washed several times in PBS and mounted on slides using Prolong Diamond Antifade mounting medium (Thermo Fisher Scientific, RRID:SCR_008452).

Microscopy: Confocal imaging was conducted using a Zeiss LSM 880 microscope (Carl Zeiss MicroImaging) equipped with Nikon Plan-Apochromat 63x/1.4 NA oil immersion objective. Images were prepared using Fiji software (Fiji, RRID:SCR_002285).

Respirometry: Mitochondrial respiration was assayed at 30°C by high-resolution respirometry using a Oxygraph-2k high-resolution respirometer (OROBOROS Instruments) using a chamber volume set to 2 mL. Calibration with air-saturated medium was performed daily. Data acquisition and analysis were carried out using Datlab software (OROBOROS Instruments). Five flies per genotype were homogenized in Respiration Buffer [120 mM sucrose, 50 mM KCl, 20 mM Tris-HCl, 4 mM KH_2PO_4, 2 mM $MgCl_2$, and 1 mM EGTA, 1 g L^{-1} fatty acid-free BSA, pH 7.2]. For coupled (state 3) assays, complex I-linked respiration was measured at saturating concentrations of malate (2 mM), glutamate (10 mM) and adenosine diphosphate (ADP, 2.5 mM). Complex II-linked respiration was assayed in Respiration Buffer supplemented with 0.15 µM rotenone, 10 mM succinate and 2.5 mM ADP. The addition of proline to the respiration buffer can increase respiration rate in insect samples but was not included here. Respiration was expressed as oxygen consumed per fly. Flies' weight was equal in all genotypes tested.

Biochemical assays: Mito-APOBEC1 flies and controls were aged to 10-12 days before collection for mitochondria isolation. 1 mL of flies per genotype were homogenized in 2 mL of STE buffer [250 mM Sucrose, 5 mM Tris, 2 mM EGTA, pH 7.4] + 1% BSA with several strokes in a Teflon-glass homogenizer at 700 rpm. Nuclei and fly body debris were pelleted by two centrifugation steps at 1,000 *g* for 5 min, 4°C and the supernatant centrifuged for 10 min at 3,000 *g*, 4 °C. After a wash in 5 mL STE buffer + 1% BSA, the mitochondrial pellet was resuspended in STE buffer

(without BSA) and centrifuged at 7,000 *g*, 4 °C for 10 min. Mitochondria pellet was resuspended in 1 mL of STE buffer for protein quantification using the Pierce BCA method (Thermo Fisher Scientific, RRID:SCR_008452). Sample preparation and Blue-Native PAGE were performed as follows: 800 μg of pelleted mitochondria were resuspended in 1.5 M aminocaproic acid, 50 mM Bis-Tris·HCl pH 7 to a final concentration of 10 mg mL^{-1} and solubilized with 4 mg digitonin per mg of protein. After 5 min incubation on ice, mitochondria were centrifuged at maximum speed (20,000 *g*) at 4°C for 30 min. Supernatant was mixed with 10 μL of Blue-Native Sample Buffer [750 mM aminocaproic acid, 50 mM Bis-Tris·HCl pH 7, 0.5 mM EDTA, 5% Serva Blue G] and 10 μL of extracts loaded for Blue-Native Gel Electrophoresis on a pre-cast NativePAGE 3-12% Bis-Tris gel (Life Technologies, RRID:SCR_008817). For in-gel complex activity, the BN-PAGE gel was incubated for several hours at room temperature in complex-specific solutions. For complex-I activity: 5 mM Tris–HCl, pH 7.4, 0.1 mg mL^{-1} NADH, and 2.5 mg mL^{-1} NTB (NitroTetrazolium Blue); for complex-II activity: 5 mM Tris–HCl, pH 7.4, 0.2 mM phenazine methasulfate, 20 mM sodium succinate and 2.5 mg mL^{-1} NTB; for complex-IV: 50 mM sodium phosphate buffer, pH 7.4, 0.5 mg mL^{-1} DAB (3,3'-Diaminobenzidine tetrahydrochloride), 24 U mL^{-1} catalase, 10 mg mL^{-1} cytochrome c (horse heart <95% purity), 75 mg mL^{-1} sucrose.

Subcellular fractionation: 200-250 adult flies (both males and females) were homogenized in 4 mL of cold 250-STM buffer [250 mM sucrose, 50 mM Tris·HCl pH 7.4, 5 mM $MgCl_2$] freshly supplied with 1 mM DTT, 1 mM PMSF, 25 μg mL^{-1} Spermine and 25 μg mL^{-1} Spermidine. The homogenate was then sieved through a 70 μm cell strainer (Corning BV) to remove gross body parts (legs and wings). An aliquot of the decanted homogenate was collected as total fraction, and the rest centrifuged at 800 *g* for 15 min at 4°C using a swing-out rotor. Supernatant was stored on

ice for further isolation of mitochondria and cytosolic fraction. Nuclei-containing pellet was washed with lysis buffer and nuclei isolated at the bottom of a 2M-STMDPS [2 M sucrose, 50 mM Tris·HCl pH 7.4, 5 mM $MgCl_2$, 1 mM DTT, 1 mM PMSF, 25 μg mL^{-1} Spermine and 25 μg mL^{-1} Spermidine] cushion after ultracentrifugation at 80,000 *g* for 30 min at 4°C (Beckman Coulter, RRID:SCR_008940 ultracentrifuge). The nuclear pellet was lysed with 50 μL of NE buffer [20 mM HEPES pH 7.9, 0.5 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 20% glycerol] for at least 30min at 4°C and nuclear fraction stored at -80°C. Mitochondria were separated from cytosol by centrifugation of the nuclei-free homogenate at 6,000 *g* for 15 min using a swing-out rotor. After a quick wash of the pellet, mitochondria proteins were extracted in 70 μL of HDP buffer [10 mM HEPES pH 7.9, 1 mM DTT, 1 mM PMSF] and stored at -80°C. Supernatant from the mitochondrial centrifugation was processed to obtain a clean cytosolic fraction by ultracentrifugation at 100,000 *g* for 1 h at 4°C (Beckman Coulter, RRID:SCR_008940 ultracentrifuge) (pellet discarded) and stored at -80°C. Samples from each fraction were quantified using Pierce BCA assay (Thermo Fisher Scientific, RRID:SCR_008452) and 15 μg of total protein lysate were loaded on a 4-20% gel (Bio-Rad Laboratories, RRID:SCR_008426), resolved via SDS-PAGE (Bio-Rad Laboratories, RRID:SCR_008426) and blotted to a nitrocellulose membrane. Histone-H3 immuno-reactivity was used to mark the nuclear fraction, ATP5A was used as a mitochondrial marker and GAPDH as a cytoplasmic marker.

Immunoblotting: For APOBEC1 expression analysis, protein samples were isolated from whole adult flies. Flies were homogenized in RIPA lysis buffer [50 mM Tris·HCl, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.5% SDS] with 1 mM PMSF and protease inhibitor mixture (Roche, RRID:SCR_001326). Protein extracts were quantified using the Pierce BCA assay (Thermo Fisher

Scientific, RRID:SCR_008452). Typically, 30 µg of protein were resolved by SDS-PAGE (Bio-Rad Laboratories, RRID:SCR_008426) and transferred onto nitrocellulose membrane using a semi-dry Transblot apparatus (Bio-Rad Laboratories, RRID:SCR_008426) according to the manufacturer's instructions. Membranes were blocked with 5% skimmed milk in TBST [Tris-buffered saline, 0.1% Tween 20] for 1 h at RT and incubated with primary antibody overnight at 4°C. After several washes in TBST, appropriate horseradish peroxidase-conjugated secondary antibodies were incubated for 1 h at RT. Detection was achieved with ECL-Plus detection kit (GE Healthcare, RRID:SCR_000004) using a ChemiDoc XRS+ molecular imager (Bio-Rad Laboratories, RRID:SCR_008426) and analyzed by Image Lab Software (Image Lab Software, RRID:SCR_014210).

For immunoblotting of native OXPHOS complexes, Blue Native gels of mitochondria-enriched samples were transferred onto PVDF membranes via wet transfer (Bio-Rad Laboratories, RRID:SCR_008426) following manufacturer's instructions. Membranes were blocked and incubated with primary and secondary antibodies as described above. Detection was achieved with ECL-Plus detection kit (GE Healthcare, RRID:SCR_000004) using an Amersham Imager 600 imaging device.

Antibodies and dyes: For immunofluorescence experiments, the following primary antibodies were used: mouse anti-ATP5A (Abcam Cat# ab14748, RRID:AB_301447; 1:500), rabbit anti-HA (Abcam Cat# ab9110, RRID:AB_307019; 1:500). Secondary antibodies were: anti-rabbit AF647 (Thermo Fisher Scientific Cat# A-21244, RRID:AB_2535812; 1:200), anti-mouse AF488 (Thermo Fisher Scientific Cat# A-11001, RRID:AB_2534069; 1:200).

For immunoblot experiments, the following antibodies were used: mouse anti-Actin (Millipore Cat# MAB1501, RRID:AB_2223041; 1:5,000), mouse anti-APOBEC1 (E-2) (Santa Cruz Biotechnology Cat# sc-166508, RRID:AB_2057252; 1:1,000), mouse anti-ATP5A (Abcam Cat# ab14748, RRID:AB_301447; 1:20,000), mouse anti-GAPDH (GeneTex Cat# GTX627408, RRID:AB_11174761; 1:2,000), rabbit anti-HA (Abcam Cat# ab9110, RRID:AB_307019; 1:2,000), rabbit anti-HistoneH3 (Abcam Cat# ab1791, RRID:AB_302613; 1:1,000), mouse anti-MTCO3 (Abcam Cat# ab110259, RRID:AB_10859925; 1:600), mouse anti-NDUFS3 (Abcam Cat# ab14711, RRID:AB_301429; 1:600), rabbit anti-Porin (Millipore Cat# PC548, RRID:AB_2257155; 1:5,000), mouse anti-Tubulin (Sigma-Aldrich Cat# T6793, RRID:AB_477585, 1:5,000), mouse anti-UQCRFS1 (Abcam Cat# ab14746, RRID:AB_301445; 1:600). Horseradish peroxidase-conjugated secondary antibodies: anti-mouse (Abcam Cat# ab6789, RRID:AB_955439; 1:5,000 to 1:20,000), anti-rabbit (Thermo Fisher Scientific Cat# G-21234, RRID:AB_2536530; 1:3,000 to 1:5,000).

Statistical analysis: Data are reported as mean ± SD, mean ± SEM or mean ± 95% CI as indicated in figure legends. For statistical analyses of lifespan experiments, a long-rank (Mantel-Cox) test was used. For behavioral analyses, a Kruskal-Wallis nonparametric test with Dunn's post-hoc correction for multiple comparisons was used. Significance in mtDNA copy number by qPCR and mitochondrial transcripts analyses was determined by a two-way ANOVA test with Tukey's post-hoc correction for multiple comparisons. Significance in Southern blot quantification and Oroboros (respirometry) experiments was determined by a one-way ANOVA test with Sidak's post-hoc correction for multiple comparisons. Significance levels are indicated in the figure legends. Unless specifically indicated, no significant difference was found between a sample and

any other sample in the analysis. Statistical analyses were performed using GraphPad Prism 7 software (GraphPad Prism, RRID:SCR_002798).

3.6 *Notes and Acknowledgements*

Acknowledgements:

This work is supported by MRC core funding (MC_UU_00015/4, MC-A070-5PSB0 and MC_UU_00015/6) and ERC Starting grant (DYNAMITO; 309742). S.A. was supported by an MRC Career Development Fellowship. A.G.D. receives support from NIHR Biomedical Research Centre pilot studies (RROI.GAAB). Work was further supported by DOD/CDMRP grant W81XWH-16-1-0579 to SRK, National Institute of Neurological Disorders and Stroke (R21NS090073) to LJP, and the Genetic Approaches to Aging Training Grant (T32AG000057) to CLS. PFC is a Wellcome Trust Principal Research Fellow (212219/Z/18/Z), and a UK NIHR Senior Investigator, who receives support from the Medical Research Council Mitochondrial Biology Unit (MC_UP_1501/2), the Medical Research Council (UK) Centre for Translational Muscle Disease (G0601943), the Evelyn Trust, and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Stocks were obtained from the Bloomington *Drosophila* Stock Center which is supported by grant NIH P40OD018537. We thank the Whitworth lab for discussions and for critical reading of the manuscript, and Aurelio Reyes for help with the mitochondrial transcription and replication analysis.

Author Contributions

A.J.W. conceived the study and supervised the work with L.J.P and S.R.K. S.A., A.S-M., E.F-V., A.G-D., J.J.L., R.T., M.J.H., E.K.S. and P.A.G. designed and performed experiments. C.S. and S.A. performed computational analyses on mtDNA mutations. A.G-D. and P.F.C. assisted with mtDNA data interpretation. T.J.N. and M.M. conceived the use of mito-APOBEC1. A.J.W. and S.A. wrote the manuscript with input from all authors.

Chapter 4. **Does Parkin select against mtDNA mutations *in vivo*?**

Throughout my thesis work, I have been intensely interested in understanding whether mitophagy selects against mtDNA mutations. In **Chapter 2**, I show my initial attempts at finding evidence of selection within the PolG$_{mut}$ fly, a transgenic *Drosophila* strain expressing a proofreading defective form of the mitochondrial DNA polymerase. I discovered that this strain has an increased mtDNA mutation frequency and a number of accompanying behavioral and molecular phenotypes. Surprisingly, I found that deleterious mtDNA mutations were overrepresented relative to a neutral model of selection in mutator flies, and my work provided no evidence in support of the model that mitophagy limits the accumulation of mtDNA mutations.

In vitro studies have shown that the mitophagy-promoting factors PINK1 and Parkin selectively target depolarized mitochondria for lysosomal degradation. Despite our earlier findings, it remains a possibility that mitophagy reduces somatic mtDNA mutations, but that the forces responsible for the overrepresentation of deleterious mtDNA mutations masks the effects of negative selection. Work in *C. elegans* demonstrates that large mtDNA deletions can be amplified through compensatory mtDNA replication, but are simultaneously eliminated through the action of *pdr-1*, the worm homolog of Parkin (*133*). Consistent with this work, I hypothesized that mitophagy selects against harmful mtDNA mutations, but that this negative selection is ordinarily overshadowed by the forces responsible for the excess of deleterious mtDNA mutations detected in mutator flies.

Importantly, all of my findings related to mtDNA selection have been made through observation of the spectrum of mtDNA mutations and through statistical analyses. In order to examine the effects at mitophagy more directly, I took a biochemical approach. Expression of our mutator polymerase should increase mitophagy in an effort to rid cells of mutation-bearing

mitochondria, and thus we would anticipate a higher rate of mitochondrial protein turnover in mutator flies. Moreover, Parkin overexpression would be anticipated to further increase the rate of turnover, and genetic perturbations of Parkin would be expected to influence mtDNA mutation frequency and pathogenicity in a predictable fashion. To test these hypotheses, I employed a proteomic technique developed in our lab that allows us compare rates of protein turnover. With this proteomic technique, we can directly compare the rates of mitochondrial protein in mutator flies with wild-type flies, as well as mutator flies overexpressing Parkin. This work is currently unpublished.

4.1 *Mutator flies do not exhibit increased mitophagy*

To test the hypothesis that mutator flies have increased mitochondrial protein turnover, we used a proteomic approach to compare the half-lives of mitochondrial proteins in mutator flies and controls. Briefly, our method involves feeding *Drosophila* a stable heavy isotope of leucine and using mass spectrometry to monitor the rates at which unlabeled proteins are degraded and replaced by labeled proteins. A key feature of our method is Topograph, software that deconvolutes the spectra resulting from overlapping isotope distributions and calculates protein half-lives by estimating the fraction of heavy isotope in the amino acid precursor pool. This proteomic approach is critical to our analysis for two reasons: First, it provides the half-lives of hundreds of proteins, yielding a more detailed picture of protein turnover than afforded either by pulse-chase radiolabeling studies or by monitoring of individual proteins. Second, experiments conducted in our lab have shown that our approach readily distinguishes the rates of turnover in mitochondrial and non-mitochondrial proteins, allowing us to differentiate alterations in mitophagy from alterations in general autophagy.

My hypothesis predicted that the half-lives of mitochondrial proteins would be shorter in mutator flies relative to controls. In contrast to my expectations, we discovered that most mitochondrial proteins displayed slower turnover in mutator flies relative to controls (Figure 4.1A), implying that mitophagy is not upregulated in mutator flies. There was also no evidence that general autophagy was affected in mutator flies, as there was no significant difference in nonmitochondrial protein turnover rates between PolG$_{mut}$ flies and control (Figure 4.1B).

In addition to estimates of protein turnover rates, our proteomics assay measures the abundance of thousands of proteins. Comparing protein abundances, we discovered that a relatively high proportion (28%) of mitochondrial proteins are significantly increased in PolG$_{mut}$ flies relative to control flies (Figure 4.2). This proportion is higher than either the proportion of mitochondrial proteins decreased in abundance in PolG$_{mut}$ flies relative to controls (5%), or the abundance of nonmitochondrial proteins increased in PolG$_{mut}$ flies (14%) (Figure 4.2). These data suggest that PolG$_{mut}$ flies display a selective increase in mitochondrial protein abundance concurrent with decreased mitochondrial protein turnover. Taken together, these data do not support the model that mitophagy is increased in PolG$_{mut}$ flies.

4.2 *Parkin overexpression does not select against mtDNA mutations in mutator flies*

A number of recent studies suggest that genetic overexpression of Parkin can facilitate mitophagic clearance of mtDNA mutations in mice, *C. elegans*, and *Drosophila*, but these interpretations come with many of the caveats outlined in **Chapter 1**. To test whether Parkin influences mtDNA mutations in PolG$_{mut}$ flies, we performed behavioral assays on PolG$_{mut}$ flies bearing homozygous *parkin* loss of function mutations (*park$_{25}$/park$_{25}$*), as well as ubiquitous overexpression of Parkin using the UAS-GAL4 system (*Da-GAL4>UAS-Parkin*). Based on the

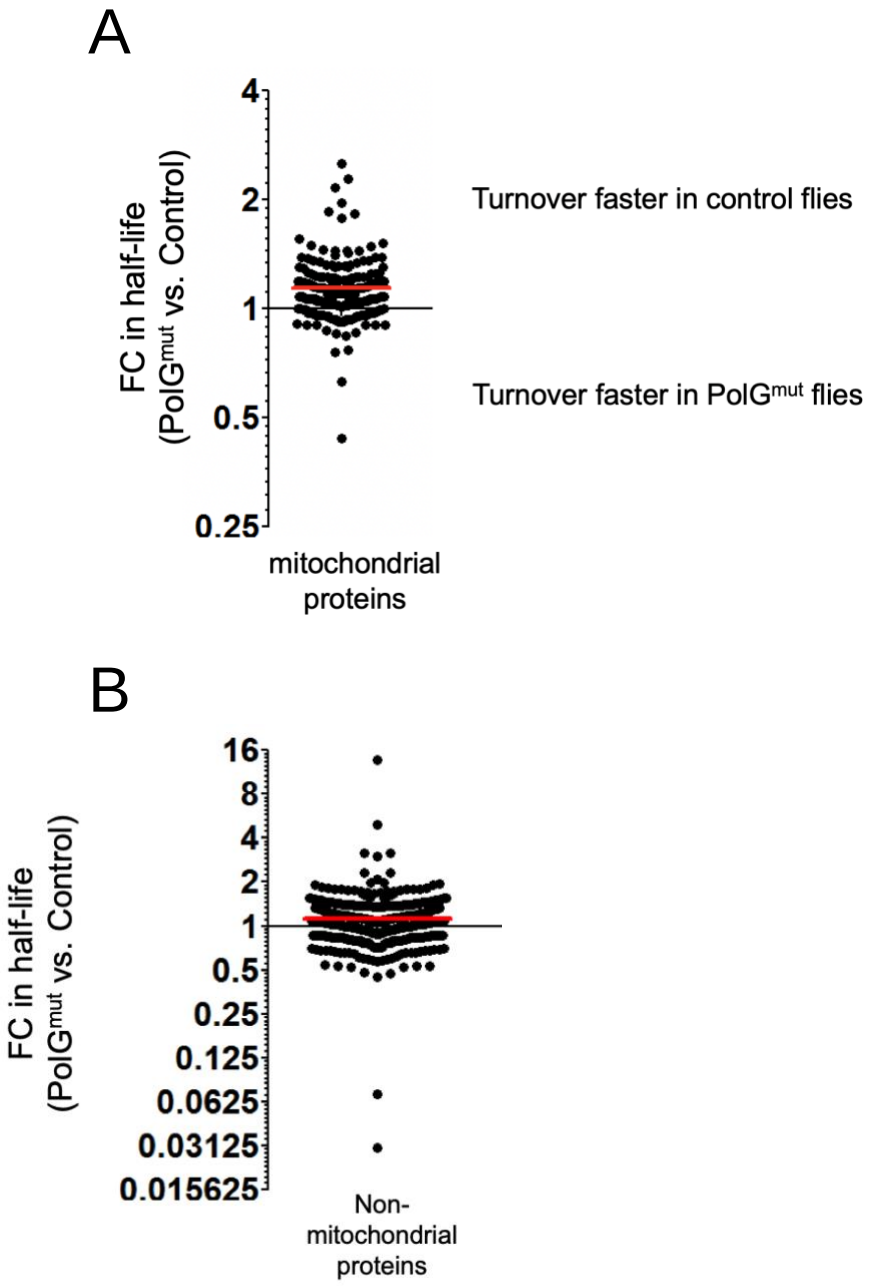


Figure 4.1. Mitochondrial protein turnover is slower in PolG$_{mut}$ flies than controls. Fold change (FC) in half-lives for (A) mitochondrial proteins (n = 228) and (B) non-mitochondrial proteins (n = 998) in PolG$_{mut}$ flies compared to controls. The x-axis represents equal protein half-lives in both genotypes, the red horizontal line represents the group mean.

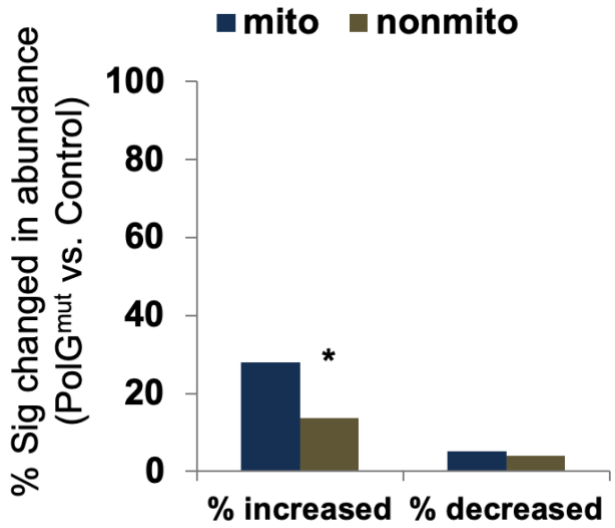


Figure 4.2. Mitochondrial proteins are selectively increased in abundance in PolG$_{mut}$ flies. Proportion of total proteins within the mitochondrial and nonmitochondrial compartments significantly increased or decreased in abundance in PolG$_{mut}$ flies relative to controls at 120 h. *$p < 0.0001$ by χ^2 test

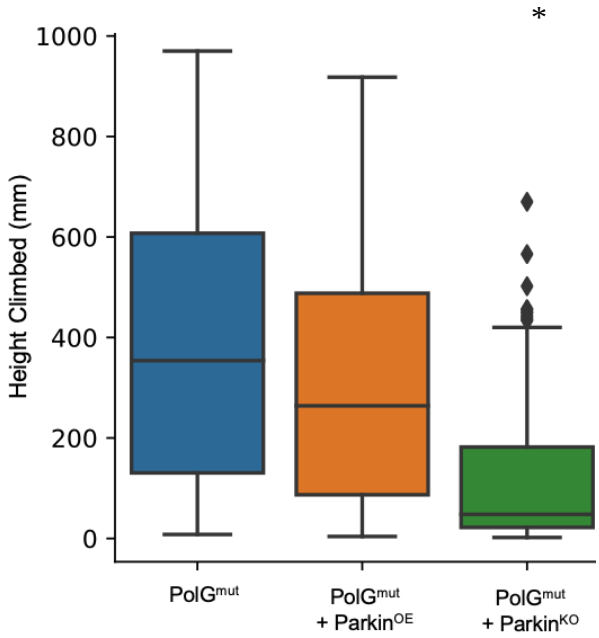
model that Parkin selects against deleterious mutations, we predicted that overexpression of Parkin should rescue mutator phenotypes, and inactivating Parkin should worsen mutator phenotypes.

To test the influence of our genetic perturbations of mitophagy, we performed assays of fly lifespan and climbing, as described in **Chapter 1**. Although Parkin overexpression did not rescue climbing at 15 or 30 days (Figure 4.3), Parkin overexpression slightly extended the lifespan of mutator flies (Figure 4.4). However, it previously been shown that ubiquitous Parkin overexpression extends lifespan in wild-type male flies by approximately 10% (*224*), comparable to the lifespan extension we observe in mutator flies, so this is likely a non-specific effect. Conversely *parkin* mutants have climbing and lifespan defects on their own (*48*). Therefore, we anticipated that introducing a parkin mutation into the mutator background would further worsen the phenotypes of mutator flies, and that was indeed what we observed (Figure 4.3, Figure 4.4).

To directly test whether Parkin influences the frequency of pathogenic mtDNA mutations, we performed Duplex Sequencing on $PolG^{mut}$ flies as described in **Chapter 2**. If the lifespan extension of $Parkin^{OE}$ flies is due to the destruction of mutant bearing mitochondria, then we hypothesized that Parkin-overexpressing flies should have a lower burden of mtDNA mutations. Conversely, knockout of *parkin* should impair mitophagic turnover of mutant mtDNA, resulting in an increased mutation frequency. Duplex sequencing of these flies revealed no significant alterations in the point mutation burden in any genotype, with $Parkin^{OE}$ flies trending towards an increased mutation frequency relative to expression of $PolG^{mut}$ alone (Figure 4.5A).

Our work in **Chapter 3** revealed that the type of mtDNA mutation in addition to its frequency is an important determinant of pathogenicity. Therefore, we tested the hypothesis that perturbations of Parkin may have affected mutation pathogenicity but not mutation frequency. This hypothesis would be valid if mitophagy selects against only the most deleterious mutations, if new

A



B

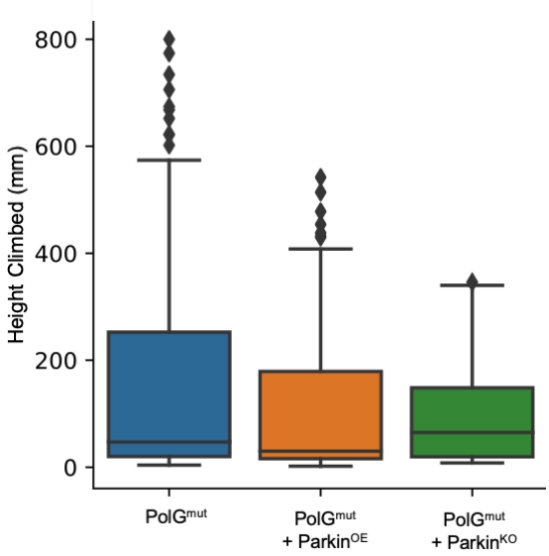


Figure 4.3. Parkin overexpression does not rescue the PolG$_{mut}$ climbing deficit. The mean distance climbed by male flies of the given genotypes in 3 seconds in the in the Rapid Iterative Negative Geotaxis assay at **(A)** 15 days and **(B)** 30 days. *$p < 0.05$ by Student's t-test, all others nonsignificant.

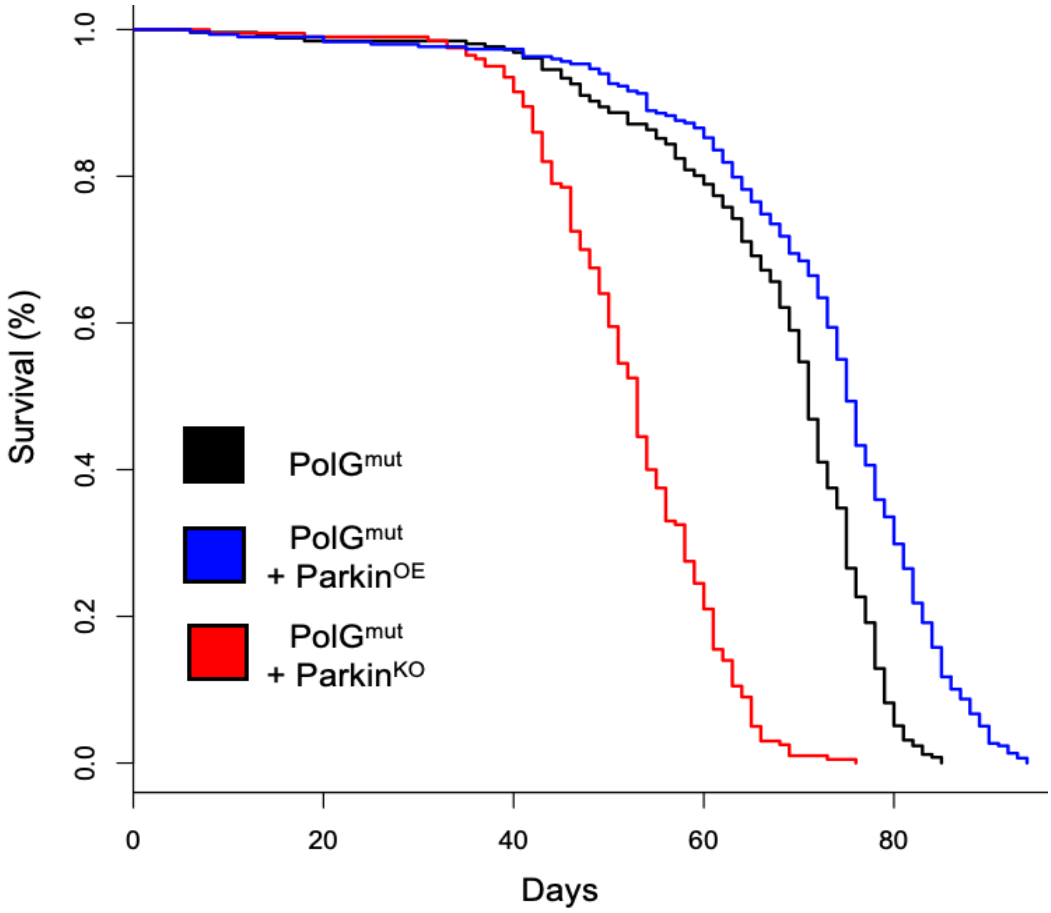


Figure 4.4. Parkin overexpression partially rescues, and Parkin knockout enhances PolG$_{mut}$ lifespan defects. The lifespan of male flies of the indicated genotypes are shown. $p < .05$ for both genotypes versus controls, determined by log-rank test.

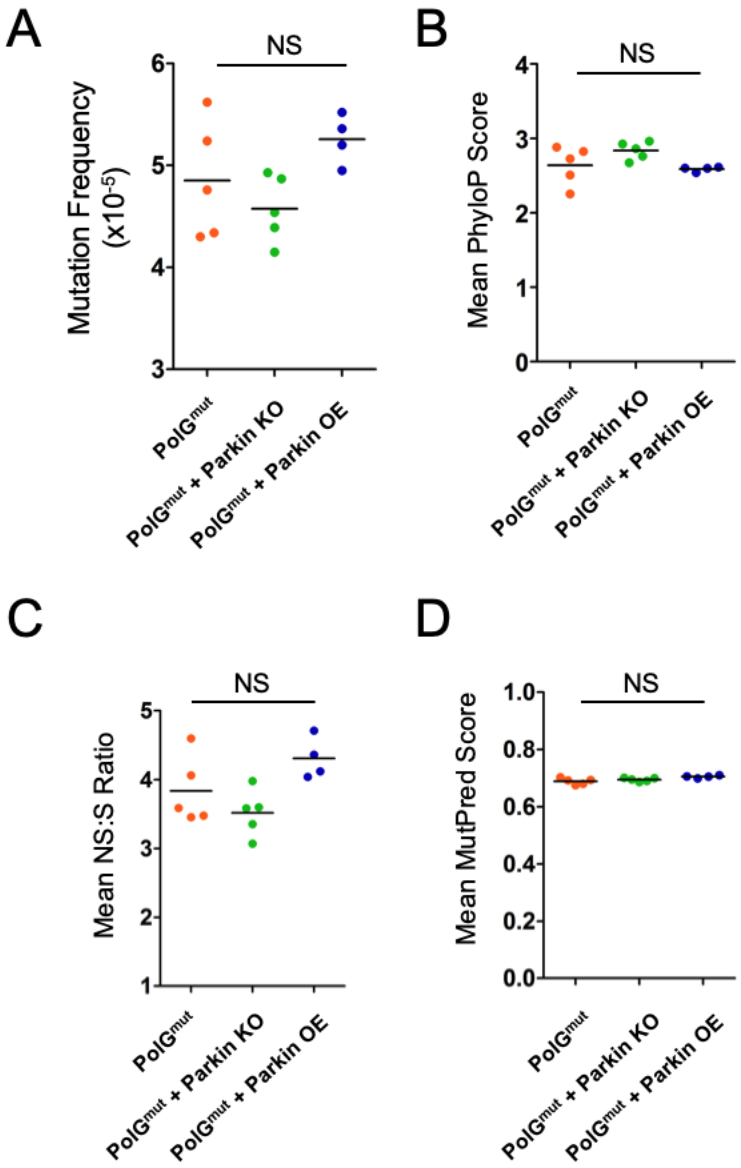


Figure 4.5. Parkin does not influence the pathogenicity of mutations in PolG$_{mut}$ flies. **(A)** Point mutation frequency, **(B)** mean PhyloP score of protein-coding mutations, **(C)** mean NS:S ratio of protein-coding mutations, and **(D)** Mean MutPred score of protein-coding mutations for 1-day-old flies of the given genotype. All comparisons are nonsignificant ($p > .05$) via Student's t-test. n = 4-5 per genotype

mutations are introduced by the PolG$_{mut}$ polymerase at the same rate that deleterious mutations are destroyed, or if replication is nonselective between wild-type and mutant mtDNA. This hypothesis predicts a shift in favor of less deleterious point mutations predominating in Parkin$_{OE}$ flies relative controls, and of deleterious mutations predominating in Parkin$_{KO}$ flies.

We used three metrics previously described in **Chapter 2** to investigate the pathogenicity of mutations. We first looked at the evolutionary conservation of mutated sites using PhyloP. We predicted a shift in mutations to lower average PhyloP in Parkin$_{OE}$ mutator flies. However, there were no significant differences in PhyloP scores between genotypes (Figure 4.5B). We next looked at the ratio of nonsynonymous to synonymous mutations (NS:S) in mutations occurring within protein-coding genes. We predicted that we would find a lower ratio in Parkin$_{OE}$ PolG$_{mut}$ flies, as deleterious NS mutations would be eliminated through mitophagy. Again, we found no significant difference between any of the three genotypes tested (Figure 4.5C). For our final test, we compared the pathogenicity of nonsynonymous protein-coding mutations using the MutPred algorithm. Our hypothesis would predict that Parkin$_{OE}$ PolG$_{mut}$ flies would harbor mutations with lower MutPred scores, indicative of less pathogenic mutations. Once more, we found no difference between genotypes (Figure 4.5D). These results suggest that Parkin does not influence the average pathogenicity of mtDNA mutations.

Previous work indicates that mutational processes have different trinucleotide signatures, and we demonstrated in our work in **Chapter 2** the trinucleotide context of mutations can be used to distinguish between mitochondrial mutagens Thus, if there are particular mutational processes that are selected against by Parkin, this should be reflected by a change to the trinucleotide signature upon perturbations of Parkin. When we compared the trinucleotide mutation signatures of flies with perturbations of Parkin to the signature of PolG$_{mut}$ flies, we discovered that their

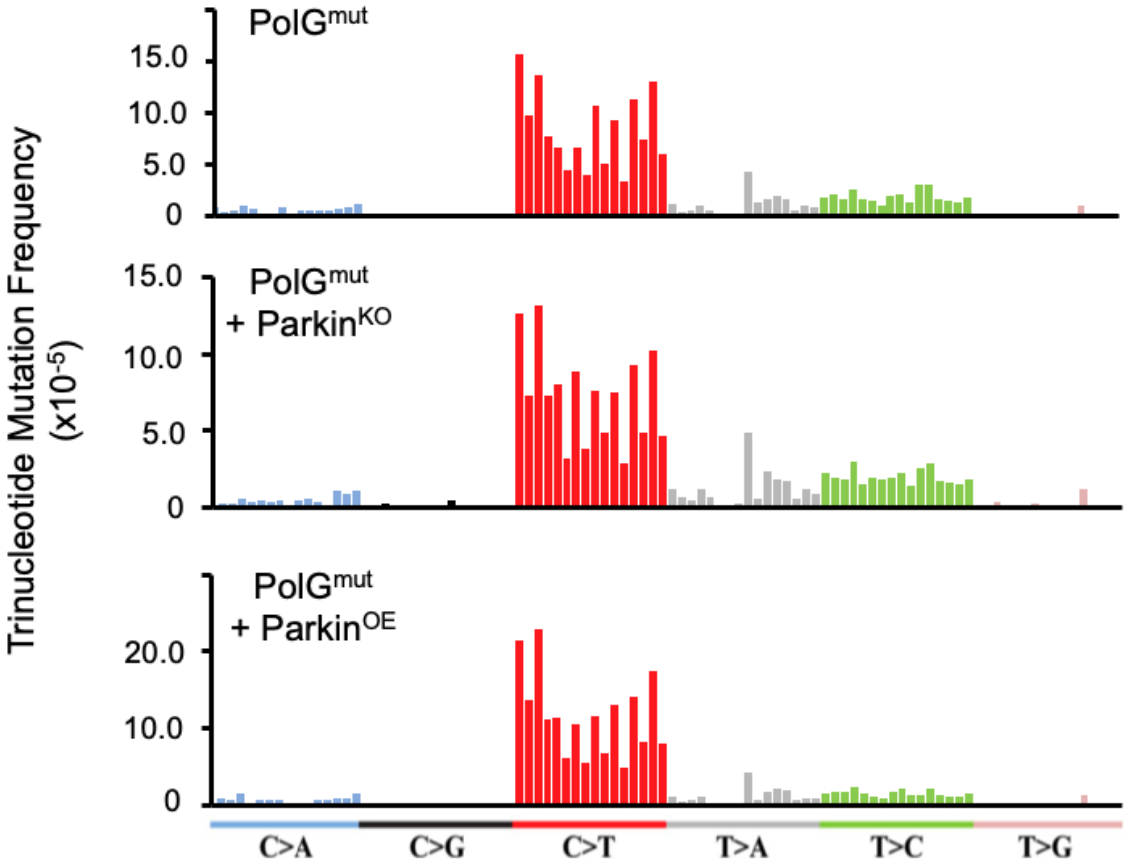


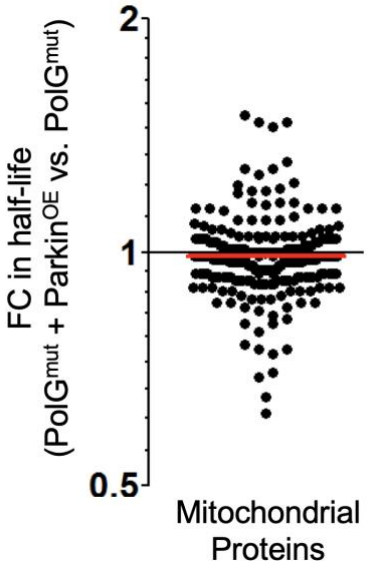
Figure 4.6. Parkin has no detectable influence on the trinucleotide context of mutations. The trinucleotide mutation frequency for each of the 96 possible trinucleotide combinations in the genome, calculated as described in Chapter 2. *n* = 4-5 per genotype

mutation spectra were all strikingly similar (Figure 4.6), implying shared mutagenic and selective processes. Taken together, these data suggest that the presence or absence of Parkin does not influence the spectrum of mtDNA mutations.

Although our data do not suggest that Parkin is capable of selecting against mtDNA mutations, this does not rule out the possibility that Parkin eliminates mitochondria damaged by mtDNA mutations without selecting against the mtDNA itself. This might occur if the mitochondrial genotype and its associated phenotype are uncoupled through fusion and fission, as outlined in **Chapter 1**. Frequent cycles of mitochondrial fusion and fission might distribute the damaged products of mtDNA mutations between mitochondria, such that mitophagic machinery cannot distinguish between a mitochondrion bearing only wild-type mtDNA from those bearing mtDNA mutations. In this instance, the induction of mitophagic pathway components might accelerate the degradation of dysfunctional mitochondria but not appreciably influence the frequency of deleterious mtDNA mutations. This effect might explain why Parkin overexpression partially rescues lifespan without affecting mutation frequency or pathogenicity. In essence, Parkin overexpression could reduce the burden of dysfunctional protein products from mtDNA mutations.

We tested this hypothesis by repeating our proteomics experiments on Parkin overexpressing PolG$_{mut}$ flies. If Parkin overexpression promotes mitochondrial turnover in the absence of mtDNA selection, we would expect to find that mitochondrial protein half-lives are reduced in these flies relative to PolG$_{mut}$ flies. Surprisingly, we discovered that Parkin overexpression did not detectably influence the rates of turnover of either mitochondrial or nonmitochondrial proteins in PolG$_{mut}$ flies (Figure 4.7, Figure 4.8A). However, two proteins that showed increased abundance in Parkin overexpressing flies are worth mentioning. One of these proteins is Parkin, thus validating that we did indeed achieve Parkin overexopression

A



B

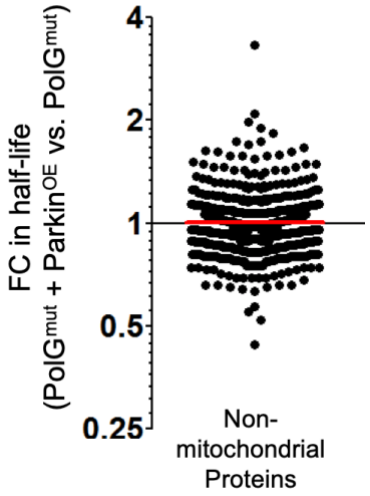


Figure 4.7. Few mitochondrial proteins are increased in turnover in PolG$_{mut}$ flies overexpressing Parkin. Fold change (FC) in half-lives for **(A)** mitochondrial proteins (*n* = 228) and **(B)** non-mitochondrial proteins (*n* = 954) in PolG$_{mut}$ flies overexpressing Parkin compared to PolG$_{mut}$ flies. The x-axis represents equal protein half-lives in both genotypes, the red horizontal line represents the group mean, points above the x-axis show faster turnover in Parkin$_{OE}$ flies. There is no significant difference between PolG$_{mut}$+Parkin$_{OE}$ and PolG$_{mut}$ in either mitochondrial or non-mitochondrial protein turnover, by nested ANOVA

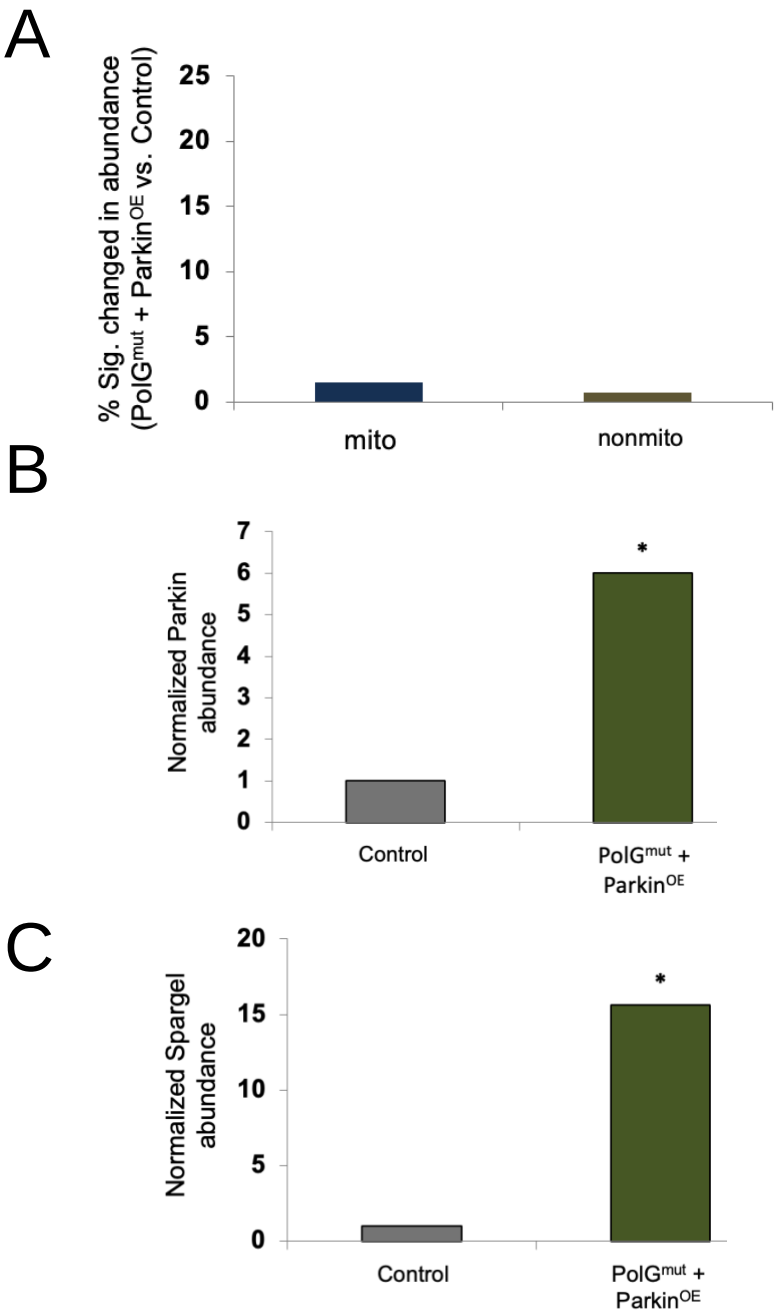


Figure 4.8. Mitochondrial biogenesis is induced in Parkin$_{OE}$ PolG$_{mut}$ flies, but there are few significant changes to protein abundance. (A) percentage of proteins, mitochondrial and non-mitochondrial, that significantly differ in abundance between control flies and PolG$_{mut}$ flies overexpressing Parkin. **(B)** Protein abundance of Parkin for the genotypes indicated. **(C)** Protein abundance of Spargel, the fly homolog of PGC-1a, for the genotypes indicated at 120 h. *$p < .05$ via Student's t-test.

(Figure 4.8B). The second protein was *Spargel*, the *Drosophila* homolog of peroxisome proliferator-activated receptor gamma coactivator 1-α (PGC-1a) (Figure 4.8C). *Spargel* is a transcription factor that serves as "the master regulator" of mitochondrial biogenesis. While preliminary, our results suggest that compensatory mitochondrial biogenesis is induced in mutator flies, in agreement with our "hitchhiker hypothesis" for the overrepresentation of deleterious mutations.

Together, our data offer no support for the hypothesis that Parkin selectively eliminates pathogenic mtDNA mutations in somatic tissues. We see no alterations in mitophagy in PolG$_{mut}$ flies, despite their harboring a high mutation burden, nor do we see upregulation of mitophagy in PolG$_{mut}$ flies overexpressing Parkin. Additionally, our sequence data shows no differences in the frequency, pathogenicity, or spectrum of mutations between PolG$_{mut}$ flies and PolG$_{mut}$ flies overexpressing Parkin. Taken together, my thesis studies did not detect any evidence that selection acts against harmful mtDNA mutations in the somatic tissues of *Drosophila*

4.3 *Notes and Acknowledgements*

The proteomic work in this chapter was performed in collaboration with Dr. Evelyn S. Vincow, of the Pallanck Lab, with mass spectrometry assistance from the MacCoss Lab at the University of Washington. Unpublished data figures in **Chapter 4** were generated using R statistical software, Microsoft Excel, and GraphPad Prism. Additional thanks to Evvie Vincow for all of her other contributions to this work, including the interpretation and graphing of proteomic data and critical reading of this chapter.

Chapter 5. **Future Directions**

5.1 *Future Experiments*

One important question that remains unresolved from my work is why deleterious mtDNA mutations are overrepresented in mutator flies. We proposed two putative models to explain this finding, but we have not yet determined the validity or relative contribution of these mechanisms. The first model is "survival of the slowest"; this model hypothesizes that mitochondria bearing deleterious mutations undergo less frequent turnover due to a diminished respiratory capacity and less oxidative damage. A second model is the "hitchhiker model"—mitochondria bearing deleterious mutations undergo compensatory mtDNA replication in an attempt to maintain function, inadvertently increasing the number of these harmful mutation-bearing genomes. Future work should test whether either the "survival of the slowest" or the "hitchhiker" models explain the overrepresentation of deleterious mtDNA mutations in mutator flies.

To distinguish between these models, we could sequence from cell populations sorted by categories of interest. One way we could test this is to correlate cellular phenotypes with the mtDNA mutations they harbor. For example, we could perform fluorescence activated cell sorting (FACS) to separate cell populations using fluorescent markers. We could separate cell populations that differ with respect to mitochondrial ROS and mitochondrial biogenesis using appropriate reporters, and then sequence the mtDNA from these cell populations. We could then determine whether the frequency and pathogenicity of mtDNA mutations differ between these cell populations. A lower frequency of deleterious mtDNA mutations in cells that express markers associated with high mitochondrial ROS would support my "survival of the slowest" model. A higher frequency of deleterious mtDNA mutations in cells that express markers associated with the induction of mitochondrial biogenesis would support my "hitchhiker" model. These studies

would provide new insights into the cellular mechanisms that promote the expansion of mtDNA mutations with age.

As I explained in my introduction, we are only now gaining an appreciation of the complexity of cell-to-cell variation in heteroplasmic mutations (for an excellent depiction of mitochondrial heterogeneity, see (*64*)). Historically we have sampled from complex tissues containing many cells of various cell types, each harboring thousands of copies of the mitochondrial genome. Sequencing from these samples can only, at best, reveal the population average of mutations. As a result, our understanding of cell-to-cell mtDNA mutation heterogeneity and its impact on mitochondrial and cellular function remains very limited. Recently, there have been exciting advances in single cell and single mitochondrion sequencing that have the potential to better reveal the cell-to-cell distributions of mutations within an organism (*225-227*). Single cell and single mitochondrion analyses will allow us to determine the distributions of mutations within an organism, to better understand how different tissue types accumulate mutations, and to detect whether mutations clonally expand within individual cells. There are many fundamental biological questions that remain in the study of mtDNA mutations. With the advent of a multitude of animal models and powerful sequencing techniques, it is a tremendously exciting time to study heteroplasmy, and the ability to detect mutations within individual cells will open many new avenues for discovery.

5.2 *Summary and conclusions*

mtDNA mutations cause a number of severe maternally transmitted diseases, and the accumulation of somatic mtDNA mutations is implicated in common diseases, including cancer, diabetes, and Parkinson's disease. Often, these mutations coexist with wild type mtDNA, a

condition known as heteroplasmy. Although the ratio of mutated to wild type mtDNA correlates with the severity of diseases associated with mtDNA mutations, the mechanisms that influence this ratio are poorly understood, as are the factors that oppose the accumulation of mtDNA mutations and their pathologies. In **Chapter 1**, I highlighted unresolved questions in the fields of mitochondrial biology and mtDNA mutation selection.

To investigate whether cells select against harmful mtDNA mutations in somatic tissues, I developed a *Drosophila* strain that expresses a proofreading-deficient mtDNA polymerase that displays up to a 55-fold elevation in mutation frequency. I used Duplex Sequencing, an innovative high-fidelity sequencing technique to detect mtDNA mutations, and applied statistical and computational techniques to study how selection acts upon somatic mtDNA mutations in mutator flies. In **Chapter 2**, I discovered that mutator flies accumulate mutations predominantly in highly conserved sites, and that mutator flies bear an excess of deleterious nonsynonymous mutations relative to a neutral mutational model.

In **Chapter 3**, I demonstrated that two fly models with elevated mtDNA mutation frequencies differ strikingly in the severity of their phenotypes as a result of differences in the spectra and clonality of their mutations. In **Chapter 4**, I showed that Parkin does not preferentially eliminate mitochondria bearing mtDNA mutations, and that mitophagy does not eliminate somatic mtDNA point mutations. Taken together, these findings help explain why pathogenic mtDNA mutations accumulate in aging and disease states and might inform the development of therapeutic targets for future treatment of these diseases

BIBLIOGRAPHY

1. M. W. Gray, G. Burger, B. F. Lang, The origin and early evolution of mitochondria. *Genome Biol* **2**, REVIEWS1018 (2001).
2. M. W. Gray, Mitochondrial evolution. *Cold Spring Harb Perspect Biol* **4**, a011403 (2012).
3. H. M. McBride, M. Neuspiel, S. Wasiak, Mitochondria: more than just a powerhouse. *Curr Biol* **16**, R551-560 (2006).
4. D. C. Wallace, Mitochondria and cancer. *Nat Rev Cancer* **12**, 685-698 (2012).
5. C. Wang, R. J. Youle, The role of mitochondria in apoptosis*. *Annu Rev Genet* **43**, 95-118 (2009).
6. K. B. Busch, A. Kowald, J. N. Spelbrink, Quality matters: how does mitochondrial network dynamics and quality control impact on mtDNA integrity? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **369**, 20130442 (2014).
7. P. F. Chinnery, G. Hudson, Mitochondrial genetics. *Br Med Bull* **106**, 135-159 (2013).
8. P. Bjorkholm, A. Harish, E. Hagstrom, A. M. Ernst, S. G. Andersson, Mitochondrial genomes are retained by selective constraints on protein targeting. *Proc Natl Acad Sci U S A* **112**, 10154-10161 (2015).
9. S. Cogliati, I. Lorenzi, G. Rigoni, F. Caicci, M. E. Soriano, Regulation of Mitochondrial Electron Transport Chain Assembly. *J Mol Biol* **430**, 4849-4873 (2018).
10. P. Mitchell, J. Moyle, Chemiosmotic hypothesis of oxidative phosphorylation. *Nature* **213**, 137-139 (1967).
11. M. P. Murphy, How mitochondria produce reactive oxygen species. *Biochem J* **417**, 1-13 (2009).
12. N. G. Larsson, Somatic mitochondrial DNA mutations in mammalian aging. *Annu Rev Biochem* **79**, 683-706 (2010).
13. M. J. Keogh, P. F. Chinnery, Mitochondrial DNA mutations in neurodegeneration. *Biochim Biophys Acta* **1847**, 1401-1411 (2015).
14. M. F. Beal, Mitochondria take center stage in aging and neurodegeneration. *Ann Neurol* **58**, 495-505 (2005).
15. D. Edgar *et al.*, Random Point Mutations with Major Effects on Protein-Coding Genes Are the Driving Force behind Premature Aging in mtDNA Mutator Mice. *Cell Metabolism* **10**, 131-138 (2009).
16. R. A. Risques, S. R. Kennedy, Aging and the rise of somatic cancer-associated mutations in normal tissues. *PLOS Genetics* **14**, e1007108 (2018).
17. V. Carelli, C. La Morgia, Clinical syndromes associated with mtDNA mutations: where we stand after 30 years. *Essays Biochem* **62**, 235-254 (2018).
18. F. Ursini *et al.*, Diversity of glutathione peroxidases. *Methods Enzymol* **252**, 38-53 (1995).
19. R. S. Esworthy, Y. S. Ho, F. F. Chu, The Gpx1 gene encodes mitochondrial glutathione peroxidase in the mouse liver. *Arch Biochem Biophys* **340**, 59-63 (1997).
20. W. A. Pryor, Oxy-radicals and related species: their formation, lifetimes, and reactions. *Annu Rev Physiol* **48**, 657-667 (1986).
21. R. De Bont, N. van Larebeke, Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169-185 (2004).

22. L. A. Esposito *et al.*, Mitochondrial oxidative stress in mice lacking the glutathione peroxidase-1 gene. *Free Radic Biol Med* **28**, 754-766 (2000).
23. S. E. Schriner *et al.*, Extension of Murine Life Span by Overexpression of Catalase Targeted to Mitochondria. *Science* **308**, 1909-1911 (2005).
24. Y. C. Jang *et al.*, Overexpression of Mn superoxide dismutase does not increase life span in mice. *J. Gerontol. A Biol. Sci. Med. Sci.* **64**, 1114-1125 (2009).
25. V. I. Perez *et al.*, The overexpression of major antioxidant enzymes does not extend the lifespan of mice. *Aging Cell* **8**, 73-75 (2009).
26. V. I. Perez *et al.*, Is the oxidative stress theory of aging dead? *Biochim Biophys Acta* **1790**, 1005-1014 (2009).
27. Y. Zhang *et al.*, Mice deficient in both Mn superoxide dismutase and glutathione peroxidase-1 have increased oxidative damage and a greater incidence of pathology but no reduction in longevity. *J Gerontol A Biol Sci Med Sci* **64**, 1212-1220 (2009).
28. F. Foury, J. Hu, S. Vanderstraeten, Mitochondrial DNA mutators. *Cell Mol Life Sci* **61**, 2799-2811 (2004).
29. J. Lebeau, T. K. Rainbolt, R. L. Wiseman, Coordinating Mitochondrial Biology Through the Stress-Responsive Regulation of Mitochondrial Proteases. *Int Rev Cell Mol Biol* **340**, 79-128 (2018).
30. W. Voos, W. Jaworek, A. Wilkening, M. Bruderek, Protein quality control at the mitochondrion. *Essays In Biochemistry* **60**, 213-225 (2016).
31. K. J. Lin *et al.*, The Overcrowded Crossroads: Mitochondria, Alpha-Synuclein, and the Endo-Lysosomal System Interaction in Parkinson's Disease. *Int J Mol Sci* **20**, (2019).
32. V. Azzu, M. D. Brand, Degradation of an intramitochondrial protein by the cytosolic proteasome. *J Cell Sci* **123**, 578-585 (2010).
33. R. D. Martinus *et al.*, Selective induction of mitochondrial chaperones in response to loss of the mitochondrial genome. *Eur J Biochem* **240**, 98-103 (1996).
34. A. M. Nargund, M. W. Pellegrino, C. J. Fiorese, B. M. Baker, C. M. Haynes, Mitochondrial Import Efficiency of ATFS-1 Regulates Mitochondrial UPR Activation. *Science* **337**, 587-590 (2012).
35. Amrita M. Nargund, Christopher J. Fiorese, Mark W. Pellegrino, P. Deng, Cole M. Haynes, Mitochondrial and Nuclear Accumulation of the Transcription Factor ATFS-1 Promotes OXPHOS Recovery during the UPRmt. *Molecular Cell* **58**, 123-133 (2015).
36. A. Melber, C. M. Haynes, UPR mt regulation and output: a stress response mediated by mitochondrial-nuclear communication. *Cell Res* **28**, 281-295 (2018).
37. Christopher J. Fiorese *et al.*, The Transcription Factor ATF5 Mediates a Mammalian Mitochondrial UPR. *Current Biology* **26**, 2037-2043 (2016).
38. A. Sugiura, G. L. McLelland, E. A. Fon, H. M. McBride, A new pathway for mitochondrial quality control: mitochondrial-derived vesicles. *EMBO J* **33**, 2142-2156 (2014).
39. G.-L. McLelland, S. A. Lee, H. M. McBride, E. A. Fon, Syntaxin-17 delivers PINK1/parkin-dependent mitochondrial vesicles to the endolysosomal system. *The Journal of Cell Biology* **214**, 275-291 (2016).
40. M. Alexeyev, I. Shokolenko, G. Wilson, S. LeDoux, The maintenance of mitochondrial DNA integrity--critical analysis and update. *Cold Spring Harb Perspect Biol* **5**, a012641 (2013).
41. D. A. Clayton, J. N. Doda, E. C. Friedberg, The absence of a pyrimidine dimer repair mechanism in mammalian mitochondria. *Proc Natl Acad Sci U S A* **71**, 2777-2781 (1974).

42. A. Moretton *et al.*, Selective mitochondrial DNA degradation following double-strand breaks. *PLOS ONE* **12**, e0176795 (2017).
43. S. L. Clark, Jr., Cellular differentiation in the kidneys of newborn mice studies with the electron microscope. *J Biophys Biochem Cytol* **3**, 349-362 (1957).
44. A. B. Novikoff, The proximal tubule cell in experimental hydronephrosis. *J Biophys Biochem Cytol* **6**, 136-138 (1959).
45. A. M. Tolkovsky, Mitophagy. *Biochim Biophys Acta* **1793**, 1508-1515 (2009).
46. I. Kim, S. Rodriguez-Enriquez, J. J. Lemasters, Selective degradation of mitochondria by mitophagy. *Arch Biochem Biophys* **462**, 245-253 (2007).
47. R. J. Youle, D. P. Narendra, Mechanisms of mitophagy. *Nat Rev Mol Cell Biol* **12**, 9-14 (2011).
48. J. C. Greene *et al.*, Mitochondrial pathology and apoptotic muscle degeneration in Drosophila parkin mutants. *Proceedings of the National Academy of Sciences* **100**, 4078-4083 (2003).
49. A. J. Whitworth *et al.*, Increased glutathione S-transferase activity rescues dopaminergic neuron loss in a Drosophila model of Parkinson's disease. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8024-8029 (2005).
50. I. E. Clark *et al.*, Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin. *Nature* **441**, 1162-1166 (2006).
51. A. C. Poole *et al.*, The PINK1/Parkin pathway regulates mitochondrial morphology. *Proceedings of the National Academy of Sciences* **105**, 1638-1643 (2008).
52. H. Deng, M. W. Dodson, H. Huang, M. Guo, The Parkinson's disease genes pink1 and parkin promote mitochondrial fission and/or inhibit fusion in Drosophila. *Proc Natl Acad Sci U S A* **105**, 14503-14508 (2008).
53. E. Ziviani, R. N. Tao, A. J. Whitworth, Drosophila parkin requires PINK1 for mitochondrial translocation and ubiquitinates mitofusin. *Proc Natl Acad Sci U S A* **107**, 5018-5023 (2010).
54. Y. Yang *et al.*, Pink1 regulates mitochondrial dynamics through interaction with the fission/fusion machinery. *Proc Natl Acad Sci U S A* **105**, 7070-7075 (2008).
55. H. Chen, D. C. Chan, Mitochondrial dynamics--fusion, fission, movement, and mitophagy--in neurodegenerative diseases. *Hum Mol Genet* **18**, R169-176 (2009).
56. D. F. Suen, D. P. Narendra, A. Tanaka, G. Manfredi, R. J. Youle, Parkin overexpression selects against a deleterious mtDNA mutation in heteroplasmic cybrid cells. *Proc Natl Acad Sci U S A* **107**, 11835-11840 (2010).
57. R. Allio, S. Donega, N. Galtier, B. Nabholz, Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker. *Molecular Biology and Evolution* **34**, 2762-2772 (2017).
58. N. A. Bonekamp, N.-G. Larsson, SnapShot: Mitochondrial Nucleoid. *Cell* **172**, 388-388.e381 (2018).
59. C. Kukat *et al.*, Super-resolution microscopy reveals that mammalian mitochondrial nucleoids have a uniform size and frequently contain a single copy of mtDNA. *Proceedings of the National Academy of Sciences* **108**, 13534-13539 (2011).
60. W. J. Valente *et al.*, Mitochondrial DNA exhibits resistance to induced point and deletion mutations. *Nucleic Acids Research* **44**, 8513-8524 (2016).

61. S. R. Kennedy, J. J. Salk, M. W. Schmitt, L. A. Loeb, Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLOS Genetics* **9**, e1003794 (2013).
62. L. S. Itsara *et al.*, Oxidative Stress Is Not a Major Contributor to Somatic Mitochondrial DNA Mutations. *PLOS Genetics* **10**, e1003974 (2014).
63. J. Aryaman, I. G. Johnston, N. S. Jones, Mitochondrial Heterogeneity. *Frontiers in Genetics* **9**, (2019).
64. A. Hahn, S. Zuryn, The Cellular Mitochondrial Genome Landscape in Disease. *Trends in Cell Biology* **29**, 227-240 (2019).
65. J. Hayashi *et al.*, 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).
66. A. Chomyn *et al.*, 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).
67. R. H. Hämäläinen *et al.*, 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* **110**, E3622-E3630 (2013).
68. A. Malena *et al.*, Mitochondrial quality control: Cell-type-dependent responses to pathological mutant mitochondrial DNA. *Autophagy* **12**, 2098-2112 (2016).
69. T. Ono, K. Isobe, K. Nakada, J. I. Hayashi, Human cells are protected from mitochondrial dysfunction by complementation of DNA products in fused mitochondria. *Nat Genet* **28**, 272-275 (2001).
70. K. Okamoto, P. S. Perlman, R. A. Butow, The sorting of mitochondrial DNA and mitochondrial proteins in zygotes: preferential transmission of mitochondrial DNA to the medial bud. *J Cell Biol* **142**, 613-623 (1998).
71. J. A. Enriquez, J. Cabezas-Herrera, M. P. Bayona-Bafaluy, G. Attardi, Very rare complementation between mitochondria carrying different mitochondrial DNA mutations points to intrinsic genetic autonomy of the organelles in cultured human cells. *J Biol Chem* **275**, 11207-11215 (2000).
72. M. Yoneda, T. Miyatake, G. Attardi, Complementation of mutant and wild-type human mitochondrial DNAs coexisting since the mutation event and lack of complementation of DNAs introduced separately into a cell within distinct organelles. *Mol Cell Biol* **14**, 2699-2712 (1994).
73. C. Kukat *et al.*, Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11288-11293 (2015).
74. R. A. Capaldi *et al.*, A replicating module as the unit of mitochondrial structure and functioning. *Biochim Biophys Acta* **1555**, 192-195 (2002).
75. C. Osman, T. R. Noriega, V. Okreglak, J. C. Fung, P. Walter, Integrity of the yeast mitochondrial genome, but not its distribution and inheritance, relies on mitochondrial fission and fusion. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E947-956 (2015).
76. T. E. S. Kauppila, J. H. K. Kauppila, N.-G. Larsson, Mammalian Mitochondria and Aging: An Update. *Cell Metabolism* **25**, 57-71 (2017).

77. K. Ye, J. Lu, F. Ma, A. Keinan, Z. Gu, Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10654-10659 (2014).
78. C. B. Park, N.-G. Larsson, Mitochondrial DNA mutations in disease and aging. *The Journal of Cell Biology* **193**, 809-818 (2011).
79. S. Matic *et al.*, Mice lacking the mitochondrial exonuclease MGME1 accumulate mtDNA deletions without developing progeria. *Nature Communications* **9**, 1202 (2018).
80. M. Vermulst *et al.*, Mitochondrial point mutations do not limit the natural lifespan of mice. *Nature Genetics* **39**, 540-543 (2007).
81. A. A. Zaidi *et al.*, Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees. *Proc Natl Acad Sci U S A*, (2019).
82. B. Cavadas *et al.*, Fine Time Scaling of Purifying Selection on Human Nonsynonymous mtDNA Mutations Based on the Worldwide Population Tree and Mother–Child Pairs. *Human Mutation* **36**, 1100-1111 (2015).
83. J. B. Stewart, P. F. Chinnery, The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics* **16**, 530-542 (2015).
84. S. De Fanti *et al.*, Intra-individual purifying selection on mitochondrial DNA variants during human oogenesis. *Hum. Reprod.* **32**, 1100-1107 (2017).
85. V. I. Floros *et al.*, Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. *Nat Cell Biol* **20**, 144-151 (2018).
86. W. Fan *et al.*, A mouse model of mitochondrial disease reveals germline selection against severe mtDNA mutations. *Science* **319**, 958-962 (2008).
87. H. Ma, H. Xu, P. H. O'Farrell, Transmission of mitochondrial mutations and action of purifying selection in Drosophila melanogaster. *Nature Genetics* **46**, 393-397 (2014).
88. J. H. Hill, Z. Chen, H. Xu, Selective propagation of functional mitochondrial DNA during oogenesis restricts the transmission of a deleterious mitochondrial variant. *Nature Genetics* **46**, 389-392 (2014).
89. Y. Zhang *et al.*, PINK1 Inhibits Local Protein Synthesis to Limit Transmission of Deleterious Mitochondrial DNA Mutations. *Molecular Cell*, (2019).
90. T. Lieber, S. P. Jeedigunta, J. M. Palozzi, R. Lehmann, T. R. Hurd, Mitochondrial fragmentation drives selective removal of deleterious mtDNA in the germline. *Nature*, 1 (2019).
91. L. M. Cree *et al.*, A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nature Genetics* **40**, 249-254 (2008).
92. H. Zhang, S. P. Burr, P. F. Chinnery, The mitochondrial DNA genetic bottleneck: inheritance and beyond. *Essays In Biochemistry* **62**, 225-234 (2018).
93. J. L. Elson, D. C. Samuels, D. M. Turnbull, P. F. Chinnery, Random Intracellular Drift Explains the Clonal Expansion of Mitochondrial DNA Mutations with Age. *The American Journal of Human Genetics* **68**, 802-806 (2001).
94. P. F. Chinnery *et al.*, Point mutations of the mtDNA control region in normal and neurodegenerative human brains. *Am J Hum Genet* **68**, 529-532 (2001).
95. H. Ma *et al.*, Deleterious mtDNA Mutations are Common in Mature Oocytes. *Biol Reprod*, (2019).

96. L. C. Greaves *et al.*, Comparison of mitochondrial mutation spectra in ageing human colonic epithelium and disease: absence of evidence for purifying selection in somatic mitochondrial DNA point mutations. *PLOS Genetics* **8**, e1003082 (2012).
97. Z. Chen *et al.*, Genetic mosaic analysis of a deleterious mitochondrial DNA mutation in Drosophila reveals novel aspects of mitochondrial regulation and function. *Molecular Biology of the Cell* **26**, 674-684 (2015).
98. T. Sasaki, Y. Sato, T. Higashiyama, N. Sasaki, Live imaging reveals the dynamics and regulation of mitochondrial nucleoids during the cell cycle in Fucci2-HeLa cells. *Scientific Reports* **7**, 1-12 (2017).
99. A. Rana *et al.*, Promoting Drp1-mediated mitochondrial fission in midlife prolongs healthy lifespan of Drosophila melanogaster. *Nature Communications* **8**, 1-14 (2017).
100. A. Malena, E. Loro, M. Di Re, I. J. Holt, L. Vergani, Inhibition of mitochondrial fission favours mutant over wild-type mitochondrial DNA. *Human Molecular Genetics* **18**, 3407-3416 (2009).
101. R. W. Gilkerson *et al.*, Mitochondrial autophagy in cells with mtDNA mutations results from synergistic loss of transmembrane potential and mTORC1 inhibition. *Hum Mol Genet* **21**, 978-990 (2012).
102. Y. Dai *et al.*, Rapamycin drives selection against a pathogenic heteroplasmic mitochondrial DNA mutation. *Human Molecular Genetics* **23**, 637-647 (2014).
103. S. E. Siegmund *et al.*, Low-dose rapamycin extends lifespan in a mouse model of mtDNA depletion syndrome. *Human Molecular Genetics* **26**, 4588-4605 (2017).
104. J. Bielas *et al.*, Long term rapamycin treatment improves mitochondrial DNA quality in aging mice. *Experimental Gerontology* **106**, 125-131 (2018).
105. D.-F. Suen, D. P. Narendra, A. Tanaka, G. Manfredi, R. J. Youle, Parkin overexpression selects against a deleterious mtDNA mutation in heteroplasmic cybrid cells. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11835-11840 (2010).
106. A. Trifunovic *et al.*, Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* **429**, 417-423 (2004).
107. N. Sun *et al.*, Measuring In Vivo Mitophagy. *Molecular Cell* **60**, 685-696 (2015).
108. A. M. Pickrell *et al.*, Endogenous Parkin Preserves Dopaminergic Substantia Nigral Neurons following Mitochondrial DNA Mutagenic Stress. *Neuron* **87**, 371-381 (2015).
109. J. M. Ross *et al.*, Voluntary exercise normalizes the proteomic landscape in muscle and brain and improves the phenotype of progeroid mice. *Aging Cell* **0**, e13029.
110. A. Safdar *et al.*, Amelioration of premature aging in mtDNA mutator mouse by exercise: the interplay of oxidative stress, PGC-1α, p53, and DNA damage. A hypothesis. *Current Opinion in Genetics & Development* **38**, 127-132 (2016).
111. N. P. Kandul, T. Zhang, B. A. Hay, M. Guo, Selective removal of deletion-bearing mitochondrial DNA in heteroplasmic Drosophila. *Nature Communications* **7**, 13100 (2016).
112. E. J. Hsieh *et al.*, Topograph, a software platform for precursor enrichment corrected global protein turnover measurements. *Mol Cell Proteomics* **11**, 1468-1474 (2012).
113. B. P. Woodall *et al.*, Parkin does not prevent accelerated cardiac aging in mitochondrial DNA mutator mice. *JCI Insight*, (2019).
114. J. P. Jenuth, A. C. Peterson, E. A. Shoubridge, Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. *Nature Genetics* **16**, 93 (1997).

115. T. Ni *et al.*, MitoRCA-seq reveals unbalanced cytocine to thymine transition in Polg mutant mice. *Scientific Reports* **5**, 12049 (2015).
116. M. Li, R. Schröder, S. Ni, B. Madea, M. Stoneking, Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2491-2496 (2015).
117. H. Ma, P. H. O'Farrell, Selfish drive can trump function when animal mitochondrial genomes compete. *Nature Genetics* **48**, 798-802 (2016).
118. O. M. Russell *et al.*, Preferential amplification of a human mitochondrial DNA deletion in vitro and in vivo. *Scientific Reports* **8**, 1799 (2018).
119. D. C. Wallace, Mitochondrial DNA mutations and neuromuscular disease. *Trends Genet* **5**, 9-13 (1989).
120. A. D. de Grey, How is mutant mitochondrial DNA clonally amplified? Much new evidence, still no answers. *Rejuvenation Res* **12**, 217-219 (2009).
121. Y. Michikawa, F. Mazzucchelli, N. Bresolin, G. Scarlato, G. Attardi, Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. *Science* **286**, 774-779 (1999).
122. A. Kowald, T. B. L. Kirkwood, Transcription could be the key to the selection advantage of mitochondrial deletion mutants in aging. *Proceedings of the National Academy of Sciences* **111**, 2972-2977 (2014).
123. A. Kowald, T. B. L. Kirkwood, Mitochondrial mutations and aging: random drift is insufficient to explain the accumulation of mitochondrial deletion mutants in short-lived animals. *Aging Cell* **12**, 728-731 (2013).
124. A. D. de Grey, Reactive oxygen species production in the mitochondrial matrix: implications for the mechanism of mitochondrial mutation accumulation. *Rejuvenation Res* **8**, 13-17 (2005).
125. X. Li *et al.*, Multi-regional sequencing reveals intratumor heterogeneity and positive selection of somatic mtDNA mutations in hepatocellular carcinoma and colorectal cancer. *International Journal of Cancer* **143**, 1143-1152 (2018).
126. Z. Cao, J. Wanagat, S. H. McKiernan, J. M. Aiken, Mitochondrial DNA deletion mutations are concomitant with ragged red regions of individual, aged muscle fibers: analysis by laser-capture microdissection. *Nucleic Acids Research* **29**, 4502-4508 (2001).
127. A. Brinckmann *et al.*, Regionalized Pathology Correlates with Augmentation of mtDNA Copy Numbers in a Patient with Myoclonic Epilepsy with Ragged-Red Fibers (MERRF-Syndrome). *PLOS ONE* **5**, e13513 (2010).
128. Y. Konokhova *et al.*, Failed upregulation of TFAM protein and mitochondrial DNA in oxidatively deficient fibers of chronic obstructive pulmonary disease locomotor muscle. *Skelet Muscle* **6**, 10 (2016).
129. A. Herbst *et al.*, Accumulation of Mitochondrial DNA Deletion Mutations in Aged Muscle Fibers: Evidence for a Causal Role in Muscle Fiber Loss. *J Gerontol A Biol Sci Med Sci* **62**, 235-245 (2007).
130. A. Herbst, C. J. Johnson, K. Hynes, D. McKenzie, J. M. Aiken, Mitochondrial biogenesis drives a vicious cycle of metabolic insufficiency and mitochondrial DNA deletion mutation accumulation in aged rat skeletal muscle fibers. *PLOS ONE* **8**, e59006 (2013).
131. W. Y. Tsang, B. D. Lemire, Stable heteroplasmy but differential inheritance of a large mitochondrial DNA deletion in nematodes. *Biochemistry & Cell Biology* **80**, 645-654 (2002).

132. I. Valenci, L. Yonai, D. Bar-Yaacov, D. Mishmar, A. Ben-Zvi, Parkin modulates heteroplasmy of truncated mtDNA in Caenorhabditis elegans. *Mitochondrion* **20**, 64-70 (2015).
133. B. L. Gitschlag *et al.*, Homeostatic Responses Regulate Selfish Mitochondrial Genome Dynamics in C. elegans. *Cell Metabolism* **24**, 91-103 (2016).
134. Y.-F. Lin *et al.*, Maintenance and propagation of a deleterious mitochondrial genome by the mitochondrial unfolded protein response. *Nature* **533**, 416-419 (2016).
135. M. Vermulst, J. H. Bielas, L. A. Loeb, Quantification of random mutations in the mitochondrial genome. *Methods* **46**, 263-268 (2008).
136. S. Jones *et al.*, Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci U S A* **105**, 4283-4288 (2008).
137. J. Shendure, H. Ji, Next-generation DNA sequencing. *Nat Biotechnol* **26**, 1135-1145 (2008).
138. J. Zhang *et al.*, Curcumin attenuates hepatic mitochondrial dysfunction through the maintenance of thiol pool, inhibition of mtDNA damage, and stimulation of the mitochondrial thioredoxin system in heat-stressed broilers. *J Anim Sci* **96**, 867-879 (2018).
139. M. W. Schmitt *et al.*, Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14508-14513 (2012).
140. J. H. Bielas, L. A. Loeb, Quantification of random genomic mutations. *Nat Methods* **2**, 285-290 (2005).
141. S. D. Taylor *et al.*, Targeted enrichment and high-resolution digital profiling of mitochondrial DNA deletions in human brain. *Aging Cell* **13**, 29-38 (2014).
142. S. R. Kennedy *et al.*, Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc* **9**, 2586-2606 (2014).
143. J. H. K. Kauppila *et al.*, Base-excision repair deficiency alone or combined with increased oxidative stress does not increase mtDNA point mutations in mice. *Nucleic Acids Research* **46**, 6642-6669 (2018).
144. L. C. Greaves *et al.*, Clonal expansion of early to mid-life mitochondrial DNA point mutations drives mitochondrial dysfunction during human ageing. *PLOS Genetics* **10**, e1004620 (2014).
145. A. Ameur *et al.*, Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. *PLoS Genet* **7**, e1002028 (2011).
146. J. B. Stewart *et al.*, Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLOS Biology* **6**, e10 (2008).
147. J. L. Burman, S. Yu, A. C. Poole, R. B. Decal, L. Pallanck, Analysis of neural subtypes reveals selective mitochondrial dysfunction in dopaminergic neurons from parkin mutants. *Proc Natl Acad Sci U S A* **109**, 10438-10443 (2012).
148. E. S. Vincow *et al.*, The PINK1-Parkin pathway promotes both mitophagy and selective respiratory chain turnover in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6400-6405 (2013).
149. C. L. Samstag *et al.*, Deleterious mitochondrial DNA point mutations are overrepresented in Drosophila expressing a proofreading-defective DNA polymerase γ. *PLOS Genetics* **14**, e1007805 (2018).
150. S. DiMauro, M. Hirano, Mitochondrial encephalomyopathies: an update. *Neuromuscul Disord* **15**, 276-286 (2005).
151. M. Pinto, C. T. Moraes, Mechanisms linking mtDNA damage and aging. *Free Radical Biology & Medicine* **85**, 250-258 (2015).

152. A. Bratic, N.-G. Larsson, The role of mitochondria in aging. *The Journal of Clinical Investigation* **123**, 951-957 (2013).
153. P. Tokarz, J. Blasiak, Role of mitochondria in carcinogenesis. *Acta Biochim Pol* **61**, 671-678 (2014).
154. D. C. Wallace, D. Chalkia, Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease. *Cold Spring Harb Perspect Biol* **5**, (2013).
155. A. Bratic *et al.*, Complementation between polymerase- and exonuclease-deficient mitochondrial DNA polymerase mutants in genomically engineered flies. *Nature Communications* **6**, 8808 (2015).
156. F. Foury, S. Vanderstraeten, Yeast mitochondrial DNA mutators with deficient proofreading exonucleolytic activity. *EMBO J* **11**, 2717-2726 (1992).
157. S. Vanderstraeten, S. Van den Brûle, J. Hu, F. Foury, The role of 3'-5' exonucleolytic proofreading and mismatch repair in yeast mitochondrial DNA error avoidance. *J Biol Chem* **273**, 23690-23697 (1998).
158. G. C. Kujoth *et al.*, Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* **309**, 481-484 (2005).
159. F. Martinez-Azorin *et al.*, Over-expression of the catalytic core of mitochondrial DNA (mtDNA) polymerase in the nervous system of Drosophila melanogaster reduces median life span by inducing mtDNA depletion. *J Neurochem* **105**, 165-176 (2008).
160. H. L. Baines *et al.*, Similar patterns of clonally expanded somatic mtDNA mutations in the colon of heterozygous mtDNA mutator mice and ageing humans. *Mechanisms of Ageing and Development* **139**, 22-30 (2014).
161. K. Szczepanowska, A. Trifunovic, Different faces of mitochondrial DNA mutators. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1847**, 1362-1372 (2015).
162. B. J. Ryan, S. Hoek, E. A. Fon, R. Wade-Martins, Mitochondrial dysfunction and mitophagy in Parkinson's: from familial to sporadic disease. *Trends Biochem Sci* **40**, 200-210 (2015).
163. D. Haddad, K. Nakamura, Understanding the susceptibility of dopamine neurons to mitochondrial stressors in Parkinson's disease. *FEBS Lett* **589**, 3702-3713 (2015).
164. A. H. Schapira, Mitochondrial pathology in Parkinson's disease. *Mt Sinai J Med* **78**, 872-881 (2011).
165. Y. Luo, A. Hoffer, B. Hoffer, X. Qi, Mitochondria: A Therapeutic Target for Parkinson's Disease? *Int J Mol Sci* **16**, 20704-20730 (2015).
166. R. W. Gilkerson *et al.*, Mitochondrial autophagy in cells with mtDNA mutations results from synergistic loss of transmembrane potential and mTORC1 inhibition. *Human Molecular Genetics* **21**, 978-990 (2012).
167. S. R. Yoshii, N. Mizushima, Autophagy machinery in the context of mammalian mitophagy. *Biochim Biophys Acta* **1853**, 2797-2801 (2015).
168. C. P. Lin, B. N. Danforth, How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Mol Phylogenet Evol* **30**, 686-702 (2004).
169. C. Haag-Liautard *et al.*, Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. *PLoS Biol* **6**, e204 (2008).
170. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110-121 (2010).

171. B. Li *et al.*, Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* **25**, 2744-2750 (2009).
172. A. Martín-Navarro *et al.*, Machine learning classifier for identification of damaging missense mutations exclusive to human mitochondrial DNA-encoded polypeptides. *BMC Bioinformatics* **18**, 158 (2017).
173. L. Pereira, P. Soares, V. Maximo, D. C. Samuels, Somatic mitochondrial DNA mutations in cancer escape purifying selection and high pathogenicity mutations lead to the oncocytic phenotype: pathogenicity analysis of reported somatic mtDNA mutations in tumors. *BMC Cancer* **12**, 53 (2012).
174. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer. *Nature* **500**, 415-421 (2013).
175. T. E. S. Kauppila *et al.*, Mutations of mitochondrial DNA are not major contributors to aging of fruit flies. *Proceedings of the National Academy of Sciences* **115**, E9620-E9629 (2018).
176. M. Vermulst *et al.*, DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. *Nature Genetics* **40**, 392-394 (2008).
177. T. Y. Kim *et al.*, Metabolic labeling reveals proteome dynamics of mouse mitochondria. *Mol Cell Proteomics* **11**, 1586-1594 (2012).
178. J. C. Price, S. Guan, A. Burlingame, S. B. Prusiner, S. Ghaemmaghami, Analysis of proteome dynamics in the mouse brain. *Proc Natl Acad Sci U S A* **107**, 14508-14513 (2010).
179. E. C. mod *et al.*, Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**, 1787-1797 (2010).
180. V. R. Chintapalli, J. Wang, J. A. Dow, Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat Genet* **39**, 715-720 (2007).
181. A. S. Bess, T. L. Crocker, I. T. Ryde, J. N. Meyer, Mitochondrial dynamics and autophagy aid in removal of persistent mitochondrial DNA damage in Caenorhabditis elegans. *Nucleic Acids Res* **40**, 7916-7931 (2012).
182. S. Haroon *et al.*, Multiple Molecular Mechanisms Rescue mtDNA Disease in C. elegans. *Cell Reports* **22**, 3115-3125 (2018).
183. K. Buchet, C. Godinot, Functional F1-ATPase essential in maintaining growth and membrane potential of human mitochondrial DNA-depleted rho degrees cells. *J Biol Chem* **273**, 22983-22989 (1998).
184. D. P. Narendra *et al.*, PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol* **8**, e1000298 (2010).
185. D. Tondera *et al.*, SLP-2 is required for stress-induced mitochondrial hyperfusion. *EMBO J* **28**, 1589-1600 (2009).
186. D. C. Samuels *et al.*, Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans. *PLOS Genetics* **9**, e1003929 (2013).
187. A. D. de Grey, A proposed refinement of the mitochondrial free radical theory of aging. *Bioessays* **19**, 161-166 (1997).
188. K. J. Venken, Y. He, R. A. Hoskins, H. J. Bellen, P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science* **314**, 1747-1751 (2006).

189. C. C. Chan, S. Scoggin, P. R. Hiesinger, M. Buszczak, Combining recombineering and ends-out homologous recombination to systematically characterize Drosophila gene families: Rab GTPases as a case study. *Commun Integr Biol* **5**, 179-183 (2012).
190. J. W. Gargano, I. Martin, P. Bhandari, M. S. Grotewiel, Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in Drosophila. *Exp Gerontol* **40**, 386-395 (2005).
191. C. D. Nichols, J. Becnel, U. B. Pandey, Methods to assay Drosophila behavior. *J Vis Exp*, (2012).
192. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).
193. D. M. Humphrey *et al.*, Alternative oxidase rescues mitochondria-mediated dopaminergic cell loss in Drosophila. *Hum Mol Genet* **21**, 2698-2712 (2012).
194. B. M. Zid *et al.*, 4E-BP extends lifespan upon dietary restriction by enhancing mitochondrial activity in Drosophila. *Cell* **139**, 149-160 (2009).
195. G. Pareek, R. E. Thomas, L. J. Pallanck, Loss of the Drosophila m-AAA mitochondrial protease paraplegin results in mitochondrial dysfunction, shortened lifespan, and neuronal and muscular degeneration. *Cell Death Dis* **9**, 304 (2018).
196. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
197. A. Smit, R. Hubley, P. Green. (2013-2015).
198. P. J. Cock *et al.*, Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).
199. E. H. Ahn *et al.*, Detection of Ultra-Rare Mitochondrial Mutations in Breast Stem Cells by Duplex Sequencing. *PLOS ONE* **10**, e0136216 (2015).
200. S. Andreazza *et al.*, Mitochondrially-targeted APOBEC1 is a potent mtDNA mutator affecting mitochondrial function and organismal fitness in Drosophila. *Nature Communications* **10**, 3280 (2019).
201. R. W. Taylor, D. M. Turnbull, Mitochondrial DNA mutations in human disease. *Nat Rev Genet* **6**, 389-402 (2005).
202. H. A. Tuppen, E. L. Blakely, D. M. Turnbull, R. W. Taylor, Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta* **1797**, 113-128 (2010).
203. L. Kazak, A. Reyes, I. J. Holt, Minimizing the damage: repair pathways keep mitochondrial DNA intact. *Nat Rev Mol Cell Biol* **13**, 659-671 (2012).
204. S. R. Kennedy, L. A. Loeb, A. J. Herr, Somatic mutations in aging, cancer and neurodegeneration. *Mechanisms of Ageing and Development* **133**, 118-126 (2012).
205. D. Edgar, N. G. Larsson, A. Trifunovic, Point mutations are causing progeroid phenotypes in the mtDNA mutator mouse. *Cell Metab* **11**, 1 (2010).
206. S. G. Conticello, M. A. Langlois, M. S. Neuberger, Insights into DNA deaminases. *Nat Struct Mol Biol* **14**, 7-9 (2007).
207. A. Koito, T. Ikeda, Apolipoprotein B mRNA-editing, catalytic polypeptide cytidine deaminases and retroviral restriction. *Wiley Interdiscip Rev RNA* **3**, 529-541 (2012).
208. J. D. Salter, R. P. Bennett, H. C. Smith, The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* **41**, 578-594 (2016).
209. S. K. Petersen-Mahrt, M. S. Neuberger, In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J Biol Chem* **278**, 19583-19586 (2003).

210. R. S. Harris, S. K. Petersen-Mahrt, M. S. Neuberger, RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-1253 (2002).
211. L. Yang *et al.*, Engineering and optimising deaminase fusions for genome editing. *Nat Commun* **7**, 13330 (2016).
212. A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, D. R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
213. J. L. Rubenstein, D. Brutlag, D. A. Clayton, The mitochondrial DNA of Drosophila melanogaster exists in two distinct and stable superhelical forms. *Cell* **12**, 471-482 (1977).
214. N. Navaratnam *et al.*, Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. *J Mol Biol* **275**, 695-714 (1998).
215. A. G. Lada *et al.*, Disruption of Transcriptional Coactivator Sub1 Leads to Genome-Wide Re-distribution of Clustered Mutations Induced by APOBEC in Active Yeast Genes. *PLoS Genet* **11**, e1005217 (2015).
216. B. Iyengar, J. Roote, A. R. Campos, The tamas gene, identified as a mutation that disrupts larval behavior in Drosophila melanogaster, codes for the mitochondrial DNA polymerase catalytic subunit (DNApol-gamma125). *Genetics* **153**, 1809-1824 (1999).
217. K. L. Montooth, D. M. Rand, The spectrum of mitochondrial mutation differs across species. *PLoS Biol* **6**, e213 (2008).
218. A. M. Celotto *et al.*, Mitochondrial encephalomyopathy in Drosophila. *J Neurosci* **26**, 810-820 (2006).
219. H. Xu, S. Z. DeLuca, P. H. O'Farrell, Manipulating the Metazoan Mitochondrial Genome with Targeted Restriction Enzymes. *Science* **321**, 575-577 (2008).
220. J. L. Burman *et al.*, A Drosophila model of mitochondrial disease caused by a complex I mutation that uncouples proton pumping from electron transfer. *Dis Model Mech* **7**, 1165-1174 (2014).
221. J. G. Hoekstra, M. J. Hipp, T. J. Montine, S. R. Kennedy, Mitochondrial DNA mutations increase in early stage Alzheimer disease and are inconsistent with oxidative damage. *Ann Neurol* **80**, 301-306 (2016).
222. M. T. Lin *et al.*, Somatic mitochondrial DNA mutations in early Parkinson and incidental Lewy body disease. *Ann Neurol* **71**, 850-854 (2012).
223. P. A. Gammage, J. Rorbach, A. I. Vincent, E. J. Rebar, M. Minczuk, Mitochondrially targeted ZFNs for selective degradation of pathogenic mitochondrial genomes bearing large-scale deletions or point mutations. *EMBO Mol Med* **6**, 458-466 (2014).
224. A. Rana, M. Rera, D. W. Walker, Parkin overexpression during aging reduces proteotoxicity, alters mitochondrial dynamics, and extends lifespan. *Proceedings of the National Academy of Sciences* **110**, 8638-8643 (2013).
225. L. S. Ludwig *et al.*, Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* **176**, 1325-1339.e1322 (2019).
226. J. Xu *et al.*, Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *Elife* **8**, e45105 (2019).
227. J. Morris *et al.*, Pervasive within-Mitochondrion Single-Nucleotide Variant Heteroplasmy as Revealed by Single-Mitochondrion Sequencing. *Cell Reports* **21**, 2706-2713 (2017).

VITA

Colby Luke Samstag was born and raised in Pittsburgh, Pennsylvania, where his childhood best friend was a Can of Beans. He graduated *cum laude* from the University of Pennsylvania in 2013, with a Bachelor of Arts degree in Cognitive Science. He joined the Molecular and Cellular Biology Graduate Program at the University of Washington in 2014. He earned his Doctor of Philosophy degree under the mentorship of Dr. Leo J. Pallanck in 2019.