Title

Characterizing the mitochondrial mutational landscape throughout aging

Author

Serrano, Isabel

Publication Date

2024

Peer reviewed|Thesis/dissertation

Characterizing the mitochondrial mutational landscape throughout aging

by

Isabel Serrano

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computational Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Peter H. Sudmant, Chair
Professor Noah Whiteman
Professor Priya Moorjani

Summer 2024

Characterizing the mitochondrial mutational landscape throughout aging

Abstract

Characterizing the mitochondrial mutational landscape throughout aging

by

Isabel Serrano

Doctor of Philosophy in Computational Biology

University of California, Berkeley

Professor Peter H. Sudmant, Chair

Mitochondrial DNA (mtDNA) is a high copy number genome, creating a population of mtDNA molecules within a cell. As healthy cells age, mitochondrial genomes (mt-genomes) accumulate mutations that vary in frequency. Until recent technological advancements, the characterization of the mt-genome mutational landscape has been limited to high frequency variants. With the advent of ultra-sensitive sequencing technologies, surveying low frequency mutations is now possible, improving the investigation of mutational processes. In order to design efficient medical therapies for mitochondrial dysfunction, an understanding of the factors that impact mtDNA mutation is necessary. Here, I characterize the mtDNA mutational landscape across different mitochondrial haplotypes and tissues to discern the molecular and evolutionary processes underlying the origin and trajectory of mutation in mtDNA.

To begin, I investigate how mitochondrial haplotype shapes mutation in mtDNA. Importantly, over one thousand components needed for mitochondrial function are encoded on the nuclear genome; thus, harmony between these two genomes is vital. With this in mind, this work additionally studies how mito-nuclear ancestral mismatching impacts somatic mutation. To address these questions, I employ a panel of four mouse strains that are identical in their nuclear genomes, but differ in their mitochondrial haplotypes. I use ultra-sensitive Duplex Sequencing to generate maps of mutations with an unprecedented level of depth and accuracy across multiple tissues and mitochondrial haplotypes. From these maps, I confirm that the increase in mutation frequency with age is a constant molecular phenotype. Interestingly, comparison of rodent and primate mtDNA mutational spectra provide evidence of species-specific mutational signatures likely associated with distinct life history traits. My findings contrast the established notion that the non-coding region in the mt-genome has the highest mutation frequency. For example, I identify that the Origin of Replication (light strand) consistently has a higher mutation frequency across both tissues and mitochondrial haplotypes. Moreover, I highlight a mutational hotspot in $MT\text{-}tRNA^{Arg}$ specifically in mice

with mismatching mito-nuclear ancestry. Lastly, my work is the first to identify cases in which there is a preference for mutations that align mito-nuclear ancestry within the organism's lifespan.

Secondly, I explore intra-individual variation by characterizing the tissue-specific mutational landscapes of two regions in the aging mouse brain. Employing Duplex Sequencing, I profile the cortex and cerebellum at three different time points in the mouse lifespan. These regions were chosen due to differences in their metabolic demand and distinct accumulation of deletions throughout aging. I identify mutations present in both the cortex and cerebellum, and note that a consistent feature between tissues is the inheritance of insertions and deletions in regions associated with mtDNA replication. Although the tissues have similar frequencies for shared mutations in young mice, these frequencies diverge with age likely as a result of genetic drift. Overall, the cortex has a higher average mutation frequency than the cerebellum, but does not differ in mutation rate. Examination of different mutation types shows that the cortex and cerebellum differ in mutational signatures associated with mtDNA replication and metabolic damage. Mutations associated with mtDNA replication error are more abundant in the cortex; yet, the cortex and cerebellum exhibit a similar mutation rate for this signature. Despite the cortex being more metabolically demanding, this tissue shows a decrease in metabolic damage with age, opposite the trend in the cerebellum, indicating that the tissues potentially differ in their regulation of metabolic damage.

Together, these findings identify robust properties of mtDNA, while pinpointing specific differences driven by tissue-type and mitochondrial haplotype for future considerations.

To Guadalupe Viramontes, Jesus Viramontes, and Maria Viramontes.

You are my fire, heart, and soul. You are my pillars. I would not be who I am without you.

Contents

List of Figures

List of Tables

Acknowledgments

I have always believed that with a strong community supporting me, *anything* is possible. I am so grateful for the wonderful people who have shaped me, reminded me of my strengths, picked me up, and cheered me on.

First and foremost, thank you to my family. Thank you for always supporting me. Knowing that you are always cheering me on is what gets me through the toughest times. Whenever I come home to you, I remember who I am and who I always want to be. I want to reflect what you have taught me: to be welcoming, to love loudly, and to always show up.

To my Abuelo and Abuela: even when the world seems dim you show me hope and love. You stand as my piece of evidence that change is possible, it might just take a lot of work. You have taught me the power that bringing people together and making people feel seen has. I always try to bring a piece of you to the communities I am in.

To my mom: you are my warrior queen and partner in life. The times you saved to let me travel and the effort you put to take me on college tours were all moments that taught me how to dream big. You have given me everything so that I could pursue whatever dream I could conjure up. You have been my rock reminding me to live, to take in the moment, and to not lose myself to my work. I love you, I am so thankful for all the work you have put in to get me here – your sacrifices are not lost on me. This piece of work is ours.

To my cousin, Daniel: you were my sunshine and my light. Contributing even the slightest to understanding this complex molecule has been an honor of a lifetime.

To Robert: thank you for caring for me and my mom. Thank you for always showing up when I needed you and reminding me that I was capable of all the tasks I was juggling. With you in our lives, I can take a breath and feel at ease.

To my partner, Jared: thank you for being my partner in its truest sense. Thank you for supporting my wins and for always encouraging me to pursue my dreams – even if at times I am a little scared by them.

To the wonderful friends: Tanner, Evan, Alyssa, Erin, Rosey, and Asem thank you for helping me acclimate to a new city. You made Berkeley home and helped me believe in myself and my ability to survive in science. Joseph, thank you for making our home a safe space and reminding me of my values.

To my CompBio community, both our alumni and current members: I could not imagine a better environment or group of peers to train alongside. We remind each other that we are human and mistakes happen. That we are in fact capable even when we feel like we are not. You have created a space where we can grow and where we encourage each other to have some fun, too. The people in this group make me so hopeful about the future of science. Thank you for the memories, the friendship, and the encouragement.

To Ally, Dana, Shaina, Katia, and Erick: you are my lifelong friends. Thank you for always answering the phone, laughing at my misfortunes with me, helping me form a plan when I felt overwhelmed, and for always unconditionally loving me.

To Rohit and Manju: I owe so much to both of you. You were the two welcoming, warm people who built the community in our labs. You showed me what a community in science

could look like. You bring energy and warmth to the people around you, and oftentimes a quick chai time with you was all I needed to feel grounded again. Thank you for building our birthday traditions, for always having your door open, and for your unconditional love and support. When my mom met you she said that she felt comforted knowing I had good people around me – I knew then that Berkeley would never feel lonely with you in my life.

To the Sudmant Lab: I am so proud to be your academic sibling. I have had the fortune of witnessing how you approach projects, breakdown a problem, and carve out a research program that connects to your passions. Working alongside you has been inspiring. Thank you for filling our labspace with warmth, kindness, and laughs. One of the beautiful aspects of our group is that we will always rally for each other when we meet some lows – you have done this multiple times for me and I am forever grateful for your support. You bring humanity to science. Thank you for creating a lab culture where we succeed together.

To Hannah: A special thanks to you for trusting me to be your mentor. Thank you for working with me on my third chapter and for the constant enthusiasm you brought to our project. I am so excited to see the brilliant scientist you will become.

To Kate and Xuan: you both are pivotal in keeping our program together and functional, and in creating events that bring our community together. Thank you for supporting my ideas and helping me bring them to life. Thank you for helping me navigate the complexities of graduate school.

To my thesis committee, Noah and Priya: thank you for your guidance and encouragement. You created an example of who I hope to be.

To the mentors I have gained along the way: Max, Rasmus, and John thank you for your kindness and support. You have helped me reach goals I never thought were possible. Di and Chris thank you for welcoming me into the NERDS family.

To my mentors at Fullerton: Verdugo, Suceava, Jaynes, Behseta, Cohen, Arce, Cuajungco, Brubaker, I can never thank you enough for training me, believing in me, and still being in my corner. Oh my god, I did it!

Lastly, to Peter: I will try to convey how honored I feel to be your graduate student. The last six years have been a whirlwind. Yet, despite all of the hurdles, you have accomplished so much all while maintaining compassion and excitement. You created an environment where learning is the priority. You encourage us to take risks and pursue questions that excite us – believing in us and helping us overcome any steep learning curve. If I found that I made a mistake, I would be nervous and sad to tell you, but I would leave your office feeling empowered because you never made me feel bad for the error. Rather, you reminded me each time that this is how science works and that being knowledgeable enough to find the mistake was the success. Thank you for creating this safe space and for helping me find the language to explain what my passion is. There were multiple times where my soul was beaten down or my heart was broken, either from science or the multiple life events you experienced with me. Thank you for always showing me compassion, for taking me on a coffee walk to help me clear my head, and for believing in me even when I felt like I had nothing left to give. You always reminded me of my strengths and you reminded me that I was never alone – that you would be there to work through whatever issues arose. Having

you in my corner made anything feel possible. I do not know how else to phrase this other than to say I believe so much in you as a scientist and a person. Our community has been so fortunate to have you in it and I am so grateful that I was able to play a small role in furthering your research program of what I know will be your incredible legacy. I will forever be in your corner.

Chapter 1

Mutagenesis and evolution of mitochondrial DNA

1.1 Overview

Mutations are changes to DNA that result in new alleles, providing the resource necessary for evolution and setting the stage for genetic variation. Mutations are classified as either germline or somatic. Germline mutations are those that occur in gametes and early stages of development, and are propagated to the next generation. As a result, these mutations are present in every cell of an organism and can be inherited across generations. By contrast, somatic mutations are those that occur in non-reproductive cells and arise after conception. Somatic mutations are present in a subset of cells derived from the same lineage from which the mutation first arose. Thus, germline mutations result in population and species-level genetic diversity, while somatic mutations lead to intra-individual variation (reviewed in [109]).

As healthy cells age, they accumulate somatic mutations as a result of environmental exposures (e.g. smoking or UV exposure) and cellular processes (e.g. DNA replication or metabolic processes) throughout development and aging. While somatic mutations have historically been studied in the context of cancer, technological advancements have allowed for the characterization and investigation of these mutations in normal tissues and cells [67, 1, 12, 16], and other disease contexts [69, 63]. Somatic mutations are hypothesized to play a role in both normal aging processes [64] and age-related disease (reviewed in [50]), though the magnitude of their contribution and their exact role remains unclear. This dissertation specifically focuses on mutation and the evolution of a small, but mighty genome: mitochondrial DNA.

1.2 The mitochondrial genome: the powerhouse of mutation

Mitochondria are essential for cellular energy production. Mitochondria contain their own double-stranded, circular genomes (mtDNA or mt-genomes). Mitochondrial DNA encodes information for 13 vital proteins that comprise several components of the Electron Transport Chain (ETC), two ribosomal RNA units, and 22 tRNA units. Additionally, mtDNA contains two non-coding regions where the origins of replication for the heavy strand (OriH) and the light strand (OriL) are located. The D-Loop is a long non-coding region, which contains the OriH. The mt-genome likely replicates independently of the nuclear genome and cell cycle (reviewed in [55]). While each cell contains two copies of the nuclear genome, there are hundreds to thousands of mtDNA copies in a cell. With respect to mutation, mtDNA has a 10- to 100-fold higher de novo germline mutation rate than nuclear DNA [82, 49, 103] due to its lack of protective histones, a higher replication rate, and less effective DNA damage repair mechanisms [61]. Thus, mt-genomes create a dynamic subcellular population that is both riddled with and susceptible to mutation throughout aging.

Mitochondrial haplotypes can co-exist in an individual, meaning that multiple alleles may be present in the population of mtDNA molecules. Homoplasmic mutations are those that are present on all copies of mtDNA, whereas heteroplasmic mutations exist on a subset of mtDNA molecules. Mutation frequencies can be used to intuit when and how mutations arose. Variants present at higher frequencies or in multiple tissues in an individual are likely inherited or are somatic mutations that occurred early in development. Mutations present in two tissues are either inherited or are independent somatic mutation events. Lastly, mutations present in one tissue or at low frequencies (e.g. in one mtDNA copy) are likely de novo somatic mutations that arose during aging (classifications according to [60, 3]). Tracking the exact occurrence of a mutation or detangling inherited from early developmental somatic mutations is a challenging feat.

Mutations in mtDNA must surpass a threshold frequency to have a phenotypic effect. This frequency is projected to be between 60% - 90% [83, 54]. Yet, this threshold and a mutation's impact depends on where it is in the mt-genome, what tissue it is found in, and when the mutation occurred [85]. In healthy tissues, deletions [17, 19] and point mutations [51, 4, 3, 86] increase in frequency with age. Association studies have linked mutations in mtDNA with a wide range of diseases, including: neurodegenerative, cardiovascular, psychiatric, and metabolic diseases (reviewed in [92]). In age-related diseases, mtDNA mutations exist at exacerbated frequencies, implying that they may play a role in the manifestation of these diseases (reviewed in [92, 85]). These properties of the mt-genome create a complex system to study the origin and trajectory of mutations, where the evolution of the mt-genome can occur at various scales: from population-level to tissue-specific differences.

1.3 Evolution of mtDNA in the germline

Genetic drift and selection shape the germline mitochondrial mutational landscape

Mitochondrial DNA is primarily maternally inherited. Mutational and evolutionary processes acting in oocytes define the initial mtDNA mutational landscape for the subsequent generation. Allele frequencies are altered by either genetic drift or selection. Genetic drift refers to the change in allele frequencies by chance, resulting from the randomness of allele segregation. By contrast, selection demonstrates a preference for specific alleles. A population's demographic history impacts the strength of both selection and genetic drift. A key demographic event in the population of mtDNA is the mitochondrial genetic bottleneck, which refers to the random subsetting of mtDNA during oocyte production (reviewed in [112]). This subsetting results in a drastic decrease in mtDNA genetic diversity and mtDNA copy number [22, 33], followed by the replication of the remaining mtDNA copies. This combination of events redefines the frequencies of heteroplasmic variants in individual oocytes. The mitochondrial genetic bottleneck is a consistent event across species (reviewed in [112]), though the strength of the bottleneck varies. The bottleneck's magnitude has been measured in mice [4], humans [111], and macaques [3], with mice exhibiting the weakest and macaques having the most restrictive bottleneck. As a result of this bottleneck, some heteroplasmic variants can rapidly increase in frequency, while others may be lost from the population. This process provides the opportunity for deleterious mtDNA mutations to increase in frequency by chance. Consequently, pathogenic mutations can surpass the frequency threshold needed to cause mitochondrial dysfunction and result in mitochondrial disease [92]. This bottleneck acts independently in oocytes, meaning that allele frequencies can differ between siblings [111]. Thus, although deleterious mutations can rise in frequency in one sibling, they may be present at lower frequencies in other siblings, leading to the variable severity and occurrence of mitochondrial diseases within families. Overall, the mtDNA bottleneck serves as a reshuffling of mutation frequencies across generations.

Mutagenesis and mtDNA turnover in oocytes throughout aging continues to mold the landscape initially set by the genetic bottleneck [82, 111]. Studies in both mice and macaques have directly observed an increase in de novo mutation frequencies in oocytes with age [4, 3]. This maternal age effect on mtDNA is reflected by an age-related increase in genetic divergence between mothers and their offspring [111]. Mutations associated with mtDNA replication are the predominant mutation type in oocytes [4, 3]. Given that oocytes are in mitotic arrest, this observation supports the notion that mtDNA is being consistently produced independent of the cell cycle.

To date, there has not been evidence of selection acting on de novo germline mutations. On the contrary, de novo germline mutations accumulate uniformly across mtDNA and do not show a significant signal of selection [4, 3]. *Arbeithuber et al.* posit that this may be due to their low frequency and therefore low probability of manifesting as a phenotype. Unlike de novo mutations, purifying selection is acting on higher frequency heteroplasmic variants that

are likely to be transmitted to offspring. Interestingly, the number of heteroplasmic variants that increase in frequency across a generation is equal to that which decrease in frequency, reflecting the random distribution of mitochondria during cell division [103, 3]. While heteroplasmic frequencies remain constant for pathogenic mutations across generations [103], they exist at lower frequencies than non-pathogenic mutations [111]. Likewise, nonsynonymous mutations exist at lower frequencies than synonymous mutations [111], making the case that the frequency of protein-coding and potentially harmful mutations is restricted. Interestingly, allele and site-specificity for transmitted mutations has been noted in humans. Alleles already present in human populations are preferentially transmitted compared to new alleles [103]. Moreover, variants at haplotype sites are more likely to be transmitted than those at other sites [103]. Lastly, alleles present in the non-coding D-Loop are more likely to be transmitted to offspring [103], which is reflected by the D-Loop's initially higher mutation frequency in young somatic tissues [51, 4, 3]. Within the D-Loop, however, regions that impact mtDNA replication and transcription do not retain mutations, further demonstrating that selection is shaping the distribution of variants across the molecule [33, 103]. In summary, the lower frequencies of nonsynonymous and pathogenic mutations, transmission of mutations in non-coding regions, and allele-specific transmission show that purifying selection influences the population-level mtDNA allele pool.

Mitochondrial DNA diversity at the population level

Germline mutations that fix in an oocyte have the potential to contribute to the population-level pool of mitochondrial haplotypes. Due to a lack of recombination in mtDNA, each unique fixed mutation defines a new mitochondrial haplotype [92]. As such, mtDNA haplotypes have been pivotal in tracing modern human origins out of Africa and understanding human migration patterns across the world [48]. Mitochondrial DNA genetic divergence increases with geographic distance, with distinct mtDNA haplotypes dominating across continents (reviewed in [97]). Additionally, within human populations region-specific mtDNA haplotypes have risen in frequency (reviewed in [97]). This observation suggests that specific mtDNA haplotypes thrive in particular environments and/or climates (reviewed in [97]). While a wide array of mtDNA haplotypes exist in the human population, less than 5% of variants are common (allele frequency $> 5\%$) [107]. Moreover, mutations present at frequencies $>1\%$ are restricted to 2.4% of the 16.5-kb mtDNA sequence in humans, with most of these nucleotides residing in the non-coding D-Loop [102, 103]. Taken together, these observations suggest that selection plays a role in defining and distributing mitochondrial haplotypes in populations.

In both humans and mice, mitochondrial haplotypes confer molecular phenotypes. Cybrid lines (reviewed in [105]) and conplastic mouse strains [108] place mitochondrial haplotypes on a specific nuclear background. These tools control for nuclear genomic influences and provide insight into the functional impacts mitochondrial haplotypes have. Cybrid lines have been used to demonstrate that common human haplotypes differ in mtDNA copy number, ATP production, the levels of mtDNA and mtRNA, and the amount of transcription and

translation products [97, 35, 34]. In mice, conplastic strains have shown that mitochondrial haplotypes that diverge by as few as 1-3 variants can differ in their: levels of mtDNA copy number [43], reactive oxygen species production, ATP production, cell proliferation activity [88], mitochondrial morphology, extent of glycolysis, microbiome composition, and behavior [108, 44, 90, 104]. Additionally, mitochondrial haplotypes can modulate the functional impact of other mutations. For example, the T>C mutation in *MT-ND1*, which is associated with Leber hereditary optic neuropathy, confers an increase in Complex I activity of the ETC on a specific mitochondrial haplotype (group M), while leading to a decrease in activity on other haplotypes [47]. While this study was pivotal in showing how mitochondrial haplotype impacts the effect of polymorphic alleles, it emphasizes that both mitochondrial haplotype and environmental factors determine whether these alleles are detrimental or beneficial. Thus, mitochondrial haplotypes can influence various aspects of the mitochondrial ecosystem.

Mito-nuclear interactions shape mtDNA dynamics

While the mt-genome retains information for thirteen vital proteins, over 1,000 genes necessary for mitochondrial function are encoded on the nuclear genome (reviewed in [41]). These include proteins necessary for mtDNA replication, transcription, translation, and additional elements of the ETC beyond those encoded on mtDNA [95]. This tight interplay between genomic products necessitates the co-evolution of the nuclear and mitochondrial genome [77, 6, 101].

Mito-nuclear ancestral discordance refers to the ancestral mismatching of the mitochondrial and nuclear genomes. Due to the maternal inheritance of mtDNA, mito-nuclear discordance can occur in hybrid species or admixed populations. Mito-nuclear incompatibilities, by extension, result when mito-nuclear discordance leads to a reduction in fitness. Studies across a variety of hybrid species have demonstrated the existence of mito-nuclear incompatibilities, providing evidence supporting the coevolution of these two genomes. Mito-nuclear incompatibilities have been shown in both laboratory crosses and natural populations of: fruit flies [68, 70], wasps [28, 75, 53], yeast [57, 15], marine copepods [29, 31, 30], swordtail fish [72], and teleost fish [5]. These mito-nuclear incompatibilities have been linked to changes in mtDNA copy number [27], alterations in mtDNA gene expression [27, 30], and reduced activity in the ETC complexes composed of interacting components from both the nuclear and mitochondrial genomes [28, 68]. These mito-nuclear incompatibilities have impacted embryonic development [72], led to male infertility [53], decreased fecundity [68], increased mortality rates [53], and resulted in developmental delays [68]. Pinpointing the exact genetic causes of mito-nuclear incompatibilities is a challenging feat, but recent work in swordtail fish identified one such case in genes interacting in Complex I of the ETC [72]. Providing further support that the reductions in fitness observed are due to mito-nuclear interactions, backcrosses in marine copepods that recover mito-nuclear ancestral matching have managed to restore fitness [29]. To emphasize, not all instances of mito-nuclear discordance result in mito-nuclear incompatibilities. Yet, these examples highlight the importance of maintaining harmony between the nuclear and mitochondrial genomes.

Mito-nuclear interactions have shaped features of both mitochondrial and nuclear DNA. In eastern yellow robins, where populations have distinct mitochondrial haplotypes, the hub of genetic divergence on the nuclear genome contains an overabundance of nuclear-encoded mitochondrial genes [71]. When classifying individuals based on outlier loci in this hub, the majority of individual assignments reflected the mitochondrial haplotype [71]. These findings suggest that mito-nuclear interactions shape the differentiation in the nuclear genome to align with divergence in the mitochondrial haplotype. In humans, individuals with mito-nuclear ancestral mismatching were more likely to transmit heteroplasmic variants that matched nuclear ancestry [103]. While the exact mechanism and timing of this selection for nuclear-matching heteroplasmic variants remains unclear, *Wei et al.* hypothesize that selection likely occurs in the germline based on the changes in allele frequencies that can occur as quickly as one generation. In addition to shaping the mutational landscape on both the nuclear and mitochondrial genomes, increased mito-nuclear discordance is correlated with a decrease in mtDNA copy number [110]. This correlation suggests that mito-nuclear discordance can impact mtDNA replication. A recent study provided additional evidence for this hypothesis [39]. By using large scale data from the UK Biobank and All of US consortia, changes in mtDNA copy number were mapped to nuclear-encoded genes involved in mtDNA replication [39]. Thus, the nuclear genome plays a role in regulating mtDNA copy number. Importantly, mtDNA copy number can modulate the impact and trajectories of mutations in mtDNA. Then, the decrease in mtDNA copy number with increased mito-nuclear discordance may potentially heighten the effects of heteroplasmic mutations and lead to variability in associated phenotypes. This large body of work has shown the genetic and molecular consequences of mito-nuclear interactions, emphasizing that when assessing the evolution of the mt-genome, the nuclear background should be taken into account as well.

1.4 Evolution of mtDNA in somatic tissues

Both genetic drift and selection have been shown to shape mutation frequencies in the germline. Mutations that are present on all copies of mtDNA, homoplasmic variants, create the mtDNA background for changes to occur throughout an individual's lifespan. This backdrop has been shown to influence the aggregation of new mutations [47]. Heteroplasmic variants can also be transmitted through the germline, setting the stage for intra-individual variation. Due to mitochondrial properties – such as consistent mtDNA replication, mitochondrial biogenesis, and cell division – allele frequencies have the potential to be reshuffled throughout aging. Furthermore, the creation of new mutations (de novo somatic mutations), introduces new alleles into the mt-genome population. As a result, the somatic mutational landscape is dynamic throughout aging.

Embryonic and somatic genetic bottlenecks

Heteroplasmic mutations exist in multiple tissues, with shared mutations being likely inherited. Yet, the frequencies of shared mutations can vary across tissues. When cells divide, mitochondria are distributed between the mother and daughter cells. Likewise when mitochondria are generated mtDNA is split between organelles. Thus, these processes randomly distribute mtDNA molecules, resulting in genetic drift shaping allele frequencies in somatic tissues. As in the germline, bottleneck events during embryonic and individual tissue development are hypothesized to alter the strength of drift, leading to tissue-specific frequencies. Using mtDNA from human hair, the bottleneck during embryonic development was estimated to be weaker than the one observed in the germline [7]. Additionally, allele frequencies for shared mutations were more similar between tissues that likely developed together than between tissues that are more distantly related, echoing the notion that potential subcellular demographic events create tissue-specific frequencies [111]. Contrasting these findings, a stronger bottleneck was observed in individual hair follicle development, having as strong a bottleneck as that observed in oocytes [7]. Evidence of a somatic bottleneck has also been observed in lymphocytes [94]. Both studies examine the depletion of mtDNA diversity and identify individual mechanisms acting at two different scales. With regards to genetic drift in hair, each hair structure develops from a small set of cells [7] restricting the pool of mtDNA alleles at the cellular level. In lymphocytes, however, mtDNA replication is slowed despite cell proliferation [94]. Thus, mtDNA diversity is restricted at the mtDNA level, constricting the pool of mtDNA alleles that daughter cells inherit. Both studies highlight the various mechanisms that lead to genetic bottlenecks and subsequently alter mtDNA genetic diversity within and between tissues.

For tissues that undergo similarly sized genetic bottlenecks, such as blood and buccal cells, the allele frequencies of these shared variants are almost identical even throughout aging [111]. Age-associated changes in genetic drift, however, were observed in human hair [6], macaque muscle, and macaque liver [3]. These conflicting results can be attributed to two factors. First, the studies conducted in human hair, blood, and buccal tissues sampled from the same individuals [7, 111]. Uniquely, hair develops from a small set of cells – unlike blood and buccal tissues – resulting in differences in genetic drift from the onset. The increase in strength of drift in hair, but not blood and buccal tissues, demonstrates that the developmental history of a tissue may play a role in the extent to which drift changes throughout aging. Secondly, the identification of drift in the liver and muscle, which have a more similar developmental history to buccal and blood tissues than to hair, may be due to the age differences in samples. The average age in the human study was 20 years [111], whereas the average age of macaques sampled was 23 years – with macaques having an average lifespan of 25 years [14]. Thus, the change in drift observed in macaques may reflect that these changes likely occur late in age. Overall, these studies support the existence of various somatic bottlenecks. Importantly, these works show how the developmental history of a tissue is intertwined with the mechanisms by which drift occurs and the strength that this drift has in altering allele frequencies throughout an individual's lifespan.

Tissue-specific aggregation of somatic mutations

Mutation frequencies in mtDNA can increase with age as a result of de novo somatic mutations or the replication of inherited mtDNA alleles (clonal expansion). Some of the first works that examined the increase in mutation frequencies with age focused on common deletions. These studies highlight the tissue-specific accumulation of deletions, suggesting that tissues' metabolic demands may influence the mutagenesis of mtDNA [17, 19]. With the advent of Duplex Sequencing (discussed further in Section 1.5), works have shown the constant mutagenesis of mtDNA throughout aging, and have begun to characterize the molecular processes that lead to mutation. Mutation frequencies in somatic tissues are consistently higher than in oocytes, and these frequencies increase throughout aging at a faster rate [4, 3]. In mice, young individuals have a mutation frequency on the order of 10^{-7}, while aged individuals have a frequency on the order of 10^{-6} [4]. In humans and macaques, the mutation frequencies range on the order of 10^{-6} [51, 3]. Young tissues showcase differences in initial mutation frequencies at an early age (<4.5 months in mice; <2.5 years in macaques). Closely related tissues (e.g. muscle and heart) are more alike in their frequencies than with tissues that have different physiologies (e.g. liver and kidney) [3, 86]. This observation likely reflects the timing and impact of embryonic and somatic bottlenecks. Additionally, tissues vary in their mutation rates, with highly proliferative tissues having a higher fold change in frequency (e.g. 3.5 fold change for liver) than post-mitotic tissues (e.g. 2.8 fold change for muscle) between young and aged samples [3]. This trend is consistent across a broad spectrum of tissues [86].

A long held hypothesis was that reactive oxygen species (ROS) damage was the primary culprit of mtDNA mutation and likely scaled with tissues' metabolic activities. Yet, characterization of the mtDNA mutational spectrum has shown otherwise. Across somatic tissues, the dominant mutation type is associated with mtDNA replication error (G>A/C>T and A>G/T>C), either resulting from DNA polymerase γ mistakes or deamination of cytosine during mtDNA replication [51, 4, 3, 86]. Mutations stemming from mtDNA replication error exhibit an age-associated increase in frequency [51, 4, 3, 86]. The extent to which this mutation increases throughout aging shows tissue-specific trends [3, 86]. Highly proliferative tissues have a higher abundance of these mutations as well as a higher fold change between young and aged tissues [3, 86]. This observation falls in line with the idea that tissues that undergo cell proliferation likely have continuous replication of mtDNA, assuming that cell proliferation triggers mtDNA replication. Recall, however, that mtDNA replication is not always stimulated by cell division, and the relationship between mtDNA replication and cell proliferation itself may be cell-specific [94]. Contrary to the ROS-damage hypothesis, the mutation types associated with ROS-damage, G>T/C>A and G>C/C>G, did not have an age-associated increase in frequency [86]. Interestingly, the young heart and muscle tissues had a $\sim$ 2-fold higher proportion of this mutation type compared to tissues in the Central Nervous System [86]. While not associated with aging, the tissue-specific initial burden and frequency in aged individuals relative to other mutation types brings into question whether mechanisms exist to explicitly regulate metabolic damage.

The distribution of mutations across mtDNA is consistent across tissues. In young tissues, the D-Loop has the highest mutation frequency in the mt-genome [51, 4, 3]. This trend reflects the preferred transmission of non-coding variants in the germline. In both mice and macaques, the D-Loop demonstrates an elevated mutation rate compared to other regions in the mt-genome [4, 3]. In the human brain, a substantial difference in mutation rate between regions in mtDNA is not observed however [51]. Additionally, D-Loop mutation rates are consistent across tissues [3]. Tissues did differ, however, in the accumulation of mutations in protein-coding genes, with the liver having a higher mutation rate in this region compared to muscle [3]. Thus, the muscle has a larger bias towards mutations in the D-Loop than the liver, potentially due to stronger drift in the liver. Lastly, the liver has an additional mutational hotspot in the Light Strand Origin of Replication, either as a consequence or mechanism to control mtDNA replication in this highly proliferative tissue [3]. By and large, tissues share large-scale trends, such as the overall mutational spectra and an elevated mutation rate in the D-Loop. Yet, there exist subtle tissue-specific distinctions, such as the degree of bias towards mutations in the D-Loop, and the prevalence of mtDNA replication associated errors and ROS-damage throughout aging.

Tissue-specific expansion of heteroplasmic variants

Like de novo somatic mutations, heteroplasmic mutation frequencies change in a tissue-specific manner throughout aging [46, 9, 86]. Within an individual, the dominance of a specific mitochondrial haplotype can vary between tissues [46]. These patterns are observed regardless of initial mitochondrial haplotype frequency, implying that drift cannot be the sole evolutionary process responsible for these trends. These mitochondrial haplotypes do not result in different ETC capabilities, advantages in mtDNA replication, or changes in mtDNA gene expression, hindering the identification of affected pathways [9]. Comparison of in vivo and in vitro mitochondrial haplotype frequencies show contrasting results, with one mitochondrial haplotype preferred in vivo and the other in vitro [9]. Again, this result held regardless of initial mitochondrial haplotype frequencies [9]. Importantly, the staunch difference between in vivo and in vitro studies is the level of cell proliferation, where in vitro cells are rapidly undergoing differentiation and division [9]. This difference, the lack of other changes in mtDNA dynamics and mitochondrial function, and the highly proliferative nature of the tissues studied imply that cell division may play a role in this phenomenon. Specifically, the impact cell division has on mitochondrial turnover and mtDNA replication may potentially lead to selection on mitochondrial haplotype driven by tissue physiology.

A recent study that used ultra-sensitive Duplex Sequencing to profile the mt-genome across multiple tissues surveyed the clonal expansion of heteroplasmic variants genomewide [86]. *Sanchez-Contreras et al.* found that (1) an increase in heteroplasmic frequency with age and (2) the biased aggregation of heteroplasmic variants in the D-Loop were consistent features across tissues [86]. Uniquely, the kidney and liver have higher heteroplasmic frequencies across the entire mtDNA molecule, indicating that stronger drift in these tissues leads to changes across the mt-genome or that selection is potentially acting in other tissues.

While in post-mitotic tissues the presence of heteroplasmic variants in protein-coding regions is sparse, multiple variants have as high, if not higher, frequencies than mutations in the D-Loop. Moreover, single nucleotide variants that occur in the liver and kidney are more likely to undergo clonal expansion than in other tissues [86]. Lastly, while the heart has an abundance of de novo ROS-damage mutations, there is a significant lack of ROS-damage mutations that clonally expand, suggesting the existence of a regulatory mechanism in the heart that keeps these mutations at low frequencies [86]. Thus, tissue-specific features exist both in the introduction and expansion of mutations in mtDNA throughout aging.

Selection shapes somatic mtDNA mutational landscapes

There has been increasing evidence that selection plays a role in shaping the somatic mtDNA mutational landscape. With respect to de novo mutations, signals of positive selection in macaque liver and skeletal muscle have been identified [3]. Positive selection on de novo mutations is hypothesized to occur as a mechanism for slowing mitochondrial function by effectively breaking down the ETC [3]. These signals, however, have not been recapitulated in mice [86, 4]. The lack of selection signals in mice is attributed to the minimal impact low frequency mutations can have; yet, selection can be acting on higher frequency mutations. The high presence of de novo ROS-damage mutations in the heart, but the inhibited expansion of these variants reflects this notion.

While it is complicated to disentangle developmental somatic mutations from low frequency inherited mutations, the tissue-specific preference for mitochondrial haplotype demonstrates that throughout aging selection shapes heteroplasmic frequencies [46, 9]. Tissue-specific heteroplasmic variants coupled with the allele-specificity of these variants [60] demonstrates that distinct tissue physiologies may create unique environments where a mutation can thrive. Additionally, positive selection has been detected in the human liver, acting on mutations present at a frequency $> 1\%$ [60]. Notably, $\frac{hN}{hS}$ ratios, which are akin to $\frac{dN}{dS}$ ratios and that measure the bias for protein-coding changes, are higher than $\frac{pN}{pS}$ ratios in humans [60]. The higher $\frac{hN}{hS}$ ratios demonstrate that selection may regulate the frequency of mutations present in somatic tissues, echoing the patterns seen for de novo somatic and germline mutations. More recently, a work using SCI-LITE, a sequencing method used to trace heteroplasmic frequencies at single cell resolution, has shown that the environment cells are in can lead to an increase in nonsynonymous allele frequencies [54]. Importantly, the cell's proliferative activity in the context of the cellular environment impacts allele frequencies. Seemingly, dynamic tissue physiologies create opportunities for positive selection. However, negative selection may still be at bay to prevent mutations from fixing in an individual.

1.5 Advancements in sequencing technologies are propelling studies of somatic mutation

Prior to the accessibility of next generation sequencing, analyses of somatic mutations in mtDNA were largely restricted to deletions. Using Polymerase Chain Reaction (PCR)-based methods, changes in frequencies for well-characterized deletions were studied across multiple tissues [17, 19]. These works established the age-associated and tissue-specific increase in deletions. Stepping away from PCR approaches, *Vermulst et al.* used a restriction enzyme approach to quantify the mtDNA mutation frequency throughout aging [96]. This work placed into question the role of mtDNA mutation in aging, showing that even a 500-fold higher mutational burden in mtDNA did not lead to aging phenotypes. With next generation sequencing, mtDNA could be sequenced at high coverage, providing the resolution necessary to identify heteroplasmic mutations in population-level data. Heteroplasmic variants with a frequency >1% were captured using deep resequencing of mtDNA [79]. This methodology allowed *Payne et al.* to show that low-frequency heteroplasmic mutations are common in healthy individuals [80]. Yet, deep resequencing of mtDNA is still restricted by sequencing technologies' error rates [79], which range between $\sim$ 1%-0.1% [51, 89]. Taking these error rates into account, mutations at a frequency $< 1\%$ cannot be confidently called and are typically filtered from studies [93, 82]. While these approaches allow for the study of mutations genome-wide, they cannot capture low frequency and de novo mutations, preventing a full understanding of mutagenesis in mtDNA. Recently, new sequencing technologies have been developed to address this sensitivity issue. Duplex sequencing exploits the complementary nature of the double-stranded DNA structure in order to eliminate technical errors [89]. Each duplex mtDNA fragment is tagged with a unique molecule identifier, which allows for the computational construction of duplex consensus sequences that root out first round PCR errors and sequencing artifacts. This process results in a theoretical error rate of $< 10^{-7}$ [89]. In practice, the error rate is estimated to be 2×10^{-7} errors per bp [1]. With the mtDNA mutation rate estimated to be 6×10^{-8} [66, 96], this sensitivity is able to capture de novo mutations. Duplex sequencing has been implemented to study somatic [51, 4, 3, 86] and germline mutations [4, 3] in mtDNA. These works have shown a lack of mutations associated with reactive oxygen species [51, 4, 3, 86], an age-related increase in replication associated mutations [51, 4, 3, 86], estimated a somatic mutation rate one to two orders of magnitude lower than previously observed [51], and showcased the continuous mutagenesis in oocytes throughout aging [4, 3].

Within the last few years, improvements have been made on the Duplex Sequencing protocol, allowing for an even higher sensitivity. Nanorate sequencing (Nanoseq), which changed the original Duplex Sequencing protocol to reduce false positives stemming from damage introduced during sonication and end repair, obtained an error rate of 3×10^{-9} [1]. This method provides the resolution necessary to study somatic mutations in postmitotic cells. Additionally, Hairpin Duplex Enhanced Fidelity sequencing (HiDEF-seq) utilizes Pacific Biosciences long read technologies to identify mutations made on a single strand of DNA

– before DNA replication or repair mechanisms can fix the mutation on both DNA strands [62]. HiDEF-seq ranks as the most sensitive sequencing technology with an error rate of 10^{-14} [62]. By profiling mutations present on a single strand, HiDEF-seq records information about mutational processes at their onset. Using HiDEF-seq to profile the mt-genome, *Liu et al.* demonstrate that the G>A/C>T mutational signature, usually associated with either DNA replication error or cytosine deamination, was in fact arising from the process of DNA replication directly [62]. Mutations in mtDNA have been challenging to study because of the molecule's complex characteristics. Yet, these sequencing technologies have created scalable opportunities to survey the mt-genome at fine resolution with the accuracy necessary to capture the entire spectrum of mutations. With these tools we are at the forefront of untangling and understanding the mutagenesis and mutational trajectories of this complicated genome.

1.6 Overview of Dissertation

Throughout this chapter, the various factors that impact mutation in the mt-genome have been delineated. In the second section, this dissertation explores how mitochondrial haplotype and mito-nuclear ancestral mismatching modulate somatic mutation in mtDNA. Duplex sequencing was employed to profile the mt-genomes in three tissues of four mouse strains that have identical nuclear genomes but differ in their mitochondrial haplotypes. This experimental design contributes the largest somatic mutation resource, and integrates mitochondrial haplotype and mito-nuclear ancestral mismatching in the efforts to understand mutational trajectories in mtDNA. The third section extends on the tissue-specific observations from the second section. Specifically, this section focuses on region-specific mutational differences in the aging brain. While mtDNA mutations have been associated with a wide-range of diseases, most mitochondrial diseases in adults result in neurological symptoms [65]. We characterize the mutational landscapes of the cortex and cerebellum, which have previously demonstrated distinct accumulation of deletions throughout aging [17]. Together, these sections aim to understand how mitochondrial haplotype, interactions between the two genomes, and tissue physiology impact the evolution of mtDNA throughout aging.

Chapter 2

Mitochondrial haplotype and mito-nuclear matching drive somatic mutation and selection throughout aging

This chapter is co-authored by Misa Hirose, Charles C. Valentine, Sharon Roesner, Elizabeth Schmidt, Gabriel Pratt, Lindsey Williams, Jesse Salk, Saleh Ibrahim, and Peter H. Sudmant. It has been published in Nature Ecology and Evolution, 15 February 2024
https://doi.org/10.1038/s41559-024-02338-3

2.1 Summary

Mitochondrial genomes co-evolve with the nuclear genome over evolutionary timescales and are shaped by selection in the female germline. Here, we investigate how mismatching between nuclear and mitochondrial ancestry impacts the somatic evolution of the mt-genome in different tissues throughout aging. We used ultra-sensitive Duplex Sequencing to profile ∼2.5 million mt-genomes across five mitochondrial haplotypes and three tissues in young and aged mice, cataloging ∼1.2 million mitochondrial somatic and ultra low frequency inherited mutations, of which 81,097 are unique. We identify haplotype-specific mutational patterns and several mutational hotspots, including at the Light Strand Origin of Replication, which consistently exhibits the highest mutation frequency. We show that rodents exhibit a distinct mitochondrial somatic mutational spectrum compared to primates with a surfeit of reactive oxygen species-associated G>T/C>A mutations, and that somatic mutations in protein coding genes exhibit signatures of negative selection. Lastly, we identify an extensive enrichment in somatic reversion mutations that "re-align" mito-nuclear ancestry within an organism's lifespan. Together, our findings demonstrate that mitochondrial genomes are a dynamically evolving subcellular population shaped by somatic mutation and selection

throughout organismal lifetimes.

2.2 Introduction

The mitochondrial genome (mt-genome) encodes for 13 proteins that are vital for the electron transport chain. However, more than 1,000 nuclear encoded genes are necessary for mitochondrial assembly and function [41]. The required coordination between nuclear- and mitochondrial-encoded proteins drives the coevolution of these two genomes [77, 6, 101]. This concerted evolution has been observed in both laboratory crosses and natural hybrids across a variety of species, including: fruit flies [70, 68], marine copepods [29, 31], wasps [75, 28, 53], yeast [57, 15], eastern yellow robins [71], swordtail fish [72], and teleost fish [5]. Hybrids often exhibit reduced fitness attributed to attenuated mtDNA copy number [27], mtDNA gene expression [27, 30], and OXPHOS function [29, 31, 28, 68]. In natural hybrids, the genetic ancestry at nuclear-encoded mitochondrial genes is highly differentiated between populations that have distinct mt-haplotypes [71]. In human admixed populations, mito-nuclear ancestral discordance has been shown to correlate with reductions in mtDNA copy number [110]. Additionally, de novo germline mitochondrial mutations in admixed human populations exhibit a bias towards concordance with nuclear ancestry [103]. In somatic tissues, the nuclear genome has been shown to influence the segregation of mtDNA haplotypes in a tissue-specific manner [46, 9, 8]. Together, these studies highlight the functional importance of mito-nuclear ancestral concordance. However, the impact of these genomic interactions on somatic mutation and selection has not been studied.

As healthy cells age, they accumulate nuclear and mtDNA damage as a result of environmental exposures and cellular processes [64]. mtDNA has a 10- to 100-fold higher de novo germline mutation rate than nuclear DNA [49, 82, 103] due to its lack of protective histones, a higher replication rate, and less effective DNA damage repair mechanisms [61]. Individual cells can contain hundreds to thousands of mt-genomes [11, 32], presenting a dynamic subcellular population that is highly susceptible to mutation. While the relationship between mutations in the mt-genome and aging remains unclear [85, 64], increased mtDNA damage has been associated with many aging phenotypes and several age-related diseases [98, 85]. Yet, profiling low frequency mutations in this population of mt-genomes has historically been challenging with most technologies limited to variants segregating at a frequency $> 1\%$ [89, 82]. More recently, several studies have employed Duplex Sequencing [89] to capture mutations with an error rate of $< 1 \times 10^{-7}$. These works have examined several different species and confirmed a robust age-associated increase in mtDNA somatic mutation frequency [51, 4, 3, 86]; identified that mitochondrial somatic mutations stem from the replication process either via DNA polymerase γ misincorporations or spontaneous DNA deamination of cytosine and adenine [51, 4, 3, 86]; and observed tissue-specific somatic mutation rates [3, 86].

Here, we investigate the impact of mitochondrial ancestry (mt-ancestry) and tissue-type on the somatic evolution of the mt-genome through age. To explore how mitochondrial hap-

lotype and mito-nuclear genomic interactions shape the mt-genome mutational landscape, we utilize a panel of mouse strains that have identical nuclear genomes but differ in their mitochondrial haplotypes (conplastic strains). We sample the brain, heart, and liver of these conplastic strains and the wildtype (C57BL/6J) in young and aged individuals. These tissues are physiologically distinct yet some of the most metabolically active, allowing us to study how evolutionary processes unfold in different molecular contexts. We use Duplex Sequencing to profile mt-genomes at an unprecedented level of depth and accuracy, allowing us to identify mutational hotspots and characterize the mutational spectrum. We use these results to discern signatures of selection acting to shape the mt-genome through age and confirm the existence of somatic mutations that work to realign mito-nuclear ancestry at short, within-lifespan, evolutionary timescales. Together, these findings characterize somatic evolution in the context of an organelle implicated in aging and age-related phenotypes.

2.3 Results

Duplex sequencing of 2.5 million mitochondrial genomes

To investigate how mitochondrial haplotype and mito-nuclear concordance influence the distribution of mt-genome somatic mutations, we employed a panel of conplastic mouse strains. These strains are inter-population hybrids developed by crossing common laboratory strains with the C57BL/6J (B6) mouse line [108]. Each conplastic mouse strain carries a unique mitochondrial haplotype on a C57BL/6J nuclear genomic background **(Fig. 2.1A)**. This fixed nuclear background enables us to attribute differences in somatic mutation to changes in mitochondrial haplotype. Alongside a wildtype B6 mouse, we used four conplastic mouse strains that exhibit changes in metabolic content and processes, or altered aging phenotypes **(Table A.1)**. Three of these conplastic mouse strains differ from the C57BL/6J mitochondrial haplotype by just 1-2 nonsynonymous variants, while a single strain (NZB) contains 91 variants distributed across the mt-genome **(Fig. 2.1B, Table A.1)**. We sampled brain, liver, and heart tissues from each of these mouse strains in young (2-4 months old (mo.)) and aged (15-22 mo.) individuals to examine how aging and metabolic demands shape the mutational spectrum in distinct physiological contexts. In total, three tissues were sampled from five mouse strains at two different ages over 3-4 replicates per condition, resulting in n=115 samples.

In order to capture low frequency variants and accurately portray the somatic mutational landscape, we used ultra-sensitive Duplex Sequencing to profile mt-genomes across different experimental conditions [89]. This approach works by tagging double stranded mtDNA sequences with molecular identifiers and computationally constructing a duplex consensus sequence **(Fig. 2.1C)** resulting in error rates of $\sim 2 \times 10^{-8}$. Importantly, duplex sequencing data can contain an artificial enrichment of G>T/C>T and G>C mutations resulting from DNA damage during sequencing preparation steps [1]. To exclude potential erroneous calls, we trimmed 10 bp from our duplex reads **(Supplementary Note A.1, Fig. A.1)**.

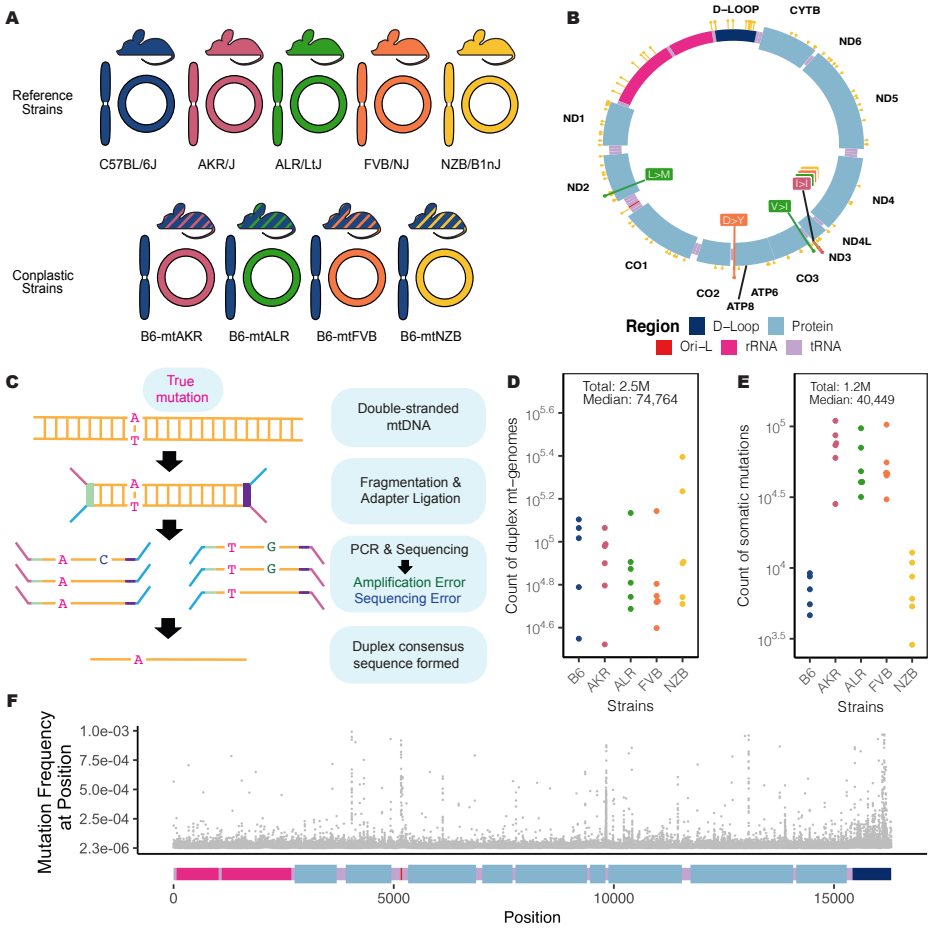


Figure 2.1: **Overview of experimental design**. (A) Maternal donors were backcrossed with a C57BL/6J mouse (blue). The result of these crossings are conplastic mouse strains (striped mice), which have identical B6 nuclear (linear) genetic backgrounds but differ by variants along their mt-genomes (circular). (B) B6-mtAKR (pink), B6-mtALR (green), and B6-mtFVB (orange) differ from wildtype by 1-3 nonsynonymous variants. B6-mtNZB (yellow) contains 91 variants distributed across the mt-genome.(C) Duplex Sequencing was used to profile mt-genomes from conplastic mice. Each double-stranded mtDNA fragment is distinctly tagged with a unique molecular barcode, allowing for the computational construction of a duplex consensus sequence. (D) The count of duplex mt-genomes sequenced was calculated for each experimental condition (n = 29 conditions; n = 4 mice for every condition with 3 mice for B6-Young-Heart). The duplex read depth at each position was aggregated across samples in a condition to quantify the duplex depth per condition. The average duplex read depth across the mt-genome was then calculated. (E) The count of variants was aggregated across samples in a condition. Mutations present at conplastic haplotype sites were filtered from the analysis. (F) The mutation frequency for each position is mapped along a linear representation of the mt-genome. Each point denotes the mutation frequency for an experimental condition at the given position in the mt-genome. Positions with a mutation frequency greater than 1×10^{-3} were excluded from this analysis.

Additionally, we analyzed the mutational spectra for atypical imbalances of complementary mutation types and compared our mutation type proportions to a previously published study [86] **(Fig. A.2, Table A.2, Fig. A.3, Table A.3)**. Using this approach, we profiled ∼2.5 million duplex mt-genomes, with a median of 74,764 duplex mt-genomes sequenced per condition (*strain x tissue x age*) **(Fig. 2.1D)**. This resource allows for the detection of somatic mutations at a frequency of 4×10^{-6}.

Duplex reads from each sample were mapped to the mouse mitochondrial genome (mm10) and variants were called using a Duplex Sequencing processing pipeline (see Methods). Mutations overlapping conplastic haplotype sites were filtered out. In total we identified 1,171,918 somatic variants, with a median of 40,449 somatic mutations per condition **(Fig. 2.1E)**. From these ∼1.2 M mutations, 81,097 mutations are unique events. These variants were distributed across the entirety of the mt-genome **(Fig.2.1F)**.

Haplotype- and tissue-specific mutational hotspots

Somatic mutations accumulate with age across both nuclear and mitochondrial genomes [64]. We observed this trend consistently across tissues and mt-haplotypes with the mutation frequency on average ∼2-fold higher in aged mice compared to young mice **(Fig. 2.2A)**. This trend was most pronounced in the liver (2.5-fold higher) and smallest in the heart (1.6-fold higher). Additionally, the heart sustained the lowest mutation frequency on average **(Fig. 2.2A)**, despite its high metabolic demands as demonstrated by its higher mitochondrial copy number **(Fig. A.4)**. Comparing age-associated mutation rates across different strains we find AKR, ALR, and NZB all exhibit strain-specific mutation rates **(Table A.4)**, p-value < 0.01: log-link regression). The FVB strain, which only differs from B6 at two sites **(Fig. 2.1B)** showed no evidence of strain-specific mutation rates compared to wildtype. AKR and ALR strains had lower mutation rates across all tissues while NZB exhibited strong increases in the brain and decreases in the heart.

We next examined the variation in mutation rates and frequencies across different regions of the mitochondrial genome **(Fig. 2.2B, C)**. The density of mutations was lowest in functional coding regions (protein coding, rRNA, and tRNA segments) while the *D-loop* exhibited a significantly higher mutation frequency in both young (6.4-fold, p-value $= 3.4 \times 10^{-9}$; two-tailed t-test) and aged (4.5-fold, p-value $= 8.6 \times 10^{-8}$; two-tailed t-test) mice. However, we further found that the light strand origin of replication (*OriL*) had an even higher average mutation frequency, 40- to 22-fold in excess of the functional coding regions in young and aged mice, respectively **(Fig. 2.2B)**. This *OriL* hotspot of mutation was most pronounced in aged wildtype B6 mice. The *OriL* was recently noted as a mutational hotspot in macaque liver, but not in oocytes or muscle in that species [3]. In contrast, we find the *OriL* to exhibit elevated mutation frequencies across the brain, heart, and liver.

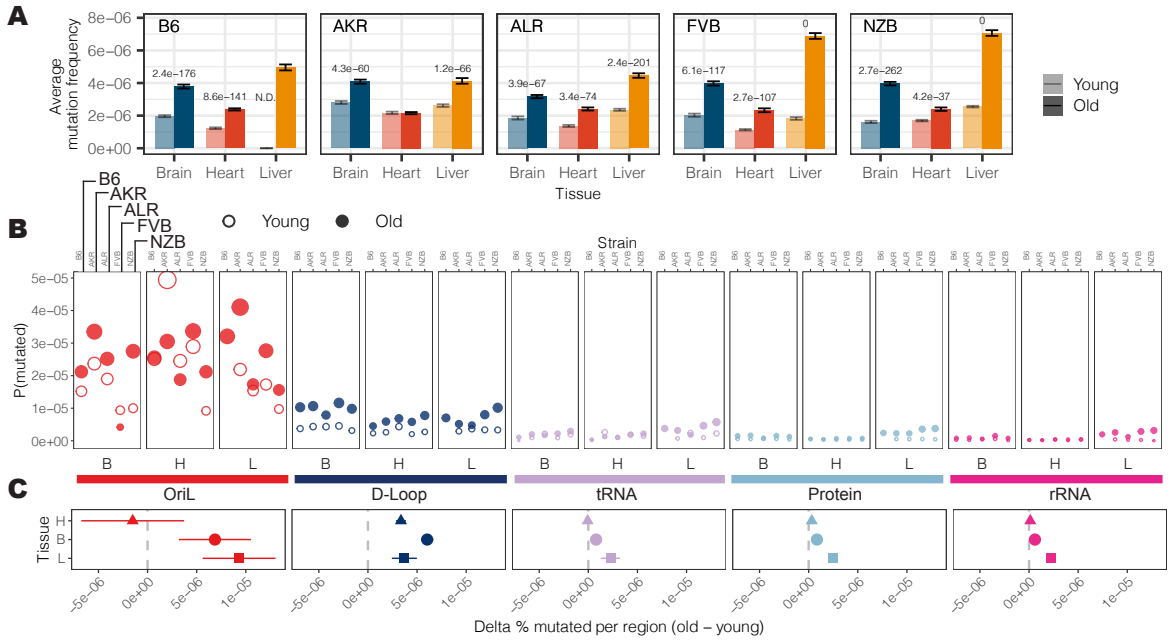


Figure 2.2: **Region-specific changes in somatic mutation frequency with age.** (A) The average mutation frequency for young (light) and aged (dark) mice in each condition (n = 29 conditions; n = 4 mice per condition with 3 mice for B6-Young-Heart). Mutation counts and duplex depth were aggregated across samples for each condition. The mutation frequency for each position along the mt-genome was calculated by dividing the total count of alternative alleles by the duplex read depth at the position. The bar denotes the average mutation frequency. Error bars denote the 95% Poisson confidence intervals. P-values indicate conditions with a significant age-associated increase in mutation frequency. The p-values were calculated from a log-linear regression and adjusted using a Bonferroni multiple hypothesis correction. (B) The mt-genome was categorized into regions: *OriL* (red), *D-Loop* (dark blue), tRNAs (purple), protein coding (light blue), and rRNAs (magenta). For each region, the probability of mutation was calculated as the total count of mutations normalized by the region length in bp multiplied by the average duplex read depth across the region. Fill indicates age group: young (hollow circles) and aged (filled circles). (C) The difference in percent bp mutated between young and aged mice for each mt-haplotype (n = 5 delta values for brain and heart; n = 4 delta values for liver). The shape denotes the average difference in percent bp mutated with age across mt-haplotypes. Error bars showcase the standard error of the mean. Positions that exceeded a mutation frequency of 1×10^{-3} were excluded from these analyses. For **Figs. 2.2B,C**, frequencies were normalized for sequencing depth across conditions at each position, and mutation counts and duplex depth were aggregated across samples in an experimental condition to calculate the percent bp mutated.

Our initial inspection of the overall distribution of mutations across the mt-genome highlighted several distinct clusters appearing at finer-scale resolution than our simple functional classification groups **(Fig. 2.1F)**. To resolve these putative mutational hotspots we quantified the average mutation frequency in 150 base-pair (bp) sliding windows over the mt-genome

independently across tissues and mt-haplotypes **(Fig. 2.3A)**. In addition to the *D-loop*, this analysis identified mutational hotspots in *OriL*, *MT-ND2*, and *MT-tRNA*^{Arg}.

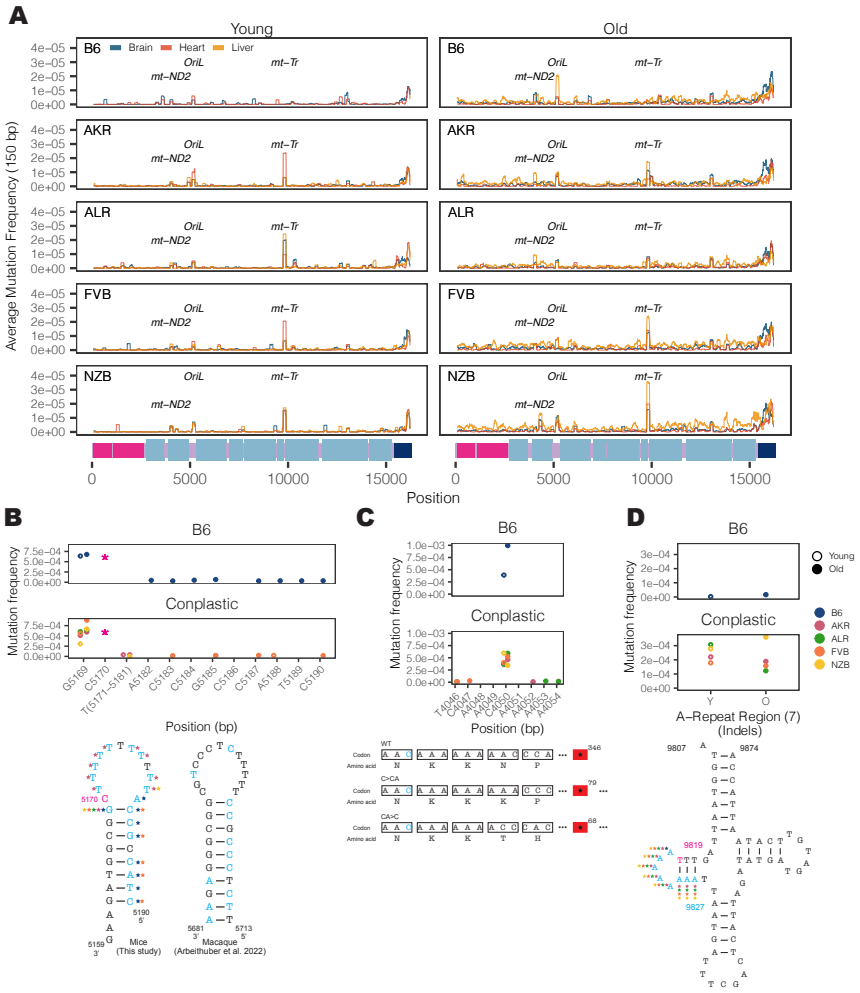


Figure 2.3: **Haplotype-specific peaks of mutation along the mt-genome.** (A) The average mutation frequency was calculated in 150-bp sliding windows for each strain and tissue independently. Regions with shared mutation frequency peaks in at least three mouse strains are labeled. (B) The high frequency region in the light strand origin of replication *(OriL)* is highlighted. For positions 5171-5181 the mutation frequency for the T-repeat region is calculated as the sum of mutations across this region divided by the sum of the duplex depth across positions. Each color denotes a different strain. The average frequency in young (left, hollow points) and aged (right, filled points) mice are compared. The schematic compares the *OriL* structure in mice to that of macaques [3]. Positions with a mutation frequency $> 1 \times 10^{-3}$ are in magenta, while positions that undergo mutations are denoted in light blue. In the macaque *OriL* diagram, variant hotspots are denoted in light blue. Stars represent strains that have mutations present at a given position. (C) The high frequency region in *MT-ND2* is highlighted. Color, shape, and calculation of the average mutation frequency are similar to those in (B). A schematic demonstrating the sequence and codon changes that result from the indels at position 4050: premature stop codons at codon 79 and 68 for C>CA and CA>C, respectively. The superscript denotes the position of the premature stop codon in the amino acid sequence. (D) The mutation frequency for *MT-tRNA*^{Arg} from bp positions 9820-9827, which is an A-repeat region. Color, shape, and calculation of the average mutation frequency are similar to those in (B). Positions that exceeded a mutation frequency of 1×10^{-3} were excluded from these analyses. Frequencies were normalized for sequencing depth across conditions at each position.

The *OriL* forms a stem loop structure that is conserved across species [100] however differs markedly in its sequence. We identified mutations throughout the loop and 3' end of the stem with the highest mutation frequencies corresponding to changes in the size of the stem loop **(Fig. 2.3B)**. Repeated mutations of the 3' end of the stem loop structure mirror those observed in macaque liver in a recent study [3], though the sequence of the stem differs between these two species. Thus, the conserved structure of the *OriL* appears to drive convergent mutational phenotypes between species.

The *MT-ND2* mutation hotspot consists of a frameshift-inducing A insertion or deletion that introduces a premature stop codon **(Fig. 2.3C)**. This stop codon reduces the length of the final protein product by more than 250 amino acids, likely severely impacting its function. This mutation increases in frequency with age across all strains except NZB. The final mutation hotspot we identified is localized to an 8 nucleotide stretch of *MT-tRNA^{Arg}* **(Fig. 2.3D)**. These nucleotides correspond to the 5' D-arm stem-loop of the tRNA and constitute primarily A insertions 2-4 nucleotides long. Computational tRNA structure predictions show that these mutations increase the size of the loop **(Fig. A.5, Fig. A.6)**. This mutational hotspot exhibits strong strain-specificity with B6 exhibiting few somatic mutations in this A-repeat stretch. The D-arm plays a critical role in creating the tertiary structure of tRNAs [81, 74] potentially contributing to its constraint in the wildtype. Together, these results highlight the emergence of distinct mutational hotspots in mitochondrial genomes occurring in age- and strain- specific contexts, and implicating both DNA and RNA secondary structures.

Replication errors and deletions distinguish aged mt-genomes

Somatic mutations are caused by various molecular processes that lead to DNA damage, each of which exhibit a distinct mutational signature [2]. To identify sources of mitochondrial somatic mutation, we categorized mutations into single nucleotide variants (SNVs), deletions (DEL) and insertions (INS) **(Fig. 2.4A)**. SNVs were further classified into the six possible substitution classes **(Fig. 2.4B)**. Overall, SNVs were the predominant somatic mutation type, with a 5-fold higher average mutation frequency than deletions and insertions in both young and aged individuals **(Fig. A.7)**. The most abundant mutation type observed was G>A/C>T, which is indicative of replication error or cytosine deamination to uracil [26, 113] **(Fig. 2.4B)**. The reactive oxygen species (ROS) damage signature of G>T/C>A [13] was the second most predominant mutation type. All somatic mutation types were used to identify two dominant mutational signatures using a multinomial bayesian inference model [36] **(Fig. 2.4C)**. These signatures explained the bulk of the variation across samples **(Fig. A.8)**.

The accumulation of somatic mutations with age is a well known phenomenon that has been hypothesized to play key roles in the etiology of lifespan [12]. We observed that 4 out of the 8 mutation types significantly increased in frequency with age (adjusted p-value < 0.01; Fisher's Exact Test with a Benjamini-Hochberg correction) **(Fig. 2.4B, Table A.5)** across all tissues, with T>A/A>T mutations additionally exhibiting age-associated accumulation

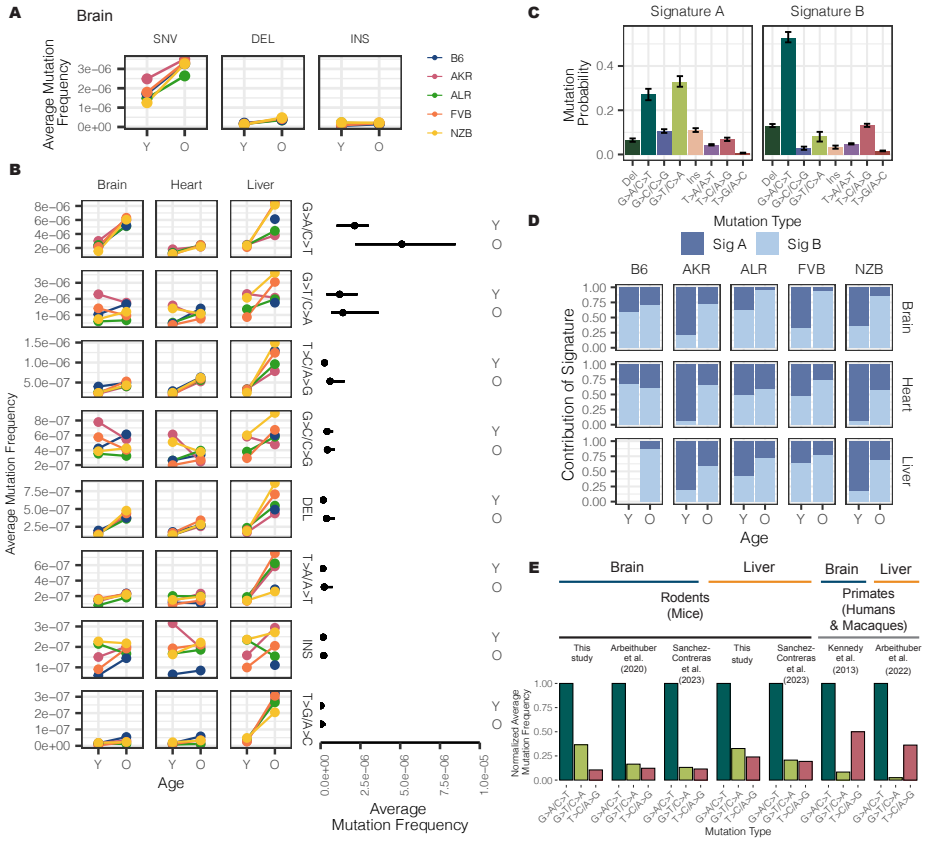


Figure 2.4: **Characterizing the mt-genome mutational landscape.** (A) The average mutation frequency is compared between mt-haplotypes for different classes of mutations: single nucleotide variants (SNVs), deletions, and insertions. The average mutation frequency was calculated as the total mutation count in a class divided by the total duplex bp depth. Each point denotes the average mutation frequency across the mt-genome for young (Y) and aged (O) mice in each strain (B6 (blue), AKR (pink), ALR (green), FVB (orange), and NZB (yellow)). The connecting line demonstrates the change in frequency from young to aged mice. Mutation counts and duplex depth were aggregated across samples for each condition to calculate the average mutation frequency. Results for the brain are featured, which showcase trends observed in the heart and liver, as well **(reference Fig. A.7)**. (B) SNVs were further classified into point mutation types. Each point denotes the average mutation frequency for the given mutation type across the mt-genome for each mouse strain (color key as in (A)). The average frequency for point mutations was calculated as the total count of mutations of the given type divided by the duplex bp depth for the reference nucleotide. The average mutation frequency for deletions and insertions was calculated as the total mutation count in these classes divided by the total duplex bp depth. Mutation types are ordered by descending average mutation frequencies. The range in frequency across experimental conditions (*strain x tissue*, n = 15 conditions for brain and heart; n = 14 conditions for liver) for each mutation type is shown. For each mutation type, the segment extends from the minimum to the maximum frequency across experimental conditions in the age group. Each point in the segment denotes the median frequency across conditions. (C) Mutational signatures were extracted from mutation type counts for each experimental condition using sigfit. Two mutational signatures were identified. The probability that a mutation type contributed to a signature is showcased. Each bar denotes the estimated probability of a mutation type comprising the signature with error bars denoting the 95% confidence interval of each estimate. (D) The presence of the mutational signatures was estimated for each condition (Signature A dark blue, Signature B light blue). The contribution of each mutational signature is compared across age. (E) The frequency of the three most abundant mutation types were compared across mice, macaques, and humans for the brain and liver. Mutation frequencies were normalized by the frequency of the G>A/C>T mutation in their respective study. All studies used Duplex Sequencing to profile mutations in the mt-genome. For the mutation frequencies in our study, we only show the mutation frequencies for B6. Refer to **Fig. A.9A** for frequencies across all strains, tissues, and young mice. For all analyses, mutation counts and duplex bp depth were aggregated across samples in an experimental condition. Positions with mutation frequencies greater than 1×10^{-3} were omitted from these analyses.

in the liver but not in brain or heart. These age-associated signatures compose *Mutational Signature B* **(Fig. 2.4C)**, which overall distinguishes aged from young samples **(Fig. 2.4D)**. The consistency of these signatures across both mutations and mt-haplotypes suggest that these mutations occur in a 'clock-like' fashion in mitochondria over lifespan.

In contrast, insertions, G>C/C>G, and G>T/C>A mutations did not exhibit consistent age-associated patterns. These mutations are represented in *Mutational Signature A* **(Fig. 2.4C)**. Of particular note, the G>T/C>A and G>C/C>G mutation patterns are associated with ROS damage. An increase in ROS damage has been hypothesized to play an important role in aging. Nonetheless, we do not find evidence of an increase in ROS-associated damage with age. Previous works in the human brain [51, 106], macaque tissues [3], and various mouse tissues [4, 86] have also observed a lack of ROS-associated damage with age.

Species-specific mitochondrial mutational signatures

While the overwhelming majority of research into somatic mutation rates and profiles has focused on humans, comparative analyses can provide insight into the evolutionary processes that have shaped mutation. To determine if mitochondrial somatic mutation exhibited species-specific patterns we compared several recent studies that Duplex Sequenced mt-genomes in multiple mouse, human, and macaque tissues [51, 4, 3, 86]. In our dataset, we found the relative magnitude of mutation frequencies to be consistent across tissues and mt-haplotypes with the G>A/C>T and G>T/C>A mutations being the most abundant **(Fig. 2.4B, Fig. A.9A)**. This signature is consistent with other duplex sequencing-based analyses of mitochondrial mutations in young and aged mouse brain, muscle, kidney, liver, eye, heart, and oocytes **(Fig. 2.4E, Fig. A.9B, Fig. A.9C)** [4, 86]. However, profiling of mitochondrial mutation signatures in several young and aged human brains noted transitions, G>A/C>T and T>C/A>G, to be the most abundant mutation signatures [51] **(Fig. 2.4E)**. We find that this pattern is also recapitulated in a recent dataset of mitochondrial somatic mutation in macaque muscle, heart, liver and oocytes **(Fig. 2.4E, Fig. A.9D)** [3]. These differences in mutational signatures were observed between primate and rodent species from duplex sequencing data generated independently by two research groups, emphasizing the reproducibility of this phenomenon. Together, these results suggest that rodent and primate mt-genomes are subject to distinct mutational processes, potentially as a cause or effect of physiological differences between these lineages.

Negative selection shapes mitochondrial mutation frequencies

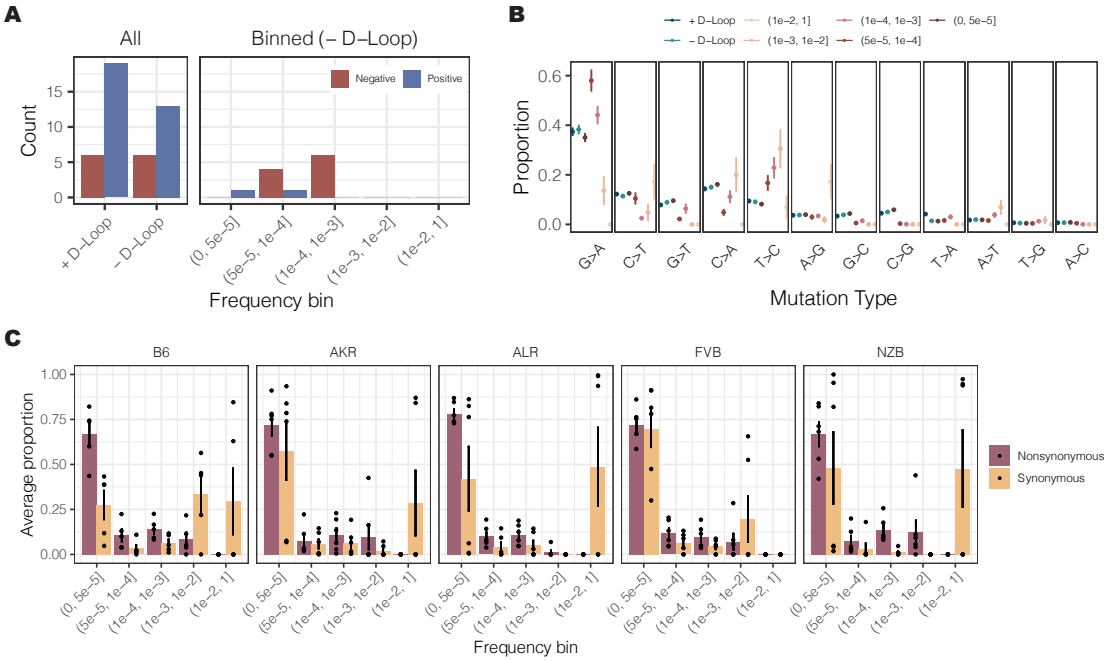


Figure 2.5: **Negative selection shapes the mt-genome.** (A) The count of genes under positive (blue) and negative (red) selection. Mutations and mutation type proportions were quantified in three different ways: (1) all mutations, without consideration of mutation frequency or mt-genome position (2) all mutations excluding the *D-Loop* (3) mutation counts binned by mutation frequency, excluding the *D-Loop*. Mutations with a frequency greater than 1×10^{-3} were excluded from analyses (1) and (2). (B) The proportion of each mutation type. Proportions are calculated as the count of mutations of each type divided by the total count of mutations in a given bin. The average mutation type proportion across experimental conditions (n = 29 conditions; *strain x tissue x age*) is shown with error bars denoting the standard error of the mean. In blue are proportions for the aggregated mutation counts with (dark blue) and without (light blue) the *D-Loop*. Mutations are ordered in descending order of average mutation frequency. (C) The frequency spectra for nonsynonymous (purple) and synonymous (orange) mutations. The bar denotes the average proportion of nonsynonymous and synonymous mutations in a bin. The proportion is calculated as the count of nonsynonymous or synonymous mutations in a frequency bin relative to the total count of nonsynonymous or synonymous mutations across all bins. The average proportion is taken across tissues and age in a strain (n = 6 except for B6, where n = 5). Each point represents the proportions that comprise the average. The error bars denote the standard error of the mean. For these analyses mutation counts were aggregated across samples in an experimental condition.

While only thirteen proteins are encoded on the mt-genome, these products are vital components of the electron transport chain. Given their importance, we investigated whether

selection was acting to shape the mutational landscape of the mt-genome. We calculated the $\frac{hN}{hS}$ statistic [60], which is akin to the $\frac{dN}{dS}$ statistic, for every gene across our 29 experimental conditions. For each condition, we also simulated the expected distribution of $\frac{hN}{hS}$ ratios using the observed mutation counts and mutational spectra **(Fig. A.10)**. *Prima facie*, the mt-genome appeared to be shaped predominantly by positive selection **(Fig 2.5A)**. However, differences in the mutation spectrum between mutational hotspots, including the *D-loop*, could bias the simulated null distribution of mutations. Indeed, mutation spectra were significantly different between *D-loop* and non-D-loop mutations, as well as among mutations at different frequencies **(Fig 2.5B)**. Furthermore, quantifying the allele frequency spectrum of nonsynonymous and synonymous mutations revealed that nonsynonymous substitutions were strongly enriched at low frequencies compared to synonymous substitutions indicating that negative selection is likely preventing these mutations from increasing in frequency **(Fig 2.5C)**.

We thus performed our $\frac{hN}{hS}$ analyses independently in different frequency bins **(Fig 2.5A)** revealing primarily signatures of negative selection. From the twelve possible cases of selection, 75% were in conplastic mice, and both cases of positive selection were in NZB mice. Genes with selective signatures include *MT-CO1* (3), *MT-CO3* (1), *MT-CYTB* (1), *MT-ND2* (1), and *MT-ND5* (6), whose protein products compose complexes I, III, and IV of the electron transport chain. Intriguingly, signatures of negative selection in *MT-ND5* were observed across all five strains in our experiment. Taken together, these results indicate that negative selection dominates the distribution of mutations in mitochondrial genomes, though this signature is predominantly found at intermediate frequencies.

Mito-nuclear mismatches drive somatic reversion mutations

Mismatching of mitochondrial and nuclear haplotypes, such as in hybrid populations, has been associated with reductions in fitness [68, 70, 29, 31, 28, 75, 53, 57, 15, 72]. The conplastic strains we employ are hybrids with mismatched nuclear and mt-genome ancestries. Since hybrids often demonstrate reduced fitness as a result of mito-nuclear discordance, we reasoned that at sites where the conplastic mt-genome differed from the B6 mt-genome (haplotype sites), there may be a preference for "reversion" mutations to the B6 allele **(Fig. 2.6A)**.

We hypothesized that if somatic selection were to favor the B6 allele, then reversions would occur at a higher rate than background mutations **(Table A.6)**. Three of the strains have several fixed haplotype sites in their mitochondrial genomes, ALR, FVB, and NZB. We find that in ALR and FVB strains, haplotype sites were among the most mutated sites in the mt-genome, with 6-fold and 7-fold higher average mutation frequency than non-haplotype sites, respectively **(Fig. 2.6B**, adjusted p-values <0.001). In NZB, which differs from B6 at 91 locations, these haplotypes sites had 122-fold higher mutation frequency than background (adjusted p-values $< 1\times10^{-53}$). The overwhelming majority (75%−100%) of these mutations are reversions to the B6 allele with the specific reversion mutation occurring significantly

more than expected by chance **(Table A.7)**. These results demonstrate the extreme selective pressures impressed by nuclear-mitochondrial matching to reintroduce the ancestral allele.

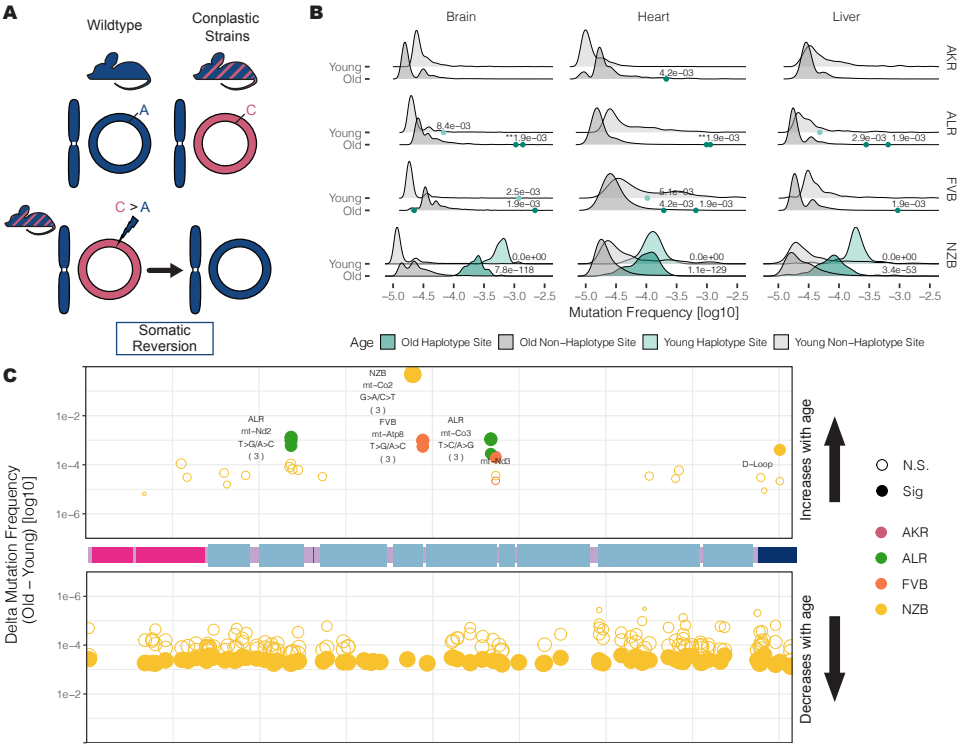


Figure 2.6: **Somatic reversion mutations segregate in the mt-genome population.** (A) A schematic that explains somatic reversion mutations. The wildtype strain (B6) has matching nuclear and mt-genome ancestries (B6 ancestry denoted in blue). Conplastic strains are hybrids with mismatching nuclear and mt-genome ancestry. Haplotype sites are positions in the mt-genome where the conplastic mt-genome differs from the B6 mt-genome. Somatic reversion mutations refer to the reintroduction of B6 ancestral alleles at haplotype sites. (B) The mutation frequency at non-haplotype sites (gray distributions) is compared to the mutation frequency at haplotype sites (green distributions or points when less than three haplotype sites exist). All mutations were included in the distribution of non-haplotype site mutation frequencies, including positions with a mutation frequency greater than 1×10^{-3}. Haplotype site mutation frequencies were corrected for potential NUMT contamination. Mutation frequencies were normalized for sequencing depth between young and aged experimental conditions. Denoted are the adjusted empirical p-value using the Benjamini-Hochberg correction. Asterisks denote the number of haplotype sites with the same p-value. For AKR, ALR, and FVB empirical p-values for haplotype sites were calculated as the count of non-haplotype sites with a higher frequency than the haplotype site divided by the total number of non-haplotype sites (one-sided test). For NZB, the distribution of mutation frequencies for haplotype sites and non-haplotype sites was compared using the Wilcoxon Rank Sum Test. (C) A map of the somatic reversion mutations along a linear mt-genome. Each point denotes the change in frequency with age (delta) for a somatic reversion mutation. The size of the point indicates the magnitude of delta and color represents the conplastic strain the somatic reversion occurs in. Empirical p-values were calculated as the count of sites with a delta greater (for sites with an increase in frequency with age) or less (for sites with a decrease in frequency with age) than the haplotype site delta divided by the total number of deltas (one-sided test). Fill denotes significance (adjusted empirical p-value < 0.02 using the Benjamini-Hochberg correction within each mt-haplotype group).

We next hypothesized that if somatic selection were to favor reversions, these B6 reversion alleles should increase in frequency with age. We quantified the relative change in reversion frequencies across strains and found that in ALR and FVB all unique fixed sites

exhibited a significant increase in the reversion allele frequency with age across all tissues **(Fig 2.6C, Table A.8**, BH adjusted p-value < 0.02). This suggests a strong benefit of these coding reversion substitutions in ALR and FVB strains. In contrast, in NZB we observed an overwhelming preference for reversion alleles to decrease in frequency with age, particularly in the brain **(Fig 2.6C, Table A.8)**. One potential explanation for this decrease in frequency compared to ALR and FVB is that the many NZB fixed substitutions exhibit epistatic interactions which manifest negatively if single reversions are not accompanied by mutations at other sites, though this is challenging to test. We did identify two cases in NZB of reversions that increase in frequency with age, a synonymous substitution in the *MT-CO2* gene, and a noncoding mutation in the *D-loop*. Together, our results demonstrate that sites which contribute to mito-nuclear mismatching are prone to elevated levels of mutation with reversion alleles preferred across multiple tissues in populations. This preference for nuclear mitochondrial matching potentially drives somatic selection increasing the frequency of these alleles with age.

2.4 Discussion

The interdependence of nuclear- and mitochondrially-encoded genes has significantly shaped patterns of diversity and constraint on these two genomes across organisms and populations. Yet, the mt-genome also exhibits exceptionally high somatic mutation rates and is thus dynamic within individuals' lifespans. Here, we explored the relationship between nuclear-mitochondrial ancestral matching and the accumulation of somatic mutations. Our results extended upon other works that employ highly sensitive Duplex Sequencing to characterize the mutational landscape of the mt-genome [51, 4, 3, 86] and generated, to our knowledge, the largest mt-genome somatic mutational catalog to date. We profiled 2.5 million mt-genomes across 4 conplastic mouse strains and the B6 wildtype, with a median of more than 74,000 mt-genomes per biological condition. Our results corroborated other studies of somatic mutation in the mt-genomes of mice, humans, and macaques, which demonstrate an age-associated increase in mutations as well as tissue-specific mutation rates. Our study design also uniquely allowed us to assess the role of mt-haplotype on mutation rates. The conplastic strains employed each exhibit distinct physiological differences which likely play a role in modulating mutations. We find cases in which mt-haplotype impacts mitochondrial mutation rate, which emphasizes that specific fixed substitutions between mitochondrial haplotypes can play a vital role in shaping mutational profiles.

Our study design allows us to compare mutational landscapes across three of the most metabolically active tissues: the brain, heart, and liver. We show that SNV mutation rates are haplotype- and tissue-specific. We also noted that while most mutations increase in frequency with age, ROS-associated mutations do not. These findings recapitulate prior studies conducted by *Arbeithuber et al. 2020* [4] and *Sanchez-Contreras et al. 2023* [86]. Several hypotheses have been proposed to explain this phenomenon ranging from potential molecular mechanisms that recognize and remove ROS-damage through aging [86] to the existence

of mitochondrial subpopulations that dictate the replication of specific, mutation-limited mtDNA molecules [20, 21]. Further work is needed to identify the ultimate mechanisms that result in such non-clocklike signatures.

Our high duplex coverage allowed us to characterize mutational hotspots throughout the mt-genome at fine-scale resolution. As expected, the *D-Loop* had a high mutation frequency compared to other regions in the genome. However, we identified three additional hotspots exhibiting either age- or haplotype-associated mutation frequencies. The first of these hotspots occurs in the *OriL*, the light strand origin of replication, which forms a DNA stem-loop structure. While the sequence of this region is highly divergent among species, the structure is conserved. Recently, *Arbeithuber et al.* [3] identified the *OriL* as a hotspot of mutation in macaque liver. We were also able to capture these hotspots in a reanalysis of duplex sequencing data from several different tissues that was recently published despite their relatively lower coverage **(Fig A.11 A-B, Table A.9)** [86]. Similarly to *Arbeithuber et al.* [3], we find mutations at the *OriL* locus to be most prominent in aged wild-type liver, however our increased sequencing depth allows us to show that this is a consistent hotspot across tissues and strains regardless of age. We identify mutations in both the stem and the loop of this structure, with mutations occurring in the same 5' end of the stem as those identified in macaque, albeit impacting completely different sequences. We further noted that there is a well supported mechanism for this specific mutational hotspot occurring on one arm of the stem loop structure due to replication priming by an RNA which is susceptible to slippage [87]. These results highlight that the conserved structure of this stem loop is sensitizing to conserved mutational patterns across species. Similarly, peaks of mutation frequency within the *D-Loop* have been narrowed to regions associated with mtDNA replication [103, 60, 84]. The mutational hotspots surrounding mtDNA origins of replication have been hypothesized to serve as a compensatory mechanism against mutation by slowing mtDNA replication [3]. We identified an additional hotspot occurring in a highly structured nucleic acid, *MT-tRNA^{Arg}*. Intriguingly, these mutations are only found in conplastic strains, suggesting that they are poorly tolerated in the wildtype. These mutations are expected to increase the size of the loop in the 5' D-arm stem-loop of this tRNA. Polymorphisms that increase the size of the D-arm in *MT-tRNA^{Arg}* have been linked with an increase in mtDNA copy number triggered by heightened reactive oxygen species production [73]. Specifically, the size of the loop, not the underlying sequence, was found to be important for mitochondrial function [74]. Together, these findings suggest that mutational hotspots in the mt-genome may potentially act as compensatory mechanisms to alter mtDNA replication and copy number.

While Duplex Sequencing approaches have allowed us to identify processes associated with DNA replication as the primary driver of somatic mutations in mitochondria [51, 4, 3], the full extent of mutational processes impacting mt-genomes remains unknown. Recent studies have determined that different species exhibit varying contributions of mutational signatures in the nuclear genomes of aging intestinal crypt cells [12]. Comparing the mutational spectra across several studies that duplex- sequence the mt-genomes of humans, mice, and macaques in various tissues, we similarly find that rodents exhibit distinct mutational profiles compared to primates. Although DNA replication error is the predominant muta-

tion type across species, rodents exhibit a surfeit of G>T mutations in contrast to primates. While studies in humans previously concluded that transitions are the most common mutation in mitochondria, we find that the abundance of the T>C marker of DNA replication error is a species-specific effect. We are not able to determine whether this finding is the result of physiological differences between these organisms, or variability in the repair pathways of these species. Of note, we find these trends hold both for somatic mutations as well as de-novo oocyte mutations in the mt-genome, suggesting that the underlying mechanisms driving these distinct profiles have potentially shaped patterns of mt-diversity across different species. Intriguingly, a very similar signature distinguishing mice and ferrets from other mammals was found using laser-capture and deep sequencing of intestinal crypts across 16 species [12]. Together, these findings thus recapitulate that distinct life history traits impact the evolution of the mt-genome across species. Compared to primates, rodents have a shorter lifespan and a substantially smaller body size. The difference in G>T mutations, which are associated with reactive oxygen species damage, and T>C mutations, a marker of DNA replication error, suggest that repair mechanisms or ROS defenses may differ between these species. Further analyses across broad ranges of species, such as those recently performed in the nuclear genome [12], are needed to inform how mitochondrial mutational processes are specifically associated with disparate life histories.

Mitochondrial genomes exist as a population inside cells, where selection can act to shape this population at various biological scales. In oocytes, strong genetic drift induced by the mitochondrial genetic bottleneck [82] and purifying selection [103, 111] have been shown to shape the transmission of mt-genome mutations across generations. In somatic tissues, the population of mt-genomes may be shaped at the cellular level, as a result of inter-cellular competition; at the inter-mitochondrial level, with mitochondrial turnover; or at the intra-mitochondrial level between mt-genomes harboring different variants. Previous studies focused on de novo mutations in mice have not identified signals of selection [4, 86]. However, it has been suggested that this may be due to the low frequencies of these variants, which prevents them from having a phenotypic effect on mitochondrial function. By contrast, studies focused on higher frequency mutations in humans (greater than 0.5% frequency) have identified signals of positive selection [60], and potential signals of negative selection [60, 59] based on differences between polymorphic and heteroplasmic nonsynonymous mutations. Furthermore, signatures of positive selection in the liver have been previously reported in both macaques [3] and humans [60]. Complementing these studies, we find that low frequency mutations, which are likely de novo somatic mutations, are not under strong selection. However, mutations at intermediate frequencies, ranging from 5×10^{-5} to 1×10^{-3}, do exhibit some signatures of negative selection. While these are likely not de novo mutations, they demonstrate a low frequency threshold for tissue-specific selection of inherited or early developmental mutations. Importantly, our results demonstrate that mitochondrial mutational spectra vary across different frequencies and mutational hotspots **(Fig. 2.5B)**, which can influence signatures of selection. Together, these findings suggest that purifying selection acts on mitochondrial mutations segregating at lower frequencies.

Our study also allowed us to examine a very specific form of somatic selection: the emer-

gence and persistence of reversion mutations that re-match the mitochondrial haplotype to its nuclear constituent. Previously, *Wei et al.* identified the preferred maternal transmission of variants that worked to re-align mitochondrial and nuclear ancestry in humans [103]. This result showcased that selection for reversions can occur in as short as one generation, emphasizing the strong influence mito-nuclear interactions have in shaping mt-genome diversity. We find selection for reversion mutations at even shorter timescales: within an organism's lifespan. These mutations are extremely abundant, much more than expected by chance, and in several cases increase in frequency with age. Importantly, the change in reversion allele frequencies with age was dependent on the genetic divergence between the mitochondrial and nuclear ancestries. That is, mice with the greatest genetic divergence from the B6-mitochondrial haplotype observed the opposite trend, a decrease in reversion frequency with age. Despite mito-nuclear ancestral mismatching, NZB mice did not showcase phenotypic differences compared to B6 [42]. Thus, understanding to what extent mito-nuclear matching influences somatic evolution remains to be explored. Altogether, these observations highlight a hitherto unexplored impact of mito-nuclear incompatibility, namely its potential role on the somatic evolution of tissues.

While we sequence to great depth across various tissues, we are unable to characterize the impact that cell-type plays on the evolution of the mt-genome. This analysis is of particular interest in tissues such as the heart and the brain which consist of both mitotic and post-mitotic cellular populations. Given that DNA replication error is a predominant age-associated mutational signature, sequencing at the single cell level will be pivotal in understanding the role that cell proliferation has on mitochondrial somatic mutation. Additionally, our study only examined the wildtype B6 mouse and conplastic strains with nuclear B6 backgrounds. Reciprocal conplastic strains, in which both mitochondrial haplotypes are placed in the context of both nuclear genomes, will allow us to parse apart the different roles that the nuclear and mitochondrial genomes have in driving distinct mutational processes. Lastly, our study emphasizes the importance of comparative somatic mutation profiling in order to discern how processes that shape mutation in the mt-genome differ with life history traits. Altogether, our findings explore somatic evolution in the context of an important cellular organelle and begin to discern the various scales at which evolutionary processes act to shape the population of mt-genomes.

2.5 Methods

Ethics Statement

Animal use and all methods were approved by the Animal Care and Use Committee (V242-7224. 122-5, Kiel, Germany). Experiments were performed in accordance with the relevant guidelines and regulations by certified personnel.

Data collection and sample preparation

The wildtype C57BL/6J (B6) and inbred mouse strains were obtained from Jackson Laboratory and maintained at the University of Lubeck. Conplastic strains B6-mtAKR, B6-mtALR, B6-mtFVB, and B6-mtNZB were generated and bred as described in [108] at the University of Lubeck. Briefly, the conplastic strains were developed by crossing female mice from AKR, ALR, FVB, and NZB mouse strains with male B6 mice. Female offspring were then backcrossed with male B6 mice. After the tenth generation of backcrossing mice were deemed conplastic mice with a B6 nuclear background and their respective maternal mitochondrial haplotypes. Samples were validated by checking for their defining haplotype sites **(Table A.1, Fig. A.12)**. The brain, heart, and liver were sampled from young (2-4 months old (mo.)) and aged (15-22 mo.) mice in each strain (n = 115). B6 young liver samples and one B6 young heart sample were omitted from the data due to possible contamination **(Fig. A.12)**. Whole tissue samples were flash frozen and processed for DNA isolation using the Qiagen DNAeasy Blood and Tissue kit [ID: 69504]. All mice in this study are female.

Sequencing

Duplex sequencing libraries were prepared by TwinStrand Biosciences (Seattle, WA) as previously described [89, 51]. Sequencing was performed at the University of California, Berkeley using Illumina NovaSeq generated 150 bp paired-end reads.

Data processing

Sequencing reads were processed into Duplex Sequencing reads and mapped to the reference mouse mitochondrial genome (mm10) using a modified version of the Duplex Sequencing processing pipeline developed by Dr. Scott Kennedy's group at the University of Washington, Seattle. The pipeline was edited to take as input bam files formed from the NovaSeq reads using bwa mem [58] and samtools [25], and is available on our github. Software versions used along with all processing steps including duplex consensus sequencing generation, mapping, and variant calling can be found in the github repository referenced below. Parameters used to run the Duplex Sequencing pipeline are provided in **Supplemental Note A.1**.

Trinucleotide Spectra

Only de novo mutations were used in this analysis and were identified by filtering for mutations with an alternative allele depth < 100 and a mutation frequency < 0.01. Each mutation is scored once to create a proxy mutation frequency for de novo mutations. The mutation fraction is the proportion of each mutation type in a given trinucleotide context divided by the total count of de novo mutations for a condition. The trinucleotide and mutation type featured represent mutations on either strand. For example, ACG represents both ACG and CGT, where either a C>T or a G>A mutation has occurred **(Fig. A.13)**.

Mutational Signature Extraction

Observed counts for each mutation type (excluding mutations at a frequency $> 1 \times 10^{-3}$) were used as input for sigfit (v 2.2), which is an R package used to identify mutational signatures. Sigfit uses Bayesian probabilistic modeling to uncover mutational signatures, as explained by *Gori and Baez-Ortega* [36]. Specifically, we use multinomial models to extract our signatures, which is akin to the traditional non-negative matrix factorization approaches. The optimal number of signatures was determined by extracting a range of 1-7 signatures and comparing the cosine similarity **(Fig. A.8)** for this range (iterations = 1000, seed = 1756). The model was then refitted with two signatures, as determined by the goodness of fit test, with 10,000 iterations (seed = 1756).

Testing for selection

The number of nonsynonymous variants per nonsynonymous sites (hN) and synonymous variants per synonymous sites (hS) were calculated as described in [60, 4].

We quantified mutations and mutation type proportions in three different ways: (1) all mutations, without consideration of mutation frequency or mt-genome position (2) all mutations excluding the *D-Loop* (3) mutation counts binned by mutation frequency, excluding the *D-Loop*. Mutation counts were calculated as the sum of the duplex alternative allele depth across all sites in the mt-genome. Counts were aggregated across replicates in experimental conditions **(Fig. A.14A)**.

For analyses where mutation counts were aggregated across frequencies, the $\frac{hN}{hS}$ statistic was quantified for each gene across the 29 experimental conditions. To test if the observed $\frac{hN}{hS}$ statistics were significant signals of selection, we used a multinomial distribution to simulate mutations for all experimental conditions using the observed mutation counts and proportions for each mutation type. The mutations were sampled across the mt-genome uniformly with replacement, and the simulated $\frac{hN}{hS}$ statistics were calculated for each gene. 10,000 simulations were conducted for each experimental condition. $\frac{hN}{hS}$ statistics that could not be calculated ($hS = 0$ or $hN = 0$) were excluded from the analysis. Mutations with a frequency $> 1 \times 10^{-3}$ were excluded from these analyses **(Fig. A.14B)**. Empirical p-values were used to determine if an observed $\frac{hN}{hS}$ statistic was significant. The empirical p-value was calculated as the number of simulated $\frac{hN}{hS}$ statistics with a more extreme value than the observed $\frac{hN}{hS}$ statistic divided by the total number of simulated statistics. These empirical p-values were multiplied by a factor of 2 in order to account for both tails of the distribution. $\frac{hN}{hS}$ statistics with a Benjamini-Hochberg adjusted p-values < 0.01 were denoted as significant, where $\frac{hN}{hS} > 1$ signified that a gene was under positive selection and $\frac{hN}{hS} < 1$ suggested a gene was under negative selection.

For analyses where mutation counts were binned by frequencies, mutations with a frequency $> 1 \times 10^{-3}$ were included. $\frac{hN}{hS}$ statistics were calculated for each bin across every experimental condition. Mutation counts were calculated as the sum of the duplex alternative allele depth across all sites in the mt-genome within the frequency bin. Counts were

aggregated across replicates in experimental conditions. The analysis described above was performed for every bin in each experimental condition. A Benjamini-Hochberg multiple hypothesis test correction was applied for each bin to maintain consistency in the number of tests corrected for between the aggregated and binned analyses.

Comparison of the null and observed mutational spectra for nonsynonymous and synonymous mutations

The null mutational spectra for synonymous and nonsynonymous mutations was calculated by quantifying the number of nonsynonymous and synonymous mutations each mutation type could produce. The proportion was calculated as the count of nonsynonymous or synonymous mutations resulting from a given mutation type divided by all possible nonsynonymous or synonymous mutations in the mt-genome **(Fig. A.15A)**. For the observed mutational spectra, the proportion of each mutation type comprising the total count of nonsynonymous and synonymous mutations was calculated **(Fig. A.15B)**.

Estimation of and correction for NUMT contamination

We originally mapped duplex paired end reads to mm10 masking the known NUMT region. We reasoned that given the 10- to 100-fold higher somatic mutation rate of the mitochondrial genome along with its several-hundred fold increased copy number with respect to the nuclear genome, that nuclear contaminating reads contribute minimally to our mutation frequencies. However, this low-level of contamination can potentially cause an issue when examining reversion mutations which are expected to be enriched for the B6 allele in the NUMT.

To estimate and correct for this NUMT contamination, duplex paired end reads were remapped to mm10 using bwa mem. For this remapping, the NUMT region in chr1 (nt24611535 to nt24616184) was unmasked. The duplex read depth at junction regions, which captured sequences 10 bp upstream and 10 bp downstream of the NUMT region in chr1 and the corresponding region in the mt-genome (nt6394 to nt11042), was calculated using *samtools depth input.in1 -b input.in2 > output*. These junction regions contain sequences unique to chr1 and chrM, allowing us to estimate the % of reads mapping to ch1 as the number of reads mapping to chr1 divided by the average duplex depth of the mt-genome calculated with the chr1 NUMT masked **(Fig. A.16A)**.

To validate this estimated contamination, we remapped the duplex paired end reads to the NZB reference genome (generated using an in-house script). The NZB mt-genome and B6 NUMT region differ by 24 SNVs. To map the duplex reads to the NUMT region we used *samtools depth input.in1 -bq 30 input.in2 > output*, setting a strict mapping quality score given that reads may differ from the regions by as few as 2 positions. Six regions across the NUMT were identified as having multiple SNVs within a read (< 130 bp apart), which we refer to as SNV clusters. We used these clusters to identify reads that mapped to chr1. The estimated % of NUMT contamination at each SNV cluster was calculated as the number of reads mapping to chr1 divided by the average duplex depth of the mt-genome calculated

with the chr1 NUMT masked. The estimated % of contamination at the junction regions was compared to the distribution of estimated % of contamination for the SNV clusters in NZB **(Fig. A.16B)**. The maximum estimated % of contamination between the junction regions was consistently equal to the median estimated % of contamination for the SNV clusters in NZB, verifying the consistency of the estimated % of NUMT contamination.

We take as a conservative measure the maximum estimated % of contamination from the junction regions ($\sim$ 0.5% contamination). The estimated chr1 read depth was then calculated as the original average mt-genome duplex depth multiplied by the estimated % of contamination. We calculated the corrected duplex mt-genome read depth in this region as the original duplex mt-genome read depth subtracted by the estimated chr1 read depth count. Likewise, the reversion allele depth is calculated as the original reversion allele count subtracted by the estimated chr1 read depth, assuming all reads from chr1 contain the B6 allele.

Statistical analysis: Testing for significance

Statistical analyses were performed using R (4.1.2). Poisson confidence intervals for average mutation frequencies were calculated using the *qchisq* function from the stats package in R (v 4.1.2). Log-link regressions were performed to determine an age-associated change in mutation frequency and mt-haplotype specific mutation rates *(glm, stats package)*. Associations between (1) mutation type and age and (2) reversion allele and haplotype site were determined via Fisher's exact tests (*fisher.text, stats* package). To compare (1) the average mutation frequencies between the *D-Loop* and other regions in the mt-genome and (2) the average mutation frequencies of simulated and observed $\frac{hN}{hS}$ statistics on a gene-by-gene basis we used two tailed t-tests (*t.test, stats* package). To determine that our simulated $log10(\frac{hN}{hS}) > 0$ altogether and in a gene-by-gene analysis we used a one-sample t-test (*t.test, stats* package). To compare (1) the average $log10(\frac{hN}{hS})$ between our simulated and observed data and (2) background to haplotype site mutation frequencies for the NZB conplastic strain, we conducted Wilcoxon-Rank Sum tests (*wilcoxon.test, stats* package). Lastly, we calculate empirical p-values to determine the significance of observed $\frac{hN}{hS}$ statistics; high haplotype site mutation frequencies for AKR, ALR, and FVB; and for changes in mutation frequency with age for all haplotype sites. The empirical p-values are calculated as the number of simulated $\frac{hN}{hS}$ statistic (for 1) or number of background sites (for 2 and 3) with a more extreme value than our observed $\frac{hN}{hS}$ statistic (for 1), haplotype site mutation frequency (for 2), or change in mutation frequency for the haplotype site (for 3), divided by the total number of simulations (for 1) or total number of background sites (for 2 and 3). Multiple hypothesis test corrections were performed using the Benjamini-Hochberg correction (*p.adjust, stats* package), adjusted p-values refer to p-values that have undergone multiple hypothesis correction.

Data Availability

All raw data has been uploaded to sequence read archive (SRA BioProject ID: PRJNA1054208). Single stranded consensus and duplex consensus sequences, and processed data has been deposited in zenodo. The zenodo repository can be accessed at the following (https://doi.org/10.5281/zenodo.7787188).

Code availability

Funding: Institute of General Medical Sciences [grant: R35GM142916] to PHS. Vallee Scholars Award to PHS. National Science Foundation Graduate Student Fellowship [grant: DGE 1752814; DGE 2146752], University of California, Berkeley Graduate Fellowship, the Rose Hills Foundation Fellowship, and the Ford Foundation Dissertation Fellowship to IMS.

2.6 Author Contributions

Conceived the experimental design: PHS. Constructed the conplastic strains and provided experimental samples: MH and SI. Sequencing design and library preparation: CCV, SA, ES, GP, LW, JS. LW and GP contributed to this project while affiliated with TwinStrand Biosciences. Processed and analyzed the data: IMS and PHS. Wrote and edited the manuscript: IMS and PHS.

2.7 Competing Interests

CCV, SA, ES, GP, LW, and JS declare they are equity holders of TwinStrand Biosciences, Inc. Additionally, CCV, ES, and JS are current employees of TwinStrand Biosciences and CCV, ES, LW, and JS are inventors on one or more Duplex Sequencing-related patents. The remaining authors declare no competing interests.

Chapter 3

Tissue-specific mtDNA mutational landscapes in the aging brain

In Chapter 2, I examine how mitochondrial haplotype and mito-nuclear ancestral mismatching influence the aggregation of somatic mutations throughout aging. By sampling the brain, heart, and liver from each conplastic mouse strain, I was also able to survey how differences in tissue physiology and metabolic demands impact mtDNA mutation. To briefly summarize the tissue-related findings in Chapter 2: the distribution of mutations along the mtDNA molecule and mutation frequencies show tissue-specific trends. Notably, the heart has the lowest mutation frequency, despite it being the most metabolically active tissue and having the highest mtDNA content. Whereas the liver has an elevated mutation frequency across mtDNA, the heart has concentrated peaks of mutation and biases mutations towards the non-coding *D-Loop*. These observations suggest that the heart and liver mutational landscapes may be shaped by different evolutionary, physiological, or molecular processes. Moreover, assessment of mutation types demonstrate that signatures associated with mtDNA replication processes and metabolic damage differ between tissues, emphasizing that tissue physiology influences mutation in the mt-genome. Throughout these analyses the heart and liver have the most contrasting mutational profiles. While the brain shows intermediate trends, it is a tissue that is tightly linked to mtDNA and mitochondria. The brain is among the most metabolically active tissues, is sensitive to mtDNA mutation [86], and is implicated in mtDNA-associated disorders [37]. In fact, most age-related mtDNA diseases have neurological symptoms [65]. Moreover, the severity of neurological disorders and the abundance of mutations in mtDNA both worsen with age, leading to the hypothesis that mtDNA mutation plays a role in neurodegenerative disease [17]. Due to this tight interplay between mtDNA mutation and the Central Nervous System, this last chapter focuses on characterizing the mtDNA mutational landscape in two regions of the brain: the cortex and cerebellum.

3.1 Summary

Mitochondrial DNA (mtDNA) encodes integral components necessary for cellular energy production. Mutations in mtDNA can vary in impact depending on when, where in the mtDNA molecule, and what tissue these mutations occur in. Throughout aging, mutations accumulate and are distributed in a tissue-specific manner, indicating that tissue physiology and metabolism shape the origin and trajectory of mutation in mtDNA. Importantly, late-onset mitochondrial diseases are generally characterized by neurological symptoms, emphasizing the relationship between the Central Nervous System and mtDNA. Here, we investigate the tissue-specific mutational landscapes in the aging brain. We employ Duplex Sequencing to profile mtDNA in the cortex and cerebellum across three different timepoints in the mouse lifespan. We profile ∼184,000 mt-genomes and catalog 81,221 variants, where 5,343 of these variants are unique mutational events. We highlight a consistent feature across tissues: the inheritance of INDELs that are predominantly located in regions associated with mtDNA replication. Yet, we show that shared mutations between these tissues have distinct frequencies throughout aging, likely due to genetic drift. We pinpoint features that demonstrate differences in metabolic demand, such as the cortex's 6-fold higher mtDNA copy number than the cerebellum, reflective of its higher metabolic demand. Additionally, the cortex has a higher average mutation frequency than the cerebellum, but does not differ in average mutation rate. Examination at the resolution of mutation type, however, uncovers differences among mutational signatures associated with replication error and metabolic damage. While the cortex has a higher frequency of replication-associated error (G>A/C>T), the tissues do not differ in mutation rate throughout aging. Despite being a more metabolically active tissue, the cortex exhibits a decrease in mutations associated with metabolic damage throughout aging, while the cerebellum shows the opposite trend. Together, these findings emphasize mtDNA mutational differences driven by tissue physiology and development.

3.2 Introduction

Mitochondria are essential for cellular energy production, and contain their own high copy number genome (mtDNA, mt-genomes) that encodes 13 vital components necessary for mitochondrial function [55]. Individual cells can contain hundreds to thousands of mtDNA molecules [11, 32, 24], with this abundance varying across cell types [76]. Mutations in this population can be present across all mtDNA molecules (homoplasmic variants) or in a subset of mtDNA (heteroplasmic variants). The frequency that a mutation needs to reach to have an impact, however, varies according to when, where in the mtDNA molecule, and what tissue the mutation arose in [86]. Both deletions [19, 17] and single nucleotide variants (SNVs) [60, 4, 3, 86] have been shown to aggregate in a tissue-specific manner. Deletions accumulate at a higher frequency in highly metabolic, post-mitotic tissues, such as the brain and muscle [19]. On the other hand, SNVs are present at higher frequencies in proliferative tissues (e.g. kidney and liver) [86]. Moreover, the distribution of mutations across the mtDNA molecule

can also vary between tissues [91]. Taken together, these studies demonstrate that tissue physiology and metabolic requirements influence the origin and trajectory of mutations in mtDNA.

Mutations in mtDNA have been associated with a variety of diseases that impact a wide range of tissues (reviewed in [37, 65]). Notably, mitochondrial diseases in children often impact multiple organ systems, while late-onset disease symptoms are typically neurological [65]. The brain is among the most metabolically active [99] and energy demanding tissues [40]. Thus, unsurprisingly, tissues that comprise the Central Nervous System (CNS) are particularly perturbed by mitochondrial dysfunction, and have a lower frequency of mutations associated with metabolic damage compared to other tissues [86]. Coinciding with the heightened susceptibility and worsening of neurological symptoms, mutations in mtDNA increase in frequency with age, leading to the hypothesis that mtDNA plays a role in neurodegenerative diseases [17, 52, 38, 10]. Supporting this hypothesis an increased prevalence of mtDNA mutations has been observed in individuals with neurodegenerative diseases [45, 18]. Additionally, genes encoded on the nuclear genome that localize to mitochondria have been recently implicated in modulating Tau aggregates, further tying mitochondrial function to neurodegenerative disease [78]. This tight interplay between the CNS and mtDNA mutation demands a full characterization of the mtDNA mutational landscape in order to understand the implications of mtDNA in aging and neurodegenerative disease.

Here, we compare the mitochondrial mutational landscapes of two brain regions, the cortex and cerebellum, across three timepoints in the mouse lifespan. These tissues were chosen due to their contrasting metabolic demands [40] and stark differences in the accumulation of deletions throughout aging [17]. We use Duplex Sequencing [89] to profile these mt-genomes at a high level of depth and accuracy, allowing us to examine differences in mutational processes. We used these profiles to survey the full gamut of mutations. We identify differences caused by genetic drift throughout aging, elevated mutation frequencies in the cortex, and distinct mutational profiles between the tissues. Together, our findings show the diverse mtDNA mutational landscapes that likely arise from differences in tissue development and physiology.

3.3 Results

The cortex has a 6-fold higher mtDNA copy number than the cerebellum

To compare mtDNA mutation within the brain, the cortex and cerebellum were sampled from C57BL/6J (B6) mice at three different ages across the mouse lifespan: young (4 months old (mo.); n = 3 mice), mid (13 mo.; n = 3 mice), and aged (20 mo.; n = 3 mice) for a total of n = 9 mice. These two tissues differed in their mtDNA content, with the cortex on average having a 6-fold higher mtDNA copy number than the cerebellum **(Fig. 3.1A)**.

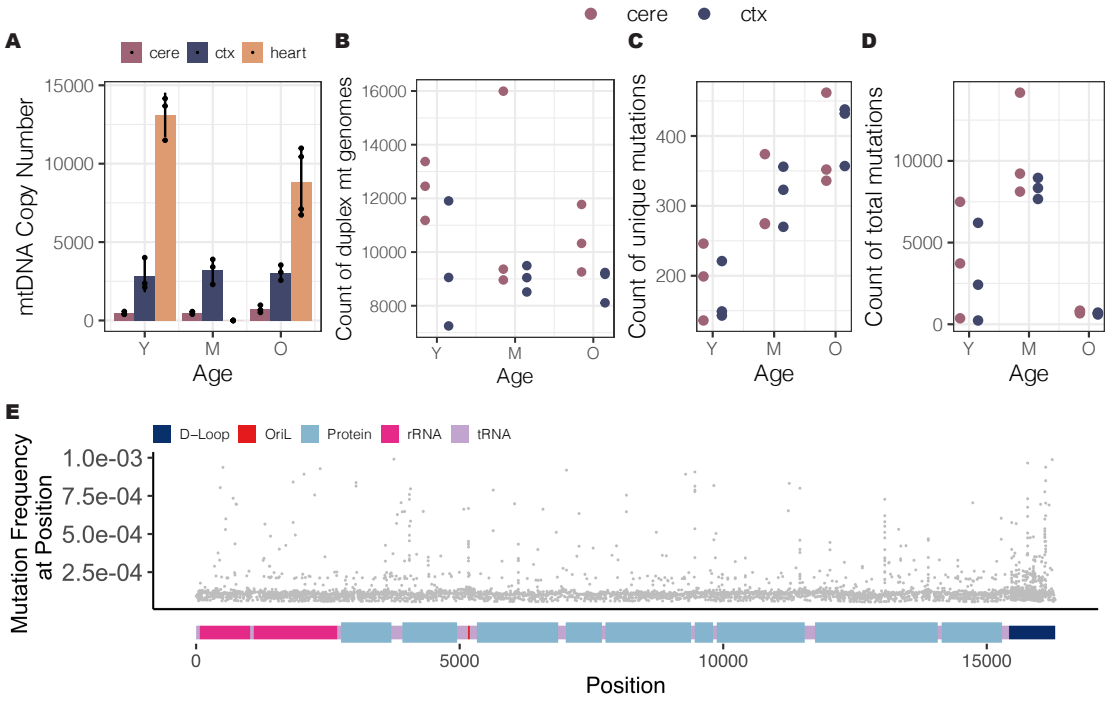


Figure 3.1: **Duplex sequencing of the cortex and cerebellum.** (A) Mitochondrial and nuclear copy numbers were quantified using qPCR. The mitochondrial to nuclear copy number ratio is used as an estimate of mitochondrial copy number in a tissue. Data for the heart is from *Serrano et al. 2024*, where heart tissues were sampled from young (2-3 mo.; n = 3 mice) and aged (18-19 mo.; n = 4 mice) B6 mice. Points denote the measure for each sample, the bar represents the average mtDNA copy number, and error bars reflect the standard deviation. (B) The count of duplex mt-genomes was calculated for each sample (n = 18 samples). The average duplex read depth is the average depth across positions in the mt-genome. (C) Mutations at each position in the mt-genome were counted once to create a proxy for unique mutational events in mtDNA. These unique mutational events were then aggregated across positions in the mtDNA molecule for each sample. (D) The alternate allele depths for the duplex consensus sequences were aggregated across positions. This count denotes the total number of alternate alleles present in the population of mtDNA molecules in a sample. (E) The mutation frequency for each position is mapped along a linear representation of the mt-genome. The mt-genome was segmented into: the non-coding *D-Loop* (dark blue), Light strand Origin of Replication (*OriL*; red), protein coding regions (light blue), rRNA coding regions (magenta), and tRNA coding regions (light purple). Each point is the mutation frequency at the position (bp) for a sample. Mutation frequency is calculated as the count of alternate alleles divided by the duplex depth at the position. Positions with a mutation frequency $> 1 \times 10^{-3}$ were excluded from this analysis. For (B) - (D) each point denotes a sample. For both the cortex and cerebellum across all three age bins, n = 3 mice, resulting in n = 18 samples.

The mtDNA copy number was on average 3,030 copies for the cortex and 560 copies for the cerebellum throughout the mouse lifespan. This difference in copy number, however,

is only significant in aged mice (paired, two-tailed t-test, adjusted p-value = 0.01 with a Benjamini-Hochberg correction). The mtDNA copy number does not significantly fluctuate with age for either cortex or cerebellum (paired, two-sided t-test between young and aged mice; p-value = 0.77 and p-value = 0.29, respectively). To place the brain in the context of other highly metabolic tissues, mtDNA copy number measurements were extracted from *Serrano et al. 2024* for young (2-3 mo.; n = 3 mice) and aged (18-19 mo.; n = 4 mice) B6 mice [91]. On average, the heart had a 2-fold and 12-fold higher mtDNA copy number than the cortex and cerebellum, respectively. These trends replicate the difference in mtDNA copy number between whole brain and heart [91].

Duplex sequencing of 184,279 mt-genomes in the cortex and cerebellum

Duplex sequencing (DupSeq) was used to profile the mt-genome. Briefly, DupSeq uses unique molecular identifiers to tag double-stranded mtDNA molecules, allowing for the computational construction of duplex consensus sequences [89]. Through this technique, first round PCR errors and sequencing artifacts are filtered, resulting in an error rate of $\sim 2 \times 10^{-7}$ errors per bp [1]. However, false variant calls may arise due to mtDNA damage during library preparation, often appearing at the end of reads [1]. To address this issue, the duplex consensus reads were trimmed an additional 10 bp **(Supplementary Note B.1)**. Using DupSeq, approximately 184,279 mt-genomes were profiled, with an average of 9,966 mt-genomes per sample **(Fig. 3.1B)**. This depth allowed us to capture mutations present at a frequency as low as 6.3×10^{-5}.

Duplex consensus reads were constructed and mapped to the mouse reference genome (mm10), and variants were called using a DupSeq processing pipeline (see Methods). In total, 5,343 unique mutational events were identified, with an average of 297 unique mutations per sample **(Fig. 3.1C)**. This resulted in a total of 81,221 variants – comprising de novo, inherited, and clonally expanded mutations – across the population of mt-genomes, with an average of 4,512 variants per sample **(Fig. 3.1D)**. After filtering high frequency positions ($> 1 \times 10^{-3}$) that likely contain inherited mutations, remaining mutations were widely distributed across the entire mtDNA molecule **(Fig. 3.1E)**.

The cortex has a 1.2-fold higher mutation frequency than the cerebellum

An age-associated increase in mutation frequencies has been observed across a multitude of tissues, including the whole brain [19, 17, 51, 4, 3, 86, 91]. Throughout aging, mutations can accumulate in the population of mt-genomes as a result of de novo somatic mutations or the replication of mutated mtDNA molecules. To compare the degree of mutation between the cortex and cerebellum, the mt-genome wide average mutation frequency was calculated **(Fig. 3.2A)**. The cortex consistently had a 1.2-fold higher mutation frequency than the

cerebellum. Indeed, tissue-specific mutation frequencies are present within the brain; however, the mutation rates are not significantly different between the cortex and cerebellum (ANOVA comparison of $mutfreq \sim age + tissue$ and $mutfreq \sim age \times tissue$; p-value = 0.08). Nonetheless, both tissues exhibit an age-associated increase in mutation frequency.

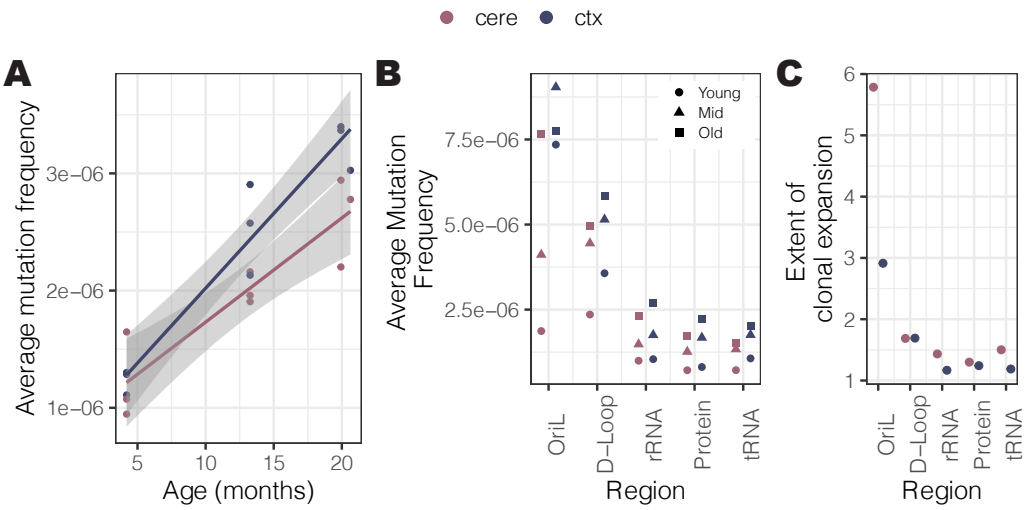


Figure 3.2: **Tissue-specific mutation frequencies.** (A) The average mutation frequency for all samples in the cortex (dark purple) and cerebellum (light purple). Each point represents a sample's average mutation frequency (n = 18 samples). (B) The mt-genome was categorized into five regions: *D-Loop*, *OriL*, protein, rRNA, and tRNA coding regions (x-axis). A de novo mutation count was calculated by counting each unique mutation once. The average de novo mutation frequency for each region was calculated as the de novo mutation count divided by the average read depth of each region multiplied by the length of each region. Each point denotes the average mutation frequency for each sample, where shape represents age (circle (young), triangle (mid), and square (aged)) and color highlights tissue (cortex (dark purple), cerebellum (light purple)). (C) The extent of allele propagation via replication of mtDNA is measured for each region in the mt-genome. A clonal mutation frequency was calculated as the sum of alternate alleles divided by the average read depth multiplied by the length of each region. The fold change between clonal and de novo mutation frequencies was calculated for each sample. The average fold change was calculated for samples across each tissue. Each point denotes the average fold change for the cortex (dark purple) and cerebellum (light purple).

De novo mutation frequencies were then measured for each region in the mt-genome to compare the distribution of mutations across the mtDNA molecule. Mutation counts and read depths were aggregated within experimental conditions (*age x tissue*) to increase power. Alternate alleles with a frequency greater than 1×10^{-3} were filtered from each sample, as these mutations were likely inherited. Remaining mutations were counted once, assuming that all other instances of this mutation are due to clonal expansion. In both

tissues, the Origin of Replication for the Light Strand (*OriL*) had the highest mutation frequency, surpassing the *D-Loop* **(Fig. 3.2B)**. This 31 bp region had a 2-fold and 8-fold higher mutation frequency in young, and a 4-fold and 3-fold higher frequency relative to RNA coding regions in aged cerebellum and cortex, respectively. The *D-Loop* had on average a 3-fold higher mutation frequency than RNA coding regions throughout aging in both tissues **(Fig. 3.2B)**. Consistent across both tissues, the *D-Loop* had a 2-fold and 1.6-fold increase in mutation frequency with age in the cerebellum and cortex, respectively. Interestingly, protein and tRNA coding regions had the lowest mutation frequency in aged mice across both tissues **(Fig. 3.2B)**. Yet, coding regions – rRNA, tRNA, and protein – had 2.4 and 2.6-fold increases in frequency with age in the cerebellum and cortex, respectively, changes similar in magnitude to the *D-Loop*. These results echo the age-related increase in average mtDNA mutation frequency, where the mutation frequency increases across all regions in the mtDNA molecule with age – with exception of the OriL in the cortex.

De novo mutation frequencies are a measure of mutagenesis in a tissue. However, abundance of these mutations is likely driven by the replication of mutated mtDNA. To compare the propagation of mutations between both tissues, the extent of clonal expansion was also measured for each region in the mt-genome. Unlike de novo mutation frequencies which counted each mutation once, the clonal mutation frequency considers all alternate allele counts. The extent of clonal expansion was then calculated as the fold change between clonal and de novo mutation frequencies. In both tissues, the *D-Loop* and *OriL* had a higher extent of expansion than coding regions **(Fig. 3.2C)**. For rRNA and tRNA, the cerebellum had a 43 and 50% increase in mutation frequency while the cortex had a 16 and 18% increase in frequency due to clonal expansion, respectively **(Fig. 3.2C)**. Overall, these results capture the elevated abundance of mutation in the cortex, and the similar distributions of mutation along the mtDNA molecule.

Evidence of genetic drift shaping distinct mutational profiles for shared mutations between tissues

Heteroplasmic variants are either inherited [80, 103] or occur in early development and aging. Determining when mutations arose is a challenging feat; however, the presence of mutations across multiple tissues can be used to classify variants as either inherited or de novo. Mutations shared between tissues are likely either inherited, or occurred before the development of independent tissues during embryogenesis. Mutations that are tissue-specific are presumably de novo somatic mutations that occurred during aging [60, 4].

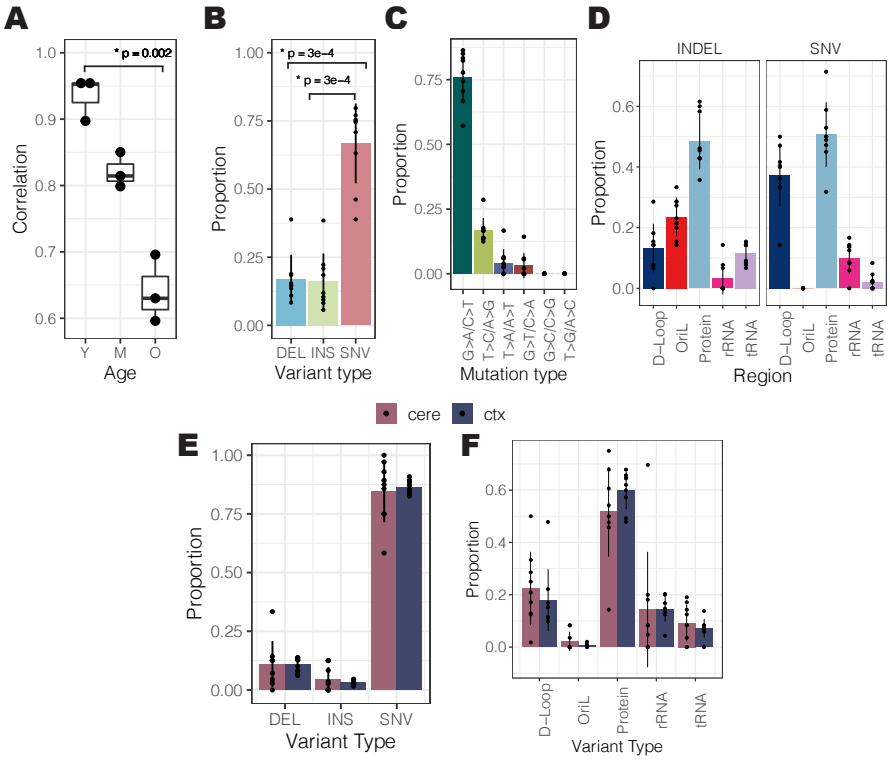


Figure 3.3: **Characterization of shared and tissue-specific mutations.** (A) The correlation in mutation frequency for shared mutations in the cortex and the cerebellum (reference **Figs.**B.1, B.2). Points denote the correlation for each sample (n = 3 mice per age group). Boxplots show the distribution of correlations in each age group. The line in each boxplot denotes the median, and the box locates the first and third quartiles. Young and aged correlations are significantly different (two-sided t-tests, p-value = 0.002). (B) Shared mutations are categorized into deletions (blue), insertions (green), and single nucleotide variants (SNVs, pink). The proportion of each category is calculated as the count of mutations in a category divided by the total number of shared mutations. SNVs are the prominent shared variant type (paired, two-sided t-tests; adjusted p-value = 3×10^{-4} for SNVs to DELs and SNVs to INS; p-values were corrected for multiple hypothesis testing using a Benjamini-Hochberg correction). (C) Shared SNVs were categorized into each mutation type. The proportion was calculated as the count of each mutation type divided by the total number of SNVs. (D) The distribution of shared INDELs and SNVs across regions of the mt-genome. The proportion is calculated as the count of mutations in each region divided by the total number of shared INDELs or SNVs. For (B) - (D), each point denotes the proportion for each sample (n = 9 mice). Bars denote the average proportion with the standard deviation shown. (E) The proportion of tissue-specific deletions, insertions, and SNVs in the cortex and cerebellum. Proportions are calculated as in (B). (F) The proportion of tissue-specific mutations in each region of the mt-genome. Proportions are calculated as the number of mutations in the region divided by the total count of tissue-specific mutations in a sample. For (E) - (F), each point denotes the proportion for each sample (n = 9 mice). Bars denote the average proportion, and the standard deviation is shown.

For each sample, variants present in both the cortex and cerebellum were identified. Across samples, 348 instances of shared mutations were found, where on average each sample had 42 shared mutations. The frequencies of these shared mutations were compared between tissues to determine if tissue-specific evolutionary processes distinctly shaped the trajectories of these mutations **(Figs. B.1, B.2)**. In young mice, the frequencies were highly correlated with an average correlation of 0.94 across samples **(Fig. 3.3A)**. Yet, this correlation significantly decreases with age, leading to an average correlation of 0.64 in aged samples (two-sided t-test; p-value = 0.002). Notably, these shared mutations do not preferentially increase in frequency in a specific tissue, suggesting that this age-associated decoupling in frequencies is driven by genetic drift **(Figs. B.1, B.2)**.

Shared mutations were predominantly single nucleotide variants (SNVs; paired, two-sided t-test adjusted p-value = 3×10^{-4} for SNVs v.s. DEL and SNVs v.s. INS using Benjamini-Hochberg correction) **(Fig. 3.3B)**. Despite this predominance, on average 33% and 38% of high frequency shared mutations were deletions and insertions (INDELs), respectively **(Fig. B.2)**. Thus, while SNVs are more likely to be shared, INDELs are present at a higher frequency in the population of mt-genomes. Further assessing SNVs, transitions comprised on average 92% of SNVs shared between the cortex and cerebellum, whereas mutational signatures of reactive oxygen species (ROS) damage accounted for on average 3%of shared mutations **(Fig. 3.3C)**. Shared SNVs and INDELs showed distinct distributions across portions of the mtDNA molecule **(Fig. 3.3D)**. While both SNVs and INDELs occurred in protein-coding regions at similar proportions (0.5 and 0.48, respectively), a higher proportion of INDELs were present in the *OriL* and tRNAs (paired, two-sided t-test; adjusted p-values = 2.1×10^{-5} and 2.4×10^{-3}, respectively using the Benjamini-Hochberg correction).

Tissue-specific mutations were identified by filtering out shared mutations and normalizing for read depth at each position across tissues in a sample. In total, 420 and 2,004 tissue-specific mutations were found in the cerebellum and cortex, respectively, across samples. As with shared mutations, tissue-specific mutations were primarily SNVs in both the cortex and cerebellum **(Fig. 3.3E)**. Tissue-specific mutations were also distributed similarly across the mtDNA molecule **(Fig. 3.3F)**, reflecting trends seen in **Fig. 3.2**.

Tissue-specific differences for replication-associated error and metabolic damage

Somatic mutations are caused by various molecular processes that lead to DNA damage, each of which exhibit a distinct mutational signature [2]. Comparing tissue-specific mutational signatures can provide insight into how the origin and trajectory of mutations is driven by distinct tissue physiologies. SNVs were the predominant class of tissue-specific mutations **(Fig. 3.3E)**. With respect to de novo mutations, the mutational signatures associated with replication errors in polymerase γ or cytosine deamination to uracil (G>A/C>T) [26, 113], and reactive oxygen species damage (G>T/C>A) [13] were the most abundant in both tissues **(Fig. 3.4A)**. On average the cortex had a higher abundance of replication

error (G>A/C>T), with a 1.5-fold higher frequency than the cerebellum in young and aged samples. Interestingly, the ROS-damage signature was 1.2-fold higher in the cortex for young tissues, but 1.8-fold higher in the cerebellum for aged tissues.

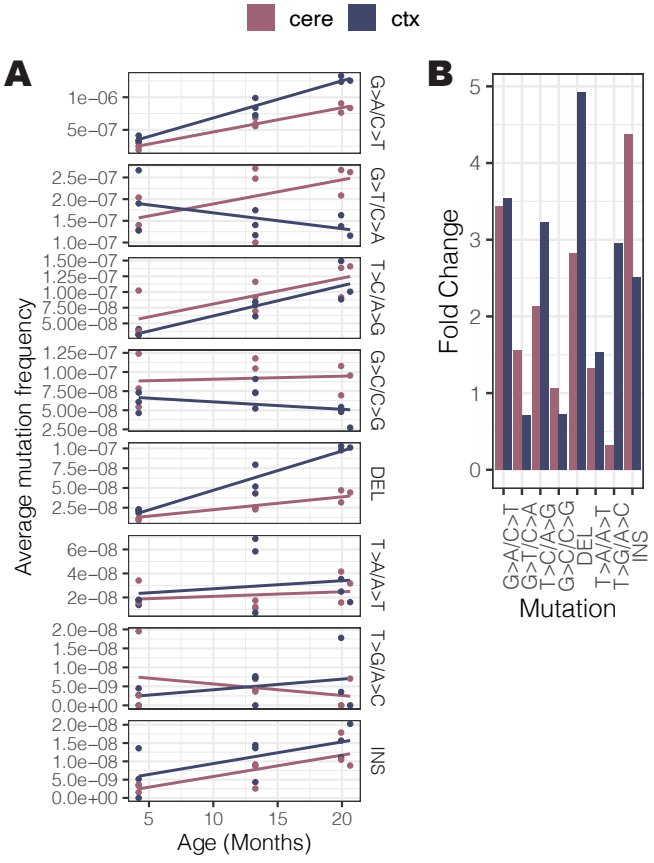


Figure 3.4: **Tissue-specific mutational signatures.** (A) A comparison of average de novo mutation frequencies for each mutation type. Each point denotes the average mutation frequency in a sample. For de novo mutation counts, each unique mutation was counted once. Average mutation frequencies were calculated by dividing the de novo mutation count by the reference duplex base depth for SNVs. For INDELs, the de novo mutation frequency was calculated by dividing the de novo mutation count by the total duplex base pair depth. Mutation types were arranged by descending average mutational frequencies. (B) For each mutation type, the average de novo mutation frequencies across samples in a condition were averaged (n = 3 mice per condition). The average mutation frequencies for aged samples were divided by the average mutation frequencies of young samples to calculate the fold change between the age groups. For this analysis, mutations shared between both tissues were excluded.

Overall, four mutation types (G>A/C>T, T>C/A>G, DEL, INS) monotonically increased in frequency throughout aging. Three mutation types had tissue-specific directional

changes (G>T/C>A, G>C/C>G, and T>G/A>C), while T>A/A>T remained constant through aging **(Fig. 3.4A)**. This suggests that replication-associated damage and INDELs accumulate in a "clock-like manner" consistently across both tissues. The G>A/C>T replication signature did not show a tissue-specific difference in fold change with age **(Fig. 3.4B)**. However, T>C/A>G and deletions had a higher fold change with age in the cortex **(Fig. 3.4B)**, implying that these mutations occur at a higher rate in the cortex than the cerebellum. The G>T/C>A signature and insertions had the inverse trend, with the cerebellum having a higher fold change with age. These findings demonstrate the tissue specificity of mutation type accumulation throughout aging, and pinpoint these differences in mutagenesis to replication and metabolic processes.

3.4 Discussion

Within an individual, the mtDNA mutational landscape varies across tissues and, with the introduction of de novo somatic mutations, continues to change throughout aging. Both deletions and SNVs show an age-associated increase in frequency across all tissues [17, 19, 51, 4, 3, 86]. Yet, the abundance and types of mutations present can diverge between tissues. These differences are likely driven by tissues' metabolic demands and subcellular mitochondrial demography, impacting both mutagenesis and evolutionary processes. While mtDNA mutation has been tied to a wide range of diseases that affect multiple tissue-systems, the CNS is of particular interest since late onset mitochondrial diseases primarily display neurological symptoms [52, 65]. Moreover, within the aging brain mitochondrial-related processes [114, 23] and mutational profiles [17, 86] have been shown to have region-specific changes. Here, we use Duplex Sequencing to accurately characterize the mtDNA mutational landscape in the cortex and cerebellum, two brain tissues that markedly differ in their metabolic demands [40] and deletion frequencies throughout aging [17]. To our knowledge, this is the first characterization of the cortex using ultra-sensitive sequencing technologies. We achieve a similar duplex depth ($\sim 10,000$X) as *Sanchez-Contreras et al.*, who have surveyed the hippocampus and retina [85]. Together, our studies have begun to catalog mtDNA mutation in healthy CNS tissues throughout aging.

At the onset, the cerebellum and cortex differ in their mtDNA copy number, with the cortex having a 6-fold higher copy number than the cerebellum. This result coincides with the cortex having a 10-fold higher glucose per neuron uptake than the cerebellum [40], demonstrative of its more intense metabolic demands. This higher copy number suggests that the replication of mtDNA may be needed to sustain the metabolic requirements in the cortex. In line with this trend, the heart has a higher mtDNA copy number than the cortex, reflecting the fact that the heart outranks the brain in metabolic activity, and further supporting the relationship between mtDNA abundance and metabolic demand [24].

While not a direct measure of the effective population size, the consensus mtDNA copy number can modulate the strength of evolutionary processes such as genetic drift and selection. The drastic difference in mtDNA copy number can consequently result in distinct

trajectories for shared mutations between tissues. Since frequencies for shared mutations did not tend to increase in one tissue over the other, genetic drift likely caused the distinct mutational trajectories throughout aging **(Figs. B.1, B.1)**. Previous works have shown that tissues that are more closely related with respect to their physiology and developmental histories tend to have higher correlations in shared mutation frequencies [3, 7]. On the other hand, *Barrett et al.* show that despite comprising the same tissue, hairs are highly divergent in shared mutation frequencies due to ongoing bottlenecks in hair follicles that limit mtDNA diversity in the source populations [7]. Our findings in conjunction with *Barrett et al.* emphasize that differences in mtDNA demography and population size are key properties that modulate the divergence in frequencies between aging tissues. In short, tissue-specific metabolic demands coincide with mtDNA abundance, resulting in variable mutation trajectories of likely inherited mutations throughout aging.

A more detailed characterization of shared mutations showed that inherited and/or variants that arose before the independent development of the cortex and cerebellum were primarily SNVs **(Fig. 3.3B)**. Despite this predominance, INDELs were present at a higher frequency in the population of mt-genomes **(Figs. B.1, B.1)**. While a germline tissue or additional tissues are needed to determine if these INDELs are inherited or de novo developmental mutations, their higher frequencies relative to other mutations likely make them inherited **(Fig.B.1)**. These INDELs are significantly more present in the *OriL* compared to SNVs **(Fig. 3.3D)**, placing INDELs in a region pertinent for mtDNA replication and transcription [55]. Additionally, tRNAs had a significantly higher proportion of INDELs than SNVs. Although not coding for origins of replications, mutations in tRNAs, specifically *MT-tRNA^{Arg}*, have been previously linked to an increase in mtDNA copy number [73]. Thus, the shared INDELs identified may play a role in mtDNA replication and copy number content. Our results align with a recent association study that surveyed mtDNA in blood cells of $\sim 275,000$ individuals, which found that 60% of individuals inherited INDELs [39]. Their results suggest that inherited INDELs serve as a calibration tool that tunes mtDNA with alleles present on the nuclear genome to confer compatibility in mtDNA replication and maintenance. Mutations in the *OriL* and *D-Loop* demonstrate the largest extent of clonal expansion **(Fig. 3.2C)**, supporting the idea that mutations in the *OriL* and *D-Loop* likely confer a replicative advantage. Despite these trends for shared INDELs, tissue-specific INDELs were not among the most prevalent de novo mutation types **(Fig. 3.4A)**. Again, this trend aligns with observations in *Gupta et al.*, where de novo somatic mutations were largely SNVs, whereas INDELs were typically inherited mutations. Contrasting their results, however, we do see a continual introduction of INDELs with age in both tissues **(Fig. 3.4A)**, a feature we would be able to capture with the sequencing sensitivity used in this study. Taken together, these results support the hypothesis that inherited INDELs play a key role in mtDNA replication and are a consistent tuning feature across tissues. Future works that utilize long read sequencing can determine if these mutated mtDNA molecules propagate the presence of other mutations.

Lastly, we honed in on tissue-specific changes in mutation abundance. The cortex had a higher average mutation frequency than the cerebellum, but did not have a higher mu-

tation rate **(Fig. 3.2A)**. Thus, these tissues have a similar propensity to overall mutation throughout aging. Additionally, the tissues had similar proportions of mutation classes **(Fig. 3.3E)**and distributions of de novo mutations across the mtDNA molecule **(Fig. 3.2B)**. Despite the cortex having similar or higher average de novo mutation frequencies across all regions of the mt-genome, the cerebellum had a larger extent of clonal expansion in the *OriL*, tRNA, and rRNA regions **(Figs. 3.2B,C)**. These results begin to show that while the cortex has a higher abundance of mutations, there may be tighter regulation on where mutations can increase in frequency along this tissue's mt-genome.

The cortex's higher mtDNA abundance and metabolic demands may imply that this tissue would be more vulnerable to mutations arising from mtDNA replication and/or metabolic processes. Indeed, higher frequencies for G>A/C>T mutations were observed for the cortex **(Fig. 3.4A)**. However, the G>A/C>T mutational signature had similar fold changes with age in the cortex and cerebellum **(Fig. 3.4B)**. Then, despite the drastic difference in mtDNA copy number content, the cortex does not introduce replication-associated errors at a faster rate than the cerebellum. Whether this observation is a result of differences in mtDNA replication rate or mechanisms regulating mutation necessitates further study. To be said, this finding dispels the notion that a high mtDNA copy number content is associated with increased replication errors. Aligning with results in other studies, there was a lack of ROS-damage associated mutational signatures [51, 4, 3, 86]. Surprisingly, there were tissue-specific directional differences in the accumulation of ROS-associated mutations (G>T/C>A) **(Fig. 3.4A)**. Although the cortex began with a higher frequency of G>T/C>A mutations, the cerebellum continually introduced ROS damage throughout aging while the cortex did not **(Fig. 3.4A,B)**. In fact, the unique presence of G>T/C>A mutations decreased, suggesting that these mutations are filtered from the population of mt-genomes in the cortex throughout aging. This distinction suggests that these two tissues differ in mechanisms that regulate ROS damage. A similar trend is seen in the heart [86]. Comparatively, the heart and cortex share many features: high metabolic demand, activity, and mtDNA copy number content, and are both post mitotic tissues. When comparing the heart to the cerebellum and hippocampus [45] – both of which have a lower mtDNA copy number than the cortex – the heart showed a higher frequency of G>T/C>A mutations in young tissues [86]. Yet, like the cortex, the heart showed a decrease in frequency for this mutation with age [86]. Departing from the cortex, however, the heart maintained a lower average mutation frequency, and lower mutation frequency for replication-associated mutations [86]. Thus, a mechanism uniquely aimed at regulating ROS-damage, but not other mutations, may be present in highly metabolic tissues.

While this work is able to categorize mutations as shared or tissue-specific, an experimental design that incorporates a broader set of tissues and oocytes is needed to more confidently record when mutations occur. Here, we conclude that shared mutations present at a higher frequency are likely inherited. We base this conclusion on the assumption that inherited mutations have the opportunity to undergo more cycles of mtDNA replication, and are among the mtDNA molecules amplified following the genetic bottleneck. However, shared mutations can also occur during embryonic development. By incorporating oocytes

into a study design, as done in *Arbeithuber et al. 2020 and 2022* [4, 3], we can distinguish inherited mutations (present in both the oocyte and somatic tissues) from early embryonic mutations (present only in somatic tissues). While we assume that tissue-specific mutations likely occurred during aging, we cannot discount the case where mitochondrial haplotypes are present in one cell but not another. This would result in cases where low frequency inherited or early embryonic mutations are present in one tissue, but not another. Given the assumption that older mutations would have more time to increase in frequency as a result of mtDNA replication, these mutations would likely manifest as high frequency, tissue-specific mutations. Our current data can be used to identify these variants, but further analyses would be needed to determine whether these are in fact uniquely inherited haplotypes or are de novo somatic mutations that rose in frequency as a result of selection. Dating these mutations and characterizing the mtDNA mutational landscape through time can provide insight into when pathogenic mutations arise and potentially aid in predicting the trajectory of these mutations throughout an individual's lifespan.

Altogether, our findings highlight consistent features of mtDNA mutation in the brain, such as the inheritance of INDELs in replication-associated regions, increased mutation frequency with age, and the distribution of de novo mutations along the mtDNA molecule. However, we find differences in the details, finding distinct accumulation of replication and metabolic associated mutations, and diverged trajectories in frequencies for mutations present in both tissues. Thus, tissue physiology and its implications on the intensity of metabolic processes and mtDNA abundance shape mutational processes in mtDNA throughout aging.

3.5 Methods

Ethics Statement

All animal procedures were approved by the UC Berkeley institutional animal care committees. Animals were housed with a 12:12 light:dark cycle with food and water available ad libitum. Experiments were performed in accordance with the relevant guidelines and regulations by certified personnel.

Tissue sampling

Mice were anesthetized with Euthasol euthanasia solution. Brains were removed, freshly dissected. All mice in this experiment are female.

Sequencing

Duplex sequencing libraries were prepared by TwinStrand Biosciences (Seattle, WA) as previously described [89, 51]. Sequencing was performed at the University of California, Berkeley

using Illumina NovaSeq generated 150 bp paired-end reads.

Data processing

Sequencing reads were processed into duplex consensus sequences and mapped to the reference mouse mitochondrial genome (mm10) using the Duplex Sequencing processing pipeline developed by Dr. Scott Kennedy's group at the University of Washington, Seattle. Software versions used along with all processing steps including duplex consensus sequence generation, mapping, and variant calling can be found in the github repository referenced below. Parameters used to run the Duplex Sequencing pipeline are provided in **Supplemental Note B.1**.

Variant Calling

Mutations were identified using two variant callers: vardict-java (1.8.0) and pysam (0.15.3). Mutation call sets were compared across the two callers. The intersection of the sets was used to construct the final call set. From the intersection, mutations that had an overabundant alternate allele depth in pysam were removed (difference in alternate allele depth > 5). Variants present in both vardict and pysam sets, but with a higher alternate allele depth in vardict, were kept. The pysam alternate allele depth for these variants was used, providing conservative mutation frequency for these mutations.

Statistical analyses: Testing for significance

Statistical analyses were performed using R (4.1.2). Two tailed-tests, and when appropriate paired t-tests, were conducted to test for differences between averages across conditions *(t.test, stats package)*. Three linear models were fit to the average mutation frequencies in each tissue: $mutfreq \sim age$, $mutfreq \sim age + tissue$, $mutfreq \sim age \times tissue$ *(glm, stats package)*. An ANOVA was then used to compare these linear models *(anova, stats package)*. Regressions were fit to data using *geom smooth, method = "lm" (ggplot2)*. The Spearman correlation was calculated for frequencies of shared mutations between cortex and cerebellum *(cor, stats package)*. Multiple hypothesis test corrections were performed using the Benjamini-Hochberg correction *(p.adjust, stats package)*, adjusted p-values refer to p-values that have undergone multiple hypothesis correction.

Chapter 4

Future Directions

4.1 Characterization of the mtDNA mutational landscape throughout aging: insights and limitations

This dissertation profiled mtDNA mutation throughout aging. Specifically, this work investigated how mitochondrial haplotype influences the accumulation and trajectory of new mutations. Moreover, I considered the constant interaction between mitochondrial and nuclear genomes by examining how mismatching between mitochondrial and nuclear ancestries further shapes these mutational landscapes. Addressing intra-individual variation, I explored how tissue physiology impacts mutational processes in mtDNA. By employing ultra-sensitive Duplex Sequencing, I generated mutational maps at an unprecedented depth and accuracy. This sequencing technology allowed me to capture low frequency mutations that likely arose during aging. Despite this resolution and sensitivity, our work demonstrates the challenges in assessing mutation in mtDNA, and the complexity underpinning mutation in this small genome. Here, I summarize our findings and expand on ideas that can help gain insight into mutation and the evolution of mtDNA.

In Chapter Two, I duplex sequenced the mt-genomes of four mouse strains bred to have identical nuclear genomes but that differ in their mitochondrial haplotypes (conplastic mouse strains). I characterized and compared the mtDNA mutational landscapes for these conplastic mouse strains and the control B6 mouse, which has matching B6 nuclear and mitochondrial ancestry. Through this study, I identified a mutational hotspot in $MT\text{-}tRNA^{Arg}$ that is uniquely present and increases in frequency in conplastic mice. This hotspot, and one found across all five mouse strains in the *OriL*, have been previously associated with mtDNA replication [74]. Examining age-related changes in mutation frequency, I found strain-specific mutation rates across all tissues, with some conplastic mice exhibiting a lower mutation frequency than wildtype. Lastly, I also found evidence of somatic reversion mutations that aim to realign the mitochondrial and nuclear ancestries specifically in conplastic mouse strains

that differ by 1-2 mutations from wildtype. These results show that mitochondrial ancestry and its interplay with nuclear ancestry influence mutation throughout aging, and shine a light on the importance of mtDNA replication in these mutational dynamics.

In Chapter Three, I further investigated tissue-specific effects observed in Chapter Two. Conclusions from both chapters demonstrated that mutation in the mt-genome primarily arises from processes associated with mtDNA replication. The cortex, which had a higher mtDNA content, does have a higher frequency of replication-associated mutations than the cerebellum. However, the heart, which had an even higher mtDNA copy number than the cortex, had a lower mutation frequency than the whole brain. Together, these findings showed that a higher mtDNA content, and presumed higher degree of replication, does not necessarily yield a higher mutation frequency. Interestingly, the frequency of mutations associated with metabolic damage decreased throughout aging in the cortex, reflecting similar trends previously observed in the heart [86]. This decrease in signatures of metabolic damage was not observed in the cerebellum, leading to the hypothesis that there may exist tissue-specific mechanisms that regulate the presence of metabolic damage in metabolically demanding tissues. Although the cortex and cerebellum differed in their aggregation of replication and metabolic associated mutations, they were alike in their inheritance of INDELs located in replication-associated regions. Yet, I found that even in the mutations that these two tissues share in common, there is a decoupling in mutation frequencies throughout aging. This phenomenon is likely due to varying strengths of genetic drift distinctly shaping the mutational landscapes in these tissues. These results demonstrate the complex relationship between tissue physiology and mtDNA mutation, calling for a deeper understanding of how mtDNA replication and metabolic processes differ between tissues.

4.2 Leveraging long read sequencing to investigate the segregation of mitochondrial haplotypes

In Chapter Three, I observed that high frequency shared mutations between the cortex and cerebellum were primarily INDELs. Compared to SNVs, INDELs were enriched in the *OriL* and tRNAs. Interestingly, INDELs specifically in *MT-tRNA*^{Arg} were a mutational hotspot for mice with mismatched mitochondrial and nuclear ancestry in Chapter Two. Importantly, the *OriL* and *MT-tRNA*^{Arg} have been shown to modulate mtDNA replication [74]. These results highlight that INDELs in replication associated regions are a potential commonality across tissues; yet, are impacted by mito-nuclear interactions. Indeed, a recent study showed that inherited INDELs are a common occurrence in human populations – albeit the frequency and the length of these INDELs differ [39]. Intriguingly, this study postulates that inherited INDELs may serve as a calibration tool with alleles in the nuclear genome to regulate mtDNA copy number [39]. These findings suggest that mitochondrial haplotypes with these INDELs have a replicative advantage, surviving the germline bottleneck and retaining their presence in the population. If so, inherited INDELs may play an important role in harmonizing the

two genomes and in driving the presence of co-segregating mutations.

Although Duplex Sequencing is highly sensitive, its use of short read sequencing prevents us from capturing the entirety of the mitochondrial haplotype. Thus, mutations segregating together cannot be identified unless they occur in the same read (< 150 bp in distance). By using long read sequencing, we can survey mtDNA molecules these INDELs inhabit. Long read sequencing of multi-generational pedigrees would provide the opportunity to measure the consistent composition of inherited INDELs (i.e. the proportion of each INDEL allele present) and track the inheritance of co-segregating alleles across generations. Moreover, by using individuals in the pedigree as different time points, we can assess how the frequency of these INDELs changes throughout aging. This characterization of INDELs would quantify the consistency of these INDELs in mtDNA. Moreover, these analyses would determine if INDELs have a replicative advantage, allowing these INDELs to serve as "vessels" that can drive the presence of co-segregating mutations on the same mitochondrial haplotype.

In addition to clarifying the trajectory of inherited INDELs, long read sequencing can allow us to study the replication processes of mtDNA. A current hypothesis concerns the existence of mitochondrial subpopulations that dictate the replication of specific, mutation-limited mtDNA molecules [20]. Long read sequencing would enable us to determine if signatures of replication-associated mutations and metabolic damage exist on distinct mitochondrial haplotypes, and determine if there are key characteristics that distinguish these haplotypes. Such characteristics would in turn allow us to identify mitochondrial haplotypes that are preferentially replicated. Overall, profiling mitochondrial haplotypes would allow us to further investigate the dynamics of mtDNA replication, and the subsequent impact on variant transmission and mechanisms regulating mtDNA mutation.

4.3 Decoupling signatures of cellular proliferation and mtDNA replication

Errors during mtDNA replication were the primary cause of mtDNA mutation across tissues and mitochondrial haplotype in my work. Moreover, these mutational signatures consistently increased in frequency with age, unlike the tissue-specific directional changes observed for metabolic damage. Yet, the relationship between mtDNA content and mutational signatures associated with replication error is not straightforward. That is, tissues with a higher mtDNA copy number do not necessarily have an abundance of replication error. The heart and cortex share key characteristics; they are both postmitotic tissues and have a high metabolic demand. Based on these similarities, I expected these tissues to share mutational profiles, especially in comparison to tissues with a lower mtDNA content. While these tissues share a decrease in metabolic damage with age (see [86] for results in the heart), the heart has a lower mutation frequency of G>A/C>T mutations compared to the whole brain. This difference in replication-associated mutation frequencies can be due to differences in: 1) cellular composition 2) mtDNA replication rates 3) mechanisms that regulate

mutation 4) some degree of all these possibilities. Mitochondrial DNA replication may occur independently of cell proliferation; thus, although both tissues are post mitotic they may differ in mtDNA replication rates. Duplex sequencing data can be used to capture differences in mtDNA replication rate. Transitions (G>A/C>T, T>C/A>G) are associated with mtDNA replication, and the strand bias of complementary mutations is likely a byproduct of mtDNA replication [51]. By comparing the degree of strand bias and the proportion of transitions to transversions for de novo somatic mutations, we can begin to uncover tissue-specific mtDNA replication processes. Importantly, sequencing bulk tissue loses the distinct characteristics of cells, including cellular energy demands and proliferative activities that can impact mtDNA replication. Using single-cell sequencing technologies that have been applied to survey mtDNA (reviewed in [76]), we can profile the cell-specific mutational spectra while also gaining insight into active cellular pathways via transcriptomic data.

The results in this dissertation congregate around the role of mtDNA replication and how this process can introduce or potentially amplify specific variants. By employing long read and single cell sequencing technologies, we may be able to uncover consistent features in the mt-genome that result in proliferative advantages, identify variants that co-segregate in the mt-genome, and determine if and how mtDNA replication coincides with cellular proliferation.

Bibliography

[1] Federico Abascal et al. "Somatic mutation landscapes at single-molecule resolution". en. In: *Nature* 593.7859 (Apr. 2021), pp. 405–410.

[2] Ludmil B. Alexandrov et al. "Signatures of mutational processes in human cancer". In: *Nature* 500.7463 (Aug. 2013), pp. 415–421. ISSN: 1476-4687. DOI: 10.1038/nature12477. URL: http://dx.doi.org/10.1038/nature12477.

[3] Barbara Arbeithuber et al. "Advanced age increases frequencies of de novo mitochondrial mutations in macaque oocytes and somatic tissues". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 119.15 (Apr. 2022), e2118740119.

[4] Barbara Arbeithuber et al. "Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues". en. In: *PLoS Biol.* 18.7 (July 2020), e3000745.

[5] Tara Z Baris et al. "Evolved genetic and phenotypic differences due to mitochondrial-nuclear interactions". en. In: *PLoS Genet.* 13.3 (Mar. 2017), e1006517.

[6] Felipe S Barreto and Ronald S Burton. "Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA". en. In: *Mol. Biol. Evol.* 30.2 (Feb. 2013), pp. 310–314.

[7] Alison Barrett et al. "Pronounced somatic bottleneck in mitochondrial DNA of human hair". In: *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375.1790 (Dec. 2019), p. 20190175.

[8] Brendan J Battersby, J C Loredo-Osti, and Eric A Shoubridge. "Nuclear genetic control of mitochondrial DNA segregation". en. In: *Nat. Genet.* 33.2 (Feb. 2003), pp. 183–186.

[9] Brendan J Battersby and Eric A Shoubridge. "Selection of a mtDNA sequence variant in hepatocytes of heteroplasmic mice is not due to differences in respiratory chain function or efficiency of replication". en. In: *Hum. Mol. Genet.* 10.22 (Oct. 2001), pp. 2469–2479.

[10] Andreas Bender et al. "High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease". In: *Nature Genetics* 38.5 (Apr. 2006), pp. 515–517. ISSN: 1546-1718. DOI: 10.1038/ng1769. URL: http://dx.doi.org/10.1038/ng1769.

[11] Daniel Bogenhagen and David A. Clayton. "The Number of Mitochondrial Deoxyribonucleic Acid Genomes in Mouse L and Human HeLa Cells". In: *Journal of Biological Chemistry* 249.24 (Dec. 1974), pp. 7991–7995. ISSN: 0021-9258. DOI: 10.1016/s0021-9258(19)42063-2. URL: http://dx.doi.org/10.1016/S0021-9258(19)42063-2.

[12] Alex Cagan et al. "Somatic mutation rates scale with lifespan across mammals". en. In: *Nature* 604.7906 (Apr. 2022), pp. 517–524.

[13] K C Cheng et al. "8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G-T and A-C substitutions." In: *Journal of Biological Chemistry* 267.1 (Jan. 1992), pp. 166–172. ISSN: 0021-9258. DOI: 10.1016/s0021-9258(18)48474-8. URL: http://dx.doi.org/10.1016/S0021-9258(18)48474-8.

[14] Kenneth L Chiou et al. "Rhesus macaques as a tractable physiological model of human ageing". en. In: *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375.1811 (Nov. 2020), p. 20190612.

[15] Jui-Yu Chou et al. "Multiple molecular mechanisms cause reproductive isolation between three yeast species". en. In: *PLoS Biol.* 8.7 (July 2010), e1000432.

[16] Sangita Choudhury et al. "Somatic mutations in single human cardiomyocytes reveal age-associated DNA damage and widespread oxidative genotoxicity". en. In: *Nature Aging* 2.8 (Aug. 2022), pp. 714–725.

[17] M Corral-Debrinski et al. "Mitochondrial DNA deletions in human brain: regional variability and increase with advanced age". In: *Nat. Genet.* 2 (Dec. 1992), pp. 324–329.

[18] Marisol Corral-Debrinski et al. "Marked Changes in Mitochondrial DNA Deletion Levels in Alzheimer Brains". In: *Genomics* 23.2 (Sept. 1994), pp. 471–476. ISSN: 0888-7543. DOI: 10.1006/geno.1994.1525. URL: http://dx.doi.org/10.1006/geno.1994.1525.

[19] G A Cortopassi et al. "A pattern of accumulation of a somatic deletion of mitochondrial DNA in aging human tissues". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 89.16 (Aug. 1992), pp. 7370–7374.

[20] Auden Cote-L'Heureu et al. "The 'Stem' and the 'Workers' of the mtDNA population of the cell. Evidence from mutational analysis". In: *bioRxiv* (Apr. 2023). DOI: 10.1101/2023.04.14.536897. URL: http://dx.doi.org/10.1101/2023.04.14.536897.

[21] Auden Cote-L'Heureux et al. "Are some mutations more equal than others?" In: *eLife* 12 (Apr. 2023). ISSN: 2050-084X. DOI: 10.7554/elife.87194. URL: http://dx.doi.org/10.7554/eLife.87194.

[22] Lynsey M Cree et al. "A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes". en. In: *Nat. Genet.* 40.2 (Feb. 2008), pp. 249–254.

[23] Stephen C. Cunnane et al. "Brain energy rescue: an emerging therapeutic concept for neurodegenerative disorders of ageing". In: *Nature Reviews Drug Discovery* 19.9 (July 2020), pp. 609–633. ISSN: 1474-1784. DOI: 10.1038/s41573-020-0072-x. URL: http://dx.doi.org/10.1038/s41573-020-0072-x.

[24] Anna Maria D'Erchia et al. "Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity". In: *Mitochondrion* 20 (Jan. 2015), pp. 13–21. ISSN: 1567-7249. DOI: 10.1016/j.mito.2014.10.005. URL: http://dx.doi.org/10.1016/j.mito.2014.10.005.

[25] Petr Danecek et al. "Twelve years of SAMtools and BCFtools". In: *GigaScience* 10.2 (Jan. 2021). ISSN: 2047-217X. DOI: 10.1093/gigascience/giab008. URL: http://dx.doi.org/10.1093/gigascience/giab008.

[26] Bruce K. Duncan and Jeffrey H. Miller. "Mutagenic deamination of cytosine residues in DNA". In: *Nature* 287.5782 (Oct. 1980), pp. 560–561. ISSN: 1476-4687. DOI: 10.1038/287560a0. URL: http://dx.doi.org/10.1038/287560a0.

[27] C K Ellison and R S Burton. "Cytonuclear conflict in interpopulation hybrids: the role of RNA polymerase in mtDNA transcription and replication". en. In: *J. Evol. Biol.* 23.3 (Mar. 2010), pp. 528–538.

[28] C K Ellison, O Niehuis, and J Gadau. "Hybrid breakdown and mitochondrial dysfunction in hybrids of Nasonia parasitoid wasps". en. In: *J. Evol. Biol.* 21.6 (Nov. 2008), pp. 1844–1851.

[29] Christopher K Ellison and Ronald S Burton. "Disruption of mitochondrial function in interpopulation hybrids of Tigriopus californicus". en. In: *Evolution* 60.7 (July 2006), pp. 1382–1391.

[30] Christopher K Ellison and Ronald S Burton. "Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 105.41 (Oct. 2008), pp. 15831–15836.

[31] Christopher K Ellison and Ronald S Burton. "Interpopulation hybrid breakdown maps to the mitochondrial genome". en. In: *Evolution* 62.3 (Mar. 2008), pp. 631–638.

[32] Roberta Filograna et al. "Mitochondrial DNA copy number in human disease: the more the better?" In: *FEBS Letters* 595.8 (Dec. 2020), pp. 976–1002. ISSN: 1873-3468. DOI: 10.1002/1873-3468.14021. URL: http://dx.doi.org/10.1002/1873-3468.14021.

[33] Vasileios I Floros et al. "Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos". en. In: *Nat. Cell Biol.* 20.2 (Jan. 2018), pp. 144–151.

[34] Aurora Gómez-Durán et al. "Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy". en. In: *Biochim. Biophys. Acta* 1822.8 (Aug. 2012), pp. 1216–1222.

[35] Aurora Gómez-Durán et al. "Unmasking the causes of multifactorial disorders: OXPHOS differences between mitochondrial haplogroups". en. In: *Hum. Mol. Genet.* 19.17 (Sept. 2010), pp. 3343–3353.

[36] Kevin Gori and Adrian Baez-Ortega. "sigfit: flexible Bayesian inference of mutational signatures". In: *bioRxiv* (July 2018). DOI: 10.1101/372896. URL: http://dx.doi.org/10.1101/372896.

[37] Laura C. Greaves and Robert W. Taylor. "Mitochondrial DNA mutations in human disease". In: *IUBMB Life* 58.3 (Mar. 2006), pp. 143–151. ISSN: 1521-6551. DOI: 10.1080/15216540600686888. URL: http://dx.doi.org/10.1080/15216540600686888.

[38] Guangyu Gu et al. "Mitochondrial DNA Deletions/Rearrangements in Parkinson Disease and Related Neurodegenerative Disorders". In: *Journal of Neuropathology and Experimental Neurology* 61.7 (July 2002), pp. 634–639. ISSN: 1554-6578. DOI: 10.1093/jnen/61.7.634. URL: http://dx.doi.org/10.1093/jnen/61.7.634.

[39] Rahul Gupta et al. "Nuclear genetic control of mtDNA copy number and heteroplasmy in humans". en. In: *Nature* 620.7975 (Aug. 2023), pp. 839–848.

[40] Suzana Herculano-Houzel. "Scaling of Brain Metabolism with a Fixed Energy Budget per Neuron: Implications for Neuronal Activity, Plasticity and Evolution". In: *PLoS ONE* 6.3 (Mar. 2011). Ed. by Matjaz Perc, e17514. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0017514. URL: http://dx.doi.org/10.1371/journal.pone.0017514.

[41] Geoffrey E Hill. "Mitonuclear Compensatory Coevolution". en. In: *Trends Genet.* 36.6 (June 2020), pp. 403–414.

[42] Misa Hirose et al. "Lifespan effects of mitochondrial mutations". In: *Nature* 540.7633 (Dec. 2016), E13–E14. ISSN: 1476-4687. DOI: 10.1038/nature20778. URL: http://dx.doi.org/10.1038/nature20778.

[43] Misa Hirose et al. "Low-level mitochondrial heteroplasmy modulates DNA replication, glucose metabolism and lifespan in mice". en. In: *Sci. Rep.* 8.1 (Apr. 2018), pp. 1–15.

[44] Misa Hirose et al. "Mitochondrial gene polymorphism is associated with gut microbial communities in mice". en. In: *Sci. Rep.* 7.1 (Nov. 2017), p. 15293.

[45] Jake G. Hoekstra et al. "Mitochondrial DNA mutations increase in early stage Alzheimer disease and are inconsistent with oxidative damage". In: *Annals of Neurology* 80.2 (July 2016), pp. 301–306. ISSN: 1531-8249. DOI: 10.1002/ana.24709. URL: http://dx.doi.org/10.1002/ana.24709.

[46] Jack P Jenuth, Alan C Peterson, and Eric A Shoubridge. "Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice". en. In: *Nat. Genet.* 16.1 (May 1997), pp. 93–95.

[47] Fuyun Ji et al. "Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and high-altitude Tibetans". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 109.19 (May 2012), pp. 7391–7396.

[48] M J Johnson et al. "Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns". en. In: *J. Mol. Evol.* 19.3-4 (1983), pp. 255–271.

[49] Hákon Jónsson et al. "Parental influence on human germline de novo mutations in 1,548 trios from Iceland". en. In: *Nature* 549.7673 (Sept. 2017), pp. 519–522.

[50] Scott R Kennedy, Lawrence A Loeb, and Alan J Herr. "Somatic mutations in aging, cancer and neurodegeneration". en. In: *Mech. Ageing Dev.* 133.4 (Apr. 2012), pp. 118–126.

[51] Scott R Kennedy et al. "Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage". en. In: *PLoS Genet.* 9.9 (Sept. 2013), e1003794.

[52] Michael J. Keogh and Patrick F. Chinnery. "Mitochondrial DNA mutations in neurodegeneration". In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1847.11 (Nov. 2015), pp. 1401–1411. ISSN: 0005-2728. DOI: 10.1016/j.bbabio.2015.05.015. URL: http://dx.doi.org/10.1016/j.bbabio.2015.05.015.

[53] T Koevoets et al. "Hybrid incompatibilities in the parasitic wasp genus Nasonia: negative effects of hemizygosity and the identification of transmission ratio distortion loci". en. In: *Heredity* 108.3 (Mar. 2012), pp. 302–311.

[54] Anna V Kotrys et al. "Single-cell analysis reveals context-dependent, cell-level selection of mtDNA". en. In: *Nature* (Apr. 2024), pp. 1–9.

[55] Nils-Göran Larsson. "Somatic mitochondrial DNA mutations in mammalian aging". en. In: *Annu. Rev. Biochem.* 79 (2010), pp. 683–706.

[56] Dean Laslett and Björn Canbäck. "ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences". In: *Bioinformatics* 24.2 (Nov. 2007), pp. 172–175. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btm573. URL: http://dx.doi.org/10.1093/bioinformatics/btm573.

[57] Hsin-Yi Lee et al. "Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species". en. In: *Cell* 135.6 (Dec. 2008), pp. 1065–1073.

[58] Heng Li. "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv* (2013). DOI: 10.48550/ARXIV.1303.3997. URL: https://arxiv.org/abs/1303.3997.

[59] Mingkun Li et al. "Detecting Heteroplasmy from High-Throughput Sequencing of Complete Human Mitochondrial DNA Genomes". In: *The American Journal of Human Genetics* 87.2 (Aug. 2010), pp. 237–249. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2010.07.014. URL: http://dx.doi.org/10.1016/j.ajhg.2010.07.014.

[60] Mingkun Li et al. "Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations". In: *Proceedings of the National Academy of Sciences* 112.8 (2015), pp. 2491–2496.

[61] A W Linnane et al. "Mitochondrial DNA mutations as an important contributor to ageing and degenerative diseases". In: *The Lancet* (1989).

[62] Mei Hong Liu et al. "DNA mismatch and damage patterns revealed by single-molecule sequencing". en. In: *Nature* (June 2024), pp. 1–10.

[63] Michael A Lodato et al. "Aging and neurodegeneration are associated with increased mutations in single human neurons". en. In: *Science* 359.6375 (Feb. 2018), pp. 555–559.

[64] Carlos López-Otín et al. "Hallmarks of aging: An expanding universe". en. In: *Cell* (Dec. 2022).

[65] William L Macken et al. "Applying genomic and transcriptomic advances to mitochondrial medicine". en. In: *Nat. Rev. Neurol.* 17.4 (Apr. 2021), pp. 215–230.

[66] L A Marcelino and W G Thilly. "Mitochondrial mutagenesis in human cells and tissues". en. In: *Mutat. Res.* 434.3 (July 1999), pp. 177–203.

[67] Iñigo Martincorena et al. "Somatic mutant clones colonize the human esophagus with age". In: *Science* 362.6417 (2018), pp. 911–917.

[68] Colin D Meiklejohn et al. "An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in Drosophila". en. In: *PLoS Genet.* 9.1 (Jan. 2013), e1003238.

[69] Michael B Miller et al. "Somatic genomic changes in single Alzheimer's disease neurons". en. In: *Nature* 604.7907 (Apr. 2022), pp. 714–722.

[70] Kristi L Montooth et al. "Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of Drosophila". en. In: *Evolution* 64.12 (Dec. 2010), pp. 3364–3379.

[71] Hernán E Morales et al. "Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates". en. In: *Nat Ecol Evol* 2.8 (Aug. 2018), pp. 1258–1267.

[72] Benjamin M Moran et al. "A lethal mitonuclear incompatibility in complex I of natural hybrids". en. In: *Nature* 626.7997 (Jan. 2024), pp. 119–127.

[73] Raquel Moreno-Loshuertos et al. "Differences in reactive oxygen species production explain the phenotypes associated with common mouse mitochondrial DNA variants". In: *Nature Genetics* 38.11 (Oct. 2006), pp. 1261–1268. ISSN: 1546-1718. DOI: 10.1038/ng1897. URL: http://dx.doi.org/10.1038/ng1897.

[74] Raquel Moreno-Loshuertos et al. "Length variation in the mouse mitochondrial tRNAArg DHU loop size promotes oxidative phosphorylation functional differences". In: *The FEBS Journal* 280.20 (Sept. 2013), pp. 4983–4998. ISSN: 1742-4658. DOI: 10.1111/febs.12466. URL: http://dx.doi.org/10.1111/febs.12466.

[75] Oliver Niehuis, Andrea K Judson, and Jürgen Gadau. "Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of Nasonia giraulti and N. vitripennis". en. In: *Genetics* 178.1 (Jan. 2008), pp. 413–426.

[76] Lena Nitsch, Caleb A. Lareau, and Leif S. Ludwig. "Mitochondrial genetics through the lens of single-cell multi-omics". In: *Nature Genetics* 56.7 (July 2024), pp. 1355–1365. ISSN: 1546-1718. DOI: 10.1038/s41588-024-01794-8. URL: http://dx.doi.org/10.1038/s41588-024-01794-8.

[77] Naoki Osada and Hiroshi Akashi. "Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex". en. In: *Mol. Biol. Evol.* 29.1 (Jan. 2012), pp. 337–346.

[78] Celeste Parra Bravo et al. "Human iPSC 4R tauopathy model uncovers modifiers of tau propagation". In: *Cell* 187.10 (May 2024), 2446–2464.e22. ISSN: 0092-8674. DOI: 10.1016/j.cell.2024.03.015. URL: http://dx.doi.org/10.1016/j.cell.2024.03.015.

[79] Brendan A I Payne et al. "Deep resequencing of mitochondrial DNA". en. In: *Methods Mol. Biol.* 1264 (2015), pp. 59–66.

[80] Brendan A I Payne et al. "Universal heteroplasmy of human mitochondrial DNA". en. In: *Hum. Mol. Genet.* 22.2 (Jan. 2013), pp. 384–390.

[81] Gary J. Quigley and Alexander Rich. "Structural Domains of Transfer RNA Molecules: The ribose 2 hydroxyl which distinguishes RNA from DNA plays a key role in stabilizing tRNA structure." In: *Science* 194.4267 (Nov. 1976), pp. 796–806. ISSN: 1095-9203. DOI: 10.1126/science.790568. URL: http://dx.doi.org/10.1126/science.790568.

[82] Boris Rebolledo-Jaramillo et al. "Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 111.43 (Oct. 2014), pp. 15474–15479.

[83] Rodrigue Rossignol et al. "Mitochondrial threshold effects". en. In: *Biochem. J* 370.Pt 3 (Mar. 2003), pp. 751–762.

[84] David C. Samuels et al. "Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans". In: *PLoS Genetics* 9.11 (Nov. 2013). Ed. by Gregory S. Barsh, e1003929. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1003929. URL: http://dx.doi.org/10.1371/journal.pgen.1003929.

[85] Monica Sanchez-Contreras and Scott R Kennedy. "The Complicated Nature of Somatic mtDNA Mutations in Aging". en. In: *Front Aging* 2 (Jan. 2022).

[86] Monica Sanchez-Contreras et al. "The multi-tissue landscape of somatic mtDNA mutations indicates tissue-specific accumulation and removal in aging". In: *Elife* 12 (Feb. 2023), e83395.

[87] Azadeh Sarfallah et al. "Mechanism of transcription initiation and primer generation at the mitochondrial replication origin OriL". In: *The EMBO Journal* 40.19 (Aug. 2021). ISSN: 1460-2075. DOI: 10.15252/embj.2021107988. URL: http://dx.doi.org/10.15252/embj.2021107988.

[88] Marianne Schauer et al. "A mutation in the NADH-dehydrogenase subunit 2 suppresses fibroblast aging". en. In: *Oncotarget* 6.11 (Apr. 2015), pp. 8552–8566.

[89] Michael W Schmitt et al. "Detection of ultra-rare mutations by next-generation sequencing". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 109.36 (Sept. 2012), pp. 14508–14513.

[90] Torsten Schröder et al. "Mitochondrial gene polymorphisms alter hepatic cellular energy metabolism and aggravate diet-induced non-alcoholic steatohepatitis". en. In: *Mol Metab* 5.4 (Apr. 2016), pp. 283–295.

[91] Isabel M Serrano et al. "Mitochondrial haplotype and mito-nuclear matching drive somatic mutation and selection throughout ageing". en. In: *Nat. Ecol. Evol.* 8.5 (May 2024), pp. 1021–1034.

[92] James B Stewart and Patrick F Chinnery. "Extreme heterogeneity of human mitochondrial DNA from organelles to populations". en. In: *Nat. Rev. Genet.* 22.2 (Sept. 2020), pp. 106–118.

[93] Nicholas Stoler et al. "Streamlined analysis of duplex sequencing data with Du Novo". en. In: *Genome Biol.* 17.1 (Aug. 2016), p. 180.

[94] Zhongjie Tang et al. "A Genetic Bottleneck of Mitochondrial DNA During Human Lymphocyte Development". en. In: *Mol. Biol. Evol.* 39.5 (May 2022).

[95] Scott B Vafai and Vamsi K Mootha. "Mitochondrial disorders as windows into an ancient organelle". en. In: *Nature* 491.7424 (Nov. 2012), pp. 374–383.

[96] Marc Vermulst et al. "Mitochondrial point mutations do not limit the natural lifespan of mice". en. In: *Nat. Genet.* 39.4 (Apr. 2007), pp. 540–543.

[97] Douglas C Wallace. "Mitochondrial DNA variation in human radiation and disease". en. In: *Cell* 163.1 (Sept. 2015), pp. 33–38.

[98] Douglas C. Wallace. "Mitochondrial Diseases in Man and Mouse". In: *Science* 283.5407 (Mar. 1999), pp. 1482–1488. ISSN: 1095-9203. DOI: 10.1126/science.283.5407.1482. URL: http://dx.doi.org/10.1126/science.283.5407.1482.

[99] ZiMian Wang et al. "Specific metabolic rates of major organs and tissues across adulthood: evaluation by mechanistic model of resting energy expenditure". In: *The American Journal of Clinical Nutrition* 92.6 (Dec. 2010), pp. 1369–1377. ISSN: 0002-9165. DOI: 10.3945/ajcn.2010.29885. URL: http://dx.doi.org/10.3945/ajcn.2010.29885.

[100] Sjoerd Wanrooij et al. "In vivo mutagenesis reveals that OriL is essential for mitochondrial DNA replication". In: *EMBO reports* 13.12 (Oct. 2012), pp. 1130–1137. ISSN: 1469-3178. DOI: 10.1038/embor.2012.161. URL: http://dx.doi.org/10.1038/embor.2012.161.

[101] Ryan J Weaver et al. "Genomic Signatures of Mitonuclear Coevolution in Mammals". en. In: *Mol. Biol. Evol.* 39.11 (Nov. 2022).

[102] Wei Wei et al. "Background sequence characteristics influence the occurrence and severity of disease-causing mtDNA mutations". en. In: *PLoS Genet.* 13.12 (Dec. 2017), e1007126.

[103] Wei Wei et al. "Germline selection shapes human mitochondrial DNA diversity". en. In: *Science* 364.6442 (May 2019).

[104] Heike Weiss et al. "The mitochondrial Atp8 mutation induces mitochondrial ROS generation, secretory dysfunction, and β-cell mass adaptation in conplastic B6-mtFVB mice". en. In: *Endocrinology* 153.10 (Oct. 2012), pp. 4666–4676.

[105] Heather M Wilkins, Steven M Carl, and Russell H Swerdlow. "Cytoplasmic hybrid (cybrid) cell lines as a practical model for mitochondriopathies". en. In: *Redox Biol* 2 (Apr. 2014), pp. 619–631.

[106] Siôn L. Williams et al. "Somatic mtDNA Mutation Spectra in the Aging Human Putamen". In: *PLoS Genetics* 9.12 (Dec. 2013). Ed. by Bennett Van Houten, e1003990. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1003990. URL: http://dx.doi.org/10.1371/journal.pgen.1003990.

[107] Kenichi Yamamoto et al. "Genetic and phenotypic landscape of the mitochondrial genome in the Japanese population". en. In: *Communications Biology* 3.1 (Mar. 2020), pp. 1–11.

[108] Xinhua Yu et al. "Dissecting the effects of mtDNA variations on complex traits using mouse conplastic strains". en. In: *Genome Res.* 19.1 (Jan. 2009), pp. 159–165.

[109] Zhi Yu et al. "Genetic variation across and within individuals". en. In: *Nat. Rev. Genet.* (Mar. 2024).

[110] Arslan A Zaidi and Kateryna D Makova. "Investigating mitonuclear interactions in human admixed populations". en. In: *Nat Ecol Evol* 3.2 (Feb. 2019), pp. 213–222.

[111] Arslan A Zaidi et al. "Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 116.50 (Dec. 2019), pp. 25172–25178.

[112] Haixin Zhang, Stephen P Burr, and Patrick F Chinnery. "The mitochondrial DNA genetic bottleneck: inheritance and beyond". en. In: *Essays Biochem.* 62.3 (July 2018), pp. 225–234.

[113] Weiming Zheng et al. "Origins of human mitochondrial point mutations as DNA polymerase -mediated errors". In: *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 599.1–2 (July 2006), pp. 11–20. ISSN: 0027-5107. DOI: 10.1016/j.mrfmmm.2005.12.012. URL: http://dx.doi.org/10.1016/j.mrfmmm.2005.12.012.

[114] Yao Zong et al. "Mitochondrial dysfunction: mechanisms and advances in therapy". In: *Signal Transduction and Targeted Therapy* 9.1 (May 2024). ISSN: 2059-3635. DOI: 10.1038/s41392-024-01839-8. URL: http://dx.doi.org/10.1038/s41392-024-01839-8.

Appendix A

Appendix of Chapter 2

A.1 Supplementary Notes

Supplementary Note 1: Input parameters for the duplex sequencing pipeline configuration file

A configuration file was created to process samples in each experimental group. The following python script was used to generate the configuration file. The file is also available at https://github.com/sudmantlab/conplastic_mt_profiling.

```
import pandas as pd
sample_names = pd.read_csv("conplastic_sample_names", sep = "\t", header = None)
sample_names.rename(columns = {0: "sample"}, inplace = True)

df = pd.DataFrame()
df["sample"] = sample_names["sample"]
df["rglb"] = sample_names["sample"]
df["rgpl"] = sample_names["sample"]
df["rgpu"] = sample_names["sample"]
df["rgsm"] = sample_names["sample"]

#hardcoded but will remain the same regardless of the directory created -- will
    change if reference and bedfile
#paths change or if a different organism is used
df["reference"] = "/global/scratch/isabel_serrano/Sudmant_Lab/
mm10_chr1NUMT_masked/mm10_NUMT_masked.fa"
df["target_bed"] =
    "/global/scratch/isabel_serrano/Sudmant_Lab/mm10_chr1NUMT_masked/mm10.bed"
df["blast_db"] = "."
df["targetTaxonId"] = "10090"
df["baseDir"] = "duplex_seq_data"
```

```
df["inbam"] = df["sample"] + ".bam"
df["mqFilt"] = 0
df["minMem"] = 3
df["maxMem"] = 200
df["cutOff"] = 0.7
df["nCutOff"] = 1
df["umiLen"] = 18
df["spacerLen"] = 0
df["locLen"] = 10
df["readLen"] = 151
df["clipBegin"] = 10
df["clipEnd"] = 0
df["minClonal"] = 0
df["maxClonal"] = 0.1
df["minDepth"] = 100
df["maxNs"] = 1
df["runSSCS"] = "FALSE"
df["recovery"] = "noRecovery_noSynLink.sh"
df.to_csv("config.csv", header=True, index=False, sep = ",")
```

A.2 Supplementary Figures

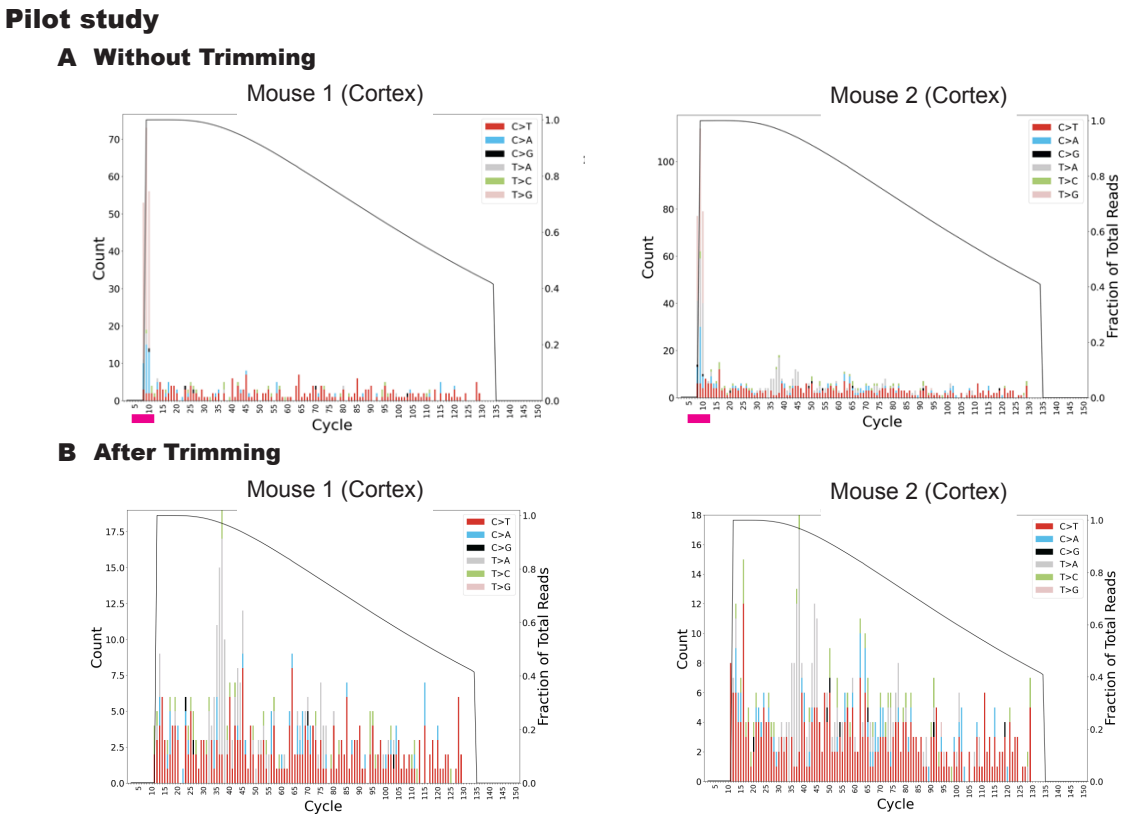


Figure A.1: **Pilot study informed trimming of 5' end of duplex reads by 10 bp in this study.** (A) Mutation calls across the read before trimming. This data is for a pilot study where the mitochondrial genomes in the cerebellum and cortex of B6 mice were duplex sequenced. The 5' end of the duplex read is highlighted in magenta. The x-axis denotes position along the read while the y-axis denotes counts of each mutation type: C>T (red), C>A (blue), C>G (black), T>A (lavender), T>C (green), T>G (pink). (B) The samples shown in (A) after trimming 10 bp from the 5' end. Note that the y-axis scalings differ across panels.

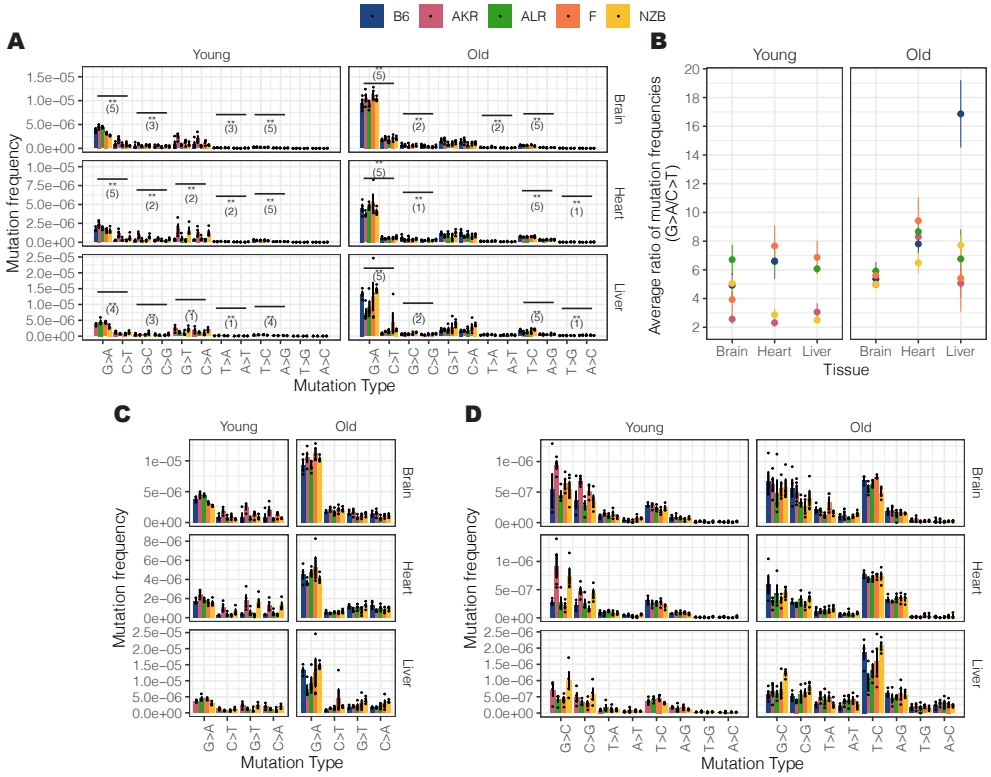


Figure A.2: **Strand asymmetry varies across haplotypes, age, and tissues.** (A) Mutation strand asymmetry is compared between the light strand (L-strand) and heavy strand. Mutations are called and referenced with respect to the L-strand. Mutation frequencies were calculated as the count of mutations of that type divided by the number of times the reference base was duplex sequenced. Error bars denote the standard error of the mean. Points represent the mutation frequency for each sample. Significance denotes strand bias for mutation types; adjusted p-value < 0.01 (two-sided Fisher's Exact test, Benjamini-Hochberg correction). The number of haplotypes with strand bias for respective mutation types is denoted in parentheses (maximum number = 5 mt-haplotypes). For exact p-values for each mutation pair comparison reference **Table A.2**. (B) The strand asymmetry for G>A vs. C>T mutation types is compared across mt-haplotypes. The ratio of G>A to C>T mutation frequencies is averaged across samples within an experimental condition. Error bars denote the standard error of the mean. (C) Frequencies for G>A, C>T, G>T, and C>A mutation types. Error bars denote the standard error of the mean. Points represent the mutation frequency for each sample. (D) Frequencies for G>A, C>G, T>A, A>T, T>C, A>G, T>G, and A>C mutation types. Error bars denote the standard error of the mean. Points represent the mutation frequency for each sample. Note for Fig. (C) and Fig. (D) the y-axis scales differ between figures and tissues to capture smaller mutation frequencies. Mutations with an alternate allele depth > 100 and a mutation frequency $> 1\%$ were excluded. Each mutation is scored once to create a proxy mutation frequency for de novo mutations. The averages calculated for each analysis were across samples in each experimental condition: (n = 4 mice, except for B6-Young-Heart n = 3 mice).

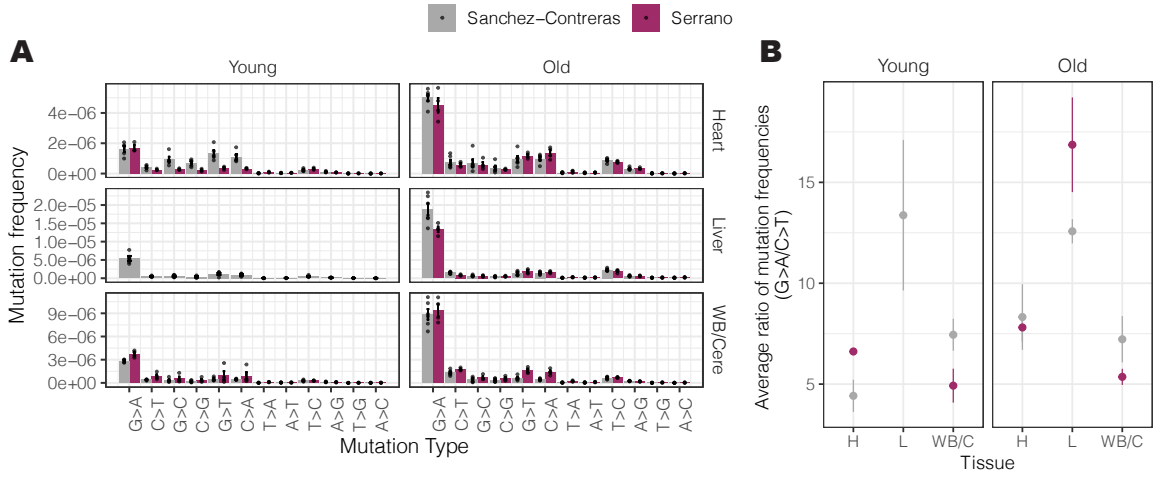


Figure A.3: **Comparison of Sanchez Contreras et al. 2023 and Serrano et al. 2023 unique mutation frequencies and strand asymmetry.** (A) Mutation frequencies in B6 mice are featured, where the whole brain (*Serrano et al.*) is compared to cerebellum (*Sanchez-Contreras et al.*). Mutations with an alternate allele depth > 100 and a mutation frequency $> 1\%$ were excluded. Each mutation is scored once to create a proxy mutation frequency for de novo mutations as described in *Sanchez-Contreras et al. 2023* [86].(A) Mutation frequencies were compared between the *Sanchez-Contreras et al.* and *Serrano et al.* studies. Mutation frequencies were calculated as the count of mutations of that type divided by the number of times the reference base was duplex sequenced. The average mutation frequency across samples in experimental conditions is plotted with error bars denoting the standard error of the mean. Points denote the mutation frequency for each sample. Corresponding statistical testing (two-sided t-test) is provided in **Table A.3**. (B) The strand asymmetry for G>A and C>T mutations is compared across studies. The ratio of G>A to C>T mutation frequencies was calculated. The error bars denote the standard error of the mean. Mice that were chemically treated in the *Sanchez-Contreras et al.* study were excluded from these analyses as chemical treatments were shown to modulate the mutational landscape. The averages calculated for each analysis were across samples in each experimental condition (young n = 5 mice *Sanchez-Contreras et al. 2023*; aged n = 6 mice *Sanchez-Contreras et al. 2023*; n = 4 mice *Serrano et al.*, except B6-Young-Heart n = 3 mice).

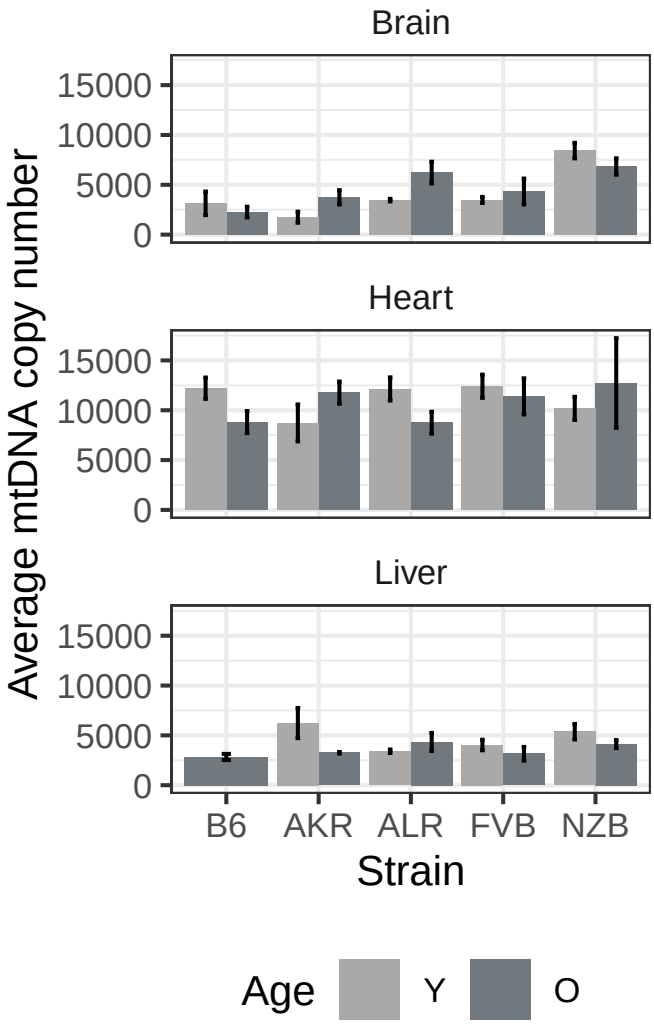


Figure A.4: **Mitochondrial copy number.** Mitochondrial and nuclear copy numbers were quantified using qPCR. The mitochondrial to nuclear copy number ratio is used as an estimate of mitochondrial copy number in a tissue. The average mitochondrial DNA copy number across samples in an experimental condition for young (light gray) and aged (dark gray) mice is featured with error bars denoting the standard error of the mean (n = 4 mice, except for B6, Young, Heart n = 3 mice).

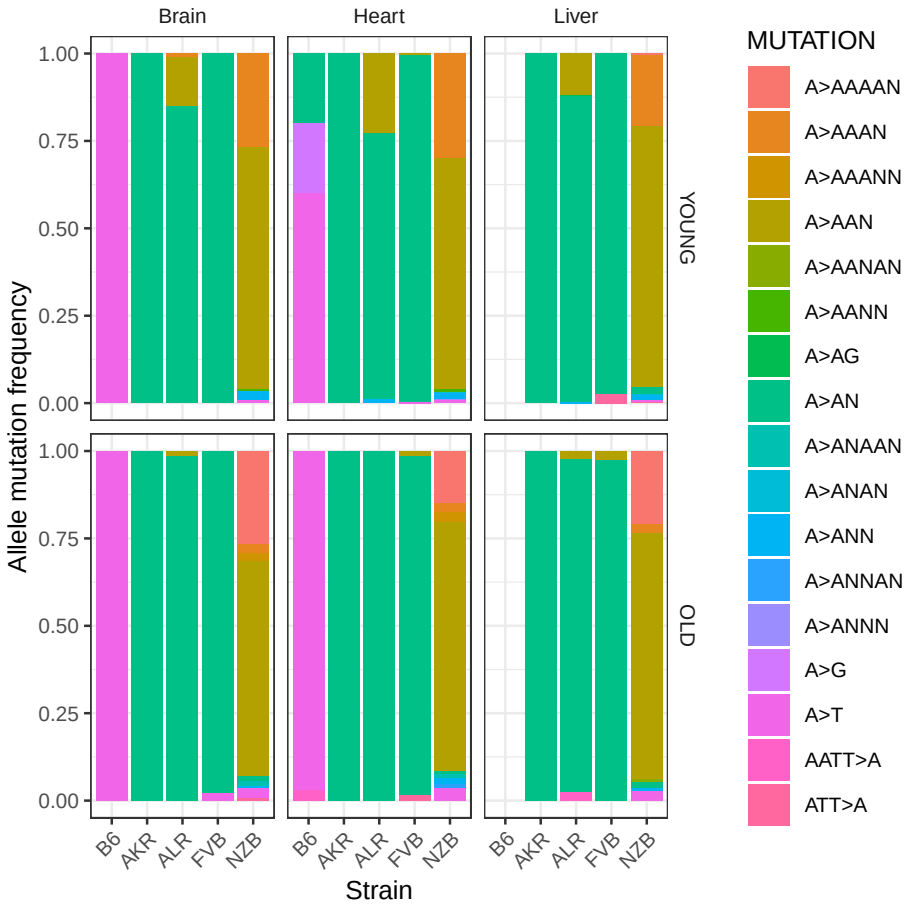


Figure A.5: **Proportion of alleles in the tRNA$^{\mathbf{Arg}}$ 8 base mutation peak region.** The proportion of each allele present in the A repeat region of the DHU arm of *mt-tRNA^{Arg}* was calculated as the total number of each mutation type divided by the total number of mutations that occur in this region. This proportion was calculated for each experimental condition. Mutations that occur at a frequency $> 1 \times 10^{-3}$ were excluded from the analysis, including the previously known heteroplasmic frequency in *mt-tRNA^{Arg}* across conplastic strains [108]. The mutation key (REF>ALT) denotes the mutation that occurs in *mt-tRNA^{Arg}* , Ns in the ALT allele represent bases where a duplex consensus could not be reached.

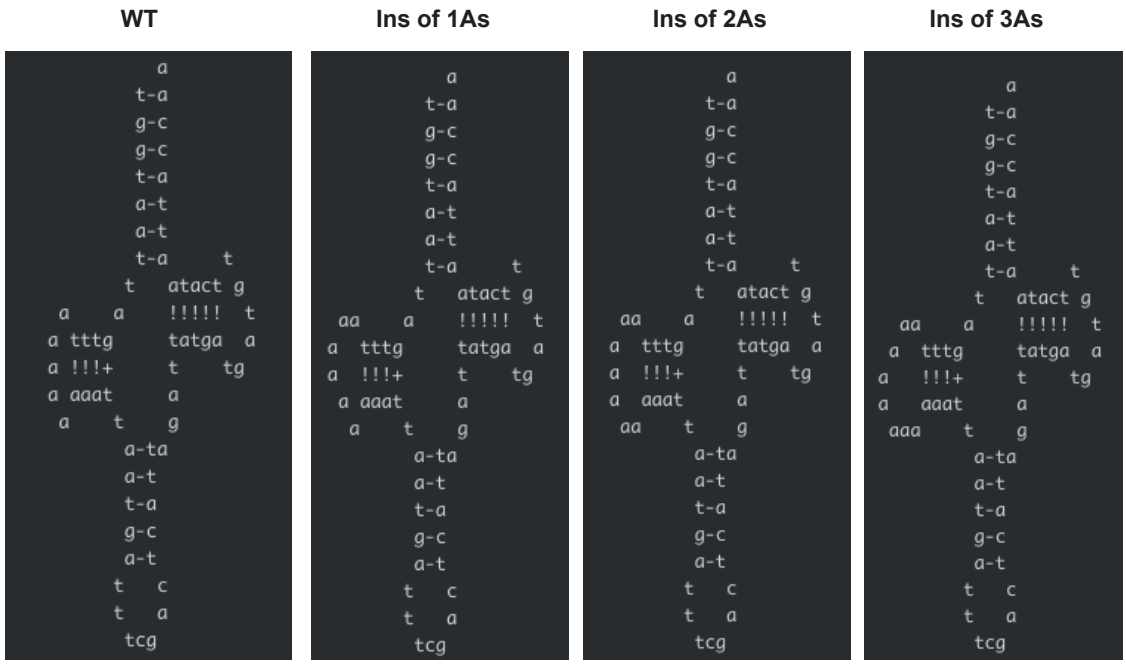


Figure A.6: **Computationally predicted changes in mt-tRNA^{Arg} structures.** Structures were created using ARWEN (v 1.2) [56]. Structures were created for the 3 most abundant mutation types across conditions **(Fig. A.5)**: 1-3 A insertion in the DHU A-repeat region. As were inserted at position 9819 since we lack the resolution to locate the exact position an A-insertion occurs.

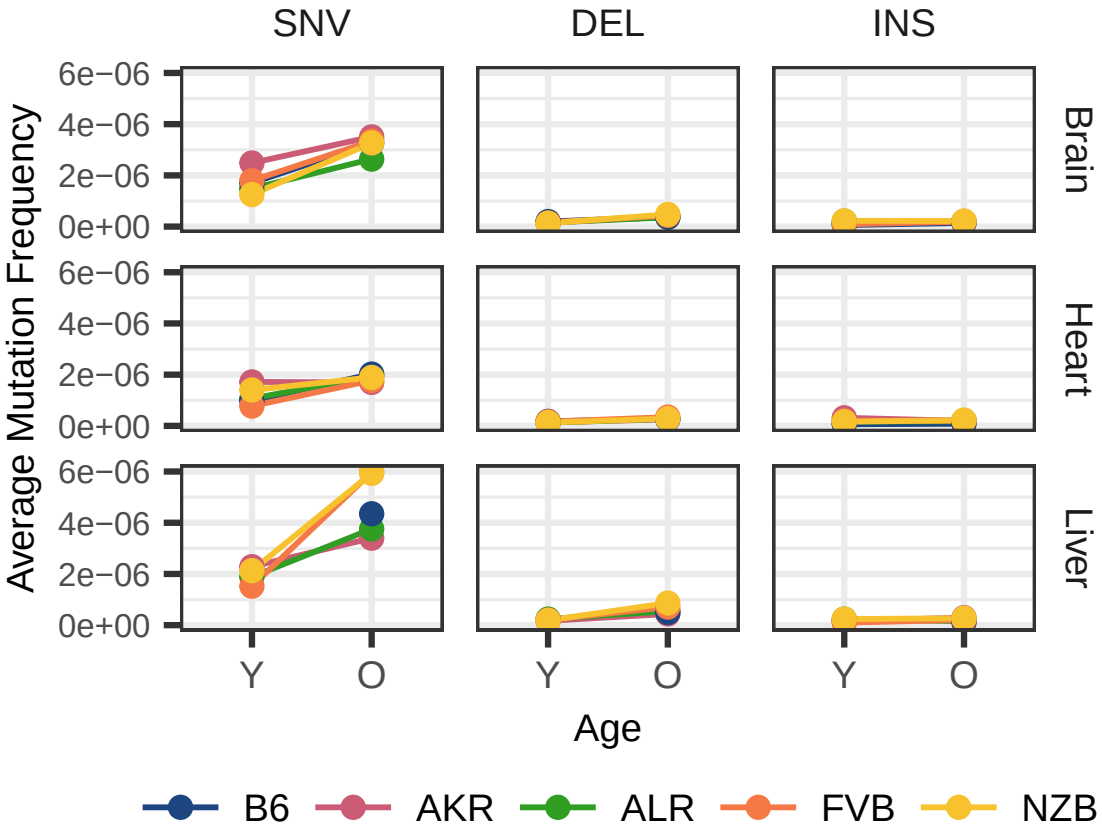


Figure A.7: **Change in SNVs, deletions, and insertions with age across all tissues.** The average mutation frequency is compared between mt-haplotypes for different classes of mutations: single nucleotide variants (SNVs), deletions, and insertions. The average mutation frequency was calculated as the sum of alternative alleles in a class divided by the total duplex bp depth. Mutations with a frequency $> 1 \times 10^{-3}$ were excluded from this analysis. Mutation counts and duplex depth were aggregated across samples in experimental conditions (n = 29 conditions; n = 4 mice, except for B6-Young-Heart n = 3 mice).

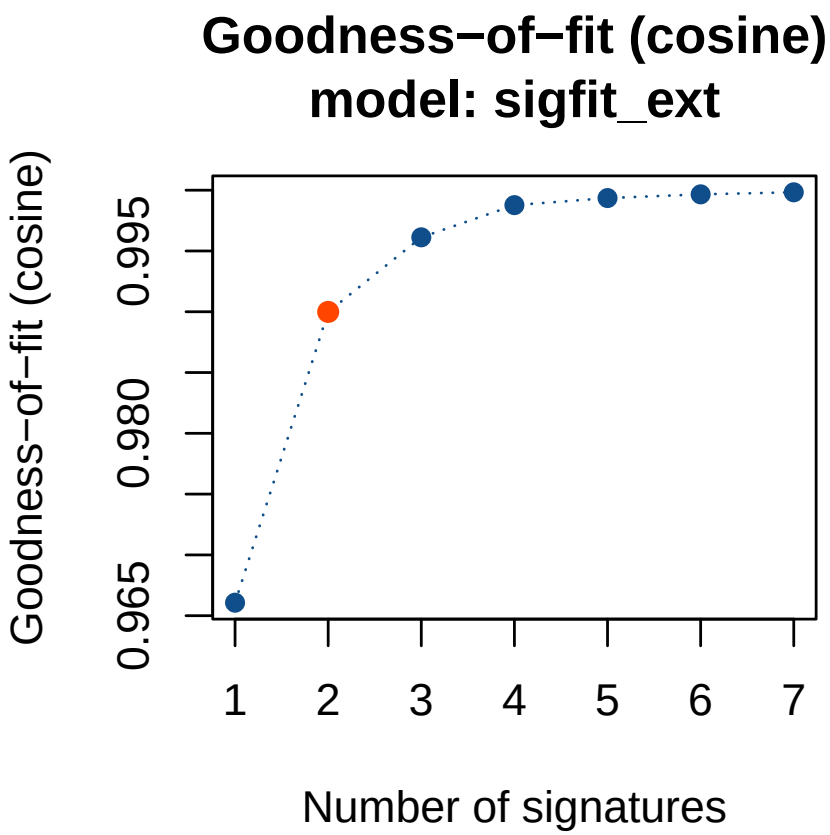


Figure A.8: **Estimation of the optimal number of mutational signatures.** Observed counts for each mutation type were used for signature extraction using sigfit. A range of signatures (1-7) was given for possible signature extraction. Sigfit highlights the best number of signatures (as shown in red; signatures = 2), based on a cosine similarity metric that compares the original catalog of mutational signatures and the sigfit inferred signatures, as previously described in [36].

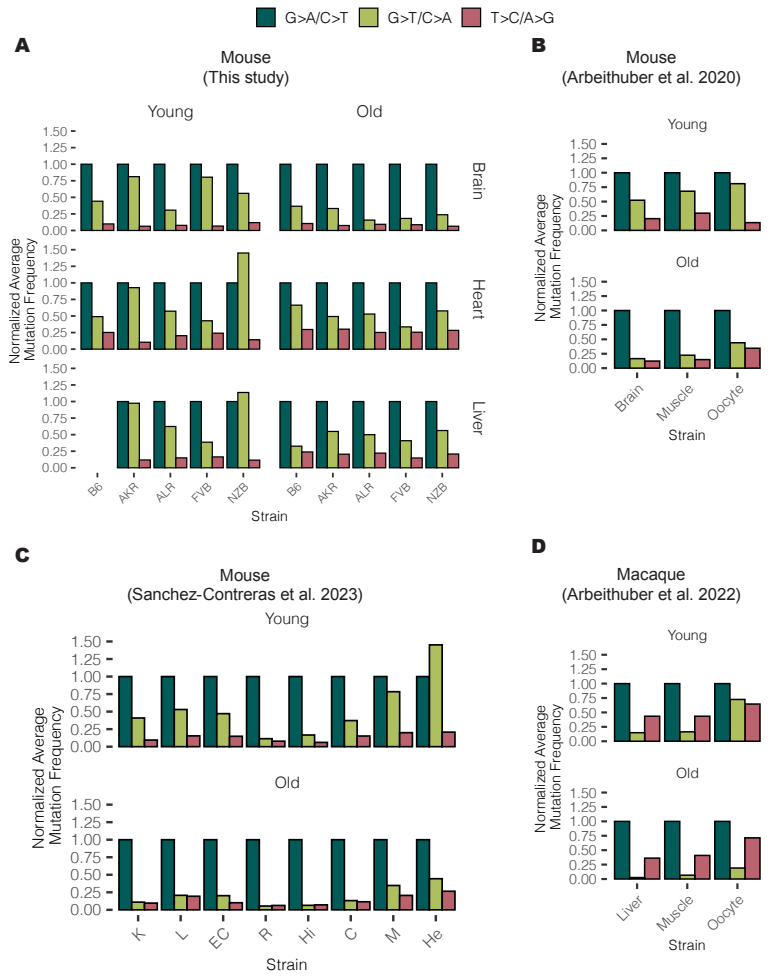
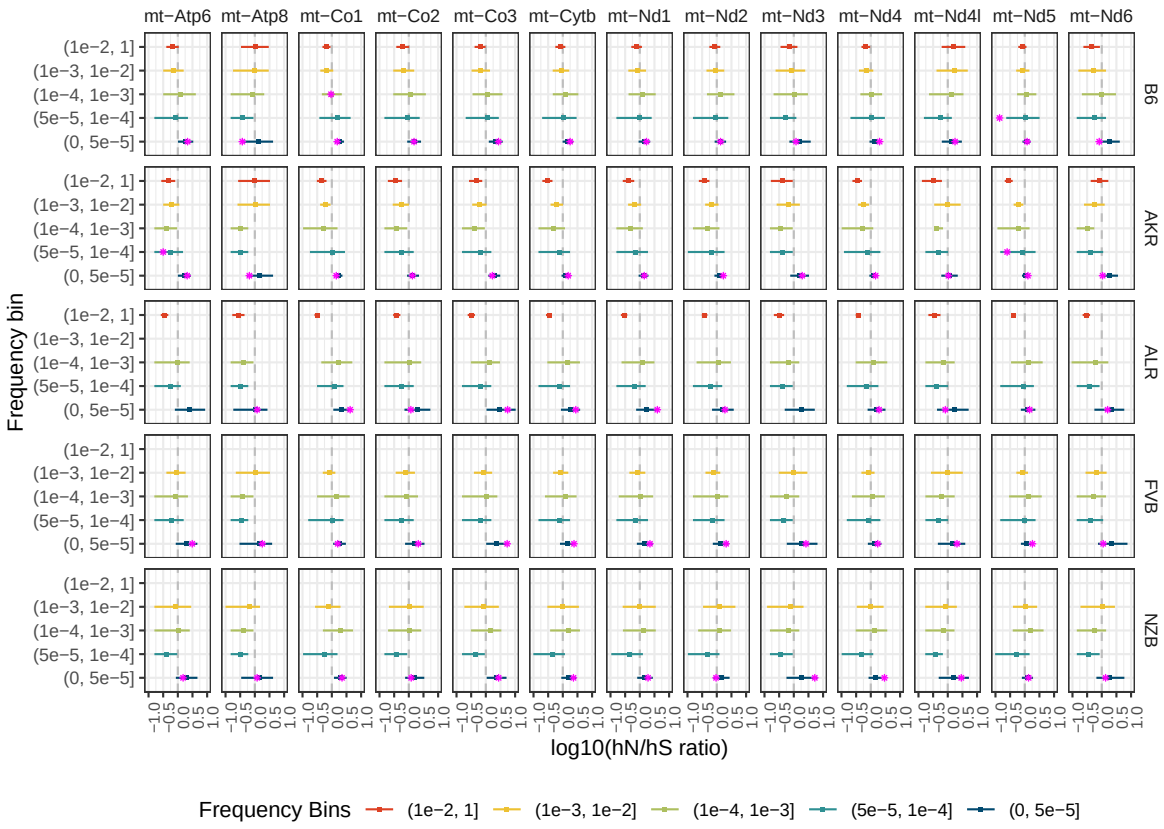


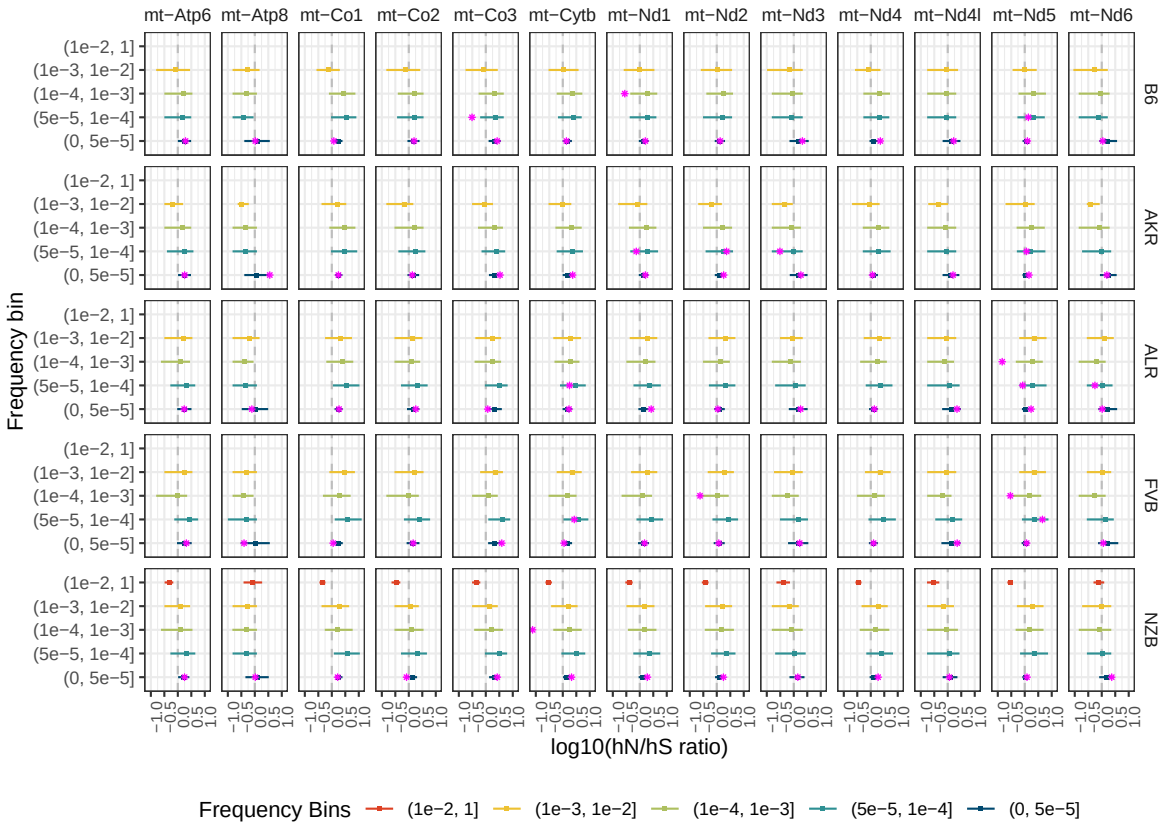
Figure A.9: **Mutation frequencies for macaque and mouse tissues.** The mutation frequency for de novo point mutations was calculated as the total count of mutations of the given type divided by the duplex bp depth for the reference nucleotide. Counts and duplex depth for complementary mutation types were aggregated. These mutation frequencies were then normalized to the G>A/C>T frequency, which is consistently the most abundant mutation type. (A) The normalized average mutation frequency for the three most abundant mutations in this study. Counts and depths across samples in an experimental condition were aggregated (n = 29 conditions, n = 4 mice per condition except B6-Young-Heart n = 3 mice). (B) The normalized mutation frequency across three mouse tissues in *Arbeithuber et al. 2020* [4] (n = 28 mice for young; n = 8 mice for old). (C) The normalized mutation frequency across mouse tissues in *Sanchez-Contreras et al. 2023* [86] (n = 5 mice for young; n = 6 mice for old). Tissue labels are as follows: kidney (K), liver (L), RPE/choroid (RC), retina (R), hippocampus (Hi), cerebellum (C), skeletal muscle (M), and heart (He). (D) The normalized mutation frequency across macaque tissues in *Arbeithuber el al. 2022* [3] (n = 9 macaques for young, n = 7 macaques for old).

Figure A.10: Position of observed $\frac{hN}{hS}$ statistics in the distribution of simulated $\frac{hN}{hS}$ statistics

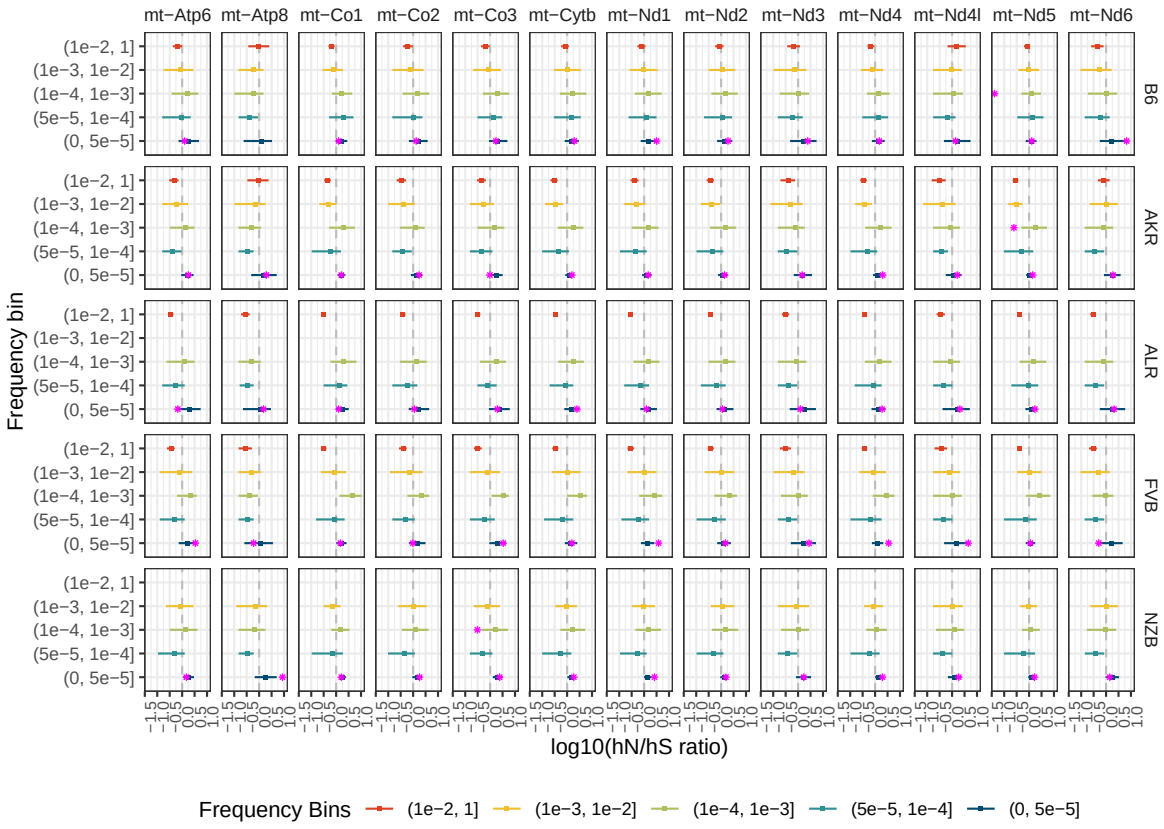
Young; Brain



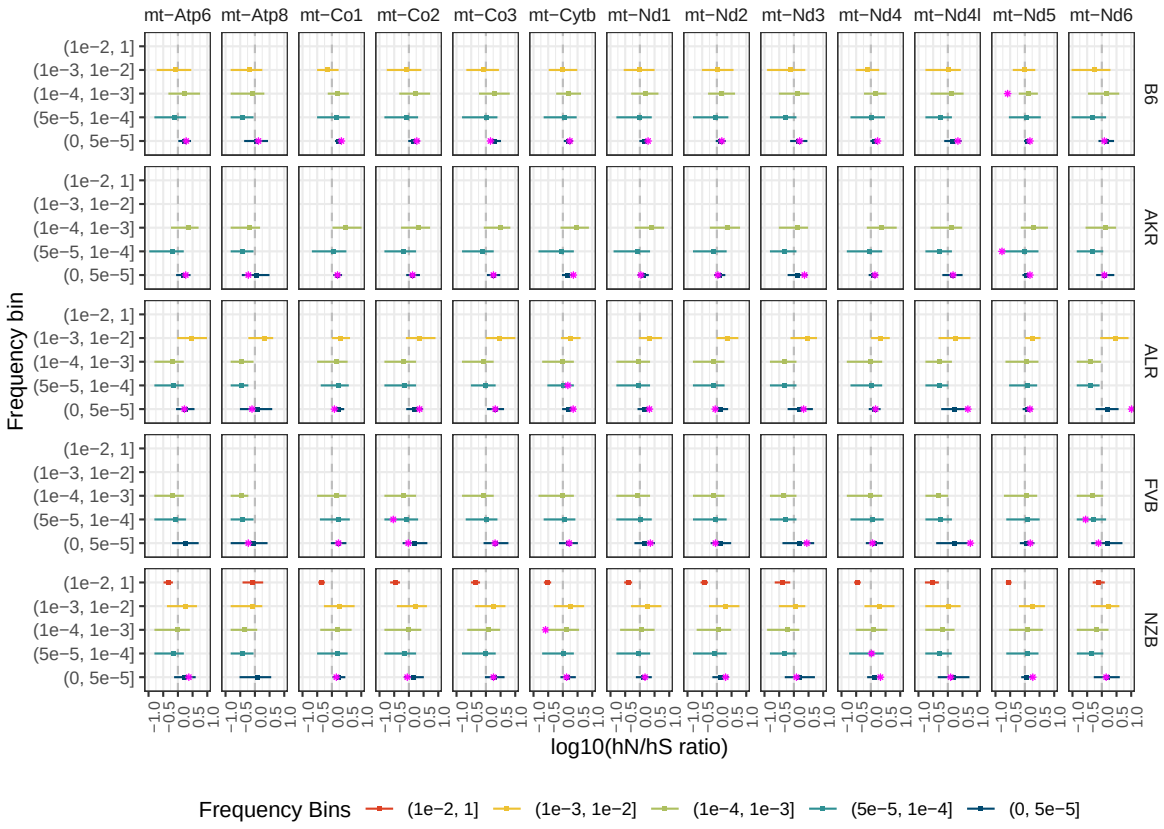
Old; Brain


mt−Atp6
mt−Atp8
mt−Co1
mt−Co2
mt−Co3
mt−Cytb
mt−Nd1
mt−Nd2
mt−Nd3
mt−Nd4
mt−Nd4l
mt−Nd5
mt−Nd6
(1e−2, 1]
(1e−3, 1e−2]
(1e−4, 1e−3]
(5e−5, 1e−4]
(0, 5e−5]
Frequency bin
B6
AKR
ALR
FVB
NZB
−1.0
−0.5
0.0
0.5
1.0
log10(hN/hS ratio)
Frequency Bins
(1e−2, 1]
(1e−3, 1e−2]
(1e−4, 1e−3]
(5e−5, 1e−4]
(0, 5e−5]

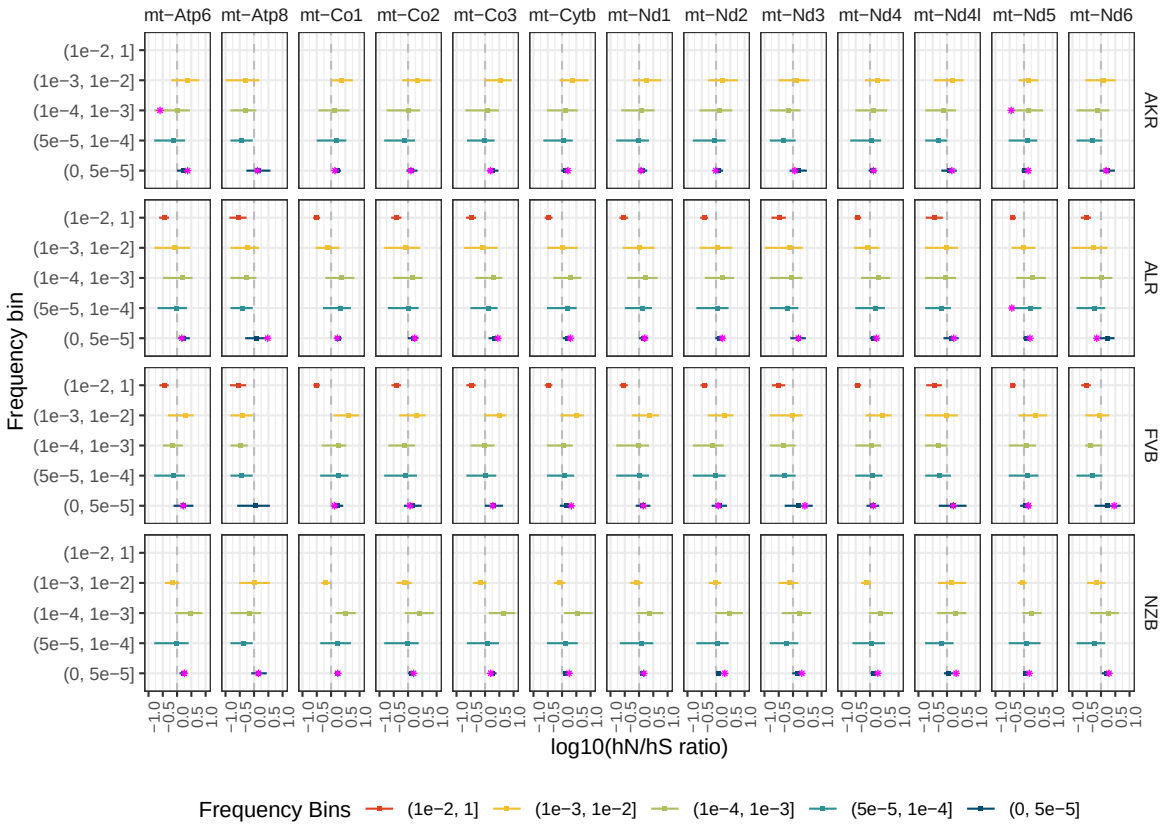
Young; Heart


mt−Atp6
mt−Atp8
mt−Co1
mt−Co2
mt−Co3
mt−Cytb
mt−Nd1
mt−Nd2
mt−Nd3
mt−Nd4
mt−Nd4l
mt−Nd5
mt−Nd6
B6
AKR
ALR
FVB
NZB
Frequency bin
(1e−2, 1]
(1e−3, 1e−2]
(1e−4, 1e−3]
(5e−5, 1e−4]
(0, 5e−5]
log10(hN/hS ratio)
Frequency Bins

Old; Heart


mt−Atp6
mt−Atp8
mt−Co1
mt−Co2
mt−Co3
mt−Cytb
mt−Nd1
mt−Nd2
mt−Nd3
mt−Nd4
mt−Nd4l
mt−Nd5
mt−Nd6
(1e−2, 1]
(1e−3, 1e−2]
(1e−4, 1e−3]
(5e−5, 1e−4]
(0, 5e−5]
B6
AKR
ALR
FVB
NZB
Frequency bin
−1.0
−0.5
0.0
0.5
1.0
log10(hN/hS ratio)
Frequency Bins
(1e−2, 1]
(1e−3, 1e−2]
(1e−4, 1e−3]
(5e−5, 1e−4]
(0, 5e−5]

Young; Liver


mt−Atp6
mt−Atp8
mt−Co1
mt−Co2
mt−Co3
mt−Cytb
mt−Nd1
mt−Nd2
mt−Nd3
mt−Nd4
mt−Nd4l
mt−Nd5
mt−Nd6
AKR
ALR
FVB
NZB
Frequency bin
(1e−2, 1]
(1e−3, 1e−2]
(1e−4, 1e−3]
(5e−5, 1e−4]
(0, 5e−5]
−1.0
−0.5
0.0
0.5
1.0
log10(hN/hS ratio)
Frequency Bins
(1e−2, 1]
(1e−3, 1e−2]
(1e−4, 1e−3]
(5e−5, 1e−4]
(0, 5e−5]

Old; Liver

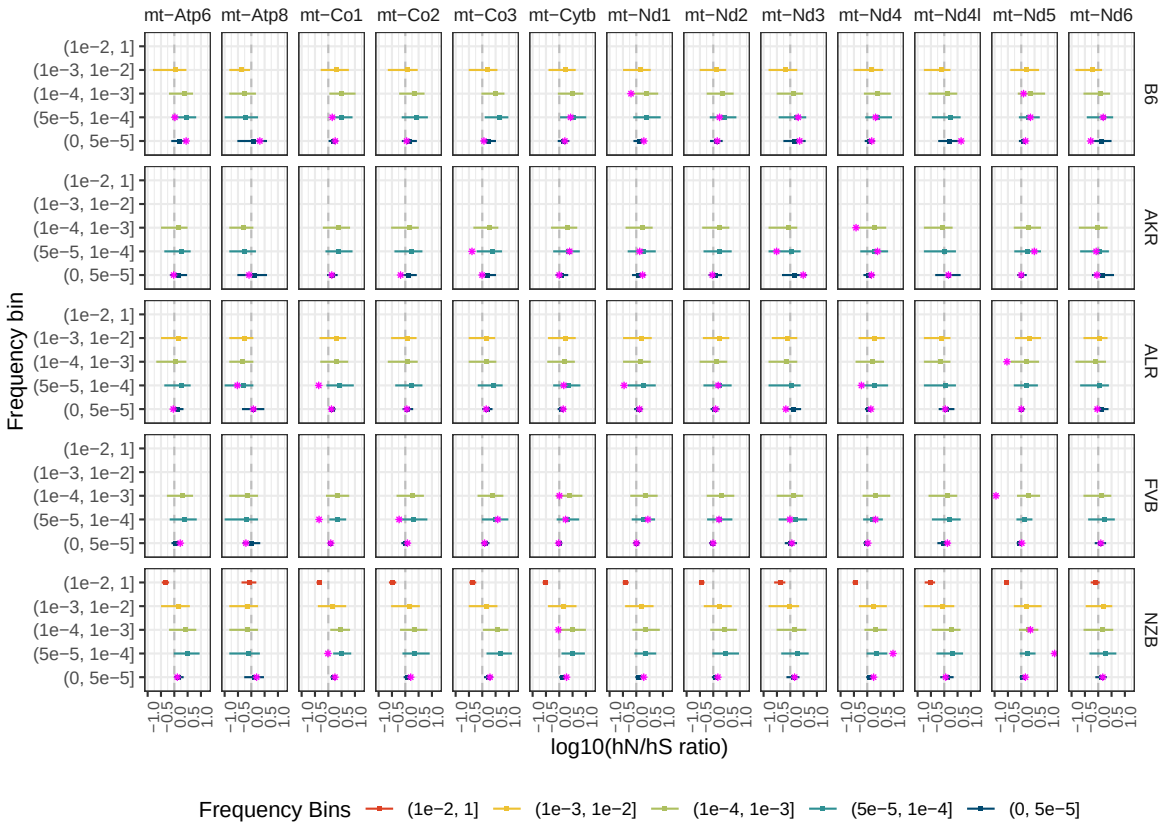


Figure A.10: **Position of observed $\frac{hN}{hS}$ statistics in the distribution of simulated $\frac{hN}{hS}$ statistics.** Simulated $\frac{hN}{hS}$ statistics were generated as described in Methods: Testing for selection. Shown are the average simulated $\frac{hN}{hS}$ statistics. The segment denotes the range in simulated ratios extending from the 2.5 percent quantile to the 97.5 percent quantile. The $\frac{hN}{hS}$ statistics are in log10 scale, where the dotted, gray vertical line denotes theoretical neutrality. The magenta asterisk indicates the observed $\frac{hN}{hS}$ statistic for each bin. There are cases where either the simulated and/or observed $\frac{hN}{hS}$ statistics could not be calculated: (1) Absent observed statistic, but simulated distribution generated: either hN or hS equals 0. Due to this, the observed ratio was excluded from this study. Note, the observed mutation proportions and counts could generate nonzero hN and hS values for a gene in a simulation. (2) Absent simulated distribution and observed ratio: there are zero mutations in the given mutation bin for the experimental condition. 10,000 simulations were performed for each frequency *bin x strain x age x tissue* combination. Simulated ratios were filtered as explained above. The mutation counts were combined across samples in experimental conditions to generate the observed mutation spectra and observed $\frac{hN}{hS}$ statistics (n = 29 conditions; n = 4 mice, except B6-Young-Heart n = 3 mice).

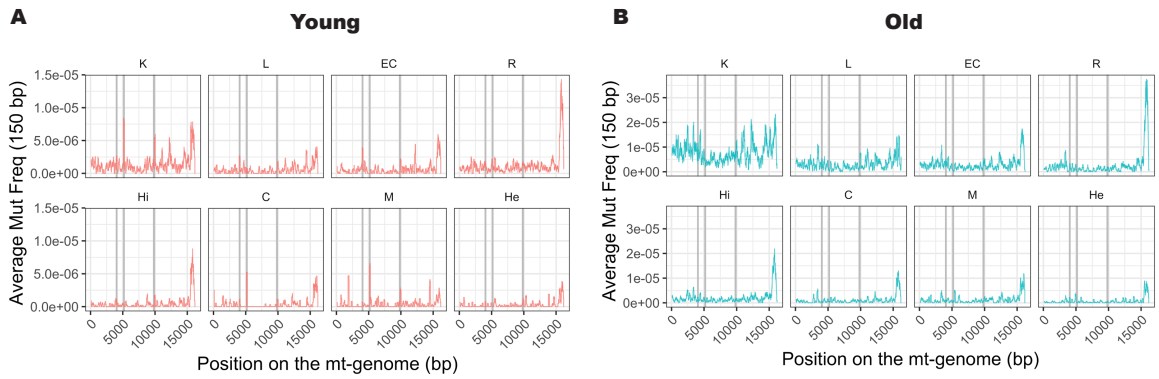


Figure A.11: **Mutational hotspots in Sanchez-Contreras et al. (2023).** In *Sanchez-Contreras et al. 2023*, B6 mice were used. (A) The average mutation frequency in a 150 bp sliding window is shown for young mice. (B) The average mutation frequency in a 150 bp sliding window is shown for aged mice. Mutation counts and depths for replicates in an experimental condition were aggregated. The gray bands denote where the *OriL* (pos 5159-5190), *MT-ND2* (pos 4030-4070), and *MT-tRNA*^{Arg} (pos 9807-9874) are located. For *MT-ND2*, the region containing the mutational hotspot in this study is highlighted. Tissue labels are as follows: kidney (K), liver (L), RPE/choroid (RC), retina (R), hippocampus (Hi), cerebellum (C), skeletal muscle (M), and heart (He).

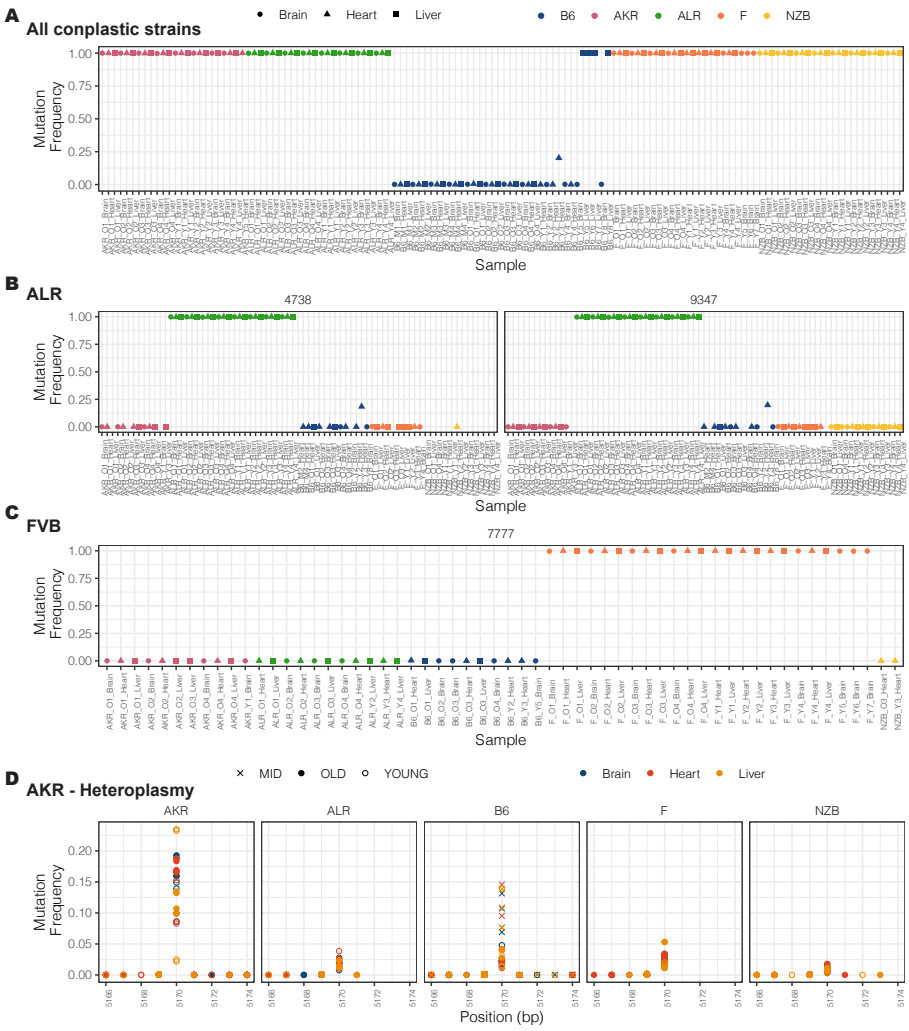
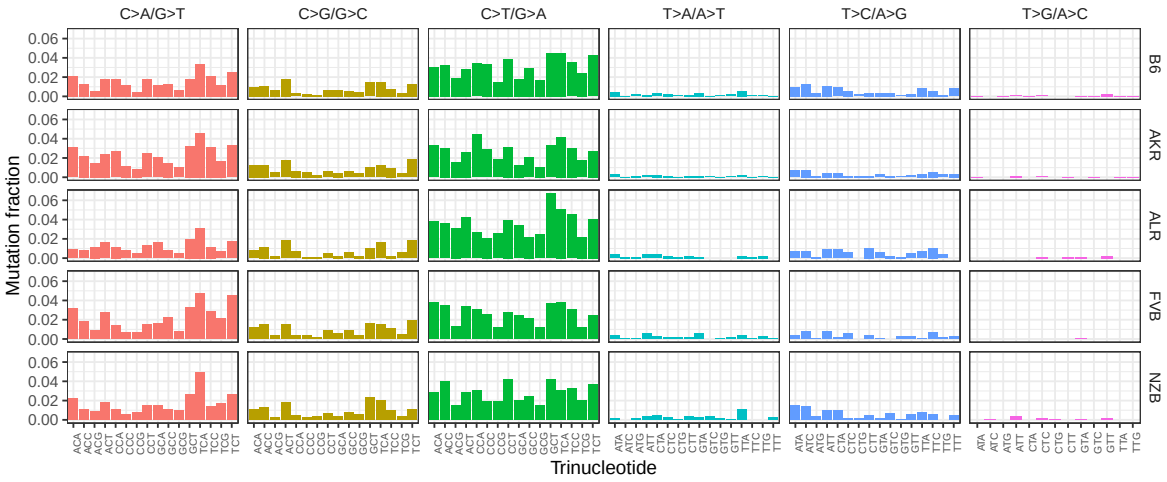


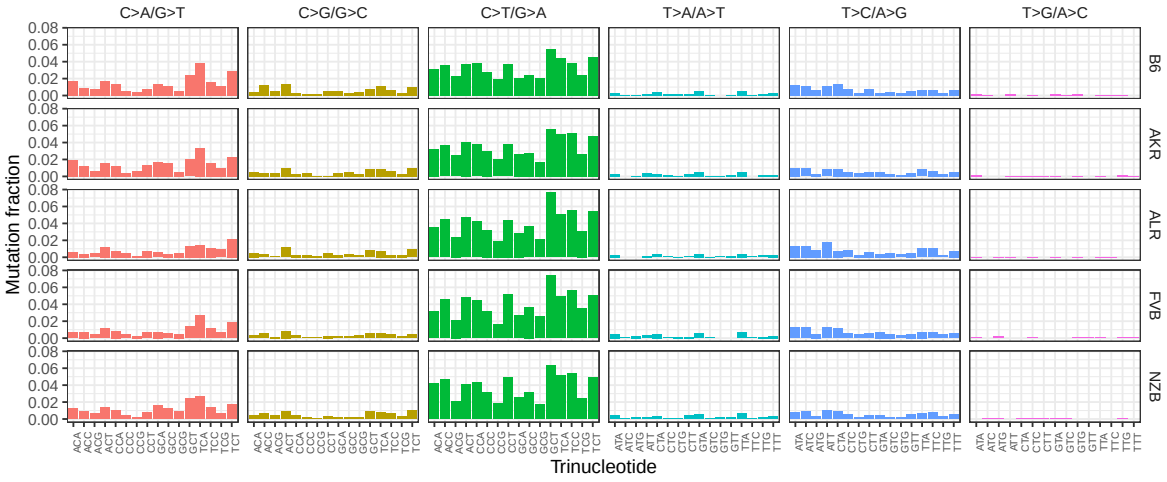
Figure A.12: **Validation of conplastic strains.** The mutation frequency, calculated as the number of alternative alleles divided by the number of duplex mt-genomes at the position, was checked for each haplotype site. (A) The mutation frequency for the T>C mutation at position 9460 in the gene *MT-ND3*. All conplastic strains differ from B6 (blue) at this position. B6 young liver samples that had this haplotype present were excluded from the study due to potential contamination. (B) The mutation frequency for ALR (green) haplotype sites positions 4738 (*MT-ND2*) and 9347 (*MT-COXIII*). The B6 young heart sample with presence of this haplotype was excluded from the study due to potential contamination. (C) The mutation frequency for the FVB (orange) haplotype site at position 7777 (*MT-ATP8*). For (A) - (C), color denotes strain (B6 (blue), AKR (pink), ALR (green), FVB (orange), NZB (yellow)) and shape denotes tissue (brain (circle), heart (triangle), liver (square)). (D) Comparison of the heteroplasmic allele frequency in the *OriL*. Shape denotes age (mid (cross), old (filled circle), young (hollow circle)) and color denotes tissue (brain (blue), heart (red), liver (yellow)).

Figure A.13: Trinucleotide Spectra.

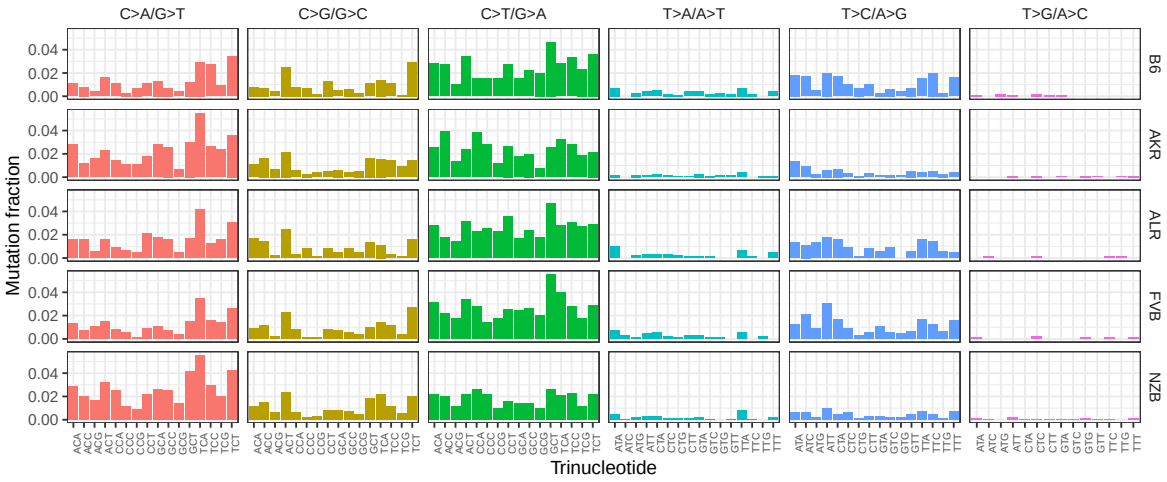
Young; Brain



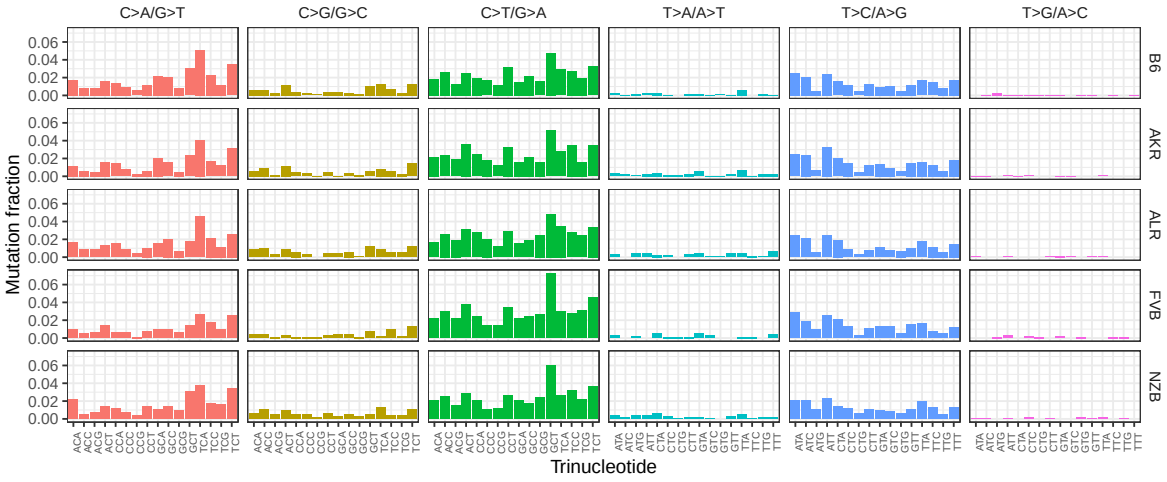
Old; Brain



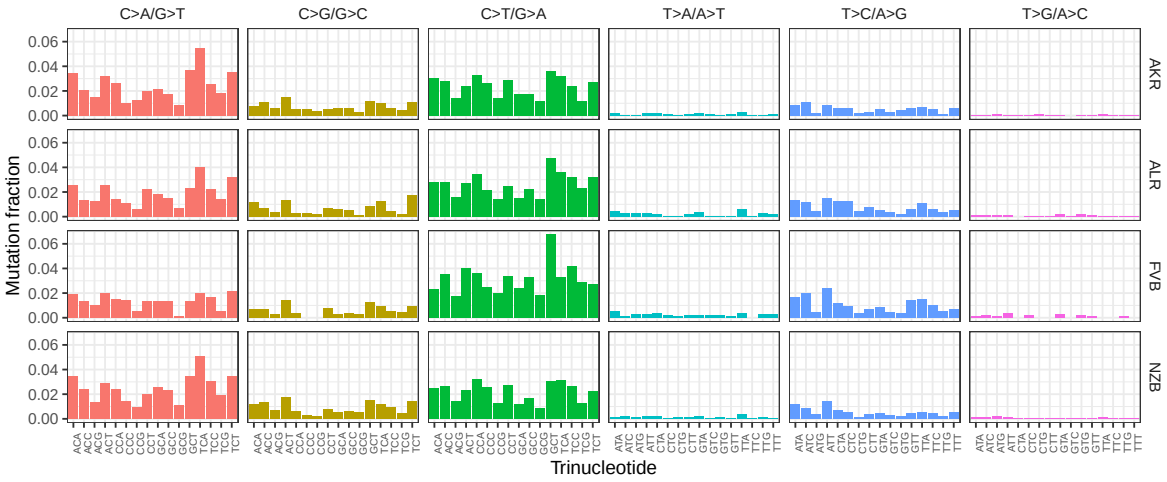
Young; Heart



Old; Heart



Young; Liver



Old; Liver

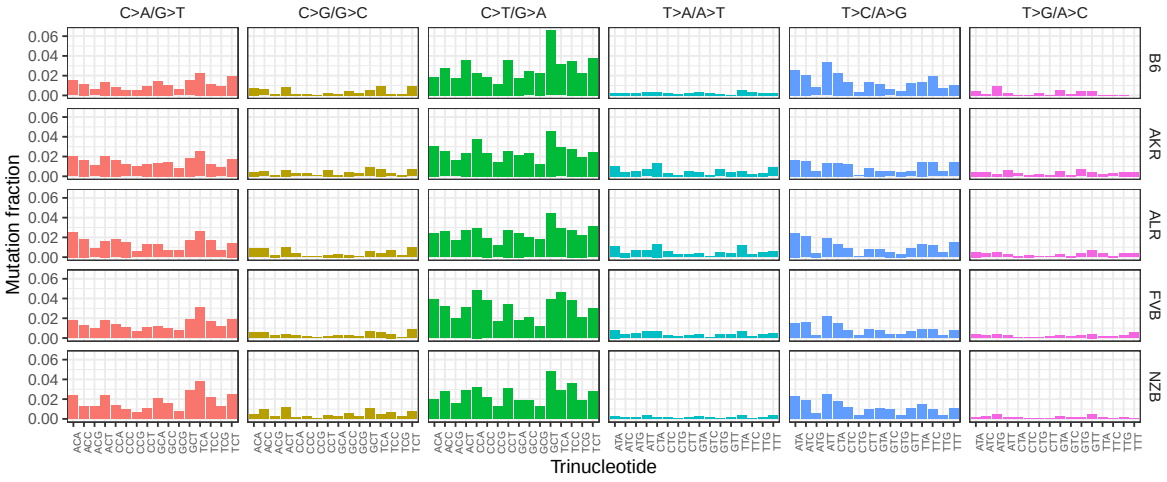


Figure A.13: **Trinucleotide Spectra** The trinucleotide spectra for all experimental conditions is depicted below. Only de novo mutations were used in this analysis and were identified by filtering mutations with 1) an alternative allele depth < 100 and 2) a mutation frequency < 0.01. Each mutation is scored once to create a proxy mutation frequency for de novo mutations. The mutation fraction is the proportion of each mutation type in a given trinucleotide context divided by the total count of de novo mutations for a condition. The trinucleotide and mutation type featured represent mutations on either strand. For example, ACG represents both ACG and CGT, where either a C>T or a G>A mutation has occurred.

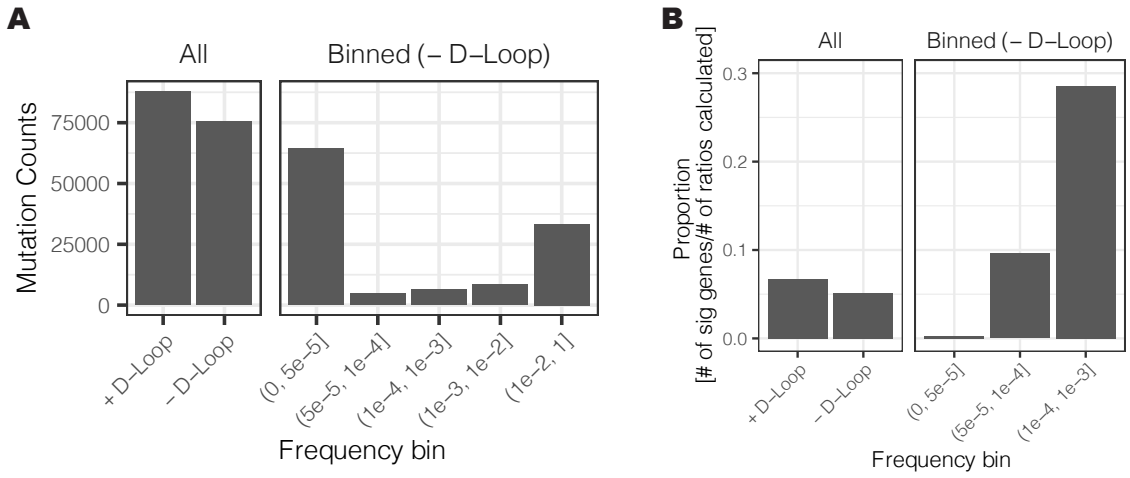


Figure A.14: **Distribution of mutation counts and proportion of genes under selection across frequency bins.** (A) Count of total mutations in each analysis and frequency bin. Mutation counts were calculated as the sum of the duplex alternative allele depth across experimental conditions. For analyses labeled "all", mutation counts were aggregated regardless of frequency. These mutation counts were used to simulate $\frac{hN}{hS}$ statistics. (B) The proportion of genes under selection in each analysis. Within each bin and the aggregated analyses, the proportion is calculated as the count of genes under selection divided by the count of observed $\frac{hN}{hS}$ statistics that could be calculated. Note: $\frac{hN}{hS}$ statistics were excluded from the study if hN or hS equaled 0. For the aggregated analyses (labeled "All"), mutations with a frequency greater than 1×10^{-3} were excluded. For the binned frequencies, mutations in the *D-Loop* were excluded.

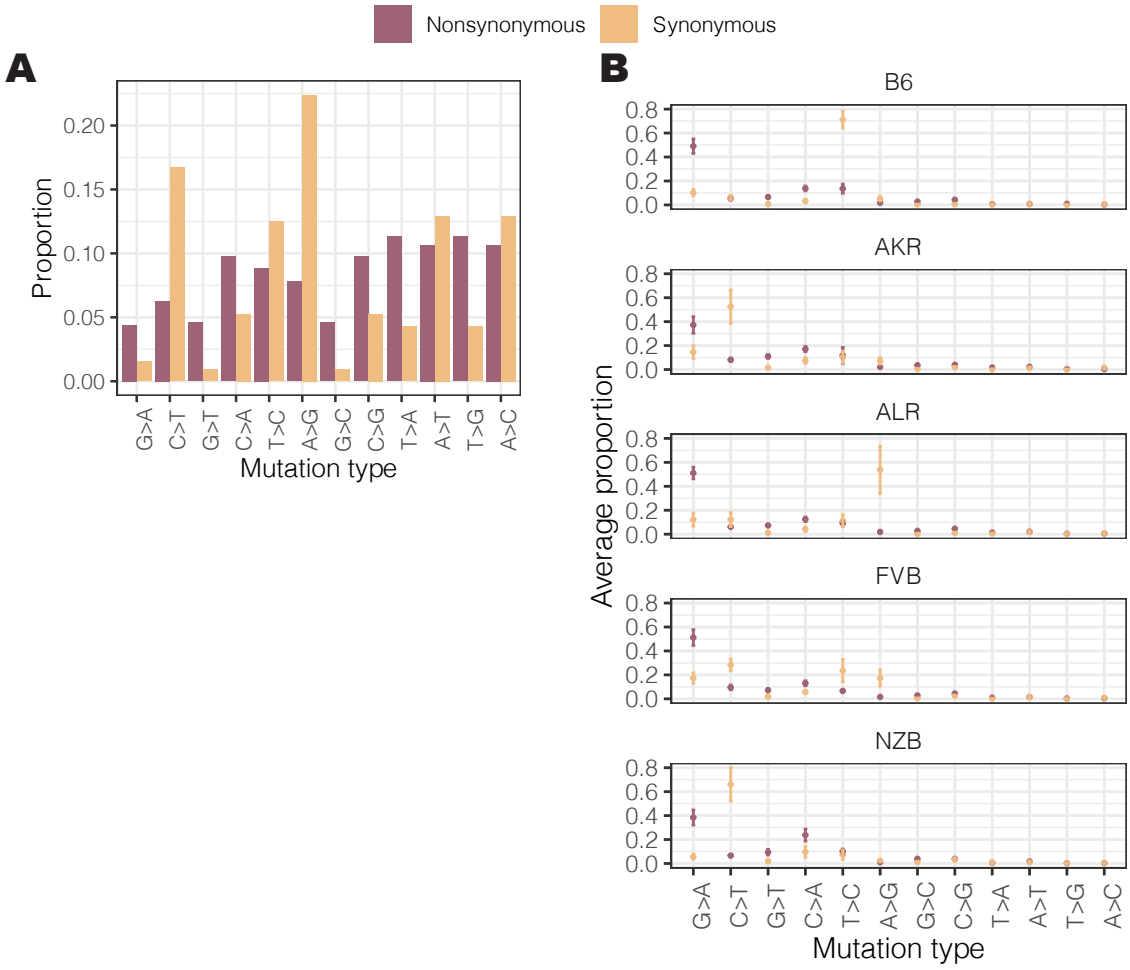


Figure A.15: **Mutational spectra for nonsynonymous and synonymous mutations.** (A) The null mutational spectra for synonymous (orange) and nonsynonymous (purple) mutations across experimental conditions. For every mutation type, the number of nonsynonymous and synonymous mutations a given mutation type can produce was calculated. The proportion was calculated as the count of nonsynonymous or synonymous mutations resulting from a given mutation type divided by all possible nonsynonymous or synonymous mutations in the mt-genome. (B) The proportion of each mutation type comprising the total count of nonsynonymous (purple) and synonymous (orange) mutations. The average proportion across *tissues x age* (n = 6 conditions, except for B6 where n = 5 conditions) for each mt-haplotype is shown with error bars denoting the standard error of the mean.

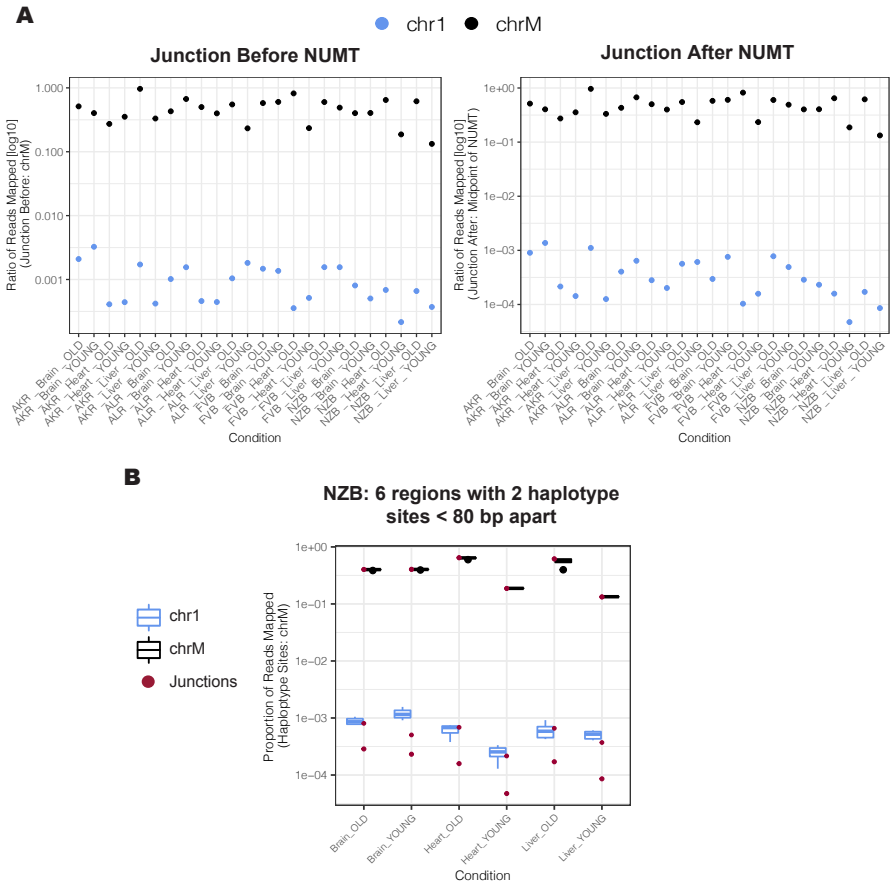


Figure A.16: **Estimation of NUMT contamination.** (A) To estimate for NUMT contamination paired end duplex sequencing reads were remapped to mm10 with an unmasked chr1 NUMT. The number of reads that map to 10 bp upstream and 10 bp downstream of the NUMT region, referred to as the junction regions, was calculated. The figures denote the number of duplex reads that map to the junction regions divided by the average duplex depth of the corresponding region in the mt-genome with the chr1 NUMT masked (blue). Shown is also the number of reads that map to the corresponding junction regions in the mt-genome divided by the average duplex depth of the corresponding region in the mt-genome with the chr1 NUMT masked (black). (B) Duplex sequencing reads for the NZB strain were remapped using a NZB reference mt-genome. Reads were aggregated across samples in each experimental condition (n = 4 mice per experimental condition *strain x age*). Six clusters of haplotype sites (regions where more than one haplotype site exists in a read) were identified. The distribution of the proportion of reads mapping the chr1 and those mapping to the mt-genome are depicted in boxplots. The center line in the boxplot highlights the median and the bounds of the box represent the first and third quartiles. The whiskers depict the maximum and minimum values defined by the sum of the third quartile and 1.5 multiplied by the interquartile range and the difference between the first quartile and 1.5 multiplied by the interquartile range, respectively. The proportion of reads mapping to chr1 and the mt-genome at the junction sites (shown in A) are represented in red.

A.3 Supplementary Tables

For **Table A.1**, amino acid changes with * were originally identified in *Yu et al., 2009* and characterized as a Met-Met amino acid change. This study verified the Ile-Ile amino acid change in our data via the UCSC Genome Browser.

Table A.1: **Variants that define the mt-haplotype and their associated phenotypes for each conplastic strain.**

Strain	**Gene (Position, REF >ALT)**	**Amino Acid Change (Citation)**	**Phenotypes**	**Phenotype citation**
B6 - mtAKR	Heteroplasmic mutations: - OriL (nt 5171, 11A >12A/13A) - tRNA Arginine (nt9820, 8A >9A) Fixed mutations: - ND3 (9460, T>C)*	- Ile-Ile*	- Shorter lifespan (only in females) - Imparied glucose metabolism (only in females) - Decrease in mtDNA copy number - Increase in expression of mtDNA genes	Hirose et al. (2018) Scientific Reports.
B6 - mtALR	Fixed mutations: - ND3 (9460, T>C)* - ND2 (4738, C>A) - COXIII (9347, G>A)	- Ile-Ile* - Leu - Met (Yu et al. (2009). Genome Research.) - Val - Ile (Yu et al. (2009). Genome Research.) Has not been documented beforehand.	*Note: In the phenotyping study the control was B6-mtAKR - Higher Complex I activity - Lower ROS production - Improved ATP production - Reduced levels SA-beta-gal (senescence marker) & senescence-related cytokines - Increased cell proliferation - Decreased activation of p38MAPK pathway - Overall, decrease in cellular senescence	Schauer et al. (2015) Oncotarget.
B6 - mtFVB	Fixed mutations: - ND3 (9460, T>C)* - Atp8 (7777, G>T)	- Ile - Ile* - Asp - Tyr (Yu et al. (2009). Genome Research)	- Exhibit more anxiety-related behavior than WT and B6-mtAKR mice - Significantly different microbiome composition relative to WT - Increase in glycolysis and decreases in OXPHOS respiration - Similar lifespan to WT - Highly sensative to metabolic stressors (e.g. change in diet) leading to more severe phenotypes - Baseline decrease in ATP production and increase in ROS - Under dietary stressors, results in non-alcoholic steatohepatitis and liver lipid accumulation - Altered mitochondria morphology	- Yu et al. (2009) Genome Research. - Hirose et al. (2017) Science Reports. - Schröder et al. (2016) Molecular Metabolism. - Weiss et al. (2012) Endocrinology.
B6 - mtNZB	Heteroplasmic mutations: - 1 synonymous mutation in mt-COX2 - 1 missense mutation in mt-ND4L Fixed mutations: - 13 missense mutations - 21 non-coding variants including 2 indels - 55 synonymous variants (Total of 91 haplotype sites)	- Annotations by Hirose et al. 2016 - Additionally annotated and verified via in-house script	- No significant phenotypic differences from WT reported - Lower disease incidence than B6-mtAKR and B6-mtFVB	- Hirose et al. (2016) Brief Communications Arising from A. Latorre-Pellicer et al. Nature (2016). - Yu et al. (2009) Genome Research.

Table A.2: **Strand bias analysis.**

Experimental Condition (Strain_Tissue_Age)	Mutation Pair	p-value	OR	Adjusted p-value
AKR_Brain_OLD	G>A_C>T	6.9E-237	4.91	0.00
AKR_Brain_YOUNG	G>A_C>T	7.6E-59	2.47	0.00
AKR_Heart_OLD	G>A_C>T	2.1E-217	8.10	0.00
AKR_Heart_YOUNG	G>A_C>T	4.7E-30	2.15	0.00
AKR_Liver_OLD	G>A_C>T	2.1E-87	4.89	0.00
AKR_Liver_YOUNG	G>A_C>T	9.9E-65	2.64	0.00
ALR_Brain_OLD	G>A_C>T	1.8E-290	6.16	0.00
ALR_Brain_YOUNG	G>A_C>T	3.7E-98	6.58	0.00
ALR_Heart_OLD	G>A_C>T	1E-170	9.06	0.00
ALR_Heart_YOUNG	G>A_C>T	1.7E-60	5.44	0.00
ALR_Liver_OLD	G>A_C>T	1.1E-176	5.36	0.00
ALR_Liver_YOUNG	G>A_C>T	1.1E-229	5.84	0.00
B6_Brain_OLD	G>A_C>T	5.2E-208	5.14	0.00
B6_Brain_YOUNG	G>A_C>T	1.3E-148	4.51	0.00
B6_Heart_OLD	G>A_C>T	1.7E-260	7.99	0.00
B6_Heart_YOUNG	G>A_C>T	2.9E-77	6.69	0.00
B6_Liver_OLD	G>A_C>T	0E+00	17.19	0.00
F_Brain_OLD	G>A_C>T	1.6E-261	5.82	0.00
F_Brain_YOUNG	G>A_C>T	6.6E-39	3.09	0.00
F_Heart_OLD	G>A_C>T	1E-107	8.67	0.00
F_Heart_YOUNG	G>A_C>T	9E-88	6.05	0.00
F_Liver_OLD	G>A_C>T	8.5E-122	2.77	0.00
F_Liver_YOUNG	G>A_C>T	1.1E-125	6.79	0.00
NZB_Brain_OLD	G>A_C>T	1.6E-307	4.94	0.00
NZB_Brain_YOUNG	G>A_C>T	3.5E-80	4.89	0.00
NZB_Heart_OLD	G>A_C>T	1.1E-94	6.58	0.00
NZB_Heart_YOUNG	G>A_C>T	6.8E-46	2.59	0.00
NZB_Liver_OLD	G>A_C>T	0E+00	7.10	0.00
NZB_Liver_YOUNG	G>A_C>T	1.3E-116	2.55	0.00
AKR_Brain_OLD	G>C_C>G	8.7E-02	1.29	0.14
AKR_Brain_YOUNG	G>C_C>G	2E-03	1.39	0.00
AKR_Heart_OLD	G>C_C>G	4.7E-01	1.13	0.58
AKR_Heart_YOUNG	G>C_C>G	3.8E-09	1.90	0.00
AKR_Liver_OLD	G>C_C>G	4.4E-03	1.82	0.01
AKR_Liver_YOUNG	G>C_C>G	7.9E-04	1.46	0.00
ALR_Brain_OLD	G>C_C>G	6.6E-01	1.09	0.75

ALR_Brain_YOUNG	G>C_C>G	9.5E-02	1.41	0.15
ALR_Heart_OLD	G>C_C>G	2.1E-02	1.47	0.04
ALR_Heart_YOUNG	G>C_C>G	7.8E-01	1.06	0.86
ALR_Liver_OLD	G>C_C>G	9.4E-01	0.98	1.00
ALR_Liver_YOUNG	G>C_C>G	8.8E-05	1.58	0.00
B6_Brain_OLD	G>C_C>G	1.8E-01	1.21	0.27
B6_Brain_YOUNG	G>C_C>G	4.8E-03	1.42	0.01
B6_Heart_OLD	G>C_C>G	2.1E-05	1.72	0.00
B6_Heart_YOUNG	G>C_C>G	7.3E-01	1.06	0.82
B6_Liver_OLD	G>C_C>G	3.9E-01	1.18	0.52
F_Brain_OLD	G>C_C>G	2.4E-05	2.23	0.00
F_Brain_YOUNG	G>C_C>G	5.9E-02	1.33	0.10
F_Heart_OLD	G>C_C>G	1.7E-01	1.45	0.25
F_Heart_YOUNG	G>C_C>G	4.6E-01	1.12	0.57
F_Liver_OLD	G>C_C>G	4.2E-01	1.12	0.53
F_Liver_YOUNG	G>C_C>G	5.6E-02	1.45	0.10
NZB_Brain_OLD	G>C_C>G	2.5E-03	1.55	0.01
NZB_Brain_YOUNG	G>C_C>G	4.7E-03	1.53	0.01
NZB_Heart_OLD	G>C_C>G	9.3E-02	1.38	0.15
NZB_Heart_YOUNG	G>C_C>G	4.8E-10	1.74	0.00
NZB_Liver_OLD	G>C_C>G	3.4E-06	1.73	0.00
NZB_Liver_YOUNG	G>C_C>G	4.6E-16	1.76	0.00
AKR_Brain_OLD	G>T_C>A	2.5E-01	1.10	0.34
AKR_Brain_YOUNG	G>T_C>A	2.7E-02	1.15	0.05
AKR_Heart_OLD	G>T_C>A	6.1E-01	1.05	0.71
AKR_Heart_YOUNG	G>T_C>A	1.4E-05	1.36	0.00
AKR_Liver_OLD	G>T_C>A	1.3E-01	1.18	0.21
AKR_Liver_YOUNG	G>T_C>A	9.1E-04	1.22	0.00
ALR_Brain_OLD	G>T_C>A	2.6E-01	1.16	0.35
ALR_Brain_YOUNG	G>T_C>A	9.4E-01	1.02	1.00
ALR_Heart_OLD	G>T_C>A	7.2E-01	1.04	0.80
ALR_Heart_YOUNG	G>T_C>A	3E-01	1.16	0.41
ALR_Liver_OLD	G>T_C>A	1.6E-01	0.88	0.24
ALR_Liver_YOUNG	G>T_C>A	4.9E-01	0.95	0.58
B6_Brain_OLD	G>T_C>A	3.7E-01	1.08	0.49
B6_Brain_YOUNG	G>T_C>A	4.2E-01	1.08	0.53
B6_Heart_OLD	G>T_C>A	4.9E-01	0.95	0.59
B6_Heart_YOUNG	G>T_C>A	8.9E-01	0.98	0.95
B6_Liver_OLD	G>T_C>A	4.3E-01	1.09	0.54

F_Brain_OLD	G>T_C>A	9.5E-01	0.98	1.00
F_Brain_YOUNG	G>T_C>A	9.6E-01	1.01	1.00
F_Heart_OLD	G>T_C>A	5.7E-01	0.90	0.67
F_Heart_YOUNG	G>T_C>A	1.3E-01	0.80	0.20
F_Liver_OLD	G>T_C>A	1E+00	1.00	1.00
F_Liver_YOUNG	G>T_C>A	7E-01	1.05	0.79
NZB_Brain_OLD	G>T_C>A	2.3E-02	1.22	0.04
NZB_Brain_YOUNG	G>T_C>A	1.9E-01	1.16	0.27
NZB_Heart_OLD	G>T_C>A	3.7E-02	1.28	0.06
NZB_Heart_YOUNG	G>T_C>A	9.3E-05	1.25	0.00
NZB_Liver_OLD	G>T_C>A	1.6E-01	1.09	0.24
NZB_Liver_YOUNG	G>T_C>A	5.6E-03	1.12	0.01
AKR_Brain_OLD	T>G_A>C	1.9E-01	1.95	0.27
AKR_Brain_YOUNG	T>G_A>C	1.8E-01	2.16	0.27
AKR_Heart_OLD	T>G_A>C	3.7E-03	12.03	0.01
AKR_Heart_YOUNG	T>G_A>C	4.1E-01	0.53	0.53
AKR_Liver_OLD	T>G_A>C	3.3E-02	0.61	0.06
AKR_Liver_YOUNG	T>G_A>C	4.3E-01	1.40	0.54
ALR_Brain_OLD	T>G_A>C	1E+00	0.96	1.00
ALR_Brain_YOUNG	T>G_A>C	1E+00	1.20	1.00
ALR_Heart_OLD	T>G_A>C	1E+00	1.20	1.00
ALR_Heart_YOUNG	T>G_A>C	1.8E-01	4.81	0.27
ALR_Liver_OLD	T>G_A>C	1.2E-03	0.55	0.00
ALR_Liver_YOUNG	T>G_A>C	6.6E-01	0.84	0.75
B6_Brain_OLD	T>G_A>C	2.1E-01	1.74	0.29
B6_Brain_YOUNG	T>G_A>C	6.4E-01	0.76	0.74
B6_Heart_OLD	T>G_A>C	4.3E-01	1.40	0.54
B6_Heart_YOUNG	T>G_A>C	4.8E-01	2.00	0.58
B6_Liver_OLD	T>G_A>C	8E-02	0.65	0.13
F_Brain_OLD	T>G_A>C	1E+00	1.00	1.00
F_Brain_YOUNG	T>G_A>C	1E+00	0.00	1.00
F_Heart_OLD	T>G_A>C	4E-01	1.68	0.52
F_Heart_YOUNG	T>G_A>C	1E+00	1.20	1.00
F_Liver_OLD	T>G_A>C	1.5E-02	0.65	0.03
F_Liver_YOUNG	T>G_A>C	2.3E-01	0.50	0.32
NZB_Brain_OLD	T>G_A>C	8.3E-01	1.10	0.90
NZB_Brain_YOUNG	T>G_A>C	1E+00	0.94	1.00
NZB_Heart_OLD	T>G_A>C	8.1E-01	0.84	0.89
NZB_Heart_YOUNG	T>G_A>C	6.1E-01	1.20	0.71

NZB_Liver_OLD	T>G_A>C	8E-03	0.58	0.02
NZB_Liver_YOUNG	T>G_A>C	5.6E-01	1.17	0.66
AKR_Brain_OLD	T>C_A>G	8.3E-13	2.65	0.00
AKR_Brain_YOUNG	T>C_A>G	9E-11	3.17	0.00
AKR_Heart_OLD	T>C_A>G	8.2E-23	2.37	0.00
AKR_Heart_YOUNG	T>C_A>G	9.6E-08	2.40	0.00
AKR_Liver_OLD	T>C_A>G	4.4E-09	2.20	0.00
AKR_Liver_YOUNG	T>C_A>G	1.2E-07	1.97	0.00
ALR_Brain_OLD	T>C_A>G	4.2E-26	3.88	0.00
ALR_Brain_YOUNG	T>C_A>G	1.9E-05	2.84	0.00
ALR_Heart_OLD	T>C_A>G	1.4E-12	2.21	0.00
ALR_Heart_YOUNG	T>C_A>G	1.4E-09	2.93	0.00
ALR_Liver_OLD	T>C_A>G	3E-22	2.53	0.00
ALR_Liver_YOUNG	T>C_A>G	4.4E-36	4.09	0.00
B6_Brain_OLD	T>C_A>G	1.5E-23	3.68	0.00
B6_Brain_YOUNG	T>C_A>G	6.1E-16	3.25	0.00
B6_Heart_OLD	T>C_A>G	1.3E-25	2.32	0.00
B6_Heart_YOUNG	T>C_A>G	5.5E-18	3.97	0.00
B6_Liver_OLD	T>C_A>G	2.3E-31	3.23	0.00
F_Brain_OLD	T>C_A>G	4.4E-24	4.00	0.00
F_Brain_YOUNG	T>C_A>G	1.8E-05	3.18	0.00
F_Heart_OLD	T>C_A>G	3.7E-08	2.14	0.00
F_Heart_YOUNG	T>C_A>G	8.1E-14	2.87	0.00
F_Liver_OLD	T>C_A>G	2.1E-28	2.67	0.00
F_Liver_YOUNG	T>C_A>G	1.4E-12	2.88	0.00
NZB_Brain_OLD	T>C_A>G	1.3E-19	3.30	0.00
NZB_Brain_YOUNG	T>C_A>G	4.9E-10	3.29	0.00
NZB_Heart_OLD	T>C_A>G	9E-10	2.18	0.00
NZB_Heart_YOUNG	T>C_A>G	7.8E-12	2.62	0.00
NZB_Liver_OLD	T>C_A>G	2.5E-57	3.53	0.00
NZB_Liver_YOUNG	T>C_A>G	2.7E-31	3.03	0.00
AKR_Brain_OLD	T>A_A>T	2.4E-02	1.61	0.04
AKR_Brain_YOUNG	T>A_A>T	9.3E-08	4.06	0.00
AKR_Heart_OLD	T>A_A>T	5.4E-03	1.75	0.01
AKR_Heart_YOUNG	T>A_A>T	2.4E-01	1.38	0.33
AKR_Liver_OLD	T>A_A>T	9.8E-02	1.32	0.15
AKR_Liver_YOUNG	T>A_A>T	2.4E-03	2.12	0.01
ALR_Brain_OLD	T>A_A>T	6.6E-02	1.58	0.11
ALR_Brain_YOUNG	T>A_A>T	2.6E-05	6.02	0.00

ALR_Heart_OLD	T>A_A>T	4.8E-01	1.20	0.58
ALR_Heart_YOUNG	T>A_A>T	1.9E-07	4.61	0.00
ALR_Liver_OLD	T>A_A>T	4.2E-01	1.11	0.53
ALR_Liver_YOUNG	T>A_A>T	6.6E-02	1.42	0.11
B6_Brain_OLD	T>A_A>T	1E-03	2.02	0.00
B6_Brain_YOUNG	T>A_A>T	2.6E-04	2.28	0.00
B6_Heart_OLD	T>A_A>T	1E-01	1.45	0.16
B6_Heart_YOUNG	T>A_A>T	5.5E-02	1.66	0.10
B6_Liver_OLD	T>A_A>T	9.6E-02	1.46	0.15
F_Brain_OLD	T>A_A>T	1.1E-05	2.57	0.00
F_Brain_YOUNG	T>A_A>T	1.1E-02	2.00	0.02
F_Heart_OLD	T>A_A>T	6.4E-03	2.62	0.01
F_Heart_YOUNG	T>A_A>T	7E-05	2.97	0.00
F_Liver_OLD	T>A_A>T	3.5E-01	1.12	0.47
F_Liver_YOUNG	T>A_A>T	5.5E-02	1.75	0.10
NZB_Brain_OLD	T>A_A>T	8.6E-01	1.04	0.93
NZB_Brain_YOUNG	T>A_A>T	2.1E-01	1.37	0.30
NZB_Heart_OLD	T>A_A>T	1.5E-02	1.90	0.03
NZB_Heart_YOUNG	T>A_A>T	2E-01	1.29	0.28
NZB_Liver_OLD	T>A_A>T	7.8E-01	1.06	0.86
NZB_Liver_YOUNG	T>A_A>T	7.2E-03	1.52	0.01

A two-sided Fisher's Exact test was used to determine if there was an association between strand and mutation type. p-values were adjusted using a Benjamini-Hochberg correction for multiple hypothesis testing. Mutation counts and the count of duplex bases sequenced for the reference base were aggregated across samples in an experimental condition.

Table A.3: **Comparison of frequencies for each mutation type between Serrano et al. and Sanchez-Contreras et al.**

Condition	**Mutation type**	**p-value**	**statistic**	**adjusted p-value (BH)**
Brain_OLD	G>T	0.0056	4.57	0.087
Brain_YOUNG	T>A	0.0038	5.34	0.087
Heart_YOUNG	C>G	0.0058	-4.26	0.087
Liver_OLD	C>T	0.0016	-4.66	0.087
Heart_YOUNG	G>T	0.0074	-4.51	0.088
Heart_YOUNG	C>A	0.0092	-4.55	0.092
Brain_OLD	C>A	0.0134	4.78	0.100
Brain_OLD	T>A	0.0164	4.69	0.100
Brain_YOUNG	G>A	0.0183	3.88	0.100
Liver_OLD	G>A	0.0169	-3.13	0.100
Liver_OLD	T>A	0.0147	4.20	0.100
Heart_YOUNG	G>C	0.0224	-3.49	0.112
Brain_OLD	C>T	0.0283	2.69	0.122
Brain_OLD	T>G	0.0286	2.94	0.122
Brain_OLD	A>T	0.0322	2.68	0.129
Heart_YOUNG	T>A	0.0759	3.05	0.285
Brain_OLD	A>G	0.0810	-2.14	0.286
Heart_YOUNG	C>T	0.0885	-2.05	0.295
Brain_YOUNG	A>C	0.0935	1.94	0.295
Brain_YOUNG	C>T	0.1121	2.20	0.312
Brain_YOUNG	T>G	0.1145	2.21	0.312
Heart_YOUNG	A>G	0.1108	-2.01	0.312
Heart_OLD	A>C	0.1602	1.59	0.356
Heart_OLD	C>A	0.1428	1.73	0.356
Heart_OLD	T>C	0.1516	-1.59	0.356
Liver_OLD	A>C	0.1599	1.75	0.356
Liver_OLD	T>C	0.1548	-1.64	0.356
Heart_OLD	T>A	0.1792	1.66	0.384
Liver_OLD	G>T	0.1980	1.45	0.410
Brain_OLD	G>C	0.2138	1.38	0.428
Brain_OLD	C>G	0.2324	1.36	0.450
Heart_YOUNG	A>T	0.2549	1.26	0.463
Heart_YOUNG	T>C	0.2502	1.31	0.463
Brain_OLD	A>C	0.2677	1.33	0.469

Brain_OLD	T>C	0.3151	1.08	0.469
Brain_YOUNG	A>G	0.3358	-1.03	0.469
Brain_YOUNG	C>G	0.3075	1.16	0.469
Heart_OLD	A>T	0.3310	1.05	0.469
Heart_OLD	G>T	0.3135	1.09	0.469
Heart_YOUNG	A>C	0.2834	1.34	0.469
Heart_YOUNG	T>G	0.3112	1.12	0.469
Liver_OLD	C>A	0.3293	1.04	0.469
Liver_OLD	C>G	0.2751	1.18	0.469
Heart_OLD	G>A	0.3850	-0.97	0.503
Liver_OLD	A>T	0.3849	0.97	0.503
Liver_OLD	T>G	0.3860	0.97	0.503
Heart_OLD	C>G	0.4304	-0.85	0.538
Heart_OLD	C>T	0.4265	-0.84	0.538
Brain_YOUNG	C>A	0.4620	0.84	0.566
Brain_YOUNG	G>T	0.4804	0.80	0.576
Brain_YOUNG	G>C	0.5178	0.70	0.609
Liver_OLD	G>C	0.5658	-0.61	0.653
Brain_YOUNG	A>T	0.6073	0.54	0.688
Brain_OLD	G>A	0.6333	0.50	0.704
Heart_YOUNG	G>A	0.7230	0.37	0.789
Heart_OLD	G>C	0.7645	-0.31	0.819
Heart_OLD	T>G	0.8214	0.23	0.865
Heart_OLD	A>G	0.8718	0.17	0.902
Brain_YOUNG	T>C	0.8889	0.15	0.904
Liver_OLD	A>G	0.9225	0.10	0.923

A two-sided t-test was conducted on each mutation type. p-values were corrected for multiple hypothesis testing using the Benjamini-Hochberg correction.

Table A.4: **Strain-specific mutation rates.**

Brain

term	estimate	std. error	p-value
Intercept	-13.14	0.016	<2e-16
ageOLD	0.66	0.023	<2e-16
strainAKR	0.36	0.023	<2e-16
strainALR	-0.06	0.031	0.06075
strainFVB	0.03	0.029	2.74E-01
strainNZB	-0.20	0.027	5.39E-13
ageOLD:strainAKR	-0.28	0.032	<2e-16
ageOLD:strainALR	-0.12	0.038	0.00165
ageOLD:strainFVB	0.02	0.037	0.64891
ageOLD:strainNZB	0.24	0.035	2.63E-12

Heart

term	estimate	std. error	p-value
Intercept	-13.61	0.022	<2e-16
ageOLD	0.66	0.026	<2e-16
strainAKR	0.57	0.028	<2e-16
strainALR	0.11	0.032	0.000805
strainFVB	-0.08	0.030	7.99E-03
strainNZB	0.33	0.026	<2e-16
ageOLD:strainAKR	-0.67	0.035	<2e-16
ageOLD:strainALR	-0.10	0.040	0.016975
ageOLD:strainFVB	0.06	0.042	0.187257
ageOLD:strainNZB	-0.32	0.037	<2e-16

Liver

term	estimate	std. error	p-value
Intercept	-12.85	0.016	<2e-16
ageOLD	0.46	0.026	<2e-16
strainALR	-0.10	0.021	6.05E-07
strainFVB	-0.36	0.028	<2e-16
strainNZB	-0.02	0.018	1.94E-01
ageOLD:strainALR	0.18	0.034	7.87E-08
ageOLD:strainFVB	0.87	0.037	<2e-16
ageOLD:strainNZB	0.56	0.031	<2e-16

For each tissue the following Poisson regression was ran: $glm(total\_mut\_count\ age * strain + offset((log(total\_duplex\_depth\_bp))$, family = poisson(link = "log")). Intercept is set for **ageYOUNG** and **strainB6** for heart and brain. B6 data omitted for Liver and

intercept is set for FVB Young. FVB was chosen due to its lack of haplotype specificity compared to WT in heart and liver.

Table A.5: **Two-sided fisher's exact test for age-associated changes in mutation type-specific frequencies.**

Mutation Type	**p-value**	**OR**	**adjusted p-value (BH correction)**
G>A	1E-242	1.528	4.14E-242
G>T	5.6E-286	0.566	4.51E-285
T>C	7.6E-58	1.399	1.22E-57
G>C	5.8E-142	0.521	1.55E-141
DEL	1.7E-47	1.354	2.26E-47
T>A	1.8E-01	1.041	1.77E-01
T>G	1.3E-22	1.839	1.5E-22
INS	1.4E-129	0.550	2.84E-129

The mutation counts for strain and tissues were aggregated across age for each mutation type. B6 Liver counts were excluded from this analysis since the old mutation counts did not have a young mutation count complement.

Table A.6: **Comparison of reversion mutation frequency to genome-wide background mutation frequency.**

NZB

Strain	Age	Tissue	Statistic	p-value	adjusted p-value (BH)	Sig.
NZB	Old	Brain	1253319	6.5E-118	7.8E-118	SIG
NZB	Young	Brain	1449240	0E+00	0E+00	SIG
NZB	Old	Heart	1243366	7.3E-130	1.1E-129	SIG
NZB	Young	Heart	1284941	0E+00	0E+00	SIG
NZB	Old	Liver	1224888	3.4E-53	3.4E-53	SIG
NZB	Young	Liver	1374159	0E+00	0E+00	SIG

AKR, ALR, FVB

Strain	Tissue	Age	Position	p-value	adjusted p-value (BH)	Sig.
FVB	Brain	OLD	7777	1.8E-04	1.9E-03	SIG
FVB	Brain	YOUNG	7777	6.1E-04	2.5E-03	SIG
FVB	Heart	OLD	7777	3.7E-04	1.9E-03	SIG
FVB	Heart	OLD	9460	1.4E-03	4.2E-03	SIG
FVB	Heart	YOUNG	7777	1.8E-03	5.1E-03	SIG
FVB	Liver	OLD	7777	3.7E-04	1.9E-03	SIG
ALR	Brain	OLD	4738	2.5E-04	1.9E-03	SIG
ALR	Brain	OLD	9347	2.5E-04	1.9E-03	SIG
ALR	Brain	YOUNG	4738	3.3E-03	8.4E-03	SIG
ALR	Heart	OLD	4738	3.1E-04	1.9E-03	SIG
ALR	Heart	OLD	9347	3.1E-04	1.9E-03	SIG
ALR	Liver	OLD	4738	4.3E-04	1.9E-03	SIG
ALR	Liver	OLD	9347	8E-04	2.9E-03	SIG
AKR	Heart	OLD	9460	1.3E-03	4.2E-03	SIG
FVB	Brain	OLD	9460	4E-02	5.3E-02	N.S.
FVB	Brain	YOUNG	9460	7.3E-02	8.8E-02	N.S.
FVB	Heart	YOUNG	9460	5.1E-03	1.1E-02	N.S.
FVB	Liver	OLD	9460	2.2E-01	2.2E-01	N.S.
FVB	Liver	YOUNG	7777	1.4E-02	2.3E-02	N.S.
FVB	Liver	YOUNG	9460	1.4E-02	2.3E-02	N.S.
ALR	Brain	OLD	9460	3.9E-02	5.3E-02	N.S.
ALR	Brain	YOUNG	9347	5.6E-02	6.9E-02	N.S.
ALR	Brain	YOUNG	9460	5.6E-02	6.9E-02	N.S.
ALR	Heart	OLD	9460	1E-01	1.1E-01	N.S.
ALR	Heart	YOUNG	4738	9.8E-03	1.8E-02	N.S.
ALR	Heart	YOUNG	9347	9.8E-03	1.8E-02	N.S.

ALR	Heart	YOUNG	9460	9.8E-03	1.8E-02	N.S.
ALR	Liver	OLD	9460	1.6E-01	1.6E-01	N.S.
ALR	Liver	YOUNG	4738	4.6E-03	1.1E-02	N.S.
ALR	Liver	YOUNG	9347	2.1E-02	3.1E-02	N.S.
ALR	Liver	YOUNG	9460	2.1E-02	3.1E-02	N.S.
AKR	Brain	OLD	9460	1.5E-01	1.6E-01	N.S.
AKR	Brain	YOUNG	9460	3.9E-02	5.3E-02	N.S.
AKR	Heart	YOUNG	9460	1.2E-01	1.3E-01	N.S.
AKR	Liver	OLD	9460	9.4E-02	1.1E-01	N.S.
AKR	Liver	YOUNG	9460	1.2E-02	2.1E-02	N.S.

To compare the reversion mutation frequency per position to background mutation frequency, we calculated the empirical p-values for haplotype sites in AKR, ALR, and FVB (one-sided test) and compared the average mutation frequency for the reversion and background distributions using a Wilcoxon Rank Sum Test for NZB.

Table A.7: **Two sided fisher's exact tests for the association of the haplotype site and the B6 reversion alleles. P-values were adjusted using the Benjamini-Hochberg correction.**

Strain	Tissue	Age	Position	Conplastic Allele	B6 Allele	p-value	OR	adjusted p-value (BH)	Sig.
AKR	Heart	OLD	9460	G	A	1.4E-22	Inf	2.5E-21	SIG
ALR	Brain	OLD	4738	T	G	9.91E-268	Inf	7.1E-266	SIG
ALR	Brain	OLD	9347	T	C	2.93E-150	Inf	1.7E-148	SIG
ALR	Heart	OLD	4738	T	G	3.03E-166	Inf	1.9E-164	SIG
ALR	Heart	OLD	9347	T	C	3.28E-125	Inf	1.7E-123	SIG
ALR	Liver	OLD	4738	T	G	3.97E-74	Inf	1.6E-72	SIG
ALR	Liver	OLD	9347	T	C	5.43E-28	Inf	1.3E-26	SIG
FVB	Brain	OLD	7777	T	G	8.61E-301	4.5E+15	7E-299	SIG
FVB	Heart	OLD	7777	T	G	3.42E-65	Inf	1.3E-63	SIG
FVB	Liver	OLD	7777	T	G	7.92E-104	Inf	3.5E-102	SIG
NZB	Brain	OLD	54	T	C	1.74E-08	Inf	4.7E-08	SIG
NZB	Brain	OLD	1518	T	C	2.38E-05	Inf	5E-05	SIG
NZB	Brain	OLD	1589	T	C	6.13E-04	Inf	1.2E-03	SIG
NZB	Brain	OLD	2339	T	C	9.48E-07	Inf	2.2E-06	SIG
NZB	Brain	OLD	2524	T	C	3.56E-09	Inf	1E-08	SIG
NZB	Brain	OLD	2797	T	C	3.56E-09	Inf	1E-08	SIG
NZB	Brain	OLD	2839	T	C	7.3E-10	Inf	2.2E-09	SIG
NZB	Brain	OLD	2933	T	C	9.48E-07	Inf	2.2E-06	SIG
NZB	Brain	OLD	3931	T	C	4.74E-06	Inf	1E-05	SIG
NZB	Brain	OLD	4122	T	C	3.86E-08	Inf	1E-07	SIG
NZB	Brain	OLD	4275	T	C	1.06E-05	Inf	2.3E-05	SIG
NZB	Brain	OLD	4407	T	C	4.74E-06	Inf	1E-05	SIG
NZB	Brain	OLD	4731	T	C	1.06E-05	Inf	2.3E-05	SIG
NZB	Brain	OLD	4884	G	T	3.07E-14	Inf	1.5E-13	SIG
NZB	Brain	OLD	4902	G	T	8.43E-12	3.12E+01	3.1E-11	SIG
NZB	Brain	OLD	5462	T	C	4.74E-06	Inf	1E-05	SIG

NZB	Brain	OLD	5929	T	C	3.86E-08	Inf	1E-07	SIG
NZB	Brain	OLD	7545	G	A	0E+00	Inf	0E+00	SIG
NZB	Brain	OLD	11842	T	C	2.12E-06	Inf	4.8E-06	SIG
NZB	Brain	OLD	11845	T	C	2.12E-06	Inf	4.8E-06	SIG
NZB	Brain	OLD	11932	G	T	2.57E-12	Inf	9.7E-12	SIG
NZB	Brain	OLD	12352	T	C	1.91E-07	Inf	4.6E-07	SIG
NZB	Brain	OLD	12574	T	A	1.25E-08	Inf	3.4E-08	SIG
NZB	Brain	OLD	13003	T	C	4.25E-07	Inf	1E-06	SIG
NZB	Brain	OLD	13443	T	C	2.12E-06	Inf	4.8E-06	SIG
NZB	Brain	OLD	13688	T	C	4.25E-07	Inf	1E-06	SIG
NZB	Brain	OLD	14210	T	C	1.91E-07	Inf	4.6E-07	SIG
NZB	Brain	OLD	14641	T	C	8.58E-08	Inf	2.1E-07	SIG
NZB	Brain	OLD	14737	T	C	7.86E-09	Inf	2.2E-08	SIG
NZB	Brain	OLD	15498	T	A	4.11E-08	Inf	1.1E-07	SIG
NZB	Brain	OLD	15548	T	C	1.2E-04	Inf	2.5E-04	SIG
NZB	Brain	OLD	15577	T	A	1.45E-03	7.31E+00	2.8E-03	SIG
NZB	Brain	OLD	15587	T	C	6.13E-04	Inf	1.2E-03	SIG
NZB	Brain	OLD	15602	T	C	5.35E-05	Inf	1.1E-04	SIG
NZB	Brain	OLD	15916	T	C	5.02E-08	3.09E+01	1.3E-07	SIG
NZB	Brain	OLD	16016	G	T	6.49E-22	Inf	1.1E-20	SIG
NZB	Heart	OLD	2339	T	C	2E-03	Inf	3.9E-03	SIG
NZB	Heart	OLD	4902	G	T	2.63E-04	3.14E+01	5.3E-04	SIG
NZB	Heart	OLD	5462	T	C	4.83E-03	Inf	9.3E-03	SIG
NZB	Heart	OLD	7545	G	A	0E+00	Inf	0E+00	SIG
NZB	Heart	OLD	11932	G	T	2.6E-03	Inf	5.1E-03	SIG
NZB	Heart	OLD	12352	T	C	4.83E-03	Inf	9.3E-03	SIG
NZB	Heart	OLD	13003	T	C	1.43E-04	Inf	2.9E-04	SIG
NZB	Heart	OLD	13688	T	C	1.43E-04	Inf	2.9E-04	SIG
NZB	Heart	OLD	15916	T	C	4.83E-03	Inf	9.3E-03	SIG
NZB	Heart	OLD	16016	G	T	1.32E-07	Inf	3.2E-07	SIG
NZB	Liver	OLD	7545	G	A	0E+00	Inf	0E+00	SIG
NZB	Liver	OLD	11932	G	T	1.32E-04	Inf	2.7E-04	SIG

NZB	Liver	OLD	12574	T	A	7.79E-04	Inf	1.5E-03	SIG
NZB	Liver	OLD	16016	G	T	9.63E-41	Inf	3.1E-39	SIG
ALR	Brain	YOUNG	4738	T	G	8.84E-12	Inf	3.2E-11	SIG
ALR	Heart	YOUNG	4738	T	G	1.03E-04	Inf	2.1E-04	SIG
ALR	Liver	YOUNG	4738	T	G	1.23E-16	Inf	9.8E-16	SIG
FVB	Brain	YOUNG	7777	T	G	4.6E-118	Inf	2.2E-116	SIG
FVB	Heart	YOUNG	7777	T	G	4.21E-41	Inf	1.4E-39	SIG
NZB	Brain	YOUNG	54	T	C	1.08E-18	Inf	1.2E-17	SIG
NZB	Brain	YOUNG	1352	G	A	1.47E-14	5.12E+01	7.8E-14	SIG
NZB	Brain	YOUNG	1518	T	C	3.18E-22	Inf	5.3E-21	SIG
NZB	Brain	YOUNG	1589	T	C	3.59E-21	Inf	5.2E-20	SIG
NZB	Brain	YOUNG	1821	G	A	1.26E-13	Inf	5.8E-13	SIG
NZB	Brain	YOUNG	2200	G	A	1.65E-10	Inf	5.2E-10	SIG
NZB	Brain	YOUNG	2339	T	C	1.82E-20	Inf	2.3E-19	SIG
NZB	Brain	YOUNG	2524	T	C	1.82E-20	Inf	2.3E-19	SIG
NZB	Brain	YOUNG	2765	G	A	1.98E-12	Inf	7.6E-12	SIG
NZB	Brain	YOUNG	2766	G	A	1.96E-12	Inf	7.6E-12	SIG
NZB	Brain	YOUNG	2797	T	C	1.54E-16	Inf	1.2E-15	SIG
NZB	Brain	YOUNG	2813	G	A	9.76E-13	Inf	3.9E-12	SIG
NZB	Brain	YOUNG	2839	T	C	2.93E-17	Inf	2.6E-16	SIG
NZB	Brain	YOUNG	2933	T	C	1.54E-16	Inf	1.2E-15	SIG
NZB	Brain	YOUNG	3193	G	A	1.26E-13	Inf	5.8E-13	SIG
NZB	Brain	YOUNG	3259	G	A	2.48E-13	Inf	1.1E-12	SIG
NZB	Brain	YOUNG	3421	G	A	3.35E-14	Inf	1.6E-13	SIG
NZB	Brain	YOUNG	3466	G	A	9.76E-13	Inf	3.9E-12	SIG
NZB	Brain	YOUNG	3598	G	A	8.67E-11	Inf	2.9E-10	SIG
NZB	Brain	YOUNG	3691	G	A	6.9E-12	Inf	2.5E-11	SIG
NZB	Brain	YOUNG	3931	T	C	2.93E-17	Inf	2.6E-16	SIG
NZB	Brain	YOUNG	4122	T	C	7.13E-22	Inf	1.1E-20	SIG
NZB	Brain	YOUNG	4275	T	C	1.82E-20	Inf	2.3E-19	SIG
NZB	Brain	YOUNG	4323	G	A	1.36E-15	Inf	8.4E-15	SIG
NZB	Brain	YOUNG	4407	T	C	1.08E-18	Inf	1.2E-17	SIG

NZB	Brain	YOUNG	4705	G	A	6.48E-14	Inf	3.1E-13	SIG
NZB	Brain	YOUNG	4731	T	C	2.46E-18	Inf	2.6E-17	SIG
NZB	Brain	YOUNG	4770	G	A	9.76E-13	Inf	3.9E-12	SIG
NZB	Brain	YOUNG	4884	G	T	2.74E-26	1.38E+02	6.5E-25	SIG
NZB	Brain	YOUNG	4902	G	T	3.66E-31	Inf	1E-29	SIG
NZB	Brain	YOUNG	5462	T	C	2.1E-19	Inf	2.6E-18	SIG
NZB	Brain	YOUNG	5551	G	A	1.96E-12	Inf	7.6E-12	SIG
NZB	Brain	YOUNG	5929	T	C	2.6E-24	Inf	5.1E-23	SIG
NZB	Brain	YOUNG	6040	G	A	7.28E-16	Inf	4.8E-15	SIG
NZB	Brain	YOUNG	6406	T	C	1.87E-15	Inf	1.1E-14	SIG
NZB	Brain	YOUNG	6469	G	A	1.58E-11	4.04E+01	5.5E-11	SIG
NZB	Brain	YOUNG	6574	T	C	1.08E-18	Inf	1.2E-17	SIG
NZB	Brain	YOUNG	6619	T	C	1.54E-16	Inf	1.2E-15	SIG
NZB	Brain	YOUNG	6784	T	C	1.87E-15	Inf	1.1E-14	SIG
NZB	Brain	YOUNG	7410	G	A	1.15E-09	Inf	3.4E-09	SIG
NZB	Brain	YOUNG	7545	G	A	0E+00	1.64E+04	0E+00	SIG
NZB	Brain	YOUNG	7869	T	C	3.53E-16	Inf	2.4E-15	SIG
NZB	Brain	YOUNG	8438	G	A	5.02E-07	Inf	1.2E-06	SIG
NZB	Brain	YOUNG	8466	G	A	5.02E-07	Inf	1.2E-06	SIG
NZB	Brain	YOUNG	8567	T	C	1.13E-10	Inf	3.6E-10	SIG
NZB	Brain	YOUNG	8857	G	A	3.69E-12	Inf	1.4E-11	SIG
NZB	Brain	YOUNG	8863	T	C	1.28E-17	Inf	1.2E-16	SIG
NZB	Brain	YOUNG	9136	G	A	8.4E-09	Inf	2.3E-08	SIG
NZB	Brain	YOUNG	9151	G	A	1.64E-08	Inf	4.4E-08	SIG
NZB	Brain	YOUNG	9390	G	A	4.84E-06	2.15E+01	1.1E-05	SIG
NZB	Brain	YOUNG	9529	T	C	4.78E-11	Inf	1.6E-10	SIG
NZB	Brain	YOUNG	9580	T	C	2.9E-13	Inf	1.3E-12	SIG
NZB	Brain	YOUNG	9598	G	A	1.65E-10	Inf	5.2E-10	SIG
NZB	Brain	YOUNG	9984	T	C	8.12E-16	Inf	5.3E-15	SIG
NZB	Brain	YOUNG	10546	T	C	1.87E-15	Inf	1.1E-14	SIG
NZB	Brain	YOUNG	10582	G	A	2.43E-11	Inf	8.3E-11	SIG
NZB	Brain	YOUNG	10951	T	G	7.79E-26	Inf	1.8E-24	SIG

NZB	Brain	YOUNG	11842	T	C	3.53E-16	Inf	2.4E-15	SIG
NZB	Brain	YOUNG	11845	T	C	5.61E-18	Inf	5.6E-17	SIG
NZB	Brain	YOUNG	11932	G	T	2.17E-28	Inf	5.6E-27	SIG
NZB	Brain	YOUNG	12352	T	C	3.53E-16	Inf	2.4E-15	SIG
NZB	Brain	YOUNG	12574	T	A	2.11E-24	Inf	4.3E-23	SIG
NZB	Brain	YOUNG	12694	G	A	3.35E-14	Inf	1.6E-13	SIG
NZB	Brain	YOUNG	12834	G	A	1.37E-17	Inf	1.3E-16	SIG
NZB	Brain	YOUNG	12889	G	A	1.37E-17	Inf	1.3E-16	SIG
NZB	Brain	YOUNG	13003	T	C	7.13E-22	Inf	1.1E-20	SIG
NZB	Brain	YOUNG	13443	T	C	3.53E-16	Inf	2.4E-15	SIG
NZB	Brain	YOUNG	13611	G	A	3.35E-14	Inf	1.6E-13	SIG
NZB	Brain	YOUNG	13688	T	C	8.08E-21	Inf	1.1E-19	SIG
NZB	Brain	YOUNG	13780	G	A	1.74E-14	Inf	9.1E-14	SIG
NZB	Brain	YOUNG	13781	G	A	9.14E-15	Inf	5E-14	SIG
NZB	Brain	YOUNG	13836	G	A	1.96E-12	Inf	7.6E-12	SIG
NZB	Brain	YOUNG	13982	G	A	1.29E-11	Inf	4.6E-11	SIG
NZB	Brain	YOUNG	14185	G	A	2.18E-16	Inf	1.6E-15	SIG
NZB	Brain	YOUNG	14210	T	C	6.37E-23	Inf	1.2E-21	SIG
NZB	Brain	YOUNG	14362	G	A	3.91E-16	Inf	2.7E-15	SIG
NZB	Brain	YOUNG	14641	T	C	5.78E-24	Inf	1.1E-22	SIG
NZB	Brain	YOUNG	14737	T	C	5.61E-18	Inf	5.6E-17	SIG
NZB	Brain	YOUNG	15498	T	A	3.01E-17	Inf	2.7E-16	SIG
NZB	Brain	YOUNG	15548	T	C	3.53E-16	Inf	2.4E-15	SIG
NZB	Brain	YOUNG	15577	T	A	4.28E-15	Inf	2.4E-14	SIG
NZB	Brain	YOUNG	15587	T	C	1.25E-13	Inf	5.8E-13	SIG
NZB	Brain	YOUNG	15602	T	C	2.31E-14	Inf	1.2E-13	SIG
NZB	Brain	YOUNG	15656	G	A	1.65E-10	Inf	5.2E-10	SIG
NZB	Brain	YOUNG	15916	T	C	9.29E-20	Inf	1.2E-18	SIG
NZB	Brain	YOUNG	16016	G	T	5.15E-35	Inf	1.5E-33	SIG
NZB	Brain	YOUNG	16267	G	A	7.28E-16	Inf	4.8E-15	SIG
NZB	Brain	YOUNG	16271	G	A	1.36E-15	Inf	8.4E-15	SIG
NZB	Heart	YOUNG	54	T	C	3.11E-06	Inf	6.9E-06	SIG

NZB	Heart	YOUNG	1352	G	A	1.37E-08	3.67E+01	3.7E-08	SIG
NZB	Heart	YOUNG	1518	T	C	4.75E-08	Inf	1.2E-07	SIG
NZB	Heart	YOUNG	1589	T	C	1.34E-06	Inf	3.1E-06	SIG
NZB	Heart	YOUNG	1821	G	A	7.77E-09	Inf	2.2E-08	SIG
NZB	Heart	YOUNG	2200	G	A	3.85E-06	Inf	8.5E-06	SIG
NZB	Heart	YOUNG	2339	T	C	2.52E-07	Inf	6E-07	SIG
NZB	Heart	YOUNG	2524	T	C	3.24E-10	Inf	9.9E-10	SIG
NZB	Heart	YOUNG	2765	G	A	5.11E-11	1.83E+01	1.7E-10	SIG
NZB	Heart	YOUNG	2766	G	A	1.34E-11	4.96E+01	4.7E-11	SIG
NZB	Heart	YOUNG	2797	T	C	8.98E-09	Inf	2.5E-08	SIG
NZB	Heart	YOUNG	2813	G	A	7.07E-10	4.22E+01	2.1E-09	SIG
NZB	Heart	YOUNG	2839	T	C	1.7E-09	Inf	5E-09	SIG
NZB	Heart	YOUNG	2933	T	C	1.34E-06	Inf	3.1E-06	SIG
NZB	Heart	YOUNG	3193	G	A	1.73E-07	Inf	4.2E-07	SIG
NZB	Heart	YOUNG	3259	G	A	3.52E-10	Inf	1.1E-09	SIG
NZB	Heart	YOUNG	3421	G	A	5.04E-07	1.22E+01	1.2E-06	SIG
NZB	Heart	YOUNG	3466	G	A	2.18E-08	Inf	5.8E-08	SIG
NZB	Heart	YOUNG	3598	G	A	4.48E-11	Inf	1.5E-10	SIG
NZB	Heart	YOUNG	3691	G	A	9.75E-11	4.59E+01	3.2E-10	SIG
NZB	Heart	YOUNG	3931	T	C	3.24E-10	Inf	9.9E-10	SIG
NZB	Heart	YOUNG	4122	T	C	5.21E-12	Inf	1.9E-11	SIG
NZB	Heart	YOUNG	4275	T	C	2.28E-12	Inf	8.7E-12	SIG
NZB	Heart	YOUNG	4323	G	A	1.2E-14	Inf	6.5E-14	SIG
NZB	Heart	YOUNG	4407	T	C	6.19E-11	Inf	2.1E-10	SIG
NZB	Heart	YOUNG	4705	G	A	1.5E-03	1.47E+01	2.9E-03	SIG
NZB	Heart	YOUNG	4770	G	A	6.88E-04	Inf	1.4E-03	SIG
NZB	Heart	YOUNG	4884	G	T	4.09E-05	Inf	8.5E-05	SIG
NZB	Heart	YOUNG	4902	G	T	9.47E-05	Inf	2E-04	SIG
NZB	Heart	YOUNG	5462	T	C	7.19E-06	Inf	1.5E-05	SIG
NZB	Heart	YOUNG	5551	G	A	6.14E-08	Inf	1.5E-07	SIG
NZB	Heart	YOUNG	5929	T	C	1.19E-11	Inf	4.2E-11	SIG
NZB	Heart	YOUNG	6040	G	A	1.55E-15	Inf	9.5E-15	SIG

NZB	Heart	YOUNG	7545	G	A	0E+00	Inf	0E+00	SIG
NZB	Heart	YOUNG	8567	T	C	2.62E-03	Inf	5.1E-03	SIG
NZB	Heart	YOUNG	9390	G	A	6.88E-04	Inf	1.4E-03	SIG
NZB	Heart	YOUNG	11842	T	C	4.75E-08	Inf	1.2E-07	SIG
NZB	Heart	YOUNG	11845	T	C	1.09E-07	Inf	2.7E-07	SIG
NZB	Heart	YOUNG	11932	G	T	7.63E-06	Inf	1.6E-05	SIG
NZB	Heart	YOUNG	12352	T	C	1.42E-10	Inf	4.6E-10	SIG
NZB	Heart	YOUNG	12574	T	A	6.52E-13	Inf	2.7E-12	SIG
NZB	Heart	YOUNG	12694	G	A	9.75E-11	4.59E+01	3.2E-10	SIG
NZB	Heart	YOUNG	12834	G	A	2.62E-13	Inf	1.2E-12	SIG
NZB	Heart	YOUNG	12889	G	A	2.63E-10	4.4E+01	8.2E-10	SIG
NZB	Heart	YOUNG	13003	T	C	3.24E-10	Inf	9.9E-10	SIG
NZB	Heart	YOUNG	13443	T	C	4.75E-08	Inf	1.2E-07	SIG
NZB	Heart	YOUNG	13611	G	A	2.77E-09	Inf	7.9E-09	SIG
NZB	Heart	YOUNG	13688	T	C	7.43E-10	Inf	2.2E-09	SIG
NZB	Heart	YOUNG	13780	G	A	3.66E-08	3.49E+01	9.6E-08	SIG
NZB	Heart	YOUNG	13781	G	A	3.52E-10	Inf	1.1E-09	SIG
NZB	Heart	YOUNG	13836	G	A	3.52E-10	Inf	1.1E-09	SIG
NZB	Heart	YOUNG	13982	G	A	2.63E-10	4.4E+01	8.2E-10	SIG
NZB	Heart	YOUNG	14185	G	A	9.86E-10	Inf	2.9E-09	SIG
NZB	Heart	YOUNG	14210	T	C	4.75E-08	Inf	1.2E-07	SIG
NZB	Heart	YOUNG	14362	G	A	7.31E-13	Inf	3E-12	SIG
NZB	Heart	YOUNG	14641	T	C	1.09E-07	Inf	2.7E-07	SIG
NZB	Heart	YOUNG	14737	T	C	4.75E-08	Inf	1.2E-07	SIG
NZB	Heart	YOUNG	15498	T	A	2.33E-06	Inf	5.2E-06	SIG
NZB	Heart	YOUNG	15548	T	C	7.19E-06	Inf	1.5E-05	SIG
NZB	Heart	YOUNG	15577	T	A	6.69E-06	2.73E+01	1.5E-05	SIG
NZB	Heart	YOUNG	15587	T	C	5.82E-07	Inf	1.4E-06	SIG
NZB	Heart	YOUNG	15602	T	C	2.52E-07	Inf	6E-07	SIG
NZB	Heart	YOUNG	15656	G	A	2.77E-09	Inf	7.9E-09	SIG
NZB	Heart	YOUNG	15916	T	C	3.91E-09	Inf	1.1E-08	SIG
NZB	Heart	YOUNG	16016	G	T	2.17E-08	Inf	5.8E-08	SIG

NZB	Heart	YOUNG	16267	G	A	3.66E-08	3.49E+01	9.6E-08	SIG
NZB	Heart	YOUNG	16271	G	A	2.04E-12	Inf	7.8E-12	SIG
NZB	Liver	YOUNG	54	T	C	4.31E-15	Inf	2.4E-14	SIG
NZB	Liver	YOUNG	1352	G	A	3.78E-17	Inf	3.1E-16	SIG
NZB	Liver	YOUNG	1518	T	C	5.77E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	1589	T	C	2.61E-09	Inf	7.5E-09	SIG
NZB	Liver	YOUNG	1821	G	A	4.47E-17	6.34E+01	3.6E-16	SIG
NZB	Liver	YOUNG	2200	G	A	7.88E-10	Inf	2.3E-09	SIG
NZB	Liver	YOUNG	2339	T	C	1.57E-13	Inf	7.1E-13	SIG
NZB	Liver	YOUNG	2524	T	C	5.77E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	2765	G	A	3.67E-12	1.78E+01	1.4E-11	SIG
NZB	Liver	YOUNG	2766	G	A	1.03E-14	Inf	5.6E-14	SIG
NZB	Liver	YOUNG	2797	T	C	2.68E-10	Inf	8.3E-10	SIG
NZB	Liver	YOUNG	2813	G	A	1.02E-15	5.84E+01	6.4E-15	SIG
NZB	Liver	YOUNG	2839	T	C	1.82E-12	Inf	7.2E-12	SIG
NZB	Liver	YOUNG	2933	T	C	5.78E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	3193	G	A	7.8E-22	8.08E+01	1.2E-20	SIG
NZB	Liver	YOUNG	3259	G	A	7.11E-16	2.24E+01	4.8E-15	SIG
NZB	Liver	YOUNG	3421	G	A	3.12E-17	3.42E+01	2.7E-16	SIG
NZB	Liver	YOUNG	3466	G	A	9.31E-16	Inf	6E-15	SIG
NZB	Liver	YOUNG	3598	G	A	1.09E-13	5.1E+01	5.1E-13	SIG
NZB	Liver	YOUNG	3691	G	A	2.3E-14	Inf	1.2E-13	SIG
NZB	Liver	YOUNG	3931	T	C	1.02E-12	Inf	4.1E-12	SIG
NZB	Liver	YOUNG	4122	T	C	5.78E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	4275	T	C	3.76E-11	Inf	1.3E-10	SIG
NZB	Liver	YOUNG	4323	G	A	5.69E-21	Inf	8.1E-20	SIG
NZB	Liver	YOUNG	4407	T	C	5.78E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	4705	G	A	3.78E-17	Inf	3.1E-16	SIG
NZB	Liver	YOUNG	4731	T	C	3.27E-12	Inf	1.2E-11	SIG
NZB	Liver	YOUNG	4770	G	A	4.63E-15	Inf	2.6E-14	SIG
NZB	Liver	YOUNG	4884	G	T	3.87E-17	6.84E+01	3.2E-16	SIG
NZB	Liver	YOUNG	4902	G	T	4.57E-14	3.03E+01	2.2E-13	SIG

NZB	Liver	YOUNG	5462	T	C	5.78E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	5551	G	A	3.42E-18	Inf	3.5E-17	SIG
NZB	Liver	YOUNG	5929	T	C	4.53E-14	Inf	2.2E-13	SIG
NZB	Liver	YOUNG	6040	G	A	1.26E-20	Inf	1.7E-19	SIG
NZB	Liver	YOUNG	7545	G	A	0E+00	4.5E+15	0E+00	SIG
NZB	Liver	YOUNG	8438	G	A	2.27E-04	1.62E+01	4.6E-04	SIG
NZB	Liver	YOUNG	8466	G	A	1.15E-05	2.12E+01	2.4E-05	SIG
NZB	Liver	YOUNG	8567	T	C	2.7E-05	Inf	5.7E-05	SIG
NZB	Liver	YOUNG	9136	G	A	8.8E-09	Inf	2.4E-08	SIG
NZB	Liver	YOUNG	9151	G	A	7.6E-05	1.06E+01	1.6E-04	SIG
NZB	Liver	YOUNG	9390	G	A	2.3E-14	Inf	1.2E-13	SIG
NZB	Liver	YOUNG	9529	T	C	1.57E-04	Inf	3.2E-04	SIG
NZB	Liver	YOUNG	9580	T	C	3.48E-07	Inf	8.3E-07	SIG
NZB	Liver	YOUNG	9598	G	A	3.94E-09	Inf	1.1E-08	SIG
NZB	Liver	YOUNG	10582	G	A	5.4E-06	2.24E+01	1.2E-05	SIG
NZB	Liver	YOUNG	10951	T	G	1.89E-11	Inf	6.6E-11	SIG
NZB	Liver	YOUNG	11842	T	C	1.69E-08	Inf	4.5E-08	SIG
NZB	Liver	YOUNG	11845	T	C	6.24E-08	Inf	1.6E-07	SIG
NZB	Liver	YOUNG	11932	G	T	1.55E-14	Inf	8.2E-14	SIG
NZB	Liver	YOUNG	12352	T	C	4.53E-14	Inf	2.2E-13	SIG
NZB	Liver	YOUNG	12574	T	A	3.35E-29	Inf	9.1E-28	SIG
NZB	Liver	YOUNG	12694	G	A	9.35E-18	6.59E+01	9.2E-17	SIG
NZB	Liver	YOUNG	12834	G	A	3.93E-25	Inf	8.3E-24	SIG
NZB	Liver	YOUNG	12889	G	A	1.54E-18	Inf	1.7E-17	SIG
NZB	Liver	YOUNG	13003	T	C	3.48E-16	Inf	2.4E-15	SIG
NZB	Liver	YOUNG	13443	T	C	2.01E-11	Inf	6.9E-11	SIG
NZB	Liver	YOUNG	13611	G	A	2.08E-15	Inf	1.2E-14	SIG
NZB	Liver	YOUNG	13688	T	C	5.77E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	13780	G	A	2.22E-15	5.72E+01	1.3E-14	SIG
NZB	Liver	YOUNG	13781	G	A	3.78E-17	Inf	3.1E-16	SIG
NZB	Liver	YOUNG	13836	G	A	3.1E-19	Inf	3.7E-18	SIG
NZB	Liver	YOUNG	13982	G	A	3.78E-17	Inf	3.1E-16	SIG

NZB	Liver	YOUNG	14185	G	A	3.04E-15	3.05E+01	1.7E-14	SIG
NZB	Liver	YOUNG	14210	T	C	7.14E-11	Inf	2.4E-10	SIG
NZB	Liver	YOUNG	14362	G	A	6.9E-19	Inf	8E-18	SIG
NZB	Liver	YOUNG	14641	T	C	8.53E-12	3.91E+01	3.1E-11	SIG
NZB	Liver	YOUNG	14737	T	C	1.89E-16	Inf	1.4E-15	SIG
NZB	Liver	YOUNG	15498	T	A	2.71E-15	8.16E+01	1.6E-14	SIG
NZB	Liver	YOUNG	15548	T	C	2.15E-06	Inf	4.8E-06	SIG
NZB	Liver	YOUNG	15577	T	A	1.58E-13	7.22E+01	7.1E-13	SIG
NZB	Liver	YOUNG	15587	T	C	2.79E-05	Inf	5.8E-05	SIG
NZB	Liver	YOUNG	15602	T	C	2.15E-06	Inf	4.8E-06	SIG
NZB	Liver	YOUNG	15656	G	A	1.4E-07	1.16E+01	3.4E-07	SIG
NZB	Liver	YOUNG	15916	T	C	5.77E-13	Inf	2.4E-12	SIG
NZB	Liver	YOUNG	16016	G	T	7.78E-43	1.7E+02	2.8E-41	SIG
NZB	Liver	YOUNG	16267	G	A	2.33E-22	Inf	4E-21	SIG
NZB	Liver	YOUNG	16271	G	A	1.77E-25	Inf	3.9E-24	SIG
AKR	Brain	OLD	9460	G	A	1E+00	0E+00	1E+00	N.S.
AKR	Liver	OLD	9460	G	A	1E+00	0E+00	1E+00	N.S.
ALR	Brain	OLD	9460	G	A	1E+00	0E+00	1E+00	N.S.
ALR	Heart	OLD	9460	G	A	1E+00	0E+00	1E+00	N.S.
ALR	Liver	OLD	9460	G	A	1E+00	0E+00	1E+00	N.S.
FVB	Brain	OLD	9460	G	A	2.1E-01	0E+00	3.3E-01	N.S.
FVB	Heart	OLD	9460	G	A	1.02E-01	Inf	1.7E-01	N.S.
FVB	Liver	OLD	9460	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	1352	G	A	2.36E-01	Inf	3.6E-01	N.S.
NZB	Brain	OLD	1821	G	A	2.36E-01	Inf	3.6E-01	N.S.
NZB	Brain	OLD	2200	G	A	2.3E-01	Inf	3.5E-01	N.S.
NZB	Brain	OLD	2765	G	A	3.83E-02	Inf	6.8E-02	N.S.
NZB	Brain	OLD	2766	G	A	5.88E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	2813	G	A	1.07E-01	4.75E+00	1.8E-01	N.S.
NZB	Brain	OLD	3193	G	A	7.08E-01	2.37E+00	9.6E-01	N.S.
NZB	Brain	OLD	3259	G	A	1.53E-01	Inf	2.5E-01	N.S.
NZB	Brain	OLD	3421	G	A	1.45E-01	Inf	2.4E-01	N.S.

NZB	Brain	OLD	3466	G	A	1.45E-01	Inf	2.4E-01	N.S.
NZB	Brain	OLD	3598	G	A	1.46E-01	Inf	2.4E-01	N.S.
NZB	Brain	OLD	3691	G	A	5.83E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	4323	G	A	3.85E-01	Inf	5.6E-01	N.S.
NZB	Brain	OLD	4705	G	A	9.16E-02	Inf	1.5E-01	N.S.
NZB	Brain	OLD	4770	G	A	1.53E-01	Inf	2.5E-01	N.S.
NZB	Brain	OLD	5551	G	A	5.88E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	6040	G	A	1.45E-01	Inf	2.4E-01	N.S.
NZB	Brain	OLD	6406	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	6469	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	6574	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	6619	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	6784	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	7410	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	7869	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	8438	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	8466	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	8567	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	8857	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	8863	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	9136	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	9151	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	9390	G	A	1E+00	Inf	1E+00	N.S.
NZB	Brain	OLD	9460	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	9529	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	9580	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	9598	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	9984	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	10546	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	10582	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	10951	T	G	1E+00	0E+00	1E+00	N.S.
NZB	Brain	OLD	12694	G	A	2.54E-02	Inf	4.6E-02	N.S.

NZB	Brain	OLD	12834	G	A	3.71E-02	Inf	6.6E-02	N.S.
NZB	Brain	OLD	12889	G	A	6.47E-03	Inf	1.2E-02	N.S.
NZB	Brain	OLD	13611	G	A	3.71E-02	Inf	6.6E-02	N.S.
NZB	Brain	OLD	13780	G	A	2.32E-01	3.76E+00	3.6E-01	N.S.
NZB	Brain	OLD	13781	G	A	5.83E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	13836	G	A	2.37E-02	Inf	4.3E-02	N.S.
NZB	Brain	OLD	13982	G	A	5.88E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	14185	G	A	5.83E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	14362	G	A	1.58E-01	4.15E+00	2.6E-01	N.S.
NZB	Brain	OLD	15656	G	A	5.88E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	16267	G	A	5.83E-02	Inf	1E-01	N.S.
NZB	Brain	OLD	16271	G	A	5.88E-02	Inf	1E-01	N.S.
NZB	Heart	OLD	54	T	C	2.84E-02	Inf	5.1E-02	N.S.
NZB	Heart	OLD	1352	G	A	3.59E-01	Inf	5.3E-01	N.S.
NZB	Heart	OLD	1518	T	C	6.9E-02	Inf	1.2E-01	N.S.
NZB	Heart	OLD	1589	T	C	2.84E-02	Inf	5.1E-02	N.S.
NZB	Heart	OLD	1821	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	2200	G	A	3.59E-01	Inf	5.3E-01	N.S.
NZB	Heart	OLD	2524	T	C	6.9E-02	Inf	1.2E-01	N.S.
NZB	Heart	OLD	2765	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	2766	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	2797	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	2813	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	2839	T	C	4.09E-01	Inf	6E-01	N.S.
NZB	Heart	OLD	2933	T	C	6.9E-02	Inf	1.2E-01	N.S.
NZB	Heart	OLD	3193	G	A	3.59E-01	Inf	5.3E-01	N.S.
NZB	Heart	OLD	3259	G	A	3.59E-01	Inf	5.3E-01	N.S.
NZB	Heart	OLD	3421	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	3466	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	3598	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	3691	G	A	6.03E-01	Inf	8.2E-01	N.S.
NZB	Heart	OLD	3931	T	C	1.68E-01	Inf	2.6E-01	N.S.

NZB	Heart	OLD	4122	T	C	1.17E-02	Inf	2.2E-02	N.S.
NZB	Heart	OLD	4275	T	C	1.68E-01	Inf	2.6E-01	N.S.
NZB	Heart	OLD	4323	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	4407	T	C	2.84E-02	Inf	5.1E-02	N.S.
NZB	Heart	OLD	4705	G	A	5.9E-01	Inf	8.1E-01	N.S.
NZB	Heart	OLD	4731	T	C	1.17E-02	Inf	2.2E-02	N.S.
NZB	Heart	OLD	4770	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	4884	G	T	1.89E-02	Inf	3.5E-02	N.S.
NZB	Heart	OLD	5551	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	5929	T	C	1.17E-02	Inf	2.2E-02	N.S.
NZB	Heart	OLD	6040	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	6406	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	6469	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	6574	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	6619	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	6784	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	7410	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	7869	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	8438	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	8466	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	8567	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	8857	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	8863	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	9136	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	9151	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	9390	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	9460	G	A	1.93E-01	0E+00	3E-01	N.S.
NZB	Heart	OLD	9529	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	9580	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	9598	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	9984	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	10546	T	C	1E+00	0E+00	1E+00	N.S.

NZB	Heart	OLD	10582	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	10951	T	G	1E+00	0E+00	1E+00	N.S.
NZB	Heart	OLD	11842	T	C	1.17E-02	Inf	2.2E-02	N.S.
NZB	Heart	OLD	11845	T	C	2.84E-02	Inf	5.1E-02	N.S.
NZB	Heart	OLD	12574	T	A	1.9E-02	Inf	3.5E-02	N.S.
NZB	Heart	OLD	12694	G	A	6.03E-01	Inf	8.2E-01	N.S.
NZB	Heart	OLD	12834	G	A	3.67E-01	Inf	5.4E-01	N.S.
NZB	Heart	OLD	12889	G	A	3.67E-01	Inf	5.4E-01	N.S.
NZB	Heart	OLD	13443	T	C	1.17E-02	Inf	2.2E-02	N.S.
NZB	Heart	OLD	13611	G	A	3.59E-01	Inf	5.3E-01	N.S.
NZB	Heart	OLD	13780	G	A	5.9E-01	Inf	8.1E-01	N.S.
NZB	Heart	OLD	13781	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	13836	G	A	1E+00	9.56E-01	1E+00	N.S.
NZB	Heart	OLD	13982	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	14185	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	14210	T	C	1.68E-01	Inf	2.6E-01	N.S.
NZB	Heart	OLD	14362	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	14641	T	C	1.17E-02	Inf	2.2E-02	N.S.
NZB	Heart	OLD	14737	T	C	1.17E-02	Inf	2.2E-02	N.S.
NZB	Heart	OLD	15498	T	A	5.11E-02	Inf	9E-02	N.S.
NZB	Heart	OLD	15548	T	C	6.9E-02	Inf	1.2E-01	N.S.
NZB	Heart	OLD	15577	T	A	6.29E-01	1.71E+00	8.5E-01	N.S.
NZB	Heart	OLD	15587	T	C	1.68E-01	Inf	2.6E-01	N.S.
NZB	Heart	OLD	15602	T	C	2.84E-02	Inf	5.1E-02	N.S.
NZB	Heart	OLD	15656	G	A	6.03E-01	Inf	8.2E-01	N.S.
NZB	Heart	OLD	16267	G	A	1E+00	Inf	1E+00	N.S.
NZB	Heart	OLD	16271	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	54	T	C	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	1352	G	A	1E+00	1.7E+00	1E+00	N.S.
NZB	Liver	OLD	1518	T	C	5.27E-02	Inf	9.2E-02	N.S.
NZB	Liver	OLD	1589	T	C	3.03E-01	Inf	4.6E-01	N.S.
NZB	Liver	OLD	1821	G	A	5.82E-01	Inf	8.1E-01	N.S.

NZB	Liver	OLD	2200	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	2339	T	C	1.65E-01	Inf	2.6E-01	N.S.
NZB	Liver	OLD	2524	T	C	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	2765	G	A	5.93E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	2766	G	A	5.82E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	2797	T	C	1.65E-01	Inf	2.6E-01	N.S.
NZB	Liver	OLD	2813	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	2839	T	C	3.03E-01	Inf	4.6E-01	N.S.
NZB	Liver	OLD	2933	T	C	3.03E-01	Inf	4.6E-01	N.S.
NZB	Liver	OLD	3193	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	3259	G	A	1E+00	1.13E+00	1E+00	N.S.
NZB	Liver	OLD	3421	G	A	2.14E-01	2.83E-01	3.3E-01	N.S.
NZB	Liver	OLD	3466	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	3598	G	A	5.93E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	3691	G	A	2.21E-01	Inf	3.4E-01	N.S.
NZB	Liver	OLD	3931	T	C	5.58E-01	Inf	7.9E-01	N.S.
NZB	Liver	OLD	4122	T	C	5.58E-01	Inf	7.9E-01	N.S.
NZB	Liver	OLD	4275	T	C	3.03E-01	Inf	4.6E-01	N.S.
NZB	Liver	OLD	4323	G	A	5.93E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	4407	T	C	5.58E-01	Inf	7.9E-01	N.S.
NZB	Liver	OLD	4705	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	4731	T	C	1E+00	1.15E+00	1E+00	N.S.
NZB	Liver	OLD	4770	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	4884	G	T	1E+00	0E+00	1E+00	N.S.
NZB	Liver	OLD	4902	G	T	5.19E-01	1.66E+00	7.5E-01	N.S.
NZB	Liver	OLD	5462	T	C	5.58E-01	Inf	7.9E-01	N.S.
NZB	Liver	OLD	5551	G	A	1E+00	1.13E+00	1E+00	N.S.
NZB	Liver	OLD	5929	T	C	5.58E-01	Inf	7.9E-01	N.S.
NZB	Liver	OLD	6040	G	A	3.5E-01	Inf	5.2E-01	N.S.
NZB	Liver	OLD	6406	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Liver	OLD	6469	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Liver	OLD	6574	T	C	1E+00	0E+00	1E+00	N.S.

NZB	Liver	OLD	6619	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	6784	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	7410	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	7869	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	8438	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	8466	G	A	2.21E-01	0E+00	3.4E-01	3.4E-01	N.S.
NZB	Liver	OLD	8567	T	C	5.36E-01	Inf	7.7E-01	7.7E-01	N.S.
NZB	Liver	OLD	8857	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	8863	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	9136	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	9151	G	A	4.88E-02	0E+00	8.7E-02	8.7E-02	N.S.
NZB	Liver	OLD	9390	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	9460	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	9529	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	9580	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	9598	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	9984	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	10546	T	C	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	10582	G	A	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	10951	T	G	1E+00	0E+00	1E+00	1E+00	N.S.
NZB	Liver	OLD	11842	T	C	5.27E-02	Inf	9.2E-02	9.2E-02	N.S.
NZB	Liver	OLD	11845	T	C	9.21E-02	Inf	1.5E-01	1.5E-01	N.S.
NZB	Liver	OLD	12352	T	C	5.71E-02	Inf	1E-01	1E-01	N.S.
NZB	Liver	OLD	12694	G	A	2.13E-01	Inf	3.3E-01	3.3E-01	N.S.
NZB	Liver	OLD	12834	G	A	1E+00	Inf	1E+00	1E+00	N.S.
NZB	Liver	OLD	12889	G	A	1E+00	Inf	1E+00	1E+00	N.S.
NZB	Liver	OLD	13003	T	C	5.27E-02	Inf	9.2E-02	9.2E-02	N.S.
NZB	Liver	OLD	13443	T	C	3.03E-01	Inf	4.6E-01	4.6E-01	N.S.
NZB	Liver	OLD	13611	G	A	3.5E-01	Inf	5.2E-01	5.2E-01	N.S.
NZB	Liver	OLD	13688	T	C	1.65E-01	Inf	2.6E-01	2.6E-01	N.S.
NZB	Liver	OLD	13780	G	A	2.13E-01	Inf	3.3E-01	3.3E-01	N.S.
NZB	Liver	OLD	13781	G	A	3.5E-01	Inf	5.2E-01	5.2E-01	N.S.

NZB	Liver	OLD	13836	G	A	5.93E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	13982	G	A	5.93E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	14185	G	A	5.93E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	14210	T	C	3.03E-01	Inf	4.6E-01	N.S.
NZB	Liver	OLD	14362	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	14641	T	C	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	14737	T	C	5.58E-01	Inf	7.9E-01	N.S.
NZB	Liver	OLD	15498	T	A	1.66E-01	Inf	2.6E-01	N.S.
NZB	Liver	OLD	15548	T	C	5.36E-01	Inf	7.7E-01	N.S.
NZB	Liver	OLD	15577	T	A	1.66E-01	Inf	2.6E-01	N.S.
NZB	Liver	OLD	15587	T	C	5.36E-01	Inf	7.7E-01	N.S.
NZB	Liver	OLD	15602	T	C	5.36E-01	Inf	7.7E-01	N.S.
NZB	Liver	OLD	15656	G	A	5.82E-01	Inf	8.1E-01	N.S.
NZB	Liver	OLD	15916	T	C	5.48E-03	Inf	1E-02	N.S.
NZB	Liver	OLD	16267	G	A	1E+00	Inf	1E+00	N.S.
NZB	Liver	OLD	16271	G	A	5.82E-01	Inf	8.1E-01	N.S.
AKR	Brain	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
AKR	Heart	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
AKR	Liver	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
ALR	Brain	YOUNG	9347	T	C	1E+00	0E+00	1E+00	N.S.
ALR	Brain	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
ALR	Heart	YOUNG	9347	T	C	1E+00	0E+00	1E+00	N.S.
ALR	Heart	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
ALR	Liver	YOUNG	9347	T	C	1E+00	0E+00	1E+00	N.S.
ALR	Liver	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
FVB	Brain	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
FVB	Heart	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
FVB	Liver	YOUNG	7777	T	G	1E+00	0E+00	1E+00	N.S.
FVB	Liver	YOUNG	9460	G	A	3.28E-01	0E+00	4.9E-01	N.S.
NZB	Brain	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	4731	T	C	6.11E-03	Inf	1.2E-02	N.S.
NZB	Heart	YOUNG	6406	T	C	1E+00	0E+00	1E+00	N.S.

NZB	Heart	YOUNG	6469	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	6574	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	6619	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	6784	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	7410	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	7869	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	8438	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	8466	G	A	1E+00	9.18E-01	1E+00	N.S.
NZB	Heart	YOUNG	8857	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	8863	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	9136	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	9151	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	9529	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	9580	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	9598	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	9984	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	10546	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	10582	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Heart	YOUNG	10951	T	G	1.07E-01	Inf	1.8E-01	N.S.
NZB	Liver	YOUNG	6406	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Liver	YOUNG	6469	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Liver	YOUNG	6574	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Liver	YOUNG	6619	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Liver	YOUNG	6784	T	C	1.4E-01	Inf	2.3E-01	N.S.
NZB	Liver	YOUNG	7410	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Liver	YOUNG	7869	T	C	5.1E-01	Inf	7.4E-01	N.S.
NZB	Liver	YOUNG	8857	G	A	1.34E-02	9.95E+00	2.5E-02	N.S.
NZB	Liver	YOUNG	8863	T	C	7.52E-02	Inf	1.3E-01	N.S.
NZB	Liver	YOUNG	9460	G	A	1E+00	0E+00	1E+00	N.S.
NZB	Liver	YOUNG	9984	T	C	1E+00	0E+00	1E+00	N.S.
NZB	Liver	YOUNG	10546	T	C	5.1E-01	Inf	7.4E-01	N.S.

Table A.8: **Adjusted empirical p-values for the change in frequency at haplotype sites compared to background sites.**

Strain	**Tissue**	**Position**	**Delta (Old - Young Freq)**	**Change**	**p-value**	**adjusted p-value (BH)**	**Sig**
ALR	Brain	4738	1.31E-03	Increase	3.41E-03	3.41E-03	SIG
ALR	Brain	9347	1.06E-03	Increase	3.41E-03	3.41E-03	SIG
ALR	Heart	4738	9.85E-04	Increase	6.32E-04	1.9E-03	SIG
ALR	Heart	9347	1.14E-03	Increase	6.32E-04	1.9E-03	SIG
ALR	Liver	4738	6E-04	Increase	1.2E-03	2.4E-03	SIG
ALR	Liver	9347	2.81E-04	Increase	2.81E-03	3.41E-03	SIG
FVB	Brain	7777	1.01E-03	Increase	4.77E-03	7.95E-03	SIG
FVB	Brain	9460	2.22E-05	Increase	8.54E-01	8.54E-01	N.S.
FVB	Heart	7777	5.54E-04	Increase	1.06E-03	2.65E-03	SIG
FVB	Heart	9460	1.94E-04	Increase	1.17E-02	1.46E-02	SIG
FVB	Liver	7777	9.35E-04	Increase	5.61E-04	2.65E-03	SIG
NZB	Brain	54	-3.73E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	1352	-5.41E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	1518	-5.46E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	1589	-5.61E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	1821	-4.23E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2200	-4.03E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2339	-4.37E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2524	-3.39E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2765	-3.65E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2766	-4.32E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2797	-2.91E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2813	-3.15E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2839	-2.62E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	2933	-3.27E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	3193	-4.24E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	3259	-3.75E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	3421	-4.34E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	3466	-4.1E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	3598	-2.87E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	3691	-3.36E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	3931	-3.56E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4122	-4.05E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4275	-4.45E-04	Decrease	6.69E-03	1.74E-02	SIG

NZB	Brain	4323	-5.63E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4407	-4.43E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4705	-5.55E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4731	-5.61E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4770	-4.94E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4884	-4.67E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	4902	-4.52E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	5462	-4.64E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	5551	-3.68E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	5929	-5.05E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	6040	-4.99E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	6406	-4.23E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	6469	-4.85E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	6574	-5.17E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	6619	-4.55E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	6784	-4.91E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	7410	-3.82E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	7545	4.98E-01	Increase	0E+00	0E+00	SIG
NZB	Brain	7869	-5.53E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	8438	-3.25E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	8466	-3.19E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	8567	-3.82E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	8857	-5.07E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	8863	-5.39E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	9136	-3.93E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	9151	-3.71E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	9390	-4.36E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	9529	-4.67E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	9580	-5.54E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	9598	-6.03E-04	Decrease	4.02E-03	1.74E-02	SIG
NZB	Brain	9984	-5.04E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	10546	-5.71E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	10582	-5.44E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	10951	-3.28E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	11842	-4.17E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	11845	-4.99E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	11932	-5.46E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	12352	-2.61E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	12574	-4.32E-04	Decrease	6.69E-03	1.74E-02	SIG

NZB	Brain	12694	-2.75E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	12834	-5.17E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	12889	-3.66E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13003	-4.14E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13443	-3.15E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13611	-3.96E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13688	-4.49E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13780	-4.46E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13781	-4.87E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13836	-2.76E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	13982	-2.71E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	14185	-5.13E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	14210	-5.24E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	14362	-4.08E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	14641	-4.69E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	14737	-2.6E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	15498	-4.12E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	15548	-5.72E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	15577	-3.88E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	15587	-4.84E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	15602	-4.42E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	15656	-2.47E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	15916	-6.23E-04	Decrease	4.02E-03	1.74E-02	SIG
NZB	Brain	16016	-4.95E-04	Decrease	6.69E-03	1.74E-02	SIG
NZB	Brain	16267	-7.47E-04	Decrease	4.02E-03	1.74E-02	SIG
NZB	Brain	16271	-7.21E-04	Decrease	4.02E-03	1.74E-02	SIG
NZB	Heart	54	-2.01E-05	Decrease	6.53E-01	6.95E-01	N.S.
NZB	Heart	1352	6.57E-06	Increase	9.99E-01	9.99E-01	N.S.
NZB	Heart	1518	-5.7E-05	Decrease	1.39E-01	2.04E-01	N.S.
NZB	Heart	1589	-2.26E-05	Decrease	5.45E-01	5.87E-01	N.S.
NZB	Heart	1821	-6.31E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	2200	1.11E-04	Increase	3.31E-02	8.48E-02	N.S.
NZB	Heart	2339	3.08E-05	Increase	2.08E-01	2.63E-01	N.S.
NZB	Heart	2524	-8.52E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	2765	-1.22E-04	Decrease	9.9E-02	1.84E-01	N.S.
NZB	Heart	2766	-1.16E-04	Decrease	9.9E-02	1.84E-01	N.S.
NZB	Heart	2797	-1.51E-04	Decrease	8.91E-02	1.84E-01	N.S.
NZB	Heart	2813	-1.55E-04	Decrease	8.91E-02	1.84E-01	N.S.
NZB	Heart	2839	-1.14E-04	Decrease	9.9E-02	1.84E-01	N.S.

NZB	Heart	2933	-2.33E-05	Decrease	5.25E-01	5.71E-01	N.S.
NZB	Heart	3193	4.63E-05	Increase	8.95E-02	1.84E-01	N.S.
NZB	Heart	3259	1.56E-05	Increase	8.56E-01	8.71E-01	N.S.
NZB	Heart	3421	-3.59E-05	Decrease	2.77E-01	3.21E-01	N.S.
NZB	Heart	3466	-2.14E-05	Decrease	5.74E-01	6.16E-01	N.S.
NZB	Heart	3598	-5.25E-05	Decrease	1.58E-01	2.18E-01	N.S.
NZB	Heart	3691	-3.21E-05	Decrease	3.47E-01	3.91E-01	N.S.
NZB	Heart	3931	-1.07E-04	Decrease	9.9E-02	1.84E-01	N.S.
NZB	Heart	4122	-7.59E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	4275	-1.3E-04	Decrease	8.91E-02	1.84E-01	N.S.
NZB	Heart	4323	-1.53E-04	Decrease	8.91E-02	1.84E-01	N.S.
NZB	Heart	4407	-8.38E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	4705	8.04E-05	Increase	4.6E-02	1.15E-01	N.S.
NZB	Heart	4731	1.07E-04	Increase	3.39E-02	8.6E-02	N.S.
NZB	Heart	4770	6.56E-05	Increase	5.65E-02	1.37E-01	N.S.
NZB	Heart	4884	-2.93E-05	Decrease	3.76E-01	4.21E-01	N.S.
NZB	Heart	4902	6.11E-05	Increase	6.21E-02	1.5E-01	N.S.
NZB	Heart	5462	3.27E-05	Increase	1.81E-01	2.37E-01	N.S.
NZB	Heart	5551	-1.67E-05	Decrease	7.62E-01	7.96E-01	N.S.
NZB	Heart	5929	-6.93E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	6040	-1.19E-04	Decrease	9.9E-02	1.84E-01	N.S.
NZB	Heart	7545	4.84E-01	Increase	0E+00	0E+00	SIG
NZB	Heart	8567	-5.05E-05	Decrease	1.58E-01	2.18E-01	N.S.
NZB	Heart	9390	-6.01E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	9460	3.61E-05	Increase	1.48E-01	2.16E-01	N.S.
NZB	Heart	11842	-1.32E-05	Decrease	7.92E-01	8.16E-01	N.S.
NZB	Heart	11845	-3.54E-05	Decrease	2.97E-01	3.4E-01	N.S.
NZB	Heart	11932	-2.28E-05	Decrease	5.45E-01	5.87E-01	N.S.
NZB	Heart	12352	-3.07E-05	Decrease	3.66E-01	4.12E-01	N.S.
NZB	Heart	12574	-4.54E-05	Decrease	1.88E-01	2.45E-01	N.S.
NZB	Heart	12694	-3.63E-05	Decrease	2.77E-01	3.21E-01	N.S.
NZB	Heart	12834	-1.03E-05	Decrease	7.92E-01	8.16E-01	N.S.
NZB	Heart	12889	-3.24E-06	Decrease	9.31E-01	9.39E-01	N.S.
NZB	Heart	13003	3.45E-05	Increase	1.65E-01	2.21E-01	N.S.
NZB	Heart	13443	-1.89E-05	Decrease	6.83E-01	7.23E-01	N.S.
NZB	Heart	13611	2.73E-05	Increase	2.52E-01	3E-01	N.S.
NZB	Heart	13688	5.85E-05	Increase	6.37E-02	1.52E-01	N.S.
NZB	Heart	13780	-7.52E-06	Decrease	8.22E-01	8.43E-01	N.S.
NZB	Heart	13781	-3.64E-05	Decrease	2.77E-01	3.21E-01	N.S.

NZB	Heart	13836	-1.54E-05	Decrease	7.72E-01	8.03E-01	N.S.
NZB	Heart	13982	-1.33E-04	Decrease	8.91E-02	1.84E-01	N.S.
NZB	Heart	14185	-7.96E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	14210	-1.06E-04	Decrease	9.9E-02	1.84E-01	N.S.
NZB	Heart	14362	-9.69E-05	Decrease	1.19E-01	1.96E-01	N.S.
NZB	Heart	14641	-1.7E-05	Decrease	7.62E-01	7.96E-01	N.S.
NZB	Heart	14737	-1.85E-05	Decrease	7.33E-01	7.72E-01	N.S.
NZB	Heart	15498	-4.85E-06	Decrease	8.61E-01	8.72E-01	N.S.
NZB	Heart	15548	-2.82E-05	Decrease	3.96E-01	4.39E-01	N.S.
NZB	Heart	15577	3E-05	Increase	2.2E-01	2.73E-01	N.S.
NZB	Heart	15587	-1.01E-04	Decrease	1.19E-01	1.96E-01	N.S.
NZB	Heart	15602	-2.68E-05	Decrease	4.26E-01	4.7E-01	N.S.
NZB	Heart	15656	8.83E-06	Increase	9.99E-01	9.99E-01	N.S.
NZB	Heart	15916	-7.42E-05	Decrease	1.29E-01	1.96E-01	N.S.
NZB	Heart	16016	2.16E-05	Increase	3.85E-01	4.28E-01	N.S.
NZB	Heart	16267	-1.15E-04	Decrease	9.9E-02	1.84E-01	N.S.
NZB	Heart	16271	-1.85E-04	Decrease	8.91E-02	1.84E-01	N.S.
NZB	Liver	54	-2.82E-04	Decrease	7.62E-02	1.73E-01	N.S.
NZB	Liver	1352	-6.28E-05	Decrease	2.29E-01	2.74E-01	N.S.
NZB	Liver	1518	-6.67E-05	Decrease	2.29E-01	2.74E-01	N.S.
NZB	Liver	1589	-9.09E-05	Decrease	1.71E-01	2.26E-01	N.S.
NZB	Liver	1821	-1.32E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	2200	-1.29E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	2339	-1.12E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	2524	-1.65E-04	Decrease	9.52E-02	1.84E-01	N.S.
NZB	Liver	2765	-1.17E-04	Decrease	1.33E-01	1.98E-01	N.S.
NZB	Liver	2766	-1.14E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	2797	-8.91E-05	Decrease	1.9E-01	2.46E-01	N.S.
NZB	Liver	2813	-1.79E-04	Decrease	8.57E-02	1.84E-01	N.S.
NZB	Liver	2839	-1.31E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	2933	-1.16E-04	Decrease	1.33E-01	1.98E-01	N.S.
NZB	Liver	3193	-1.66E-04	Decrease	9.52E-02	1.84E-01	N.S.
NZB	Liver	3259	-1.05E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	3421	-1.23E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	3466	-1.27E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	3598	-7.33E-05	Decrease	2.1E-01	2.63E-01	N.S.
NZB	Liver	3691	3.68E-05	Increase	2.14E-01	2.67E-01	N.S.
NZB	Liver	3931	-1.47E-04	Decrease	1.14E-01	1.96E-01	N.S.
NZB	Liver	4122	-1.28E-04	Decrease	1.24E-01	1.96E-01	N.S.

NZB	Liver	4275	-9.63E-05	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	4323	-1.68E-04	Decrease	9.52E-02	1.84E-01	N.S.
NZB	Liver	4407	-1.67E-04	Decrease	9.52E-02	1.84E-01	N.S.
NZB	Liver	4705	-2.12E-04	Decrease	7.62E-02	1.73E-01	N.S.
NZB	Liver	4731	-2.1E-04	Decrease	7.62E-02	1.73E-01	N.S.
NZB	Liver	4770	-1.86E-04	Decrease	8.57E-02	1.84E-01	N.S.
NZB	Liver	4884	-2.23E-04	Decrease	7.62E-02	1.73E-01	N.S.
NZB	Liver	4902	-1.22E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	5462	-1.6E-04	Decrease	1.05E-01	1.9E-01	N.S.
NZB	Liver	5551	-1.2E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	5929	-1.47E-04	Decrease	1.14E-01	1.96E-01	N.S.
NZB	Liver	6040	-1.12E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	7545	4.83E-01	Increase	0E+00	0E+00	SIG
NZB	Liver	8438	-7.35E-05	Decrease	2.1E-01	2.63E-01	N.S.
NZB	Liver	8466	-6.42E-05	Decrease	2.29E-01	2.74E-01	N.S.
NZB	Liver	8567	-5.21E-05	Decrease	2.57E-01	3.02E-01	N.S.
NZB	Liver	8857	-3.49E-05	Decrease	3.24E-01	3.67E-01	N.S.
NZB	Liver	8863	-1.95E-05	Decrease	4.95E-01	5.44E-01	N.S.
NZB	Liver	9136	-1.16E-04	Decrease	1.33E-01	1.98E-01	N.S.
NZB	Liver	9151	-4.4E-05	Decrease	2.95E-01	3.4E-01	N.S.
NZB	Liver	9390	-2.96E-04	Decrease	6.67E-02	1.57E-01	N.S.
NZB	Liver	9529	-9.86E-05	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	9580	-1.74E-04	Decrease	8.57E-02	1.84E-01	N.S.
NZB	Liver	9598	-1.6E-04	Decrease	1.05E-01	1.9E-01	N.S.
NZB	Liver	10582	-9.59E-05	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	10951	-3.57E-05	Decrease	3.24E-01	3.67E-01	N.S.
NZB	Liver	11842	-3.65E-06	Decrease	8.29E-01	8.47E-01	N.S.
NZB	Liver	11845	-1.88E-05	Decrease	5.24E-01	5.71E-01	N.S.
NZB	Liver	11932	-5.21E-05	Decrease	2.57E-01	3.02E-01	N.S.
NZB	Liver	12352	-6.25E-05	Decrease	2.29E-01	2.74E-01	N.S.
NZB	Liver	12574	-9.09E-05	Decrease	1.71E-01	2.26E-01	N.S.
NZB	Liver	12694	-7.4E-05	Decrease	2.1E-01	2.63E-01	N.S.
NZB	Liver	12834	-1.99E-04	Decrease	8.57E-02	1.84E-01	N.S.
NZB	Liver	12889	-1.37E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	13003	-8.44E-05	Decrease	2E-01	2.57E-01	N.S.
NZB	Liver	13443	-1.06E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	13611	-6.36E-05	Decrease	2.29E-01	2.74E-01	N.S.
NZB	Liver	13688	-1.04E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	13780	-5.47E-05	Decrease	2.57E-01	3.02E-01	N.S.

NZB	Liver	13781	-9.4E-05	Decrease	1.71E-01	2.26E-01	N.S.
NZB	Liver	13836	-1.16E-04	Decrease	1.33E-01	1.98E-01	N.S.
NZB	Liver	13982	-1.02E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	14185	-1.03E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	14210	-1.35E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	14362	-1.36E-04	Decrease	1.24E-01	1.96E-01	N.S.
NZB	Liver	14641	-1.57E-04	Decrease	1.14E-01	1.96E-01	N.S.
NZB	Liver	14737	-1.64E-04	Decrease	9.52E-02	1.84E-01	N.S.
NZB	Liver	15498	-1.6E-04	Decrease	1.05E-01	1.9E-01	N.S.
NZB	Liver	15548	-1.07E-04	Decrease	1.62E-01	2.18E-01	N.S.
NZB	Liver	15577	-1.42E-04	Decrease	1.14E-01	1.96E-01	N.S.
NZB	Liver	15587	-6.66E-05	Decrease	2.29E-01	2.74E-01	N.S.
NZB	Liver	15602	-8.96E-05	Decrease	1.9E-01	2.46E-01	N.S.
NZB	Liver	15656	-6.22E-05	Decrease	2.29E-01	2.74E-01	N.S.
NZB	Liver	15916	-1.54E-05	Decrease	6.19E-01	6.61E-01	N.S.
NZB	Liver	16016	4.06E-04	Increase	5.49E-03	1.74E-02	SIG
NZB	Liver	16267	-4.59E-04	Decrease	4.76E-02	1.17E-01	N.S.
NZB	Liver	16271	-4.74E-04	Decrease	4.76E-02	1.17E-01	N.S.

Empirical p-values were calculated as the count of sites with a delta greater (for sites with an increase in frequency with age) or less (for sites with a decrease in frequency with age) than the haplotype site delta divided by the total number of deltas (one-sided test). Haplotype sites with a p-value < 0.02 were labeled significant. P-values were adjusted using the Benjamini-Hochberg correction for multiple hypothesis correction within each mt-haplotype group.

Table A.9: **Comparison of duplex depth and mutation counts between Serrano et al. and Sanchez-Contreras et al.**

	Serrano et al.	**Sanchez-Contreras et al. (mice not treated with chemical intervention – data used for Fig 1 - 4)**	**Sanchez-Contreras et al. (all mice, including those with chemical interventions)**
Total duplex bp sequenced	40.4 billion bp	14.7 billion bp	27.9 billion bp
Average duplex depth per sample	21, 551X	10,314X	10,125X
Total de novo mutation count	81,097 mutations	40, 478 mutations	89, 330 mutations
Total de novo SNV count	71, 416 mutations	35, 026 mutations	77, 194 mutations

Identifying de novo somatic mutations: filtered for variants that had an alternate allele depth of < 100 and mutation frequency $< 1\%$ to exclude variants segregating at a high heteroplasmic frequency. Each mutation type occurring at a given position is counted once.

Appendix B

Appendix of Chapter 3

B.1 Supplementary Notes

Supplementary Note 1: Input parameters for the duplex sequencing pipeline configuration file

```
import pandas as pd
import numpy as np
import re


names_list = #the list of our forward and reverse fq for each sample
# Extract the sample name and replicate number from the list of file names
names = []
for elem in names_list:
    names.extend(re.findall(r"mice_\d{1,2}_\w{3,4}-\d{1}", elem))

# Remove empty elements
names = [name for name in names if name]

# Extract unique names
final_names = []
for name in names:
    if name not in final_names:
        final_names.append(name)


data_list = []
for sample in final_names:
    data = f"{sample},{sample},{sample},{sample},
```

```
    {sample},/global/scratch/users/hannahmarieaguilar
    /tissue_specific_mt_profiling/rawdata/scripts/
    mm10.fa,/global/scratch/users/hannahmarieaguilar
    /tissue_specific_mt_profiling/rawdata/scripts/
    output/mm10.bed,NONE,NONE,10090,output,
    {sample}_R1.fastq.gz,{sample}_R2.fastq.gz,0,3,200,0.7,1,18,0,10,
    151,ANNNNNNNNAGATCGGAAGAG,10,0,0,0.1,100,1,
    noRecovery.sh,10,N,GT,none,TRUE,0"
    data_split = data.split(",")
    data_list.append(data_split)


header = ["sample", "rglb", "rgpl", "rgpu", "rgsm", "reference", "target_bed",
    "maskBed", "blast_db", "targetTaxonId",
          "baseDir", "in1", "in2", "mqFilt", "minMem", "maxMem", "cutOff",
             "nCutOff", "umiLen", "spacerLen",
          "locLen", "readLen", "adapterSeq", "clipBegin", "clipEnd", "minClonal",
             "maxClonal", "minDepth", "maxNs",
          "recovery", "cluster_dist", "cm_outputs", "cm_sumTypes", "cm_filters",
             "runSSCS", "rerun_type"]

df = pd.DataFrame(data_list, columns=header)

# Save DataFrame to a CSV file
df.to_csv("final_config.csv", index=False)
```

B.2 Supplementary Figures

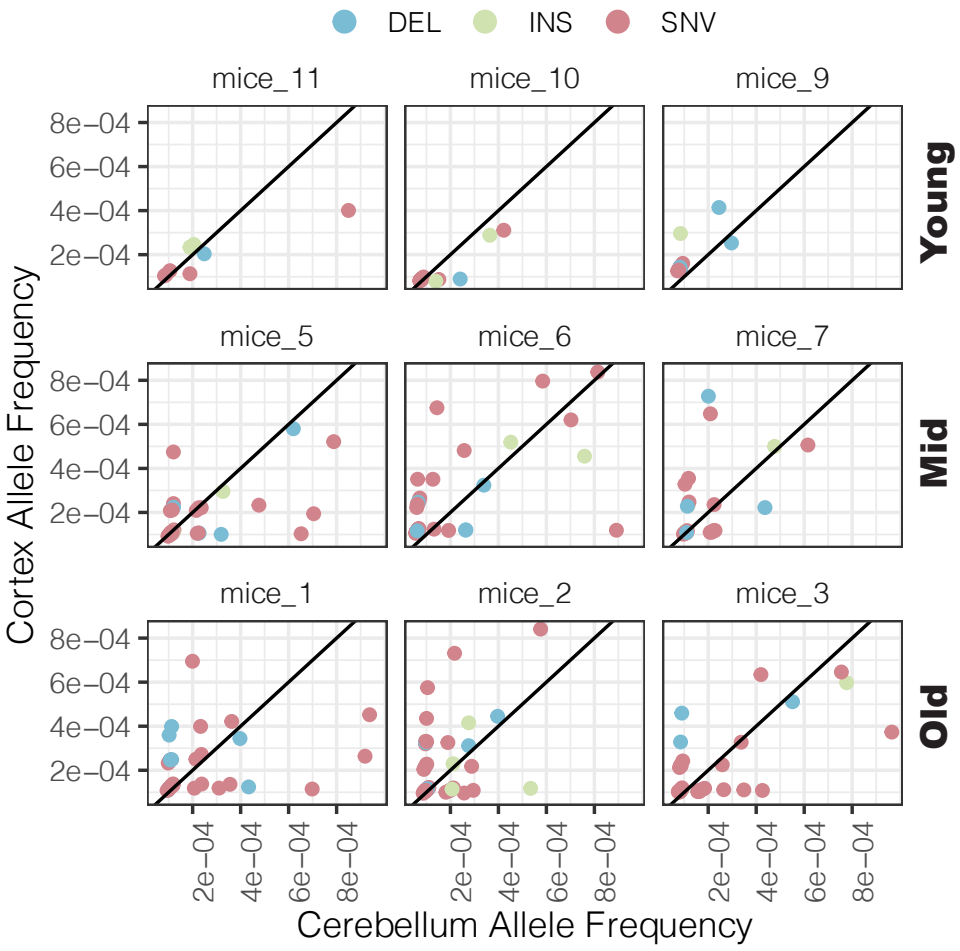


Figure B.1: **Comparison of frequencies ($< 1 \times 10^{-3}$) for shared mutations between the cortex and cerebellum.** Each point is an allele present in both the cortex and cerebellum. Color highlights the mutation class (deletions (blue), insertions (green), SNVs (red)). The mutation frequency is calculated as the alternative allele count divided by the read depth at each position. The solid, black line denotes the $x = y$ line. Samples are ordered by age.

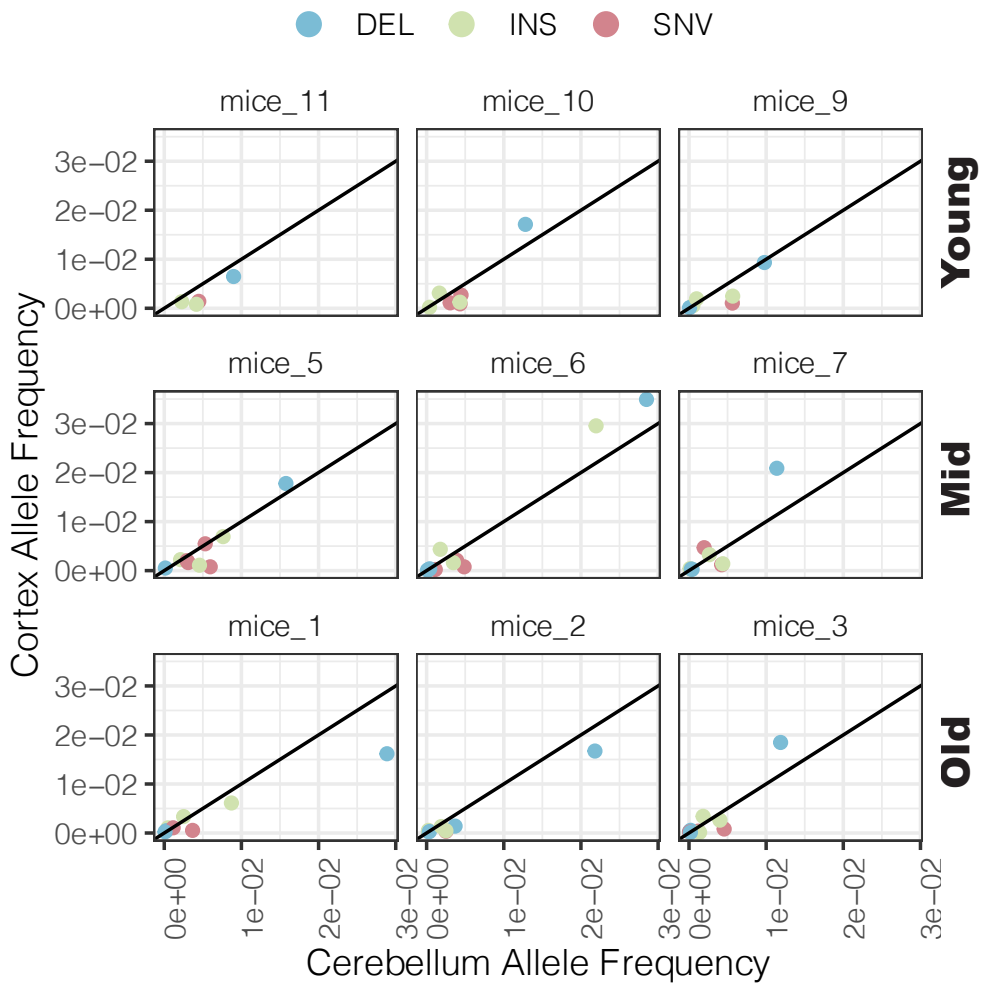


Figure B.2: **Comparison of frequencies ($> 1 \times 10^{-3}$) for shared mutations between the cortex and cerebellum.** Each point is an allele present in both the cortex and cerebellum. Color highlights the mutation class (deletions (blue), insertions (green), SNVs (red)). The mutation frequency is calculated as the alternative allele count divided by the read depth at each position. The solid, black line denotes the $x = y$ line. Samples are ordered by age.