# The companion dog: a powerful model for understanding somatic mutation, disease and aging

Ashley Paynter

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Daniel Promislow, Chair

Rosana Risques

Scott Kennedy

Program Authorized to Offer Degree:

Biology

University of Washington

## Abstract

The Domesticated Dog: An Unexpected and Powerful Tool for Aging Research

Ashley Paynter

Chair of the Supervisory Committee:
Daniel Promislow
Department of Laboratory Medicine & Pathology; Department of Biology

This dissertation discusses the significance and importance of utilizing the domesticated dog as both a resource for large data and genomic analysis. I first reviewed the uses of canine data, outlining the advantages and disadvantages of various datasets for research. I next utilized Duplex Sequencing technology to identify rare mutations in healthy dogs. I also investigated the age-related presence of mutational signatures based on the Catalog of Somatic Mutations in Cancer (COSMIC). This dissertation will help the reader understand the potential of the companion dog model in general, as well as the usefulness of new technologies in DNA sequencing to understand aging.

# TABLE OF CONTENTS

# LIST OF FIGURES



# LIST OF SUPPLEMENTARY FIGURES

# LIST OF TABLES

# ACKNOWLEDGEMENTS

I would like to acknowledge all of the hard-working members of University of Washington who made getting this degree possible including: Daniel Promislow, Scott Kennedy, Rosana Risques, Maria Partida-Aguilar**,** Jeanne Fredrickson, Brendan Korhn, Megan Nyugen, Alice Berger, Kelly Harris, Barbara Wakimoto, Sandi Shrager and Ben Kerr. I would also like to thank the collaborators who made this process go smoothly: Audrey Ruple, and Noah Snyder-Mackler. Without your support and commitment, I would not have been able to complete this dissertation. I would also like to thank the National Science Foundation, the National Institutes of Health, and the American Kennel Club for the funding that made pursuing this degree and publishing these data possible. I would also like to that my mother, Leslie Paynter.

# DEDICATION

I would like to dedicate this dissertation to my late father, Brian Paynter, whose spirit reminded me to keep fighting for this degree and my dog, Honey, whose loving nature and emotional support got me through the hardest days in the past 6 years.

# Chapter 1. INTRODUCTION

## 1.1 THE DOMESTICATED DOG AS A MODEL

The domesticated dog (*Canis lupus familiaris*) is generally known as a much beloved companion to human beings. There is a rich literature on the domestication of the dog, and many potential hypotheses for why humans and wolves developed such a bond, and how wolves eventually became the pets that we adore so much. Specifically, Galibert defined domestication as a long process where humans have crossed individuals with desirable behavioral and physiological traits to create new offspring that possess those traits, isolated from wild populations (Galibert et al., 2011). In dogs, these factors along with reproductive isolation, artificial selection, founder effects, bottlenecks and inbreeding have allowed for traits to become fixed within breeds (Frantz et al., 2020). Dogs are known as the first domesticated animal species, and early evidence of their existence dates back as far as 33,000 years ago (Freedman et al., 2014; Perri et al., 2021). It is unknown what exactly caused the desire to domesticate the dog. A leading hypothesis is that humans may have bonded with wolves due to the protection and safety they provided. Others hypothesize that dogs may have been scavengers, following the scraps of nomadic humans (Galibert et al., 2011). Li hypothesized that during domestication, dogs developed enhanced excitatory synaptic plasticity, allowing for a closer relationship with humans (Li et al., 2014). Regardless of the selective forces that ultimately led to dog domestication, the genetics of the dog remain fascinating to scientists.

Dogs are the most phenotypically diverse animal species on the planet (Wayne & vonHoldt, 2012). This has been driven in part by the human desire to create this diversity. Specifically, during the Medieval and Renaissance periods and the last 200 years, there

was a significant effort to cross dogs of different types, eventually leading to the creation of distinctive breeds (Galibert et al., 2011). Currently, over 400 breeds exist worldwide (Wilcox & Walkowicz, 1995). The strong selective pressure of domestication created breeds that varied greatly in size, shape, color, and behavior. Additionally, breeds vary in functionality; many breeds are known to perform certain behavioral tasks such as hunting, herding, and guarding, and to provide emotional and physical support.

Recent advances in genomic sequencing have allowed us to further understand the evolutionary history and genetic variation of the dog. Previously, molecular work on dogs consisted primarily of mitochondrial DNA analysis (Villa et al., 1999, Leonard et al., 2002). Further efforts to understand dog diversification and domestication used genome wide SNP and haplotype data analysis (Vonholdt et al., 2010). Additionally, a foundational study in understanding the genetics of breeds genotyped microsatellites on purebred dogs across 85 breeds. It was found that breeds had distinct microsatellite genotypes, and that within breeds, there was considerable genetic homogeneity (Parker et al., 2004). Genetic diversity between dog breeds is significantly greater than genetic diversity within dog breeds (Lindblad-Toh et al., 2005; Parker, 2012). Several studies have shown that morphological variation between breeds can be further understood through genome wide association studies (GWAS) (Karlsson et al., 2007 ; Sutter et al., 2007; Cadieu et al., 2009). Additionally, recent studies have shown that dog breeds vary greatly in disease risk and longevity (Michell, 1999; Proschowsky et al., 2003; Fleming et al., 2011; Vaysse et al., 2011; Jimenez, 2016; Wiles et al., 2017). For example, some dog breeds live twice as long as other dog breeds (Lewis et al., 2018; Yordy et al., 2020). Given the enormous variation in disease risk (Fleming et al., 2011), longevity (

Lewis et al., 2018) and rate of aging (Kraus et al., 2013) among breeds, dogs provide an ideal model to better understand the underlying genetics of pathogenesis.

The domesticated dog is a valuable model for research not only due to the fascinating variation in phenotype and genotype observed, but also because it addresses some of the limitations of alternate models. Most aging research is currently is done with organisms housed in the lab, including yeast, nematode worms, fruit flies and rodents (Fontana et al., 2010). Unlike these species, dogs are longer lived, and could provide a powerful tool to understand the biology of aging (Ruple et al., 2022). Similar to humans, dog mortality increases exponentially with age (Hoffman et al., 2018). While dogs live longer than traditional laboratory model organisms, they live shorter than humans, making them a great resource to see patterns of aging that might take decades to observe in humans (Kraus et al., 2013). Additionally, environmental factors such as pollution have been proven to be associated with accelerated aging (Schmidt, 2019), but it is nearly impossible to replicate the same environment to which humans are exposed in a laboratory environment. Therefore, having a model organism exposed to the same environment as humans offers a more meaningful comparison to the human experience. Unlike many model organisms, dogs suffer from many of the same naturally occurring age-related diseases that humans experience with age (Hoffman et al., 2018). This includes cancer, renal disease, endocrine disease, obesity, cognitive decline and cancer (Hoffman et al., 2018). As mentioned earlier, the unique genetic architecture of breeds allows scientists to study the underlying genetic drivers of disease and aging. Lastly, dogs have a healthcare system that is second only to humans in its sophistication (Ruple et al., 2022). This allows researchers to obtain extensive longitudinal health and longevity

data in dogs (Ruple et al., 2022). The importance of these datasets and their potential for groundbreaking research is outlined in chapter 1.

## 1.2 AGING AND CANCER IN DOGS

Arguably, one of the most relevant uses for the dog model is in understanding aging and cancer. Risk of disease diagnosis and mortality increases exponentially with age in dogs (Kraus et al., 2013). Some dogs age quite differently, with breeds like the Great Dane living just 6 years, to the Chihuahua who can live to 15 years of age (Fleming et al., 2011). One current open-data long-term study of dogs is the Dog Aging Project, which will follow tens of thousands of individuals (Creevy et al., 2022).

As in humans, age in dogs is the most significant risk factor for cancer, and cancer is the greatest cause of mortality in dogs (Schiffman & Breen, 2015). Each year, about 80,000 cases of lymphoma will be diagnosed in humans. Roughly the same number of cases will be diagnosed in dogs, with pathology strikingly similar to human Non-Hodgkin's Lymphoma (NHL) (Zandvliet, 2016). Canine lymphoma is nearly histologically indistinguishable from human lymphoma, and the cancer in both species responds similarly to the same types of treatment (Teske 1994). Additionally, human NHL and canine lymphoma have shared molecular aberrations and genetic commonalities (Daisuke et al., 2014). Breeds vary tremendously in the risk of lymphoma and other hematological cancers. The among-breed variation in lymphoma risk is exemplified by studies showing that Boxers have a 40-fold increase in cancer risk relative to Pomeranians (Priester & McKay, 1980). While some breeds are almost never diagnosed with lymphoma, in others, over 30% of all individuals will die of the disease (American Kennel Club). Diagnosis is costly for pet owners, and by the time lymphoma is diagnosed

in a dog, the median survival time is just 10-14 months (Zandvliet, 2016). If we could detect individuals at high risk of lymphoma before they get the disease, and diagnose individuals with the disease earlier, we could have a large positive impact on disease outcome, saving lives and health care costs.

## 1.3 SEQUENCING TECHNOLOGIES TO IDENTIFY SOMATIC MUTATION

Cancer is a genetic disease that is caused largely by nonsynonymous somatic DNA mutations in tumor suppressor genes or oncogenes (Loeb, 2016). Interestingly, mutation rate is generally higher in the somatic tissue than the germline, and accumulation of mutations is associated with cancer (Loeb, 2016). Individuals with hereditary cancers have a germline, inherited genetic variant that predisposes them to higher frequency of somatic mutation. Somatic mutations are also believed to play a role in aging. Due to current technologies, genomic variants in dog DNA can be identified, and individual tumors can be assessed for the presence of somatic mutations (Schiffman & Breen, 2015). While there have been efforts to detect disease associated mutations in dogs, little has been done to enhance early detection of these mutations. Previous studies rely on next generation sequencing of tumor tissues (Modiano et al., 2005; Karlsson et al., 2013 ; Karyadi et al., 2013; Elvers et al., 2015; Tonomura et al., 2015; Tarone et al., 2019). The goal of these prior studies is to identify high frequency cancer-mutations in dog tumors However, previous studies have shown that mutations can exist in healthy humans and increase with age (Jaiswal et al., 2014, Kakiuchi & Ogawa, 2021). In contrast to standard NGS, the technology that we use here, Duplex Sequencing, can potentially detect the somatic mutations that may be associated with disease in healthy dogs.

## 1.4 SUMMARY

This dissertation initially sought to understand the differences in disease risk of lymphoma by looking at precancerous mutations in healthy dogs. Using Duplex Sequencing technology, we discovered a wealth of evidence that somatic mutations are increasing with age in several target genes. While we did not find evidence that the mutation frequencies observed were breed specific or risk specific, we saw a strong signal of aging, and decided to focus the bulk of this dissertation on patterns of aging seen through somatic mutation in dog young to old. We hope that the research here will lay the foundation for further work to be done on the impact of somatic mutations on cancer risk.

# Chapter 2. VETERINARY BIG DATA

This chapter text comes from the manuscript: Paynter, A. N., et al. (2021). "Veterinary Big Data: When Data Goes to the Dogs." Animals 2021, Vol. 11, Page 1872 **11**(7): 1872.

## 2.1 INTRODUCTION

The era of big data has opened up many new opportunities in preventive care, chronic disease management, and treatment optimization. It has also allowed for a new model of medicine to be employed, that of personalized or precision medicine. In this model of human healthcare, decisions about medical interventions and other treatments are tailored to an individual patient based upon their predicted risk of developing disease or their predicted response to therapy (Grossglauser & Saner, 2014). This has been enabled primarily through the adoption of electronic health record systems by healthcare providers which allow for the construction of detailed longitudinal data about large populations of patients over long periods of time. These data frameworks can be interrogated to discover the combinations of risk factors that lead to disease outcomes and deliver personalized disease risk profiles for individual patients (Chawla & Davis, 2013).

Data-related problems in human healthcare such as data integration, wrangling, ease of use, and interpretability are similar to those encountered in other domains. However, there are a number of domain-specific challenges encountered in human healthcare disciplines that are unique to these data frameworks. One example of this is the use of unstructured data, such as patient notes and interpretations of diagnostic tests, which contain rich information that can provide valuable insights at both the individual patient and population levels, but the heterogeneity, variability, and diversity of these data make it difficult to access if analyzed in a controlled manner (Adnan et al., 2020). Another challenge lies in the issues of privacy and security in human healthcare which have drawn significant attention in recent years, but are especially important in

healthcare settings because of concerns related to the introduction of the HIPAA Privacy Act which declared medical information, including electronic medical records, to be protected health information covered under the Privacy Rule (Sirilla et al., 2006; Sun et al., 2020). In fact, the sheer magnitude of the size of medical big data has increased to such a level that new storage technology systems have been developed to capture, manage, and process big data; but issues with privacy control, technical vulnerabilities, security for authorization and verification, data management, and confidentiality still exist (Gupta & Rohil, 2020).

Many of these challenges surrounding medical data and use are avoided in the field of veterinary medical big data, as veterinary patient records are not considered protected health information and are thus not included in the Privacy Rule. However, naturally occurring diseases in companion animals are often similar, if not identical, to human diseases in terms of disease etiology, progression, and treatment response (Kol et al., 2015). Dogs, in particular, provide an ideal model for translational medicine as they have the most phenotypic diversity and known naturally occurring diseases of all land mammals other than humans (Starkey et al., 2005). Among those diseases, we have identified more than 400 inherited disorders in dogs that are relevant to humans which is likely due to the fact that we share ~650 Mb of ancestral genetic sequence with our dog companions (Rowell et al., 2011). In addition, dogs share both our physical and chemical environments and live in more than 63 million households in the United States (American Pet Products Association, 2020). Disease outcomes as varied as cancer (e.g. lymphoma, osteosarcoma), osteoarthritis, spinal cord disorders (e.g. thoracolumbar intervertebral disk herniation, spina bifida), eye disorders (e.g.

keratoconjunctivitis sicca), cardiomyopathies (e.g. dilated and hypertrophic cardiomyopathies), and infections – including those caused by antimicrobial resistant organisms – are all shared by our dog companions (Kol et al., 2015; McEwen & Collignon, 2018). Furthermore, the healthcare system for dogs in the United States is sophisticated and over $40 billion is spent annually on dog health care, second only to the level of healthcare received by humans (American Veterinary Medical Association, 2018; Patterson, 2000).

Thus, veterinary medical big data may very well represent an underutilized resource in medical research. Using data derived from dogs whose owners are seeking medical care for naturally occurring diseases may also have lesser ethical concerns than research involving animals in which the disease of interest is induced (Kol et al., 2015). However, despite the reduced concerns related to privacy and security, accessing veterinary medical big data is not without challenges. In this review we will describe the veterinary medical data sources available for research as well as the opportunities and challenges related to each.

## 2.2 Data Sources for Companion Animal Health Information

There are several sources of veterinary medical big data related to companion animal health information that have been utilized for research purposes. The majority of these data sources are comprised of medical records from either primary medical care centers or referral hospitals. Another data source for animal health information comes from animal insurance companies, which typically track medical claim data on privately owned animals. In the United States, approximately 2.5 million pets were insured in 2019, which was a growth of 16.7% over the total number of pets insured in the prior

year, but still only represented about 2.5% of the total pet population in the US (North American Pet Health Insurance Association). Insurance coverage rates for companion animals in the United Kingdom are approximately 25% of the total population and it is estimated that 90% of the dog population in Sweden are covered by an insurance policy (Veterinary Information Network). Other datasets have been constructed specifically for research purposes and therefore often only include animals that meet explicit inclusion criteria, but might still be useful for translational research.

### 2.2.1 *Medical Record Datasets*

One of the largest veterinary practices in the world is Banfield Pet Hospital, a privately-owned company that consists of more than 1,000 individual clinics. As of 2019, Banfield Pet Hospital had collected clinical data from more than 2.5 million dogs from 43 states in the US (2019). All Banfield hospitals use a single proprietary medical record software system to upload and centrally store electronic information related to the companion animals seen in individual hospital locations. These records include information such as laboratory test results, physical examination findings, diagnoses, treatments, procedures performed, and demographic information about the dog patients and their owners (Plant et al., 2011). In addition, these data include geocoded information about the home locations of the animals seen at Banfield hospitals. This has allowed researchers to pair electronic record data with existing ecological data to monitor infectious disease spread in animals in the U.S. (Glickman et al., 2006).

One drawback of utilizing Banfield medical data for translational medicine analyses is that these data do not consistently contain behavioral information such as type and quantity of outdoor activities undertaken, exercise history, diet composition,

and temperament of the pet (Raghavan et al., 2007). Diagnoses contained in the dataset may include both those suspected by the veterinarian and those with diagnostic test confirmation (Plant et al., 2011). However, the large number of pets with diagnoses and veterinarians employed at these clinics minimize the risk of systematic misclassification errors occurring. Perhaps most importantly, these data are owned by Banfield Pet Hospital and access to these data appear to be limited.

Another source of veterinary health information is the Veterinary Medical Database (VMDB), the oldest companion animal health database in the United States which provides veterinary medical datasets to researchers at little or no cost. This database was started in 1964 by the National Cancer Institute and includes patient data contributed from 26 university teaching hospitals in the United States and Canada, and contains over 7 million records from all species (The Veterinary Medical Databases). The VMDB diagnoses data was originally coded using the Standard Nomenclature of Veterinary Diseases and Operations (SNVDO), but has been coded in Systematized Nomenclature of Medicine- Clinical Terminology (SNOMED CT) since 1996 (The Veterinary Medical Databases). It is important to note that this hierarchical coding structure was originally created for diagnoses in human patients rather than veterinary patients, but use of these coding systems allows for searches to be conducted in either broad disease categories or using only specific diagnoses (Bartlett, 2010).

A limitation of this dataset is the narrow availability of information related to each medical record as only a coded abstract from each patient hospital visit is included in the dataset. Data contained is limited to: the institution where the animal was seen, the species and breed of the animal, signalment (age, weight, sex), hospital discharge date,

patient number (as assigned by the hospital), if the hospital visit was the first or a recheck appointment for a previously diagnosed condition, diagnoses, and the postal code of the client (The Veterinary Medical Databases). Additionally, disease prevalence estimates derived from this dataset are likely influenced by referral bias as veterinary teaching hospitals provide highly specialized and advanced animal care and so the underlying population from which the VMDB data is derived is biased towards sick animals and more serious and rare diseases are overrepresented in this database (Bartlett, 2010).

The Veterinary Companion Animal Surveillance System (VetCompass), which began collecting clinical data from primary practices in the United Kingdom in 2009, now holds data on more than 15 million animals collected from over 1,800 veterinary practices across the U.K., nearly a third of all U.K. veterinary practices (Radford et al., 2010). VetCompass began collecting clinical data in Australia in 2016 and pilot projects are being completed in Spain, Germany, and New Zealand. These data are actively collected and considered to be more representative of the general pet population than are data collected only from referral hospitals and geographic and temporal information are included which allows for analyses of disease trends over both space and time (Small Animal Veterinary Surveillance Network, The Veterinary Companion Animal Surveillance System Overview).

Clinical data from each participating practice are uploaded automatically to the VetCompass database and requires no additional work on the part of the hospital staff or veterinarians (The Veterinary Companion Animal Surveillance System Overview). Extraction of these data has not yet been fully automated though, which can result in

data access limitations. Another limitation is that these data do not include behavioral, environmental, or dietary information about these patients, all of which can be important factors in the etiology of disease. Open access VetCompass data are available through request, though typically only for use in academic settings (Radford et al., 2010).

### 2.2.2 *Insurance Datasets*

Data collected by veterinary medical insurance providers can be used as a secondary data source for translational analyses. One of the biggest advantages of insurance datasets is that they are large and high statistical power can be achieved for many types of analyses. These types of datasets include varying amounts of information about the total population of insured pets included in the dataset, but uniformly include detailed address information about individual owners. This allows for more robust interpretations about the impact of geographic factors on health outcomes to be made than might be possible using only medical records (McGreevy et al., 2017). However, the coding system used by individual insurance companies may allow for only one diagnostic code to be entered for each veterinary visit which can result in a reduction of data completeness (McGreevy et al., 2017). In addition, these datasets may not distinguish between death from natural causes and death from euthanasia, which should be considered when constructing time to event analyses (O'Neill et al., 2014).

Two well-established insurance databases, one from the U.K. and one from Sweden, have been utilized for epidemiological research on the causes of morbidity and mortality in insured dog populations (Egenvall et al., 2009; O'Neill et al., 2014). The dogs in the U.K. insurance dataset were thought to differ from the total population of

dogs in the U.K. in that the insurance population had overrepresentation of younger and purebred dogs (Egenvall et al., 2009). There was also a lack of histologic confirmation of some disease diagnoses noted by the investigators (Egenvall et al., 2009).

The Swedish insurance dataset has been validated and shown to have adequate agreement with medical records obtained from veterinary practices that provided medical care to the animals included in the dataset (Bonnett et al., 2005). The large proportion of dogs covered by insurance policies in Sweden is a benefit in that the insurance datasets are likely more representative of the total dog population in the country than the datasets available in the U.S. and the U.K. thereby increasing the external validity of research conducted using this population of dogs. The Swedish dataset held by the insurance company Agria has been utilized by several research teams over many decades, too, so there is a long record of open access to these data which may now provide insights into changes in disease trends over time that would not be identifiable in other datasets.

### 2.2.3 *Research Datasets*

Other datasets have been constructed to investigate health and disease trends in particular populations of dogs. One such dataset is part of the Golden Retriever Lifetime Study (GRLS), a prospective study restricted to a single cohort of approximately 3,000 Golden retriever dogs located throughout the U.S. Despite the limitations created by the small number of individual animals included in this dataset, there are some significant benefits to using these data for translational studies. For instance, these data were collected longitudinally and at regular intervals and the dataset contains information on not only the health of animals collected directly from their veterinarian and including

laboratory values, but also extensive environmental and behavioral health data collected from their owners (Dobson et al., 2002). Despite the burden placed on owners to track and report data about their dogs, this work has had a high level of study compliance and has resulted in a robust dataset that includes information about individual dogs from the time of enrollment at less than 2 years of age until their death (Egenvall et al., 1998).

An added advantage to GRLS is that biological samples have been collected from these dogs on at least an annual basis (Dobson et al., 2002). These samples include whole blood, serum, urine, hair and toenail clippings, feces, and tissue samples all of which were stored in a biobank at the time of collection and are available for interrogation by investigators (Dobson et al., 2002). These samples offer a unique opportunity to investigate environmental exposures over the lifetime of the dogs and their relationships to multiple disease outcomes, which could lead to insights that will benefit both animal and human health.

Another research platform, Dogslife, is collecting data directly from owners about the health and lifestyle of their U.K. Kennel Club registered Labrador retrievers (Guy et al., 2015). The study population is restricted to purebred Labrador retriever dogs born on or after January 1, 2010 and all data collection is performed using a website interface (Guy et al., 2015). Owners of enrolled dogs are requested to complete a questionnaire about their dog's health and welfare monthly for the first year of the dog's life and at 3 month intervals thereafter and the compliance with questionnaire requirements is reasonably high, with nearly 80% of the data entries being complete at the time of the first descriptive evaluation of the cohort (Guy et al., 2015). By December

31, 2013 there were more than 4,000 dogs included in this dataset and enrollment is ongoing (Ruple et al., 2021).

## 2.3 The Dog Aging Project

A long-term longitudinal study of companion dogs, the Dog Aging Project (DAP), was initiated in 2019 with the express purpose of collecting data that could be used in translational medical research. Dogs of any breed, size, and age are eligible for participation in DAP, though enrollment is currently limited to the 50 U.S. states. This project is an Open Science study and investigators unaffiliated with the project are able to access these data as well as propose ancillary studies that can build upon the foundation created by the DAP team. It is anticipated that the total number of dogs enrolled in this project will exceed 100,000 as enrollment is ongoing and there is no limit to the total number of participants that will be included in the project. Thus, this dataset represents the largest primary data source of health information in dogs collected explicitly for use in translational medical research.

### 2.3.1 *Health and Life Experience Survey*

Comprehensive information about the dog's health and environment is collected at the time of enrollment in the study through the Health and Life Experience Survey (HLES). This survey is administered through an online portal and is separated into ten sections. Data collected through this instrument includes detailed information about the dog's breed, age, sex, neuter status, behavior, diet, use of medications, types and amount of physical activity, the indoor and outdoor environments the dog is regularly exposed to, and the dog's comprehensive health history. The initial HLES data is

considered the baseline dataset for each participant in the project. HLES data is updated annually in order to identify changes related to either the dog or their environment as well as collect information about any new medical diagnoses made within the previous year.

### 2.3.2 *Other Types of Data Collected*

In addition to the HLES data, all participants are asked to upload copies of the dog participant's electronic medical record. These records can be used to validate information provided by human participants about their dog's overall health, specific diagnoses reported, and demographic information. In addition, laboratory values and other items of research interest may be extracted from the medical records and used to help inform specific research objectives.

Approximately 10,000 dogs included in DAP will have whole genome DNA sequencing (WGS) performed. This will add tremendous value to this dataset as genome wide association studies (GWAS) performed in dog populations have been shown to be able to identify functional variants associated with specific morphologic traits as well as disease risk and dog behavior (Clements et al., 2013). WGS has also been utilized to identify rare hereditary diseases in dogs as well as isolate variants located in orthologs of human cancer susceptibility genes (Clements et al., 2013; Plassais et al., 2019; Pugh et al., 2015; Woolley et al., 2020).

Approximately 1,500 of the dogs that have whole genome sequencing data collected will also have information related their metabolome, methylome, microbiome, and rest/activity cycles collected and included in the dataset. Five hundred of these dogs will also have physical examinations and diagnostic testing performed by

veterinary cardiologists and results from echocardiogram, electrocardiogram, and blood pressure measurement tests will be included in the DAP dataset.

These additional data inputs, sequencing, and physiologic measurements will enrich the data available for translational research using these subsets of DAP dogs. Ancillary studies conducted using other DAP dogs may further supplement the data available in the dataset in the future. These ancillary studies will likely be focused on specific research questions and the additional data available will be dependent upon the needs of the investigators.

## 2.4 The Future of Companion Animal health Data Collection

The increasing frequency with which companion dogs are being used to study disease outcomes highlights the importance of medical data collected within this population. Each of the datasets described herein provide valuable resources for research. However, use of large-scale epidemiological studies in which data is collected longitudinally is ideal for this work as these data allow for temporal analyses to be conducted and for incidence and prevalence of diseases to be estimated.

Secondary datasets, like those belonging to insurance companies, are useful for this work but primary datasets do offer some advantages. Because these datasets are constructed with the explicit intent to utilize data for research purposes there are often both subjective and objective measurements included in the same dataset. This helps to give a more complete understanding of the dog's health than does data related only to illness (e.g., insurance claims data). Additionally, the increased amount of data collected related to the dog's environment and potential environmental risk exposures is increased in primary datasets.

Collection of health data across multiple breeds of dogs is ideal in that the increased genetic heterogeneity in a more outbred population of dogs will allow for identification of genes related to a greater number of disease outcomes. In datasets where genomic sequencing is not included, individual dog breeds can be utilized as a proxy for specific genetic information. In these cases, outcomes with varying prevalence in different breeds of dogs can be identified as having a potential genetic or heritable component to the disease etiology.

The initiation of dog health studies, such as DAP, mark a turning point in the utilization of dogs as a model for human health and disease. These large-scale datasets constructed with the express intent to understand the biological and environmental determinants of health and disease in companion dog populations in order to apply those findings to other species will become increasingly valuable to medical research as the datasets grow in terms of content and population.

## 2.5 Conclusions

Use of medical data collected in dog populations will continue to be a rich source of information that can be used to inform our understanding health, longevity, and disease outcomes.

# Chapter 3. CANINE SOMATIC MUTATIONS AND AGING USING DUPLEX SEQUENCING

## 3.1 INTRODUCTION

Aging is defined as the age-related loss of function in an individual with time. Mutations are at the center of most basic evolutionary theories of aging. Our understanding of mutations initially started with early theories of germline mutation. Mutations in the germline allow for genetic diversity which propels evolution through natural selection. Thus, evolutionary change would not be possible without mutation (Baer et al., 2007), (Lynch et al., 2016). But mutation is not only central to adaptation. It is also thought to be central to the deleterious phenotypes associated with aging. The first widely accepted evolutionary theory of aging was put forward by Sir Peter Medawar. Medawar's mutation accumulation hypothesis states that early in life, selection gets rid of harmful mutations, but that later in life the strength of selection declines, such that late-acting deleterious mutations cannot be purged, and over evolutionary time, these damaging mutations will rise in frequency (Turan et al., 2019). A few years later, Williams further expanded on this by presenting his antagonistic pleiotropy theory. He hypothesized that an allele that is detrimental late in life might actually be favored by selection if it has pleiotropic effects that are beneficial early in life (hence, 'antagonistic') (Williams, 1957). Williams' theory further explains why the aging phenotype may be associated with loss of function. A trait essential to the early fitness of the organism can, with age, become deleterious later in life. These two earlier theories on germline mutations focus on the origins in aging over evolutionary time. But

in later work, scientists focused on the damage that somatic mutations might cause to tissues, and the entire organism, over a lifetime. It is this concept of somatic mutations that stimulated the work that we will discuss here.

Not long after the work of Medawar and Williams, Leslie Orgel suggested that molecular damage might provide the specific mechanism of aging (Orgel, 1963). He proposed the error catastrophe theory of aging, which argues that aging is due to the errors that exist in protein synthesis. Orgel stated that the mistakes that occur in protein translation lead to an accumulation/cascade of errors with age, which eventually leads to illness and mortality. Orgel's theory hypothesized that errors post translation were the cause of aging. While Williams and Medawar previous work focused entirely on germline mutations that are passed from generation to generation, and so exist in every single cell in the body, Orgel turned his attention to errors that occur within the lifetime of an organism, in somatic tissue. This laid the groundwork for later understanding of decreased maintenance of somatic tissue with age. Somatic mutations are not passed along to the next generation, and the focal point of this chapter.

Thomas Kirkwood was later inspired by Orgel's general idea of somatic damage accumulation, but from the perspective of DNA damage and mutation, not protein. One proposed reason why somatic mutations persist is that there are evolutionary limitations to the maintenance of the soma (Kirkwood, 1977). Kirkwood developed the disposable soma theory to apply evolutionary genetics to Orgel's previous hypothesis. It states that organisms have an evolutionary trade-off between the energy that can be divided between growth, reproduction and DNA repair and maintenance. Organisms will not invest in maintaining the soma endlessly at the cost of reproduction, because an

organism only has a limited amount of resources to allocate to its various cellular processes (Kirkwood, 1977). Too much of an investment in one of these variables can lead to detrimental effects to another. Thus, Kirkwood would argue that somatic mutations persist despite mechanisms for DNA repair because too much somatic maintenance would sacrifice the organism's ability to grow and reproduce. This could explain how, unlike with germline mutations, somatic mutations are not under as much selective pressure, especially if they occur late in life. Rather, there is a vested interest in somatic maintenance only to the extent that it promotes the fitness of the germline (Kirkwood, 1977). Therefore, even though individuals have mechanisms for repairing DNA damage, starting from the age of reproduction, the strength of selection will decline, and somatic mutations should increase in frequency. This may explain why somatic mutations occur and accumulate in higher frequency than germline mutations (Lynch, 2010).

Additionally, young individuals have mechanisms to avoid the effects of DNA damage such as cellular senescence and apoptosis, but these mechanisms can be detrimental later in life because of their impact on tissue functionality (Campisi, 2008, Coppe, 2010 #212). Reflecting on early hypotheses about germline mutation and later work on somatic mutations, it is clear that age impacts the regulation of healthy function of a trait, and that an older individual will be more susceptible to dysregulation of cellular function. These theories of aging, in concordance with our knowledge of somatic mutation theory and disease, underscore the need for greater understanding the relationship between somatic mutation frequency and longevity.

Somatic mutations are thought to play a major role in aging and age-related disease (Kennedy et al., 2012). However, we need more effort to understand the degree to which variation in somatic mutation might explain variation in aging and disease risk among individuals. It is known that some diseases of early aging such as progeroid syndromes are caused by inherited germline mutations that lead to increased somatic mutation (Vijg & Suh, 2013). These syndromes are characterized by premature symptoms of normal aging, as exemplified by Werner's syndrome (Goto, 1997). Studies have also shown that the aging phenotype is promoted by random events such as somatic DNA damage and mutation (Kenyon, 2010). This study investigates how somatic mutation frequency changes with age, in a new and informative model organism: dogs.

Traditional model organisms such as rats and mice have limitations and caveats when studying aging and disease, mainly because they do not have the same naturally occurring diseases as humans (Hoffman et al., 2018). Thus, we decided to utilize the canine model as a tool to better understand somatic variation with aging. Dogs allow a novel opportunity for scientists to investigate diseases and aging (Ruple et al., 2022). They are exposed to the same environment as humans and have many of the same naturally occurring diseases and extensive veterinary healthcare records (Hoffman et al., 2018). Dogs also age faster than humans (7-10x) and they live shorter, meaning that the effects of aging can be investigated in a more realistic timeframe (Hoffman et al., 2018). These factors make dogs a powerful tool to better understand the biology of aging and enable translational research.

Although domesticated dogs have existed for many thousands of years (Boyko, 2011), most of the among-breed variation has been shaped over the last 200 years. During this time, humans have artificially selected for specific behavioral and morphological traits (Lindblad-Toh et al., 2005). Dogs have more phenotypic variation than any other animal species, for example in size and behavior (Boyko et al., 2010). At the same time, this process of domestication has led to extreme variation in life expectancy and disease risk among breeds (Fleming et al., 2011 ; Kraus et al., 2013; Marsden et al., 2016 ; Rowell et al., 2011 ). Though unintentional, this breed-specific variation in disease risk makes it possible for us to assess the genetic underpinnings of canine disease and lifespan. In particular, the unique genetics of the dog has enabled us to identify many of the genetic variants associated with cancer, which is a leading cause of death in dog populations (e.g., Karlsson et al., 2013). This variation can also help us explore a key underlying genetic cause of disease in dogs and other species: age. As previous work suggests, the unique genetic architecture of dog breeds provides us with the ability to explore the underlying mechanisms that set the stage for the onset of disease (Davis & Ostrander, 2014).

We wanted to target early somatic mutations to see how they change with age in the blood of healthy dogs. Detecting these somatic mutations in healthy tissue has proven to be difficult because of how rare these mutations are (Vijg & Dong, 2020). When mutations are at low frequency the noise in sequencing is much greater than the signal, which leads to an overwhelming amount of false positives. However, cancer-linked mutations have been previously reported in blood in a subset of elderly individuals without cancer, a phenomenon referred to as clonal hematopoiesis of

indeterminant potential (CHIP) (Genovese et al., 2014; Jaiswal et al., 2014; Steensma et al., 2015; Xie et al., 2014). In one previous study of mutations associated with hematological malignancies in humans, detectable somatic mutations were found in young healthy individuals and then rose in frequency with age (Jaiswal et al., 2014). These studies were performed with conventional Next Generation Sequencing (NGS) workflows, which limits the practical detection of mutations to ≳2-5%. These modern sequencing platforms have an error rate on the order of ~1 in 500. However, the initial frequency of cancer mutations in healthy people will be low (Kakiuchi & Ogawa, 2021). To address this discrepancy, we utilized Duplex Sequencing (Kennedy et al., 2014; Schmitt et al., 2012) to identify somatic mutations in dogs. We relied on this more sensitive and accurate sequencing technology to filter out the noise of sequencing error which would overwhelm the underlying signal of rare but potentially pathogenic mutations, and results in an almost certain overestimation of the prevalence of these mutations (Salk et al., 2018). It is currently the most accurate sequencing method available and provides the ability to detect mutations at frequencies $<10^{-7}$, two orders of magnitude lower than other high accuracy sequencing methods (Salk et al., 2018). This technique has been used to study mutations in multiple human tissues, including ovarian, breast, and cervical tissues of healthy people (Ahn et al., 2015; Krimmel et al., 2016; Salk et al., 2019). Prior studies using Duplex Sequencing have detected low frequency mutations in cancer driver genes and age-related clonal expansions in human samples (Matas et al., 2022, Ghezelayagh et al., 2022)

Although we are not focused primarily on disease, rather aging and somatic mutation, our ability to successfully use this technology to detect rare somatic mutations

in healthy dogs will pave the way for future studies to investigate how those mutations may impact disease risk. Clonal expansions have been proven to be associated with cancer driver mutations, even in non-cancerous tissue (Kakiuchi & Ogawa, 2021). Thus, we targeted multiple cancer-associated targets in our gene panel. Arguably, we could have chosen any genes to assess aging, but since our initial efforts were to identify differences in disease risk, we proceeded accordingly. Results from this work will bring us an important step closer to developing the dog as a powerful model to understand somatic mutations and aging.

We used Duplex Sequencing to quantify rare de novo somatic mutations in dog blood (Kennedy et al., 2014; Schmitt et al., 2015 ). We hypothesized that mutation frequency would increase with age. Previous studies have observed increased mutation frequency with age in human peripheral blood (Akiyama et al., 1995; Grist et al., 1992; Trainor et al., 1984). Additionally, previous Duplex Sequencing studies have found increased mutation frequency with age in mtDNA (Kennedy et al., 2013) and DNA in various organs (Salk et al., 2017). We investigate whether other variables such as breed, gene, and sex impact mutation frequency at different ages. While this paper is focused on aging rather than cancer, age is a risk factor for cancer and we sought to identify how mutations in cancer genes were changing over time, to identify the earliest stages of selected mutation. Lastly, we sought to investigate if there was evidence of an aging signature in dogs similar to the mutational signature that we see in human COSMIC data (Catalog of Somatic Mutations in Cancer (Alexandrov et al., 2018; Alexandrov et al., 2013; Alexandrov & Stratton, 2014). Testing this hypothesis allowed us not only to see if not only the presence and accumulation of mutations has increased

with age, but also to ask if there are age-related trends in the processes that underlie mutagenesis. This study tested the general hypothesis that somatic mutations, as measured by Duplex Sequencing, show an increase with age. We also hypothesized that genes associated with cancer might show relatively strong associations with age, due to positive selection on non-synonymous mutations, compared to genes not associated with cancer. Finally, we hypothesized that dogs would recapitulate the specific signature of mutations observed with age in humans such as SBS1 and SBS5. We assess the age-related impact of somatic mutation in a non-human model with a new sequencing technology that will allow us to better understand the accumulation of mutations in healthy individuals with age.

## 3.2 METHODS

### 3.2.1 *Source Population*

Our target population focused on eight dog breeds, including 20 Golden Retrievers, 12 Labrador Retrievers, 22 Boxers, 11 Samoyeds, 17 Alaskan Malamutes, 13 German Shepherds and 14 Dalmatians, ages 6 months to 15 years old. Body size can contribute significantly to differences in disease risk and longevity in canines (Yordy et al., 2020). Thus, we selected breeds that have relatively similar same body size and lifespan. The average lifespan for these breeds is about 10-13 years (Kraus et al., 2023). We had 50 males and 58 females, as well as one individual where sex was not recorded. This included 31 fixed and 26 intact females, as well as one where status was not recorded. This also included 21 fixed and 26 intact males, as well as three where status was not recorded.

We only recruited dogs who were healthy as far as the owners were aware. Dog owners were recruited through social media, dog shows, veterinary clinics, and dog social venues (play parks, dog-friendly businesses) in Washington, Oregon, Florida, Colorado and Idaho. Collection occurred over the course of 18 months. Advertisements were placed in vet clinics and through clinic email to owners of the target breeds, through social media, word of mouth, and direct contact by business owners. Additionally, several local and national breed clubs put out announcements to their membership, who in turn reached out to our team to volunteer. All the work described here was carried out under University of Washington IACUC approval (PROTO201600009 : 4359-02: Healthy Aging in Companion Dogs, approved 2019/03/05, refer to supplement to see IACUC and consent form).

### 3.2.2 *Blood Sample Collection and Quality Control*

After recruitment, blood was drawn by licensed veterinarians or veterinary technicians on site. Blood was collected one time from a peripheral vein (cephalic, jugular or saphenous according to phlebotomist preference) of each unsedated dog using a needle and syringe. About 10mL of blood was collected for each dog and stored in EDTA on ice for up to 48 hours. Samples were prepared as they arrived in the lab, with most samples delivered within a few hours of collection and few samples received within 24-48 hours of collection. Prior to the start of the bulk of our collection, as quality control to determine whether delivery time impacted DNA quality, blood was prepared (as described in the next section) at both 24 and 48 hours from blood collection time. The samples were then tested for DNA quality using Agilent 4200 Genomic TapeStation

tapes. Our results showed that blood taken from both time periods had quality DNA (DIN scores >9).

### 3.2.3 *DNA Isolation and Sample Preparation*

After collection, all blood samples were delivered to the Promislow lab at the University of Washington on ice and kept at 4° C. Upon receipt, samples were centrifuged in SepMate-15 tubes with Lymphoprep density gradient medium according to manufacturer's protocols (StemCell Technologies) to isolate peripheral blood mononuclear cells. Cells were washed 2 times in 2% FBS-PBS solution in 15mL conical tubes. We counted the extracted cells in a hemocytometer and considered a minimum of 1 million cells/mL sufficient to proceed with DNA extraction. DNA isolation was performed by adding about 2 million cells (or 2mL) to spin columns using the Zymo Quick-DNA/RNA Microprep Plus Kit D7005 (Zymo Research, Lot ZRC204206). DNA was quantified using Qubit dsDNA HS Assay Kit (ThermoFisher Scientific, Lot Q32854).

### 3.2.4 *Probe Design*

We sought to identify somatic mutations in each of seven genes, including *PTEN*, *EZH2*, *FBXW7*, *POT1*, *TRAF3*, *SETD2*, *KRAS*, and *RHO* (See Supplementary Table 1). All genes were selected because of their previous mention in the literature in relation to dog and human lymphoma which was the initial disease we were looking to investigate before we decided to have this paper focus on aging (Morin et al., 2010, McDonald et al., 2018, Smith et al., 2020, Fanelli et al., 2022, Coyle et al., 2022). *RHO* was selected as a control gene as it should not be selected for after the development of the eye. For each gene, we designed probes targeting the coding exons, using the CanFam3.1 dog

genome assembly (Integrated DNA Technologies, 2022), making a total of 289 probes (Supplementary Table 1). These probes (~29kb gene panel) were pooled at a concentration of 0.75pmol/$\mu$L in TE.

### 3.2.5 *Library Preparation*

To prepare libraries for each sample, 500ng of genomic DNA were prepared using the commercially available Duplex Sequencing Kit (TwinStrand Biosciences, Seattle, WA), following manufacturer's instructions. Sonication, end-repair, A-tailing and ligation to Duplex Adapters were performed. After PCR, a capture was performed with two consecutive hybridizations of our probe panel. We used the Agilent 4200 TapeStation with HS D1000 tapes for library fragment size confirmation. DNA quantification of libraries was performed with Qubit dsDNA HS Assay kit (ThermoFisher Scientific).

### 3.2.6 *Sequencing*

Indexed libraries were pooled and sequenced on an Illumina NovaSeq6000 with 150 paired-end reads at Azenta Life Sciences (South Plainfield, NJ) and UW Medicine Virology (Seattle, WA). Approximately 80,000,000 read-pairs were allocated per sample. Sequencing reads were processed using Duplex Sequencing pipeline for error correction (v2.1.3); GitHub - Kennedy-Lab-UW/Duplex-Seq-Pipeline: A standalone end-to-end data analysis pipeline for Duplex Sequencing) and analyzed as described previously (Kennedy et al., 2014). VariantCall Format (VCF) files were converted to MAF files, which were then post processed with R version 4.1.2.

### 3.2.7 *Mutational Analysis*

For each sample, mutation frequency was calculated by dividing the number mutations in that sample by the total number of Duplex nucleotides sequenced. Mutation burden was calculated by summing the number of mutant duplex reads at that site and dividing by the total number of Duplex nucleotides sequenced at that site. Each duplex read corresponds to an original DNA molecule and thus mutation burden informs about the number of different mutations identified as well as the size of the clones. Large clones were defined as clones which had 2 or more duplex reads, and large clone mutation frequency was calculated by dividing the number of mutations present as large clones by the number of nucleotides sequenced. SNPs were filtered by removing variants with a variant allele frequency higher than 0.4. Our target for mean coding depth per dog was 6,000x, and all samples met this target except nine which were excluded. Single mutations with a depth below 1000x were also excluded from our analysis. Mutational analysis was performed in R (version 4.2.0) and packages used are listed in Supplementary Table 3.

### 3.2.8 *Mutational Signature analysis*

To test for the presence of specific mutational signatures, we first queried human Catalogue of Somatic Mutations in Cancer (COSMIC) data (Tate et al., 2019) (v95, accessed on May 15th, 2023). Mutational signature analysis was performed with SigFit (Kucab et al., 2019). Our methods were adapted from previous literature, with the following changes: (1) Instead of just looking for COSMIC SBS1, we also looked for SBS5 as another age-related signature, and (2) Our list of contexts incorporated the number of duplex reads covering each context, instead of merely the number of

positions in the target region of CanFam3 that covered each context (Cagan et al., 2022). Specifically, we selected SBS1 and SBS5 because of their known association to with aging (Alexandrov et al., 2015).

### 3.2.9 *Statistical Analysis*

When looking at single predictors, we used non-parametric models, which do not make any assumptions about the nature of the frequency distributions of the data. For cases where we wanted to explore multiple predictors there is not a non-parametric option, and so we used linear regression models. I was interested in comparing models with age, gene and breed and the interactions between these variables so all of these models went into the common framework of a linear regression. We analyzed the effect of age, gene, breed, sex and intact status on mutation frequency and mutation burden using general linear models. Total mutation frequency and mutation burden were square-root transformed to allow the data to satisfy the assumptions of normality needed in our model. Tests for specific genes or specific mutation types were done with a Spearman's correlation. Some of our measures of mutation frequency were zero inflated including those for *KRAS*, *PTEN* and for large clone data. In these cases, we converted clone frequency to a binary 'presence/absence' variable, treating any non-zero mutation frequencies as having values of 1 (i.e., mutations present), and applied logistic regression. All multiple comparison data was corrected using the Bonferroni correction method.

## 3.3 RESULTS

### 3.3.1 *Sample Sequencing Data*

Blood was collected from 109 dogs and DNA was extracted from peripheral blood monocytes. In total we analyzed genomic data from 19 Golden Retrievers, 12 Labrador Retrievers, 21 Boxers, 11 Samoyeds, 10 Alaskan Malamutes, 13 German Shepherds and 14 Dalmatians. PBMC DNA was analyzed with Duplex Sequencing using a panel of 8 genes (see section 3.2.4 in Methods). After filtering out samples with a sequencing depth lower than 6000x, 9 dogs were removed from our data. In total, data was comprised of 100 dogs, with an average depth of 11030x (± 2017.7 S.D.), (absolute minimum 6157x, maximum 15252x) (Figure 1).

### 3.3.2 *Effect of Age on Mutation Frequency and Mutation Burden*

Across all genes and all dogs, the mean coding mutation frequency was $1.8 \times 10^{-7}$ ($\pm 7.4 \times 10^{-8}$ S.D.) across all samples. The mean coding mutation burden was $2.06 \times 10^{-7}$ ($\pm 1.34 \times 10^{-7}$) across all samples. We next determined the effect of age on mutation frequency and mutation burden. After square-root transforming MF and MB to meet the assumptions of normality essential for a linear model, we found that both coding mutation frequency and coding mutation burden increased significantly with age ($F_{1,98} = 18.39$, $P < 4.2 \times 10^{-5}$; $F_{1,98} = 18.68$, $P < 3.7 \times 10^{-5}$, respectively) (Figure 2). For non-coding sequences, mutation frequency increased significantly with age ($F_{1,98} = 4.17$, $P = .044$) (Supplementary Figure 2).

We next sought to determine if the effects of age on mutation frequency and burden differed among genes and among breeds. We first examined the effect of breed.

In a linear model that included effects of age, breed, and the interaction between the two on coding mutation frequency, neither breed nor age-by-breed interaction were significant ($P > 0.05$) (Figure 3). For the specific breeds examined here, we do not see differences in mean mutation among breeds, nor do we detect variation in the rate at which mutations accumulate with age.

We next asked if mutation rates differ among genes. We first removed KRAS and PTEN, as the mutation frequency in these genes was considerably lower than in all other genes, with large numbers of zeros (see above). In a linear model that included age, gene and the interaction between the two, we found significant differences in mutation frequency among genes ($F_{5,581} = 7.04$, $P = 2.1 \times 10^{-6}$), but no evidence in support of an interaction effect, testing the hypothesis that the effect of age on mutation frequency differed among genes ($F_{5,576} = 7.04$, $P > 0.05$). Next, we performed a Spearman's correlation to look at our data from specific genes. Figure 4a shows a significant effect of age on four of the eight genes we examined, with our raw p-values including *EZH2* (Spearman's $\rho = 0.35$, $P = 0.00033$), *FBXW7* (Spearman's $\rho = 0.29$, $P = 0.0034$), *SETD2* (Spearman's $\rho = 0.26$, $P = 0.0088$) and *TRAF3* (Spearman's $\rho = 0.36$, $P = 0.00027$). These trends were significant even after multiple comparison correction for the 6 genes. Similarly, for mutation burden we observed a significant effect of gene ($F_{5,581} = 5.16$, $P < 0.0001$) but did not observe a significant age-by-gene interaction was not ($F_{5,576} = 1.196$, $P > 0.05$).

To address the large number of 0 values in PTEN and KRAS, we carried out a logistic regression to examine the effects of age in the presence of mutations. We found that mutation presence in KRAS increased with age ($z = 2.1$, $P = 0.037$), while age did

not have a significant effect on PTEN mutations ($z = 0.02$, $P > 0.05$) (Figure 4b). Finally, we asked whether the load of mutations varied by sex or sterilization status in this population. Neither factor influenced either mutation frequency or mutation burden ($P > 0.05$ in all cases).

### 3.3.3 *Large Clone Mutation Analysis*

To better understand clonal expansion of our mutations, we next wanted to test how the presence of large clones (mutations with 2 or more reads) changed with age. Large clones were relatively rare, such that for most genes, most individuals had a value of 0. Accordingly, we used a logistic regression model to test for the effect of age on presence of large clones. We found that large clone mutation presence increased significantly with age ($z = 2.56$, $P = 0.011$) (Figure 5a). An ANOVA of the logistic model of age and gene showed a significant difference among genes in the mean level of large clones ($F_{7,779} = 4.1$, $P = 0.0002$), although there was not a significant gene-by-age interaction ($F_{7,772} = 1.75$, $P > 0.05$) (Figure 5b).

### 3.3.4 *Mutation Type Analysis*

We categorized coding mutations into five mutation types, including indel, missense, nonsense, silent and splice site mutations. We then compared the different types of mutations among three distinct age bins of dogs, including young (less than 4 years old), adolescent (greater than 4 or equal to 8 years), and adult (8 years or older) (Figure 6a). We also visualized proportion of different mutation types across different breeds by age bracket (Supplementary Figure 3). Figure 6b shows the proportion of mutation type by gene. To test for differences among genes in the distribution of

mutation types, we performed a Fisher's exact test. The relative frequency of different mutation types was significantly more different among genes than we would expect by chance (P = 0.0005) (See Supplementary Figure 4).

We sought to determine which, if any, mutation types increased with age. We created a vector for mutation type frequency that included the total sum of that mutation type in that individual, divided by the total nucleotides sequenced. Figure 6c shows the distribution of mutation frequency data by mutation type. Using a general linear model, we found that the frequency of all mutation types increased significantly with age (Figure 6d). Indels were most significantly increasing with age (Spearman's $\rho = 0.4$, $P = 3.5 \times 10^{-5}$) followed by: silent mutations (Spearman's $\rho = 0.37$, $P = 0.00014$), missense mutations (Spearman's $\rho = 0.33$, $P = 0.00078$), splice site mutations (Spearman's $\rho = 0.3$, $P = 0.0022$), and nonsense mutations (Spearman's $\rho = 0.25$, $p = 0.014$). These data were significant even after multiple comparison correction.

### 3.3.5 *Mutational Signature Analysis*

Next, we wanted to look at the effect of age on mutational spectra and mutational signatures. Figure 7a shows the normalized frequency of mutations by spectra (by trinucleotide context), organized by age bracket. Visual inspection of this figure highlighted a particularly strong pattern of mutations, including C→T mutations, resembling a mutational signature known as SBS1 which is caused by the spontaneous deamination of 5-methylcytosine (COSMIC). We next looked at how mutation spectra changed with age using Spearman's correlation and found that three of the mutation spectra increased significantly with age including C→T ($\rho = 0.63$, $P = 2.9 \times 10^{-12}$), T→A ($\rho = 0.37$, $P = 6 \times 10^{-4}$) and T→C ($\rho = 0.42$, $p = 4.7 \times 10^{-5}$). We extracted two signatures

from our mutational data (coding and non-coding) including of SBS1 and SBS5 which are both known clock-like signatures (Alexandrov et al., 2015). We found that both SBS1 and SBS5 mutations signatures increased with age in both mutation count ($F_{2,215}$ = 36.86, P = 1.99 x $10^{-8}$) and proportion ($F_{2,215}$ = 3.4, P = 0.00083) respectively (Figure 7c).

## 3.4 Discussion

Duplex Sequencing is highly sensitive and thus able to detect low frequency somatic mutations in healthy individuals (Kennedy et al., 2014; Schmitt et al., 2012). This technology has the potential to help scientists better investigate and identify extremely rare somatic mutations. Through studying how these mutations change with age, we could potentially better understand their association with disease risk, predisposition and preventative medicine.

This study varies from others that utilize Duplex Sequencing in that we used the DNA from dog peripheral blood monocytes. As mentioned earlier, dogs provide a powerful tool to better understand aging and somatic mutation. These mutational patterns may take decades to pick up on in humans, with differences not being as clear or distinct human to human as they would be when comparing individuals in different breeds. The reasons for this are due to the genetic similarity within breeds, as well as the short lifespan of a canine versus a human. Thus, the identification of aging-related increase in mutations in dogs could later be investigated in human populations to help us understand the genetics of aging.

Modern sequencing efforts have changed the way scientists measure somatic mutations. Before large-scale gene sequencing began, somatic mutations were hard to

detect unless they had significantly clonally expanded. Thus, identifying mutations in healthy tissue was a difficult task. Duplex Sequencing has improved our ability to detect somatic mutations in healthy tissue that otherwise would be lost in the noise of standard NGS sequencing, as explained above.

One of the most profound results of this study shows that mutation frequency increases significantly with age. Our data also show that mutation burden and the presence of large clones is increasing with age. This is consistent with previous studies that show evidence of clonal expansion in healthy tissues (Kakiuchi & Ogawa, 2021). Though some genes have a higher baseline mutation frequency, our cross-sectional data are consistent with a model that mutations in different genes accumulate at a constant rate, however no one gene is accumulating mutations faster. Additionally, there is not sufficient evidence to suggest that there are breed-specific differences in mutation rates in the genes investigated. Initially, this study sought to investigate whether there were breed-specific differences in the rate of mutation of these genes. We had hypothesized that the cancer risk differences that are recorded/known for these breeds could potentially be explained by differences in mutation accumulation in our selected genes. Our data, however, do not support this hypothesis. It is possible that other genes that are associated with lymphoma could be selected. Initially, RHO was selected as our control gene because we do not expect there to be selection for it to change with age, as it usually functions early in the development of the eye. Our data show the frequency of mutations in RHO increase with age, which is consistent with the idea that the mutation frequency increases with age even in genes that are not oncogenes or tumor suppressor genes.

Even though certain breeds and certain genes may be more mutated than others we do not have enough evidence to suggest that mutations are accrued faster in some breeds or some genes. Thus, no breed or gene is accumulating mutations faster than another. However, it is important to note that our dogs were mostly above the age of 1, except for two dogs aged 0.50 and 0.88 years. Dog biologists suggest that by the age of one year, most dogs are already adults (Da Costa et al., 2022). It is possible that during the first year of a dog's life, there are certain patterns in mutation accumulation across these genes and breeds that we are unable to detect in dogs mostly above the age of one. Future studies should focus on mutation phenomena in very young dogs under the age of one. Expanding the age range would be particularly informative for studies assessing underlying genetic factors that contribute to differences in disease risk in dogs.

Our data not only showed that total coding mutation frequency increases significantly with age, but also that certain mutation types increase significantly with age, most notably C→T mutations. These mutations occur at CpG sites, and are known to be associated with aging. C→T mutations are highly associated with aging and various mutational signatures (Alexandrov et al., 2013). Previous papers have suggested that the mutagenic exposures and certain abnormal cell processes can be detected through the analysis of mutational signatures (Alexandrov et al., 2015; Alexandrov et al., 2018; Alexandrov et al., 2013; Alexandrov & Stratton, 2014). Mutational signature analysis allows us to analyze the biological processes that caused these somatic mutations. Thus, a mutational signature can explain the underlying molecular process that led to the mutation. Mutational signature analysis allows us to

identify these signatures in our mutational data (Alexandrov et al., 2013). We investigated the presence of SBS1 and SBS5, both mutational signatures that are characterized as “clock-like signatures” in the COSMIC database (Alexandrov et al., 2015). Mutations with these signatures are associated with age, and many cancers. We saw an increase with age in SBS1, which is known as the clock-like signature of aging associated with deamination of 5-methylcytosine, and an increase in SBS5 which is also categorized as age-associated, though the underlying molecular change is unknown. It is however important to note that while dogs and humans have highly orthologous genomes, COSMIC data are based on analysis of human cancers.

Ongoing efforts to characterize and sequence canine cancers will undoubtedly lead to more comprehensive canine cancer databases and a better understanding of mutational signatures, enabling more in-depth analysis of the underlying cause of mutagenesis in dogs. Additionally, most signature decomposition analysis occurs on whole genome or whole exome data, not targeted genes as we used in this study. Our use of a very limited panel of genes is due to the high sequencing demands of Duplex Sequencing. A main requirement to perform mutational signature analysis is a large number of mutations per sample (>200) (Koh et al., 2021). While our study with Duplex Sequencing identified an average of about 61 mutations per sample, those mutations corresponded to a very small portion of the dog genome (approximately 29KB of the 2.4GB dog genome). To properly determine the mutational signatures involved it is critical to consider the sequence composition and the selective effects operating in the target genes. Additional computational efforts are needed. In short, future research efforts could significantly improve the analysis and interpretation of signatures by

targeting more mutations and relying on mutational signatures previously identified in the dog. Mutational signature analysis is particularly relevant to our dog model, as dogs are exposed to the same environments as humans and there are many signatures that explain/encode environmental mutagens. The investigation of mutational signatures in dogs could shed light on environmental determinants of cancer as well as other processes associated with disease risk.

There are numerous potential caveats to the work we have presented here. First, the analysis we present here is based on a cross-sectional cohort. Longitudinal study of the same individuals throughout the course of their lifespan would provide a more powerful way to investigate mutation accumulation. This is because it is likely that different individuals start out with a different baseline mutation count, which may change with age differently from another individual. By using cross-sectional data we are assuming that individuals in the same breed age similarly, thus relying on the homogeneity of individuals in the same breed when comparing young dogs to older dogs. It would be much better to do a longitudinal study where we follow individuals over their entire lifespan. We also have no way of knowing if the dogs we studied will eventually get sick, which would be possible with longitudinal data. Additionally, there is no way to know what potentially mutagenic exposures different individuals have been subject to that may make some individuals in the same breed have completely different mutational backgrounds, as mutagenic exposure data are unavailable. Also, we rely on the owner's reporting of the dog's health. The individuals in the study could have cancer or precancer that is not yet detected. This would be particularly informative for studies

assessing underlying genetic factors that contribute to differences in disease risk in dogs.

It is possible that the differences in mutation accumulation that explain variation in disease risk are caused by other genes, non-mutational processes (ie: epigenetic), or are clearer when comparing other breeds or younger dogs. The breeds we selected are of similar size (relative to the huge range observed in the species) and have relatively similar life expectancies (though they do differ considerably in frequency of at least some kinds of cancer). Future studies will benefit from analyzing somatic mutations in dogs of very different life expectancies, to test the idea that shorter-lived dogs might have higher rates of accumulation of somatic mutations as was done in Cagan et al., 2022. If so, it would be an excellent opportunity to identify the mechanisms that might underlie this variation. Next, it is possible that with these sample sizes, we don't have sufficient power to detect subtle differences in rate of mutation accumulation across breeds or genes. Future studies should investigate these potential underlying mechanisms of somatic mutation accumulation. And finally, the power of Duplex Sequencing lies in part in our ability to sequence a single gene at extreme depth. The trade-off is that we are only able to analyze a handful of genes. Other genes in the genome might show clear among-breed differences in mutation frequency. Our initial goal, as stated above, was to assess risk differences in lymphoma, and so our target genes were selected because of their association with cancer. Future studies should look at other target genes associated with aging and more neutral areas of the genome, to get a better understanding of age-related mutational patterns and selection.

This project incorporated a new technology and a new non-human model to better understand somatic mutations and how they change with age. This paper marks the first of what are likely many to come that investigate mutations in dogs using Duplex Sequencing methods. We are confident that future research will help scientists answer some of the most long-awaited questions about mutational processes, predisposition and translational models.

## 3.5 FIGURES

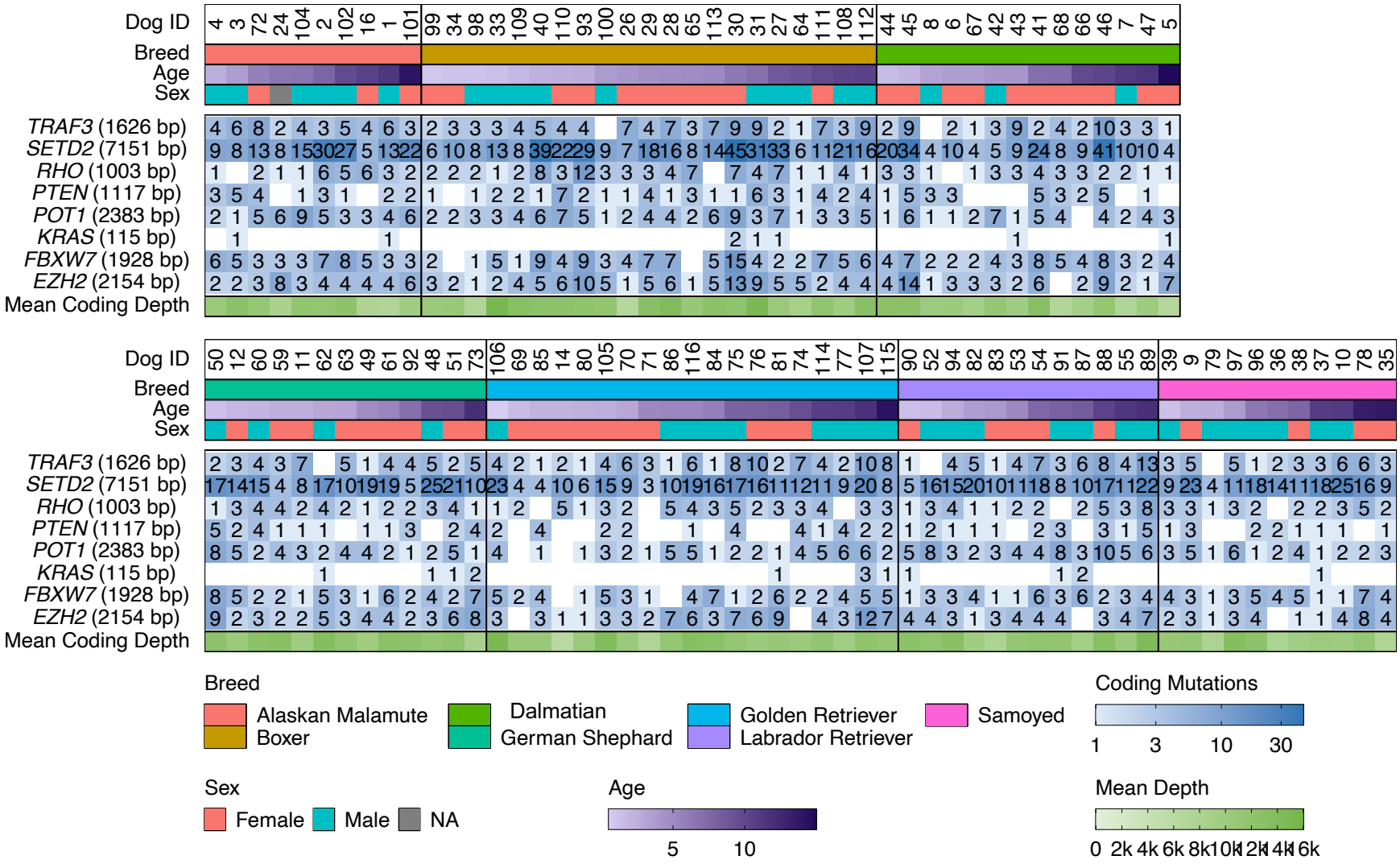



**Figure 1. Demographic information and basic sequencing data for 109 dogs.**

Individual dog information including breed, age and sex, including sequencing data. Lower portions of the figure show per-gene mutation count and homogeneity of duplex sequencing depth across sampled individual.

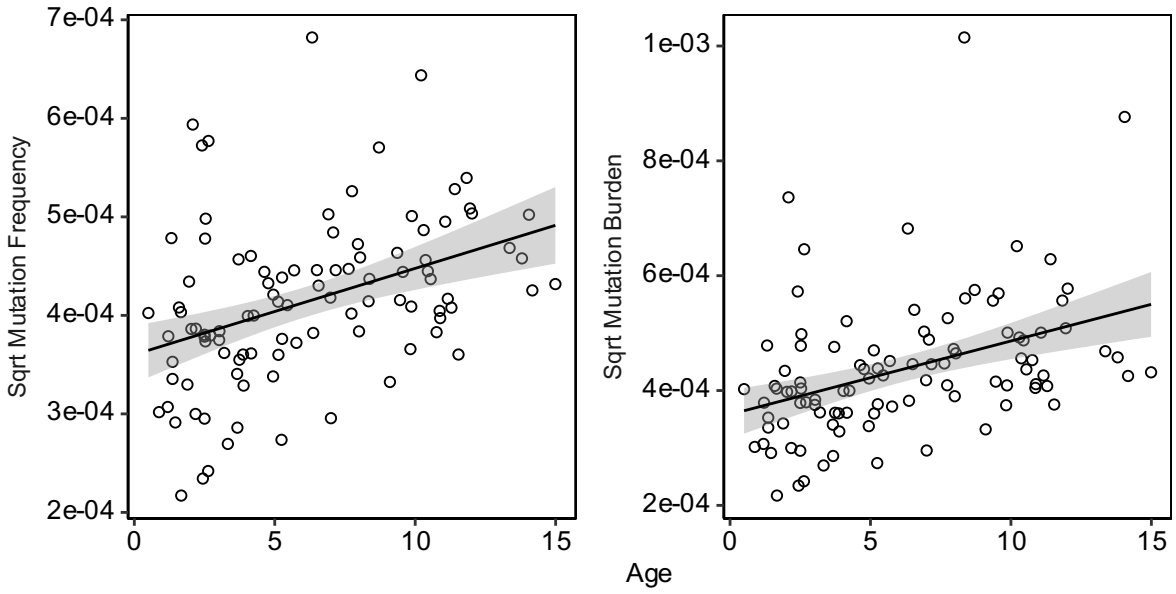



**Figure 2. Coding mutation frequency and mutation burden by age in 109 dogs.**

There is an age-related increase in both coding mutation frequency ($F_{1,98} = 18.39$, $P < 4.2 \times 10^{-5}$) and coding mutation burden ($F_{1,98} = 18.68$, $P < 3.7 \times 10^{-5}$).

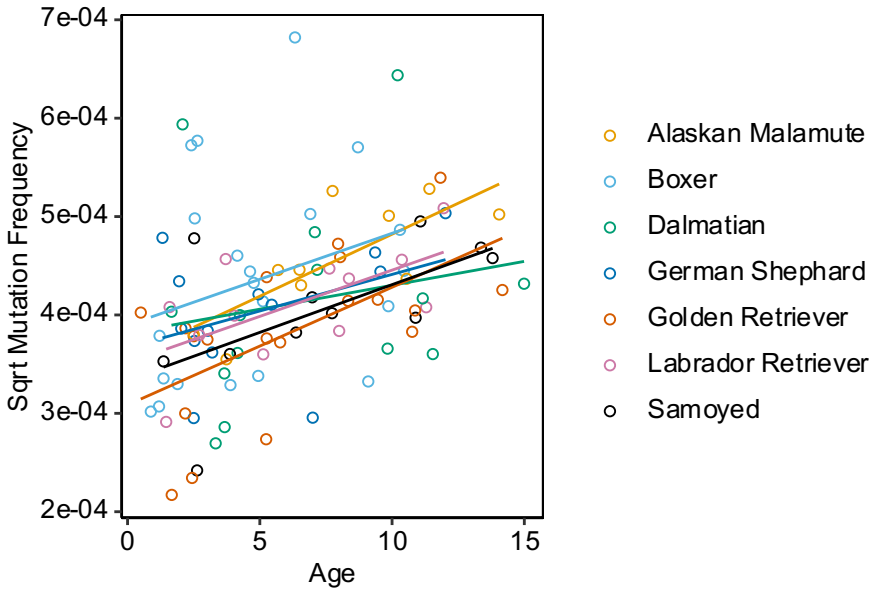



**Figure 3. Effect of breed on mutation frequency.**

There is effect of age on coding mutation frequency, but we see no effect of breed ($F_{6,86} = 1.53$, $P > 0.05$) or breed-by-age interaction ($F_{6,86} = 1.53$, $P > 0.05$).

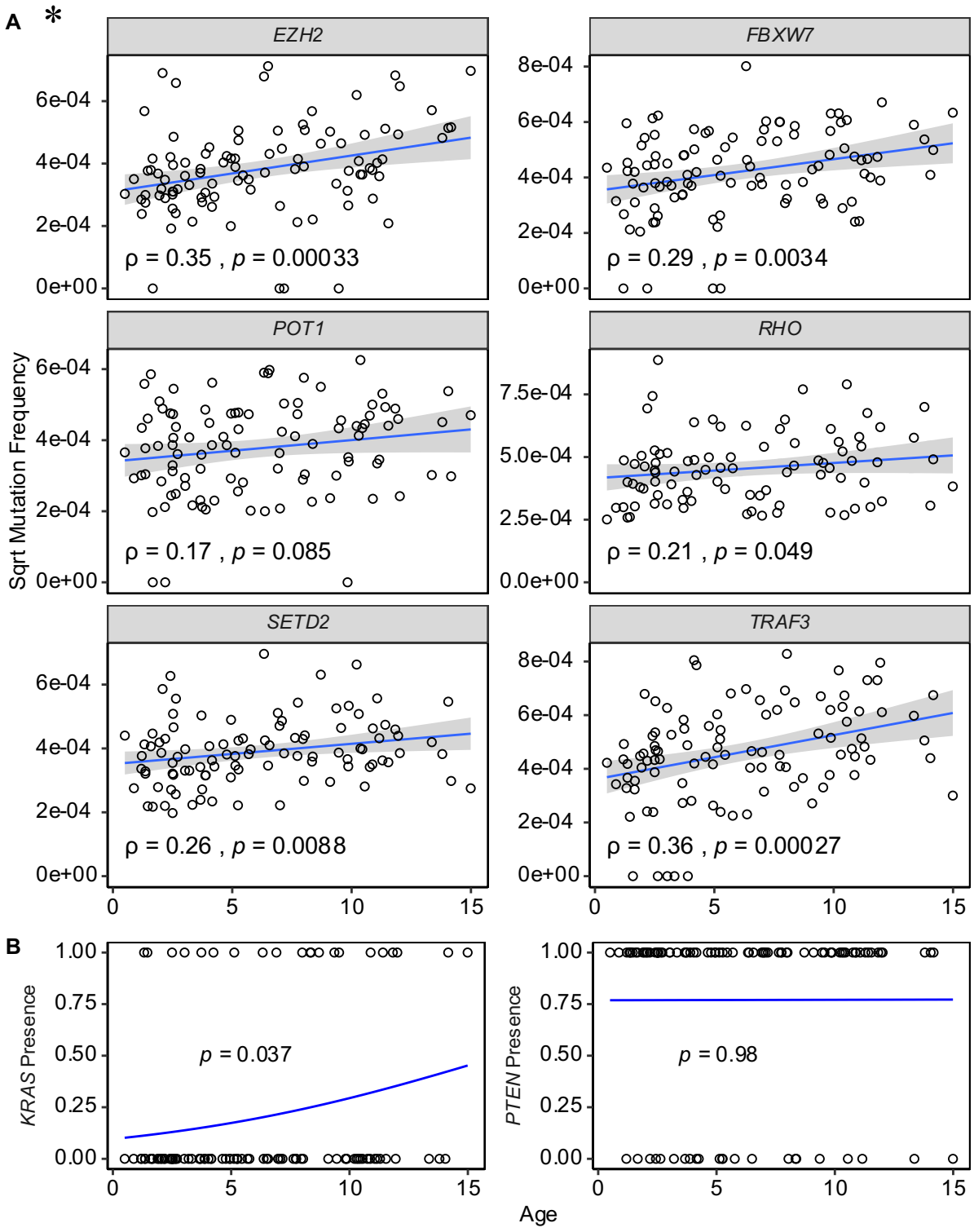



**Figure 4. Effect of gene on mutation frequency with age.**

A) For each gene, the linear regression of mutation frequency by age is shown. KRAS and PTEN data contained many zeros and were treated as qualitative variables (see panel B). B) Logistic regression of KRAS and PTEN showed that the mutations in both genes significantly increase with age. (z = 2.1, P = 0.04) and (z = 0.02, P > 0.05), respectively.

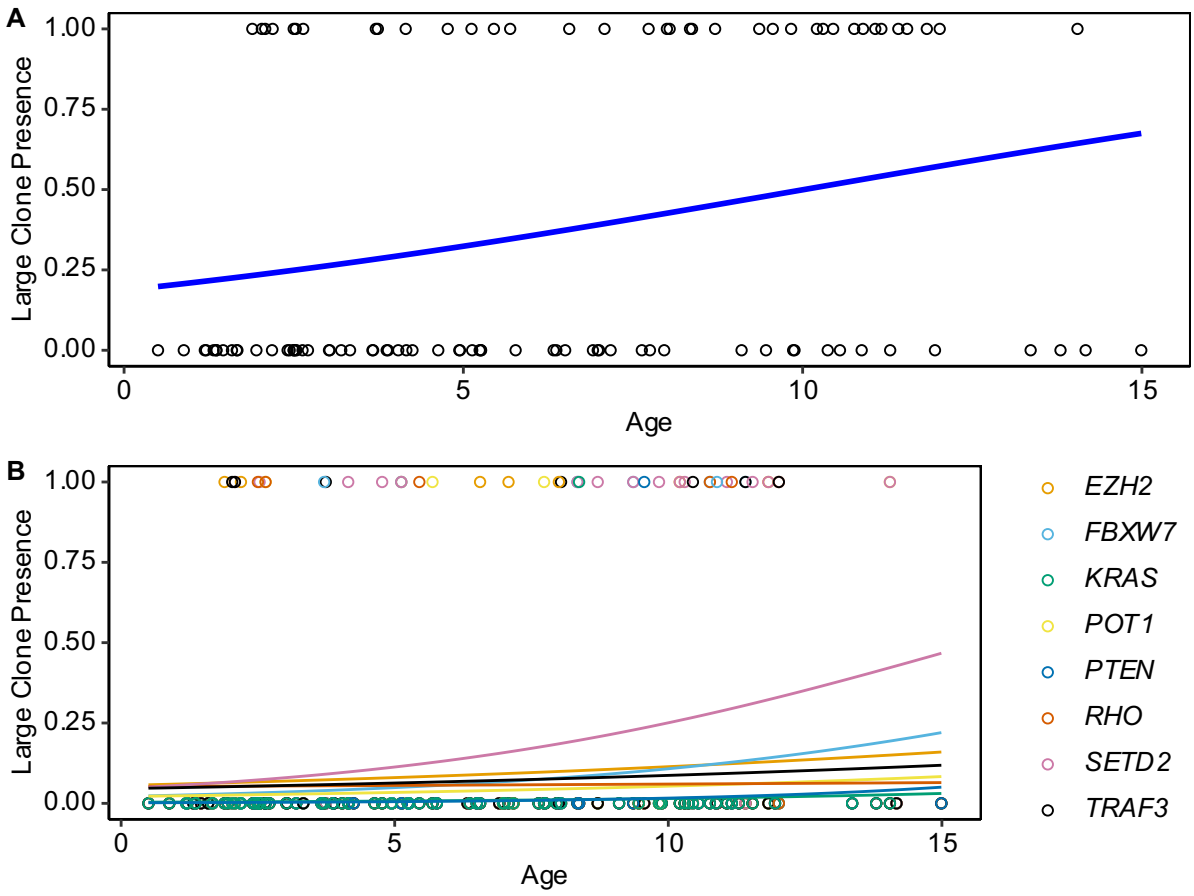



**Figure 5. Logistic model of presence of large clone mutation frequency and age.**

A) Large clone mutation presence increases significantly with age (z = 2.56, P = 0.011). B) There is also a significant effect of gene on mutation frequency ($F_{7,779} = 4.1$, P = 0.0002), but there is no significant age by gene interaction ($F_{7,772} = 1.75$, P > 0.05).

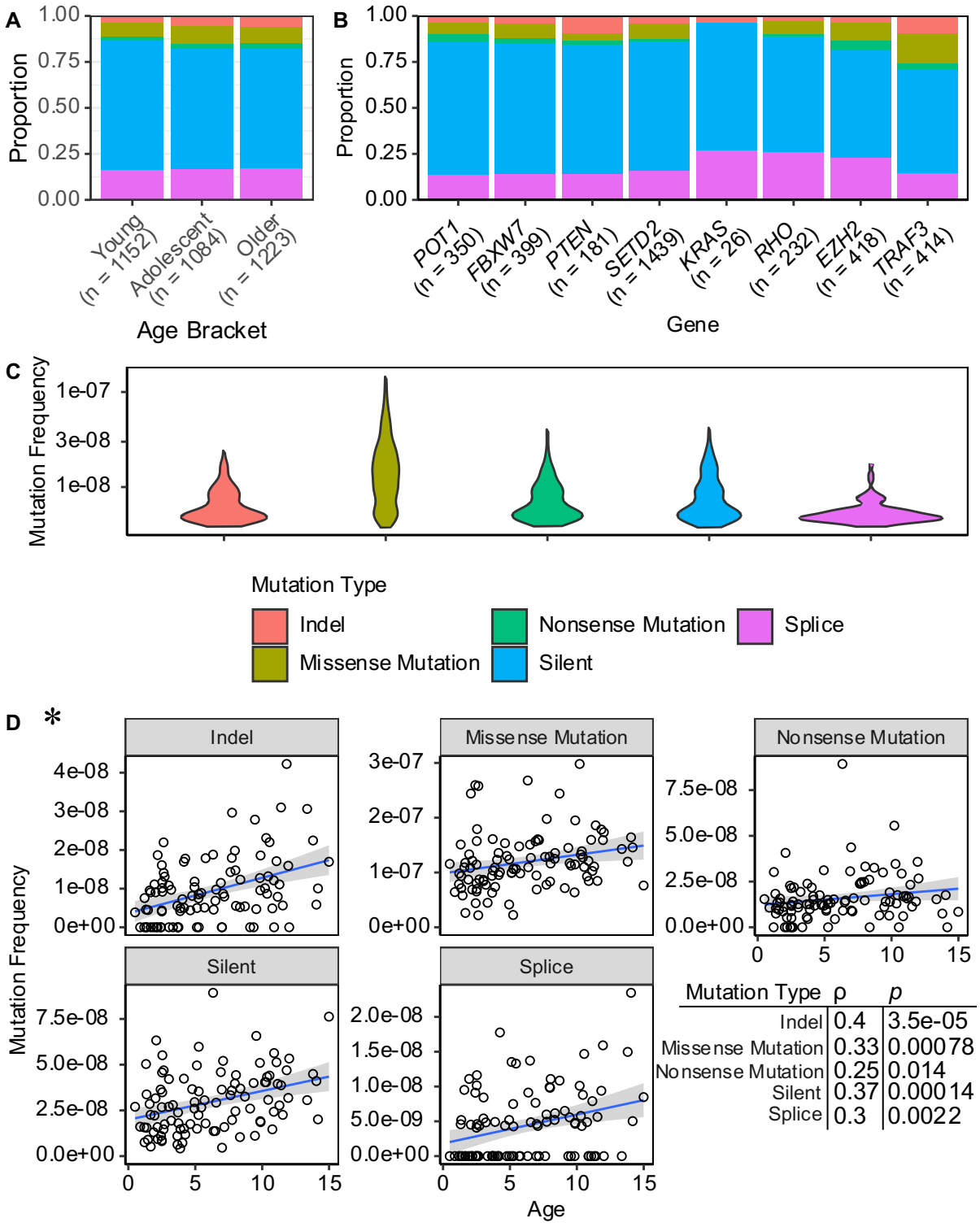



**Figure 6. Mutation type analysis.**

A) Mutation type by age bracket: young (0-4), adolescent (>4-8) and older (>8). B) Mutation type distribution by gene. The relative frequency of different mutation types was significantly more different among genes than we would expect by chance (P = 0.0005). C) Different mutation types have varying frequencies and distributions. D) All mutation types increase significantly with age.

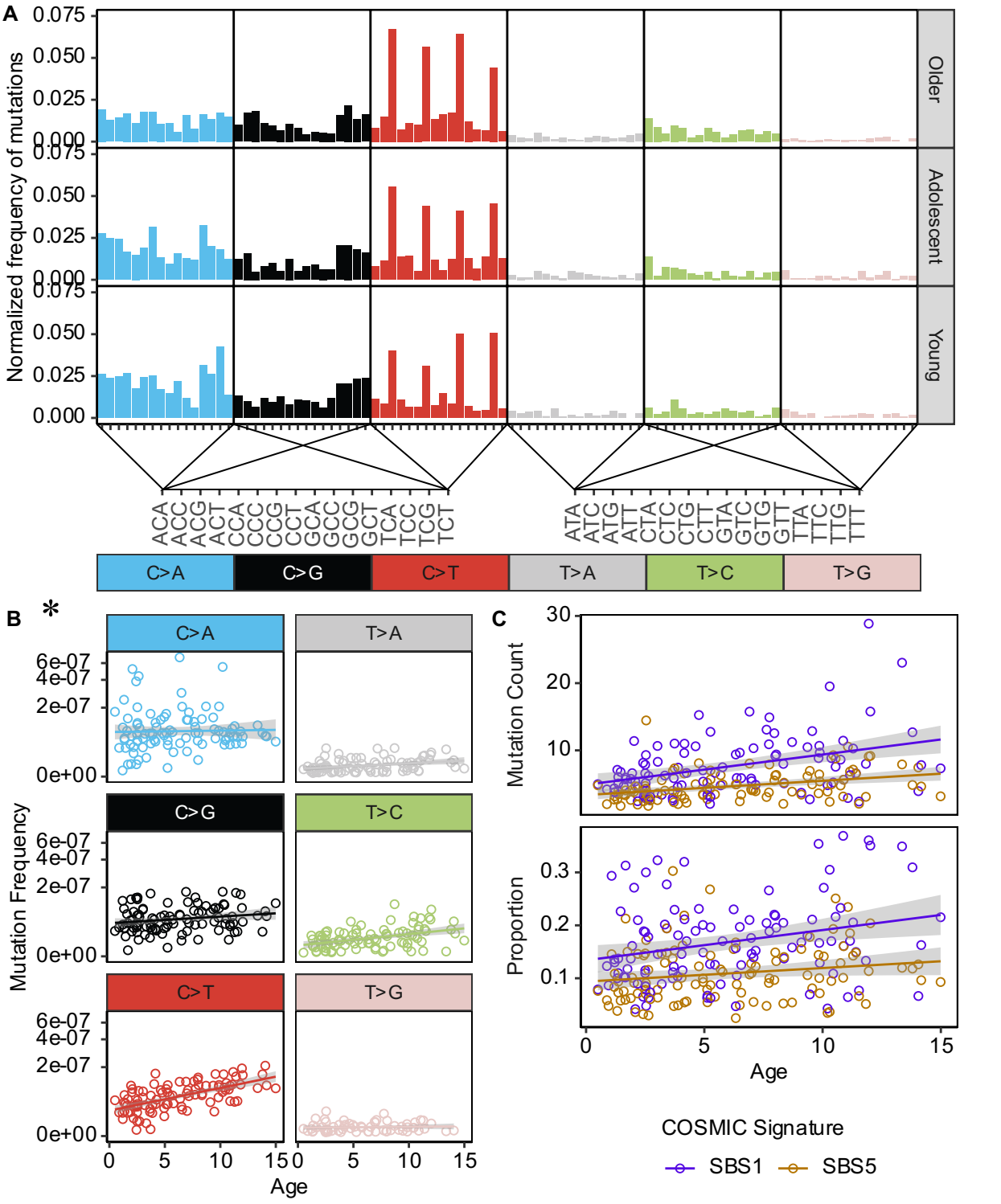
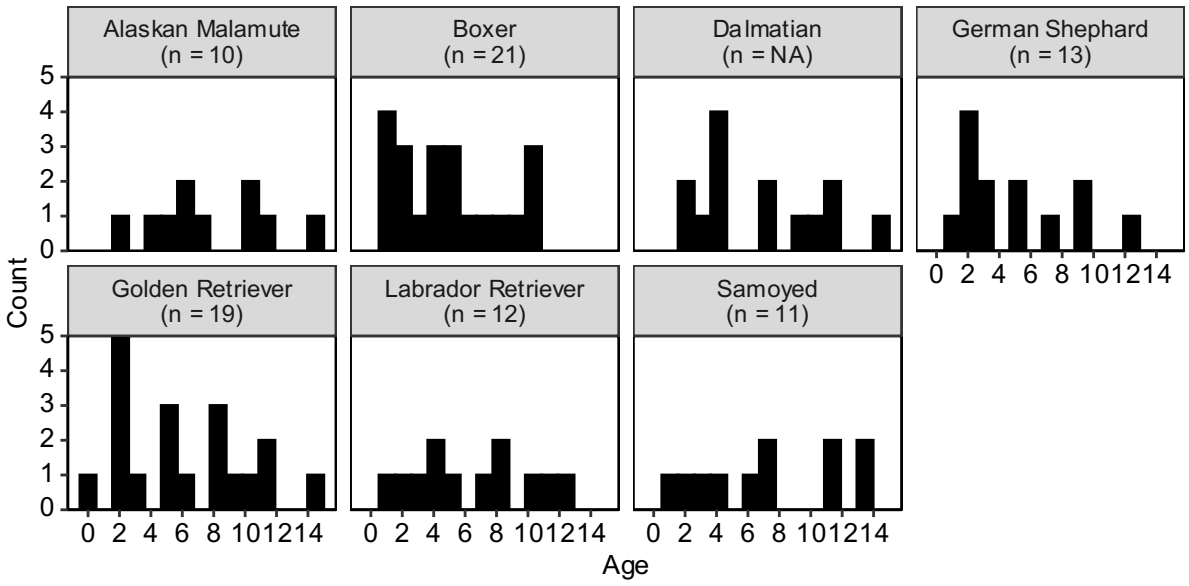



**Figure 7. Mutational spectra and mutational signature analysis.**

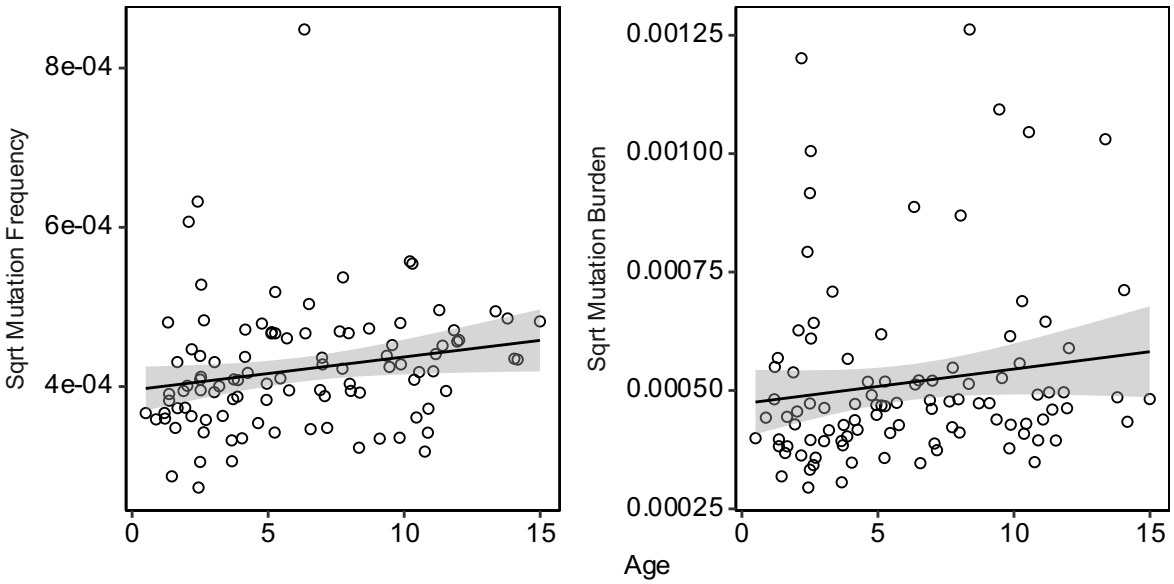
A) The normalized frequency of mutations each spectra for each trinucleotide context by age bracket. B) The increase in mutation frequency with age varies by spectra. C>T, T>A and T>C mutations are significantly increasing with age. C) Both SBS1 and SBS5 mutational signatures increase with age both in ($F_{2,215}$ = 36.86, P = 1.99 x $10^{-8}$) and proportion ($F_{2,215}$ = 3.4, P = 0.00083).

## 3.6 SUPPLEMENTARY FIGURES



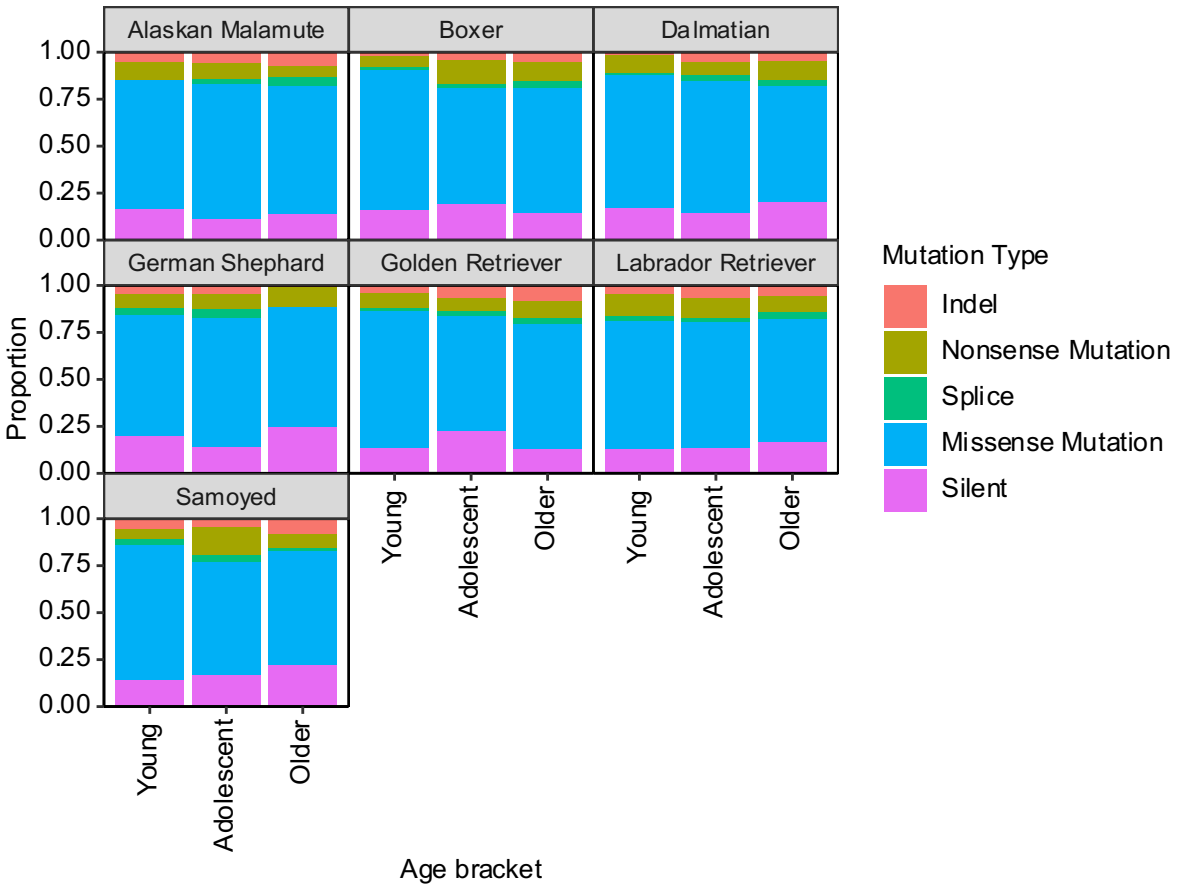

**Supplementary Figure 5. Age distributions**

Age histograms for study cohort, showing the distribution of ages from 6 months to 15 years old in seven different dog breeds.



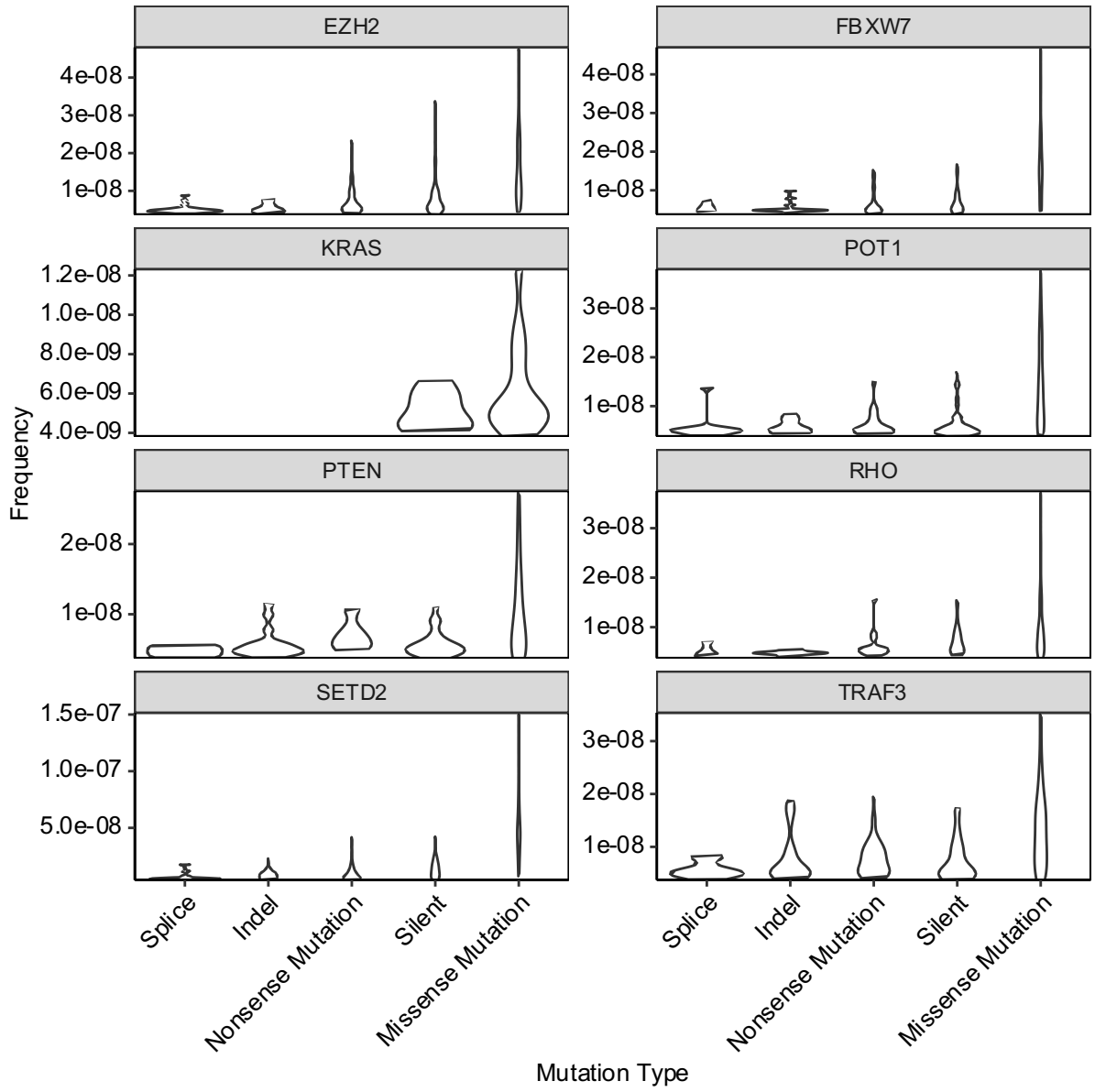

**Supplementary Figure 6. Non-coding mutation frequency and mutation burden.**

Mutation frequency increased significantly with age ($F_{1,98} = 4.17$, $P = .044$), and mutation burden was not significantly associated with age ($F_{1,98} = 2.13$, $P > 0.05$).



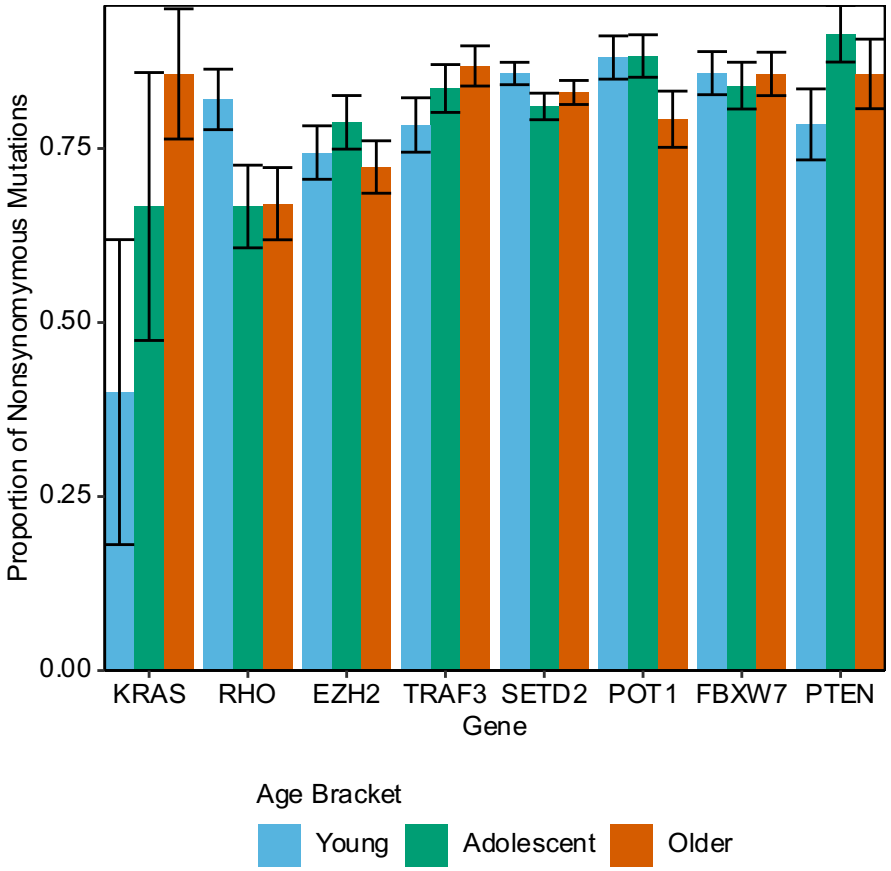

**Supplementary Figure 7. Mutation type by breed by age bracket.**

Proportion of mutation types in our 7 dog breeds, organized by age bracket.




**Supplementary Figure 8. Mutation types by gene.**

Fisher's test showed that the relative frequency of different mutation types was significantly more different among genes than we would expect by chance using a Fisher's exact test (P = 0.0005).




**Supplementary Figure 9. Proportion of non-synonymous mutations by gene by age bracket.**

## 3.7 TABLES

**Table 1. Probe Designs**

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| TRAF3 | 2 | 1 | chr8 | 70761798 | 70761918 | AAATGGAGTCAAGTAAAAAGATGGACTCCCCCGGCACACTGCAG ACTAACCCACCGCTAAAGCTGCACCCTGATCGCAGCGCCGGGAC GTCCATTTTCGTCCCCGAACAAGGAGGTTACA |
| TRAF3 | 2 | 2 | chr8 | 70761918 | 70762038 | AAGAAAGGTTTGTGAAGACCGTGGAGGACAAGTACAAGTGTGAG AAGTGCCGCCTGGTGCTGTGTAACCCGAAGCAGACTGAGTGTGG ACACCGGTTCTGTGAGACCTGCATGGCTGCCG |
| TRAF3 | 3 | 1 | chr8 | 70763943 | 70764063 | CGTGGCCGCAGCTCCAGTGAACACAGGACTTCTGCTGTCAGTAG AACCCCACTTACTGATGTCTTTATTGTGTGTTTTCCCCTCAGCTCC TCAAGCCCAAAATGCACTGCGTGTCAAGAA |
| TRAF3 | 3 | 2 | chr8 | 70764003 | 70764123 | GTCTTTATTGTGTGTTTTCCCCTCAGCTCCTCAAGCCCAAAATGC ACTGCGTGTCAAGAAAGCATCATTAAAGATAAGGTATTCTTGGGG TTTTAATAAATGACCTTGTCCAATGCTTCA |
| TRAF3 | 3 | 3 | chr8 | 70764063 | 70764183 | AGCATCATTAAAGATAAGGTATTCTTGGGGTTTTAATAAATGACCT TGTCCAATGCTTCAGTCATCCGTGGCACACGGTTTGCACCCAGT GCTCTGAGCCGGCTCTGATGGTATCTCGGG |
| TRAF3 | 4 | 1 | chr8 | 70766844 | 70766964 | TGTGGAGATTAAGAAACATTAATCCTCTGAAATAACTGTCTTCCCA TTTCAGGTGTTTAAGGATAATTGCTGCAAGAGAGAAATTCTAGCT CTTCAGATCTATTGCAGGAATGAAGGTGG |
| TRAF3 | 4 | 2 | chr8 | 70766964 | 70767084 | AGGCTGCACAGAGCAGTTGACACTGGGACACCTACTGGTACGTC CTGGGAACCATGCTCCTTGCTTACCATAAGAAATATGACTCCTTG GGGTGCCTGGGTGGCTTAGTCAGTTTAGTGT |
| TRAF3 | 5 | 1 | chr8 | 70767830 | 70767950 | AGCTTTGAAGTGTTTGTTTTACAGGTGCATTTAAAGAACGATTGC CAGTTTGAAGAGCTTTCGTGTGTCCGTGCTGACTGCAAAGAAAAA GTGTTGCGAAGAGACCTGCGTGACCATGTA |
| TRAF3 | 5 | 2 | chr8 | 70767950 | 70768070 | GAAAAGGCTTGCAAGTACCGAGAGGCCACATGCCCCCACTGTAA AAGTCAGGTCCCGATGATCACACTGCAGGTGCGCTCCCCCCCCC CTCCCAGTCACCCCTACTTGCTGCTCAGTCCA |
| TRAF3 | 6 | 1 | chr8 | 70773932 | 70774052 | TCCTTTGGGTTTTCGTGGTGACATTGAAGATACTGATAGAGCCTG AAGGGAGATGGGTTTTATGTCCCAAAGTTGATGACACTTTCTTCT TTTCCTCCAGAAACATGAAGACACTGAGTG |
| TRAF3 | 6 | 2 | chr8 | 70773992 | 70774112 | TATGTCCCAAAGTTGATGACACTTTCTTCTTTTCCTCCAGAAACAT GAAGACACTGAGTGTCCCTGTGTGGTGGTGTCCTGCCCCCATAA GTGCAGTGTCCAGACCCTTCTAAGGAGTGA |
| TRAF3 | 6 | 3 | chr8 | 70774052 | 70774172 | TCCCTGTGTGGTGGTGTCCTGCCCCCATAAGTGCAGTGTCCAGA CCCTTCTAAGGAGTGAGGTAAGGAGCATTGCCAGGGCCTGTTGA GGAGTTAGTGATGTCTTTAGGACTAGCATGCA |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| TRAF3 | 7 | 1 | chr8 | 70776715 | 70776835 | GTGACAGTACAAACCAGATGCTCTCTGTGCCCTTAATATGTTTGA ACATTTTCATGCAGACGCTTGTGCCACAACTCACGTCTCTTTCCC GTTGCAGTTGAGTGCACACTTGTCAGAGTG |
| TRAF3 | 7 | 2 | chr8 | 70776775 | 70776895 | CGCTTGTGCCACAACTCACGTCTCTTTCCCGTTGCAGTTGAGTGC ACACTTGTCAGAGTGTGTCAATGCCCCCAGCACCTGTAGTTTTAA GCGCTATGGCTGCGTTTTTCAGGTCAGTAT |
| TRAF3 | 7 | 3 | chr8 | 70776835 | 70776955 | TGTCAATGCCCCCAGCACCTGTAGTTTTAAGCGCTATGGCTGCGT TTTTCAGGTCAGTATCCAACATTTGTCCTTCCCAGTCACTGACATT CTGCCATGAGAGAAAGTTACTATCTTAAC |
| TRAF3 | 8 | 1 | chr8 | 70778498 | 70778618 | TGTAATGGAATGAGGTCAGCTTAATGCACTACATCTCTTCTCGCA GGGGACAAACCAGCAGATAAAGGCCCATGAGGCCAGCTCCGCA GTGCAGCACGTCAACTTACTGAAAGAGTGGAG |
| TRAF3 | 8 | 2 | chr8 | 70778618 | 70778738 | TAACTCTCTGGAGAAAAAGGTACACCATGCTGCTTCCTGCTTTTG TTGGCTCTTAATAATTTCTTCAAATTATTTAAAGAAGAATATATTTT TTATAGTTACTAGTGGATTTGGATAAAA |
| TRAF3 | 9 | 1 | chr8 | 70782904 | 70783024 | TATTTTCCAGGTTTCCTTGCTACAGAATGAAAGCGTAGAAAAAAA CAAGAGCATACAAAGTTTGCACAATCAGATATGTAGCTTTGAAATT GAAATTGAGAGACAAAAAGAAATGCTTCG |
| TRAF3 | 9 | 2 | chr8 | 70783024 | 70783144 | AAATAATGAATCCAAAATACTTCATTTGCAGGTAAGAAGTTTAAGA CTATGGCTGGATCAAAGGGTCGAGGAGCATGTCTGTGTTCTCCT TTTGGGTATTCTCAGTGAGCTGTTTGAGCA |
| TRAF3 | 10 | 1 | chr8 | 70787980 | 70788100 | GCCTTTCTGATGCCTTTGATCTGGAAGCGAGTAATAGACAGCCAG GCAGAGAAACTGAAAGAGCTGGACAAAGAGATCCGCCCCTTCCG GCAGAACTGGGAAGAAGCAGACAGCATGAAG |
| TRAF3 | 10 | 2 | chr8 | 70788100 | 70788220 | AGCAGCGTGGAGTCCCTCCAGAACCGAGTGACCGAGCTGGAGA GCGTGGACAAGAGTGCAGGGCAGGCGGCTCGGAACACAGGTGG GGGCTGCAGGGGCAGCGTGGGGAGGCCTAGGAGG |
| TRAF3 | 11 | 1 | chr8 | 70789027 | 70789147 | GACACTTACGGGCCCTCCACCGTCTGTCTCTCCTCACCCACGGC AGGCTTGCTGGAGTCCCAGCTGAGCCGGCACGACCAGATGCTG AGCGTCCATGACATCCGCCTGGCCGACATGGAC |
| TRAF3 | 11 | 2 | chr8 | 70789147 | 70789267 | CTGCGCTTCCAGGTCCTGGAGACCGCCAGCTACAACGGCGTGCT GATTTGGAAGATCCGAGATTACAAGCGGCGGAAGCAGGAAGCCG TCATGGGGAAGACCCTGTCTCTTTACAGCCAG |
| TRAF3 | 11 | 3 | chr8 | 70789267 | 70789387 | CCTTTCTACACGGGCTATTTTGGCTATAAGATGTGTGCCAGGGTC TACCTGAACGGGGACGGCATGGGGAAGGGGACGCACTTGTCGC TGTTTTTTGTCATTATGCGTGGAGAGTATGAT |
| TRAF3 | 11 | 4 | chr8 | 70789387 | 70789507 | GCTCTGCTTCCTTGGCCATTCAAGCAGAAAGTGACGCTCATGCTC ATGGATCAGGGGTCCTCTCGGCGTCACCTGGGAGATGCGTTCAA GCCGGACCCCAACAGCAGCAGCTTCAAGAAG |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| TRAF3 | 11 | 5 | chr8 | 70789507 | 70789627 | CCCACCGGGGAGATGAATATTGCCTCTGGCTGCCCAGTGTTCGT GGCTCAAACTGTGCTAGAAAACGGGACATATATTAAAGATGATAC CATTTTTATTAAAGTCATAGTGGATACTTCG |
| TRAF3 | 11 | 6 | chr8 | 70789627 | 70789747 | GATCTGCCCGACCCCTGACCGGTCACTGGGGAGGTGGATTGAG CAGAAGGCATCTCCTCTGGGGGGTTTGAACCGGCTGTCTCCACC GAGGTCCTCGCGCTCAGAAAAGGACCTTGTGGA |
| TP63 | 1 | 1 | chr5 | 32565055 | 32565175 | CAGCACCCACACTCCCTTCCCAGGAGCTGCGATGCAAGAGCCAC AGTCAGAGCTCAATATCGACCCCCCTCTGAGCCAGGAGACATTTT CAGAATTGTGGAACCTGTAAGTGGAGGGCAG |
| TP62 | 2 | 1 | chr5 | 32564808 | 32564928 | CATCTTCCAGGCTTCCTGAAAACAATGTTCTGGTAAGGACTGGGT GTGGGGAGGTCCAGGGCGGGGGTTGGGAGCTGGGGACCTAGG GACCTGACCCTTGACTCTGGTCTCGCCCCCTAC |
| TP61 | 3 | 1 | chr5 | 32564689 | 32564809 | CAGTCTTCGGAGCTGTGCCCAGCAGTGGATGAGCTGCTGCTCCC AGAGAGCGTCGTGAACTGGCTAGACGAAGACTCAGATGATGCTC CCAGGATGCCAGCCACTTCTGCCCCCACAGCC |
| TP60 | 3 | 2 | chr5 | 32564567 | 32564687 | TGGACCGGCCCCCTCCTGGCCCCTATCATCCTCTGTCCCTTCCC CGAAGACCTACCCTGGCACCTATGGGTTCCGTTTGGGGTTCCTG CATTCCGGGACAGCCAAGTCTGTTACTTGGAC |
| TP59 | 4 | 1 | chr5 | 32563905 | 32564025 | AAGACCTGCCCCGTGCAGCTGTGGGTCAGCTCCCCACCCCCAC CCAATACCTGCGTCCGCGCTATGGCCATCTATAAGAAGTCGGAG TTCGTGACCGAGGTTGTGCGGCGCTGCCCCCAC |
| TP58 | 5 | 1 | chr5 | 32563683 | 32563803 | CTCTTAGGTCTTGCCCCTCCTCAGCATCTCATCCGAGTGGAAGG AAATTTGCGGGCCAAGTACCTGGACGACAGAAACACTTTTCGACA CAGTGTGGTGGTGCCTTATGAGCCACCCGAG |
| TP57 | 6 | 1 | chr5 | 32563351 | 32563471 | TCTACCTCAGGTTGGCTCTGACTATACCACCATCCACTACAACTA CATGTGTAACAGTTCCTGCATGGGAGGCATGAACCGGCGGCCCA TCCTCACTATCATCACCCTGGAAGACTCCAG |
| TP56 | 7 | 1 | chr5 | 32562962 | 32563082 | TGCTGGGACGCAACAGCTTTGAGGTACGCGTTTGTGCCTGTCCC GGGAGAGACCGCCGGACTGAGGAGGAGAATTTCCACAAGAAGG GGGAGCCTTGTCCTGAGCCACCCCCCGGGAGTA |
| TP55 | 8 | 1 | chr5 | 32562572 | 32562692 | TGTCTTTTCCTCCTCTCTTTCCCAGCACTGCCTCCCAGCACCAGC TCCTCTCCCCCGCAAAAGAAGAAGCCACTAGATGGAGAATATTTC ACCCTTCAGGTACCAAGGCTGGGAGGAACA |
| TP54 | 9 | 1 | chr5 | 32562105 | 32562225 | ACTGCCACAGATCCGTGGGCGTGAACGCTATGAGATGTTCAGGA ATCTGAATGAAGCCTTGGAGCTGAAGGATGCCCAGAGTGGAAAG GAGCCAGGGGGAAGCAGGGCTCACTCCAGGTG |
| TP53 | 10 | 1 | chr5 | 32561387 | 32561507 | CCACCTTGTGTCCCCAACAGCCACCTGAAGGCAAAGAAGGGGCA ATCTACCTCTCGCCATAAAAAACTGATGTTCAAGAGAGAAGGGCC CGACTCAGACTGACTTCCCCTGCTTCCTGTC |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 1 | 1 | chr20 | 41675503 | 41675623 | GCGGCCTCCTCCCCTCCCCTCCCCTCCCCTCCCCTCTTGCGCGC GTCGCCCTGCCGCGGGGAGGGGGCTCGCGTCGCCGCCTCCAG CCGCTCCCGATGAAGCAGCTGCCGCCGCAGCCAC |
| SETD2 | 1 | 2 | chr20 | 41675563 | 41675683 | GAGGGGGCTCGCGTCGCCGCCTCCAGCCGCTCCCGATGAAGCA GCTGCCGCCGCAGCCACCTCCGAAAATGGGGGATTTCTACGACC CCGAGCACCCGACCCCTGAGTAAGTATCTGCTC |
| SETD2 | 1 | 3 | chr20 | 41675623 | 41675743 | CTCCGAAAATGGGGGATTTCTACGACCCCGAGCACCCGACCCCT GAGTAAGTATCTGCTCCGCGGCTCCCCGCGGCCGCCCTCGGCC CGAGCCCGGAGCCCCGAGGGTGGCGCGTCCTGA |
| SETD2 | 2 | 1 | chr20 | 41707723 | 41707843 | TTGTAGGCATCCTAGATTTTAATAAAAACAACTTTGATTTCAATTG GTTTTAAGTATTTATGATTTCTGTTTACTCTTTTGCATGCTTTCTTG GCTGGATGGCAATAGAGAAGAAGAAAA |
| SETD2 | 2 | 2 | chr20 | 41707783 | 41707903 | TGATTTCTGTTTACTCTTTTGCATGCTTTCTTGGCTGGATGGCAAT AGAGAAGAAGAAAATGAGGTAAGTATCACTTAACTAGTTTTTCAG TAATTATTGGACTAAAAACTCAAAATTTT |
| SETD2 | 2 | 3 | chr20 | 41707843 | 41707963 | TGAGGTAAGTATCACTTAACTAGTTTTTCAGTAATTATTGGACTAA AAACTCAAAATTTTGGGGGTGGGTGGCTCAGTTGGTTAAGTGTCC AGCACTTGATTTTGGCTCAGGTCATGATC |
| SETD2 | 3 | 1 | chr20 | 41709777 | 41709897 | CTTGTTTTGTTCATTTATAGGCAAAGATTGAAAATGTGCAGAAAAC AGGTTTCATCAAAGGACCAATGTTCAAAGGTGTTGCTTCTAGTCG ATTTTTACCCAAAGGCACCAAGACAAAAG |
| SETD2 | 3 | 2 | chr20 | 41709897 | 41710017 | TTAACTTGGAGGAACAGGGACGACAGAAGGTGTCTTTCAGCTTC AGCCTTACAAAGAAAACTTTGCAGAACAGGTTTCTGACTGCACTT GGCAATGAAAAGCAAAATGATACTCCAAACT |
| SETD2 | 3 | 3 | chr20 | 41710017 | 41710137 | CCCCCGCTGGACCTCTTCAAGTAGACTTGACCCCTAAAATTAAAA TGGACATTGGAGATACCTTATCTACTACAGAAGAATCTTCCCCAC CAAAATCAAGGGTAGAATTGGGCAAAATTC |
| SETD2 | 3 | 4 | chr20 | 41710137 | 41710257 | ATTTTAAGAAACATCTGCTTCATGTGACATCCAGACCACTGCTGA CTACTGCCACAGCAGTGGCATCTCCATCACCTCCCGTAGTACCG TTACCAGCAGTCATAGCAGAATCAACAACTG |
| SETD2 | 3 | 5 | chr20 | 41710257 | 41710377 | TAGACTCACCACCCTCATCTCCACCTCCACCACCTCCACCTCCCC AGGCCACAATACCCTCACCACCAGCACCAGTAACAGAGCCAGTG GCCTTGCCACATGCATCAATCCCAGTTCTGA |
| SETD2 | 3 | 6 | chr20 | 41710377 | 41710497 | TGACAGCACCAGTAGATGTAGCAGCTAGAGCACTGAAGGAACCA CCCATTACAACTGTACCAGAATCTTCAGAAGTGGACACTAAGCAG GATGCTATATCTAATAGTTCAGAAGAACACA |
| SETD2 | 3 | 7 | chr20 | 41710497 | 41710617 | TAACTCAAAATTTGAATGAGCAAGTAGATACTCCCTCACAAAAAG AAGATTCCCATATTGGGAAGGAAGAAGAAATTCCAGATAGTTCTA AGAGTAATCTGGGCTCTAAAAAAACAGGTT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 3 | 8 | chr20 | 41710617 | 41710737 | CTAAGAAGAAATCCTCACAATCTGAAGGCACCTTTCTTGCTTCAG AATCTGATGAAGATTCTGTACGGACTTCCTCAAGTCAAAGATCAC ATGATTTAAAATTTTCAACAGGCATTGAAA |
| SETD2 | 3 | 9 | chr20 | 41710737 | 41710857 | AGGAAAGAGATTCAAAAAAGAGCTTAGCACCTTTAAAAAGTGAGG ATTTAGGGAAATCTTCACGATCTAAAACAGAGAGAGATGATAAAT ATTTTAGCTACTCAAAACTTGAAAGAGATA |
| SETD2 | 3 | 10 | chr20 | 41710857 | 41710977 | CTCGGTATATATCTTCTCGATGTAGATCAGAGAGAGAGCGAAGG CGGAGCAGATCTCGTTCTAGATCTGACAGAGGCTCCAGAACTAG TTTATCCTATTCCCGGTCGGAACGATCCCATT |
| SETD2 | 3 | 11 | chr20 | 41710977 | 41711097 | ATTACGACTCTGATCGTCGCTACCATAGGAGTTCACCTTATCGAG AGAGGGCACGCTATTCCCGGCCGTATACAGATAACAGAGCTCGA GAGAGTTCTGACTCAGAAGAGGAGTATAAGA |
| SETD2 | 3 | 12 | chr20 | 41711097 | 41711217 | AGACCTATTCAAGGCGTACCTCCTCTCATTCTTTCTCTTACAGAG ACCTAAGGACATCATCATCCTATTCTAAATCTGATCGGGACTGCA AAACTGAGTCCTCTTACTTGGAGATGGAGA |
| SETD2 | 3 | 13 | chr20 | 41711217 | 41711337 | AGAGAGGAAAGTATTCTTCTAAACTAGAAAGAGAATCCAAAAGGA CTTCAGAAAGTGAAGCAATGAAAAGATGTTGTACTCCCCCTAATG AACTGGGATTCCGACGGGGGTCGTCCTATT |
| SETD2 | 3 | 14 | chr20 | 41711337 | 41711457 | CCAAGCATGATAACAGTGCTTCCCGTTATAAATCTGCCCTTTCAA AATCTATATCCAAGTCTGATAAATTTAAAAATTCTTTCTGTTGTACA GAATTAAATGAGGAAATCAAACCGTCTC |
| SETD2 | 3 | 15 | chr20 | 41711457 | 41711577 | ATTCTTTTAATTTACAGGCTCCTTGTTCAAAAGGTAGTGAATTAAG AATGATTAGTAAAATTCCTGAAAGAGAAAAGACTGGATCTCCATC TCCATCAAATCGATTAAATGATTCACCTA |
| SETD2 | 3 | 16 | chr20 | 41711577 | 41711697 | CTTTTAAAAAGCTAGATGAATCCCTTATTTTTAAATCTGAATTTATA GGACATGATAGCCATGATAGTATTAAGGAATTAGACTCTTTATCTA AAGCAAAGAATGATCAATTAAGAAGTT |
| SETD2 | 3 | 17 | chr20 | 41711697 | 41711817 | ATTGTCCCATAGAATTAAATATAAATGGATCTCCTGGGGCAGAAT CTGATTTGGCAACATTTTGCACTTCTAAAACTGATACTGTTTTGAT GTCTTCTGATGACAGTGTGACTGGATCGG |
| SETD2 | 3 | 18 | chr20 | 41711817 | 41711937 | AGGTATCCCCTTTGGTCAAAGCATGCATGCTTCCATCAAATGGAT TTCAGAATATCAGTAGATGTAAAGAGAAAGACTTGGATGATACTT GCATGCTGCATAGTAAGTCAGGAAGCCCAT |
| SETD2 | 3 | 19 | chr20 | 41711937 | 41712057 | TTAGAGAAGCAGAACCTCCGGTATCGCCACACCAAGATAAACTCA TGTCTTTGCCAGTTATGACTATAGATTATTCCAAAACCGTAATTAA AGAACCAGTTGATATGACAGTTTCTTGCT |
| SETD2 | 3 | 20 | chr20 | 41712057 | 41712177 | GTAAAACCAAAGATTCAGATATATACTGTACTTCAAATGACAACCC TTCTTTGCATCATTCTGAAGCTGAAAAGATTGAGCCTTCAGTTATG AAGATTTCTTCAAATAGCTTTATGAATG |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 3 | 21 | chr20 | 41712177 | 41712297 | TGCATTTGAAATCAAAAACAGTAATATGTGATAACAGAACTCTGAC AGATCAGCACTCAAAATTTGCATGTGAAGAATATAAACAGAGTGT TGGTAGCACTAGTTCAGCTGCTGTTAATC |
| SETD2 | 3 | 22 | chr20 | 41712297 | 41712417 | ATTTCGATAATTTATATCAACCTATAGAGAGTTCAGGCATTGCTTC ATCTCTTCAGAGTCTTCCACCAGGAATAAAAGTGGACAGTTTGAC TCTCCTGCAATGTGGAGAGAACACCTCTC |
| SETD2 | 3 | 23 | chr20 | 41712417 | 41712537 | CAGTTTTGGATGCTGTGCTGAAGAGTAAAAGTTCAGAGTTTTTAA AGCTTGCAGAGAAAGAAACAGTAGAAGTAGGTAGCGGCCTTCCT GATTCAGGAAGGGGATTTGCTTCTTGGGAAA |
| SETD2 | 3 | 24 | chr20 | 41712537 | 41712657 | ATAGGCATAATAATGGGTTATCTGGGAAAGGTGTGCAAGAGGTTC AAGAAGAAGGGAATTCCTTATTACCTGATAGAAGAGGAAGATCAG AAATCTCTTTAGATGAAGAAGGAGGAAGAG |
| SETD2 | 3 | 25 | chr20 | 41712657 | 41712777 | GACATGCACATACTTCTGATGATTCAGAAGTTGTATTTTCTTCTTG TGACTTGAATTTAACCATGGAAGACAGCGATGGTGTAACATATAC CTTAAAATGTGACAGTAGTGGTCATGCCT |
| SETD2 | 3 | 26 | chr20 | 41712777 | 41712897 | CAGAGATTGTATCTACTGTCCATGAAGATTATTCTGGTTCTTCTGA AAGTTCAAGTGATGAAAGTGACTCTGAAGATACAGATTCTGATGA TAGCAGTATTCCAAGAAACCGTCTTCAGT |
| SETD2 | 3 | 27 | chr20 | 41712897 | 41713017 | CTGTTGTGGTAGTGCCAAAGAATTCTACTTTGCCCATGGAAGAAA CGAGTCCTTGTTCTTCTCGGAGCAGTCAGAGTTACAGACACTATT CTGACCACTGGGAAGATGAGAGATTGGAGT |
| SETD2 | 3 | 28 | chr20 | 41713017 | 41713137 | CAAGGAGACATTCATATGAGGAAAAATTTGAGAGTGTAGCAAGTA AACCCTGTTCTCAAACCGAGAAGTTCTTCCATCATAAAGGGACAG AGAAAAATCTGGAAGTTTCTTTCACACAGC |
| SETD2 | 3 | 29 | chr20 | 41713137 | 41713257 | CCAGCAGAAAACAAATAGATAACCACCTGTCTGAAATGGCTCATC CTCAGAGTGATGGGGTTGATAGTACAAGTCATACAGATGTGAAAT CTGACTTTCTAGGTCATCCAAATTCTGAGG |
| SETD2 | 3 | 30 | chr20 | 41713257 | 41713377 | AAACCACAAAAGCCAAAATAGTTTCTAGGCAGCAAGAAGAGCTGC CAGTTTATTCTTCTGATGATTTTGAGGATGTCCCAAATAAGTCTCG GCAACAAATCACTTTCCCTAACAGGCCAG |
| SETD2 | 3 | 31 | chr20 | 41713377 | 41713497 | ATAGTAGACTGGGAAAAACAGAGTTGAGTTTTTCTTCCTCTTGTG AGCTCTCCCGAGTGGATGGTTTCCACTCATCAGAAGAGCTCAGA AACCTAGGGTGGGAATTCTCTCAACAAGAAA |
| SETD2 | 3 | 32 | chr20 | 41713497 | 41713617 | AGCCTACTACCACATATCAGCAACCTGACAGCAGCTATGGAGCC TGCGGTGGACACAAGTATCAGCAAAGTGCAGAACATTATGGCGG GACACGGAATTACTGGCAAGGCAATGGCTACT |
| SETD2 | 3 | 33 | chr20 | 41713617 | 41713737 | GGGATCCAAGATCAGCAAGTAGACCTCCTGGAACTGGGGTTGTA TATGATCGAATTCAAGGGCAGGTACCAGATTCCTTAACAGATGAT CGTGAAGAGGAGGAGAATTGGGATCAACGTG |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 3 | 34 | chr20 | 41713737 | 41713857 | GAGGATCTCATTTTTCAAGCCAGTCCAATAAATTTTTTCTATCCCT TCAGAAGGACAAGGGGTCAGTGCAAGCACCTGAAATAAGCAGCA ATTCCATTAAGGACTCTTTATCTATGAACG |
| SETD2 | 3 | 35 | chr20 | 41713857 | 41713977 | AAAAGAAGGATCTTTCAAAAAACTTAGAAAAAAATGATATGAAAGA TAGAGGGCCTCCTAAAAAAAGGAGACAGGAATTGGAGAGTGATT CTGAAAGTGATGGTGAACTTCAGGACAGAA |
| SETD2 | 3 | 36 | chr20 | 41713977 | 41714097 | AGAAAGTTCGAGTGGCGGTGGAGCAGGGGGAAACGGCAGTGCC CCTAGGCTCAGCACTGGTTGGGCCTTCTTGTGTCATGGAGGACT TCAGGGACCCACAGCGATGGAAAGAATGTGCCA |
| SETD2 | 3 | 37 | chr20 | 41714097 | 41714217 | AGCAAGGGAAGATGCCTTGTTACTTTGATCTGATTGAAGAAAATG TTTATTTAACAGAAAGGTAAGTTTGCCTAATACTTTTCCAACAAAT TTTAACCACTTTTTGTTAAGAAAATTAGC |
| SETD2 | 3 | 38 | chr20 | 41714217 | 41714337 | TTTCAACAGTTTGGAATATATGAAGTCAATAGTTCATATTAGTTTT GTGATAAGATATACATTAAAAAAAATCTAAAAAGTATTTACCTATTT AGATCATGTCATGGTGTCCTTCCCTCG |
| SETD2 | 4 | 1 | chr20 | 41717650 | 41717770 | TACTAGAAAGAAGAATAAATCCCATCGGGATATTAAGCGAATGCA GTGTGAGTGTACACCTCTTTCTAAAGATGAAAGAGCTCAAGGTGA AATAGCATGTGGGGAAGATTGTCTTAATCG |
| SETD2 | 4 | 2 | chr20 | 41717770 | 41717890 | TCTCCTCATGATTGAATGGTAAGTAAATTAAAATGCTCTCATTTTG CTCAATTGTTCTATTAAATGTCTCAAAATAGTAAAAAGAATTAGTA ATGGGTCAGAATTTAACTTTTCTCCTTC |
| SETD2 | 5 | 1 | chr20 | 41719089 | 41719209 | TCAGTTCTTCTCGGTGTCCAAATGGGGATTATTGTTCCAATAGAC GATTTCAGAGAAAGCAGCACGCAGATGTGGAAGTCATACTCACA GAAAAGAAAGGCTGGGGCTTGAGAGCTGCCA |
| SETD2 | 5 | 2 | chr20 | 41719209 | 41719329 | AAGACCTTCCTTCGTAAGTTATGTTTTAATCATCTTTCACTTCATTT TTGTGTTGTTGAAGTTGGTTAGCTACATAGGAAAATAATGTGTCT GTTATACTGAAGTTTAATAATGATGATT |
| SETD2 | 6 | 1 | chr20 | 41720885 | 41721005 | TGTGTTTTCTATATAGCTCCTAGGCAGGAAGTGCTATTGTTTCCCT TTTGTTTTGCTAGGAACACCTTCGTCCTAGAATATTGTGGAGAAG TACTTGATCATAAAGAGTTTAAAGCCCGG |
| SETD2 | 6 | 2 | chr20 | 41720945 | 41721065 | AACACCTTCGTCCTAGAATATTGTGGAGAAGTACTTGATCATAAA GAGTTTAAAGCCCGGGTAAAGGAGTATGCACGAAACAAAAACAT CCACTACTATTTCATGGCCCTGAAGAATGAT |
| SETD2 | 6 | 3 | chr20 | 41721005 | 41721125 | GTAAAGGAGTATGCACGAAACAAAAACATCCACTACTATTTCATG GCCCTGAAGAATGATGAGGTGAGCAGTGTGAATGTGTTTGCTTT GAAACAGATCTGGAGGGATCCCTGGGTGGCG |
| SETD2 | 7 | 1 | chr20 | 41722064 | 41722184 | GAGGATGTAGGTCATACATTTGCCTAATGCTTCTCCTTTGCAACA CATACATGTATTTTGAGATCTCATTTTTAATTCTTAAATGGTCCTTT TCCCCAGATAATAGATGCCACCCAGAAA |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 7 | 2 | chr20 | 41722124 | 41722244 | AGATCTCATTTTTAATTCTTAAATGGTCCTTTTCCCCAGATAATAG ATGCCACCCAGAAAGGAAACTGCTCTCGTTTCATGAATCACAGCT GTGAACCAAACTGTGAGACTCAAAAAGTA |
| SETD2 | 7 | 3 | chr20 | 41722184 | 41722304 | GGAAACTGCTCTCGTTTCATGAATCACAGCTGTGAACCAAACTGT GAGACTCAAAAAGTAAGTTGAGGTACATTTAGAGTTTGGGATACC TTGTTCAAAATTGCTGTTCTTTTAATTTTA |
| SETD2 | 8 | 1 | chr20 | 41723805 | 41723925 | ATGATGATCTTTAAAGAAAAACAGATTAATCAGTTACTAATTTCTTT TCTTTTAAAAACAACATTTCAGTGGACTGTGAACGGACAACTGAG AGTCGGGTTTTTTACCACCAAACTGGTT |
| SETD2 | 8 | 2 | chr20 | 41723865 | 41723985 | ACATTTCAGTGGACTGTGAACGGACAACTGAGAGTCGGGTTTTTT ACCACCAAACTGGTTCCTTCAGGCTCAGAGTTAACATTTGACTAT CAGTTCCAAAGATATGGGTGAGTTTTGTTT |
| SETD2 | 8 | 3 | chr20 | 41723925 | 41724045 | CCTTCAGGCTCAGAGTTAACATTTGACTATCAGTTCCAAAGATAT GGGTGAGTTTTGTTTTTATTCTGTTTTGCTGCTTCTAATGGGGAAA TGGACTGTTCAGATATAAAGATTTAAGAC |
| SETD2 | 9 | 1 | chr20 | 41726907 | 41727027 | CAGAAAAGAAGCTCAGAAATGTTTCTGTGGGTCAGCTAACTGCC GAGGTTACCTGGGAGGAGAGAACAGAGTCAGTATCAGAGCAGCA GGAGGAAAAATGAAGAAGGAGCGATCCCGTAA |
| SETD2 | 9 | 2 | chr20 | 41727027 | 41727147 | GAAGGATTCAGTAAGTGTTTCATCTTTTCTTTCAACTTTTACCATA AATGTTTTAAGAAGGGAAGTAGCTTTTCTGACTGAAATTTGAAAAG TCAGACTACAGCCAGAGAAAGTCAGGAT |
| SETD2 | 10 | 1 | chr20 | 41735430 | 41735550 | ATGGCAGGTAGATGGGGAGCTAGAAGCTCTGATGGAGAATGGTG AGGGTCTCTCTGATAAAAACCAGGTGCTCAGCTTATCCCGGCTAA TGGTTAGAATTGAAACTTTGGAACAGAAACT |
| SETD2 | 10 | 2 | chr20 | 41735550 | 41735670 | TACGTGTCTTGAACTTATACAGGTGAGCTCTTTGATGTTAATTTAT TTTTGATTTCTGAATTTGAGGTGGGGCATTGAAAGTACAATCTAG AATGATGATGACAGAAGGTTACTTCTGAG |
| SETD2 | 11 | 1 | chr20 | 41736415 | 41736535 | ATTGCCCACCCGACCCCCACCCCGATATCCTGCATCCTGCTTCC CTTCCTGCATGGAAGAATACACATTCACAATCTTGCCTGAAGTCT TTTCTGGAACGTCATGGGCTCTCATTGTTAT |
| SETD2 | 11 | 2 | chr20 | 41736475 | 41736595 | ATACACATTCACAATCTTGCCTGAAGTCTTTTCTGGAACGTCATG GGCTCTCATTGTTATGGATCTGGATGGCAGAACTAGGTGATGGC CGGGAAAGTAACCAGAAGCTTCAGGAAGAGG |
| SETD2 | 11 | 3 | chr20 | 41736535 | 41736655 | GGATCTGGATGGCAGAACTAGGTGATGGCCGGGAAAGTAACCAG AAGCTTCAGGAAGAGGTTAGTCAGTTCATATTCAACTTGCTTATTT CCTAAAACTGCTATATAGATTGCTTAGGTT |
| SETD2 | 12 | 1 | chr20 | 41737881 | 41738001 | GTGATTATTTTTTTTTATTTTGTTTAATCAGATTATAAAGACTTTGG AACACTTACCCATTCCTACTAAAAATATGTTGGAAGAAAGCAAAGT ACTTCCAATTATTCAACGCTGGTCTCA |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 12 | 2 | chr20 | 41738001 | 41738121 | GACCAAAACTGCCATTCCTCAGTTGAGTGAAGGAGATGGGTATTC AAGTGAGAATACGTCACGTGCTCACACACCACTCAACACACCTGA TCCTTCCACCAAGCTGAGCACAGAAGCTGA |
| SETD2 | 12 | 3 | chr20 | 41738121 | 41738241 | CACAGACACCCCCAAGAAGCTCATGTTTCGCAGACTTAAAATTAT AAGTGAAAATAGCATGGACAGTGCAATCTCTGATGCTACCAGTGA GCTGGAAGGCAAAGATGGCAAGGAGGACCT |
| SETD2 | 12 | 4 | chr20 | 41738241 | 41738361 | TGACCAGTTAGAAAACGTCCCTATAGAGGAAGAGGAAGAGTTAC AGTCACAACAGCTACTCACACAACAGCTACCTGAATCTAAAGTTG AGAGTGAAATCACTGTGGAAGCCAGTAAGCT |
| SETD2 | 12 | 5 | chr20 | 41738361 | 41738481 | ACCCACAACTGAACCAGAGGCTGACACTGAAATAGAGCCCAAAG AGGGCAATGGCACGAAACTGGAAGAAACCATTGCTGAGGAAACC CCATCCCAAGATGAAGAGGAGGGTGTATCTGA |
| SETD2 | 12 | 6 | chr20 | 41738481 | 41738601 | TGTGGAGAGTGAGAGGAGCCAGGAACAGCCAGATAAAACAGTAG ATATAAGTGATTTGGCCACCAAGCTCCTGGATAGTTGGAAAGACC TAAAGGTAAGAAAACATTTTTGTTTTAACCT |
| SETD2 | 13 | 1 | chr20 | 41759911 | 41760031 | GTTGCCATGTGAAATTAAGCAGCTTCAACACAATCTCTGCTTCTA CTCCACACCAATAACCCTTCTCTTGGTTTATTTTTCCAGGAGGTGT ATCGAATTCCAAAGAAAAGTCAAACTGAA |
| SETD2 | 13 | 2 | chr20 | 41759971 | 41760091 | CCTTCTCTTGGTTTATTTTTCCAGGAGGTGTATCGAATTCCAAAGA AAAGTCAAACTGAAAAGGAGAACACAAGTAGGTATTCAGAGGTAC TCTGTGTATATGGCTTTACAAATACTCTT |
| SETD2 | 13 | 3 | chr20 | 41760031 | 41760151 | AAGGAGAACACAAGTAGGTATTCAGAGGTACTCTGTGTATATGGC TTTACAAATACTCTTTTGGAAATATGCAGAGGTAAATGATATGAAC TTTAGCAGACAGTATGTAATTAGGTAAAT |
| SETD2 | 14 | 1 | chr20 | 41762861 | 41762981 | CAAGTAAATTTTTAAAAACCCTTTTCCCTTAGTAACTGAACGAGGA AGGGATGCTGTTGGCTTCAGAGATCAAACAGCTGCCCCAAAGAC TCCTAACAGGTCAAGAGAGAGAGACCCAGA |
| SETD2 | 14 | 2 | chr20 | 41762981 | 41763101 | CAAGCAAACTCAAAATAAAGAGAAAAGGAAACGAAGGGGCTCTC TCTCACCACCTTCTTCTGCCTATGAGAGGGGAACAAAAAGGCCA GATGACAGGTAAGTGGCCTGTGTGAGTTATCT |
| SETD2 | 15 | 1 | chr20 | 41768925 | 41769045 | GAACATTTGTGATGCTCTTCCTATTTTCTTTCTAGATATGATACAC CAACTTCTAAAAAGAAAGTACGAATTAAAGATCGCAATAAACTTTC TACAGAGGAGCGTCGGAAGTTGTTTGAG |
| SETD2 | 15 | 2 | chr20 | 41769045 | 41769165 | CAAGAGGTGGCTCAGCGGGAGGCTCAGAAACAACAGCAGCAGA TGCAGACCCTGGGAATGACGTCTCCACTGCCCTATGACTCTCTT GGCTACAATGCCCCTCATCACCCCTTCGCTGGC |
| SETD2 | 15 | 3 | chr20 | 41769165 | 41769285 | TACCCACCAGGTTATCCCATGCAGGCCTATGTGGATCCCAGCAA CCCTAATGCTGGAAAGGTGCTCCTCCCCACACCCAGCATGGACC CCGTGTGTTCTCCTGCTCCTTACGATCATTCT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 15 | 4 | chr20 | 41769285 | 41769405 | CAACCCTTGGTGGGACATTCTACAGAACCCCTTGCTGCCCCTCC ACCTGTGCCAGTGGTGCCACATGTGGCAGCCCCTGTGGAAGTTT CCAGTTCACAGTATGTGGCCCAAAATGATGGT |
| SETD2 | 15 | 5 | chr20 | 41769405 | 41769525 | GTGGTACACCAAGACTCCAGCGTCACTGTCTTGCCAGTGCCAGC CCCAGGCCCAGTCCAAGGACAAAATTATGGTGTTTGGGATTCAAA CCAACAGTCTGTCAGTGTCCAGCAGCAGTAT |
| SETD2 | 15 | 6 | chr20 | 41769525 | 41769645 | TCTCCTGCACAGTCTCAAGCAACCATATATTATCAAGGACAAACT TGTCCAACTGTCTATGGTGTGACATCACCCTATTCACAGACAACT CCACCAATTGTACAGGTAACTAATTTTGGT |
| SETD2 | 16 | 1 | chr20 | 41776205 | 41776325 | CTTTTAGAGTTATGCCCAGCCAAGTCTCCAGTACATCCAGGGACA ACAGATTTTCACAGCTCATCCACAAGGAGTGGTGGTCCAGCCAG CCACGGCCGTGACGACAATAGTCGCGCCAGG |
| SETD2 | 16 | 2 | chr20 | 41776325 | 41776445 | GCAGCCTCAGCCCTTACAGCCAGTAAGAAATTTCCGTTAAAATGT TCATTGGCTGTGTATCTCATCTTTAGAAGGTTACTTTTGGTTTTGG GCTTTTTTTTTTTTCCACTTTGTTTTTTA |
| SETD2 | 17 | 1 | chr20 | 41778557 | 41778677 | TTCTTTTTAGCCTGAAATGGTAGTGACAAATAATCTTCTGGATCTG CCACCCCCTTCTCCTCCCAAACCAAAAACCATTGTCTTACCTCCC AACTGGAAGACAGCCCGAGATCCAGAAGG |
| SETD2 | 17 | 2 | chr20 | 41778677 | 41778797 | GAAGATTTATTACTACCATGTGATCACAAGGTAAGAAGGGCTGTG TGGGCATCCTAAGTCAGCTTTCAGTTGGTCAAAGAAGCTGCCATC ATAAGGGGAGATTTCTGAATATTGGCTTGG |
| SETD2 | 18 | 1 | chr20 | 41780979 | 41781099 | CCTGTGCCAGAGTTCTCTGTTTTCTGGCCATGACTGTGTGGCTCT TCTTAACCTTCTTCTAGGCAGACTCAGTGGGATCCTCCTACTTGG GAAAGCCCAGGAGATGATGCCAGCCTTGAG |
| SETD2 | 18 | 2 | chr20 | 41781039 | 41781159 | AGGCAGACTCAGTGGGATCCTCCTACTTGGGAAAGCCCAGGAGA TGATGCCAGCCTTGAGCATGAAGCTGAGATGGACCTGGGAACCC CAACATATGATGAAAATCCCATGAAGGTGAGT |
| SETD2 | 18 | 3 | chr20 | 41781099 | 41781219 | CATGAAGCTGAGATGGACCTGGGAACCCCAACATATGATGAAAA TCCCATGAAGGTGAGTGTGGCTTACACGTATTTCCTGATCATTTG GGACTCTGTGCTCTGTCATTTTCCTGGGCCA |
| SETD2 | 19 | 1 | chr20 | 41797489 | 41797609 | GTGTCCCATCTGGGGGAAAGAAAGGAAGGTTTTGGAAGAGCAGG GCATCACTGACCTGACCTCTTATGCCAGCACTGTTTGTTTTCTTTT GCATTTCTAGACCTCAAAAAAGCCTAAGAC |
| SETD2 | 19 | 2 | chr20 | 41797549 | 41797669 | CTCTTATGCCAGCACTGTTTGTTTTCTTTTGCATTTCTAGACCTCA AAAAAGCCTAAGACAGCAGAAGCAGACACCTCCAGTGAACTGGC TAAGAAAAGCAAAGAAGTATTCAGAAAAGA |
| SETD2 | 19 | 3 | chr20 | 41797609 | 41797729 | AGCAGAAGCAGACACCTCCAGTGAACTGGCTAAGAAAAGCAAAG AAGTATTCAGAAAAGAGGTAAGGTTTGGTGGGTATGGGCCCTAA GCTCTGCAGGCTCAAGCCTAAGCCCTTTCTTT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| SETD2 | 20 | 1 | chr20 | 41799355 | 41799475 | CTCATTATTTCCTCAGTGCAATCCTTTGCCCTCTCCGGACTTCTG AGTAGTATGTGCTCTTTCTCCCACCTACAGATGTCCCAGTTCATC GTCCAGTGCCTGAACCCTTACCGGAAACCT |
| SETD2 | 20 | 2 | chr20 | 41799415 | 41799535 | TTCTCCCACCTACAGATGTCCCAGTTCATCGTCCAGTGCCTGAAC CCTTACCGGAAACCTGACTGCAAGGTGGGAAGAATCACCACAAC AGAAGACTTCAAACACCTGGCTCGCAAGGTA |
| SETD2 | 20 | 3 | chr20 | 41799475 | 41799595 | GACTGCAAGGTGGGAAGAATCACCACAACAGAAGACTTCAAACA CCTGGCTCGCAAGGTATCCCTCTGTTCTGAAGATGTTCTGGGCA AACCATGTACCTTTCGTCCTAGAAAGTGGGTT |
| SETD2 | 21 | 1 | chr20 | 41799802 | 41799922 | CTGGCCTTAACCCCTGAAATCCTTTCCACTCTGGTCTCTCTTCCT GTGTCCTGGACAGCTGACGCACGGCGTTATGAATAAGGAGCTGA AGTACTGTAAGAACCCTGAGGACCTGGAGTG |
| SETD2 | 21 | 2 | chr20 | 41799922 | 41800042 | CAATGAGAATGTGAAACACAAAACCAAGGAGTATATTAAGAAGTA CATGCAGAAGTTTGGGGCTGTGTACAAACCCAAAGAGGACACTG AATTAGAGTGACTTGTGGGGCCAGGGTGGGA |
| RHO | 5 | 1 | chr20 | 5632085 | 5632205 | TGGCTGGGGGTAGGTGTAGGGAATGGGAGACTCCTACAATCGG CCGTAGAGTCCTTGGCAGGGCTTAGGCCGGTGCCACCTGGCTG GTCTCCGTCTTGGAGGCGCTGGCAGAGGCCTCGT |
| RHO | 5 | 2 | chr20 | 5632145 | 5632265 | GGGCTTAGGCCGGTGCCACCTGGCTGGTCTCCGTCTTGGAGGC GCTGGCAGAGGCCTCGTCGTCACCCAGTGGGTTCTTGCCACAGC AGAGGGTGGTGATCATGCAGTTCCGGAACTGGA |
| RHO | 5 | 3 | chr20 | 5632205 | 5632325 | CGTCACCCAGTGGGTTCTTGCCACAGCAGAGGGTGGTGATCATG CAGTTCCGGAACTGGAAGGCAAGGGATATGAGTCAGGGTCAGG GTCCACTAAAAGTGTGTGTAGCAAACTGGGGTG |
| RHO | 4 | 1 | chr20 | 5633105 | 5633225 | CTGCTTGTTCATCATGATATAGATGACAGGGTTGTAGATGGAGGA GGACTTGGCGAAGAACGCTGGGAGGGTCATGAAGATGGGGCCA AAGTCGGAGCCCTGGTGGGTGAAGATGTAGAA |
| RHO | 4 | 2 | chr20 | 5633226 | 5633346 | GCCACACTGGCATAGGGCACCCAGCAGATCAGGAAAGCGATGA CCATGATGATGACCATGCGGGTGACCTCCTTTTCAGCCTTCTGG GTGGTGGCCGATTCCTGCTGCTGGGCAGCTGCC |
| RHO | 3 | 1 | chr20 | 5633430 | 5633550 | AGCACAAGGATCATACCTCCTTGACTGTGAAGACGAGCTGTCCAT AGCAGAAGAATATGACAATCATGGGGATGGCGAAGTGGACCACG AACATGTAGATGACGAAGGACTCATTGTTGA |
| RHO | 3 | 2 | chr20 | 5633551 | 5633671 | TTCTGGCTTGAGTGTGTAGTAGTCGATCCCACATGAGCACTGCAT GCCCTCTGGGATGTACCTGAGCACCAGAGGGGAGTGAGACAGG AGAGGGTGTAAGGGAGACCCACGCGTGAAGGA |
| RHO | 2 | 1 | chr20 | 5634870 | 5634990 | CTCTTTTGCTCAGTGCCATTACCTGGACCAGCCAGCGAGGGGGG GTGCCGCACAGGCCAGCGCCATGACCCAGGTGAAGGCGACGCC CATGATGGCATGGTTCTCCCCAAAACGGAAGTT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| RHO | 2 | 2 | chr20 | 5634991 | 5635111 | CTCATGGGCTTACACACCACCACGTACCGCTCAATGGCCAGGAC CACCAAAGACCACAGGGCAATTTCACCTGCAAGGCAGTCAGGGA CTGATCAGTCAACACACCACCTACCTGGACCC |
| RHO | 1 | 1 | chr20 | 5637044 | 5637164 | GCCCAGTGTGGCAAAGAAGCCCTCCACATTGCATCCTGTGGGCC CGAAGACAAAGTATCCATGCAGAGAGGTGTAGAGGGTGGTGGTG AAGCCACCGAAGACCATGAAGAGGTCAGCCAC |
| RHO | 1 | 2 | chr20 | 5637165 | 5637285 | GCCAGGTTGAGCAGGATGTAGTTGAGAGGTGTACGCAGCTTCTT GTGCTGGACTGTGACGTAGAGCGTGAGGAAGTTGATGGGGAAG CCGAGCACGATCAGCAGAAACATGTAGGCAGCC |
| RHO | 1 | 3 | chr20 | 5637286 | 5637406 | GCATGGAGAACTGCCATGGCTCAGCCAGGTAGTACTGTGGGTAC TCGAAGGGGCTGCGCACCACACCCGTCTTGTTGGAGAAGGGCA CGTAGAAGTTCGGGCCCTCCGTCCCGTTCATGG |
| PTEN | 1 | 1 | chr26 | 37853085 | 37853205 | ACATGTATATGTTAGTTTGCATTGTTGTATATTTCAGACATTTCCTT AACTTAAGTACTCAGATATTTATCCCAACATTATTGCTATGGGGTT TCCTGCAGAAAGACTTGAAGGCGTATA |
| PTEN | 1 | 2 | chr26 | 37853205 | 37853325 | CAGGAACAATATTGATGATGTAGTAAGGTAAGAACTATTATTTTCT GTTTCAAATATTGATGTATTCATGCTGTATTTTCATTTAGAGAAGA TTTCTAAATCGTGGAAAATGATGTACTA |
| PTEN | 2 | 1 | chr26 | 37878188 | 37878308 | TTTCTCAATAGTATAAATGTAATTTCAAATGTTAGCTCATTTTTGTT AATGGTGGCTTTTTGTTTGTTTGTTTTGTTTTAAGGTTTTTGGATTC AAAGCATAAAAACCATTACAAGATAT |
| PTEN | 2 | 2 | chr26 | 37878248 | 37878368 | TGTTTGTTTGTTTTGTTTTAAGGTTTTTGGATTCAAAGCATAAAAAC CATTACAAGATATACAATCTGTAAGTATGTTTTTTATTTGTATGCTT GCAAATATCTTCTAAAATAACTATTA |
| PTEN | 2 | 3 | chr26 | 37878308 | 37878428 | ACAATCTGTAAGTATGTTTTTTATTTGTATGCTTGCAAATATCTTCT AAAATAACTATTAAGTGAAAGTTATTTCCTTGTAGAATCAGGTAGA GTTAAAGATACATTTTAACAGAATTGT |
| PTEN | 3 | 1 | chr26 | 37883981 | 37884101 | ATTAAAACAGTTCAGAAAATGTTTGTTACTTTTTTTTGCCAGTGCA AAAAATAACTAGGTCACATTTAAACTTTTCTTTTAGGTGTGCTGAA AGACATTATGATACCGCCAAATTTAACT |
| PTEN | 3 | 2 | chr26 | 37884041 | 37884161 | CACATTTAAACTTTTCTTTTAGGTGTGCTGAAAGACATTATGATAC CGCCAAATTTAACTGCAGAGGTAGGTATTAACTGCTGAGTAATGT ATTATGTTATATAACTTAAAACCAGTACA |
| PTEN | 3 | 3 | chr26 | 37884101 | 37884221 | GCAGAGGTAGGTATTAACTGCTGAGTAATGTATTATGTTATATAAC TTAAAACCAGTACACTAAATCTTAACAGTCATAGCAGTAATGAATA ATCTCAATTATTAAGTGAGGTAAATAAA |
| PTEN | 4 | 1 | chr26 | 37885930 | 37886050 | TTTAATAAACAGTTCAGGAAAGACTTTTTTCTAATTCTGAGGTTAT CTTTTTACCACAGTTGCACAGTATCCTTTTGAAGACCATAATCCAC CACAGCTAGAACTTATCAAACCCTTTTG |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| PTEN | 4 | 2 | chr26 | 37886050 | 37886170 | TGAAGATCTTGACCAATGGCTAAGTGAAGATGACAATCATGTTGC AGCAATTCACTGTAAAGCTGGAAAGGGACGAACTGGTGTAATGAT TTGTGCATATTTATTACATCGGGGCAAATT |
| PTEN | 4 | 3 | chr26 | 37886170 | 37886290 | TCTAAAGGCACAAGAGGCCCTAGATTTCTATGGGGAAGTAAGGA CCAGAGACAAAAAGGTAAGTTATATTTGATTTTTTTTCTTTCTTTTT TTTCTTTCTTGGAGAATTTATTGGAAAAC |
| PTEN | 5 | 1 | chr26 | 37900543 | 37900663 | CTGTCCACCAGGGAGTAACTATTCCCAGTCAGAGGCGCTATGTG TATTATTATAGCTACCTGTTAAAGAATCATCTGGATTATAGACCAG TGGCACTGTTGTTTCACAAGATGATGTTTG |
| PTEN | 5 | 2 | chr26 | 37900663 | 37900783 | AAACTATTCCAATGTTCAGTGGCGGAACTTGCAGTAAGTGCTCTA AATTTTCATCCTTCCAGTACATGGAACACTTTTCTAGACATATCTA GAAGTTTACATTAAAATTTAGAAAGAAAT |
| PTEN | 6 | 1 | chr26 | 37906294 | 37906414 | CTAACGTGTATTTAACCGTGCAGATCCTCAGTTTGTGGTCTGCCA GCTAAAGGTGAAGATCTATTCCTCCAATTCAGGACCCACACGACG GGAAGACAAGTTCATGTACTTTGAGTTCCC |
| PTEN | 6 | 2 | chr26 | 37906414 | 37906534 | TCAGCCATTGCCTGTGTGCGGTGACATCAAAGTAGAGTTCTTCCA CAAACAGAACAAGATGCTAAAAAAGGTTTGTACTTTTATTGGGAA CAATATCCCAAATCACTGTATTTTGTATAG |
| PTEN | 7 | 1 | chr26 | 37909842 | 37909962 | ATATGTTTTCTTCTTTAAAAACAATTTAAAATTTGTCTTTTTTTTTTT AAAGGACAAAATGTTTCACTTTTGGGTAAACACATTCTTCATACCA GGACCAGAGGAAACCTCAGAAAAAGT |
| PTEN | 7 | 2 | chr26 | 37909962 | 37910082 | AGAAAATGGAAGTCTATGTGATCAAGAAATTGATAGTATTTGCAG TATAGAACGTGCAGATAATGACAAGGAATATCTAGTACTCACTTTA ACAAAAAATGATCTCGACAAAGCAAATAA |
| PTEN | 7 | 3 | chr26 | 37910082 | 37910202 | AGACAAGGCCAACCGATATTTTTCTCCAAATTTTAAGGTCGGTTA AAGAAATTTTAAGGAAGTGGTGGGTGGCTCTTTGTGTGTGTGTGG GGCGGGGTTTAGGTGTCTAGCTGGTGAAGG |
| PTEN | 8 | 1 | chr26 | 37912883 | 37913003 | TTCCATTTAAACTTTGCCTTTTTTTCTCTAGGTGAAGCTGTACTTC ACAAAAACTGTAGAGGAGCCATCAAACCCGGAGGCTAGCAGTTC AACTTCTGTGACGCCAGATGTTAGTGACAA |
| PTEN | 8 | 2 | chr26 | 37913003 | 37913123 | TGAACCTGATCATTATAGATATTCTGACACCACTGACTCTGACCC AGAGAATGAACCCTTTGATGAAGATCAGCACACACAAATCACAAA AGTCTGAATTTTTTTTTAATCAAGAGGGATA |
| POT1 | 2 | 1 | chr14 | 10988226 | 10988346 | AGCAGCAGGCTGTCAACTTCCTAATTGAATTTGCTTTCGACAATA TTTTGAAGCAGATCTGATATGCCACTTGAATTAATAAAAGAAGTGA ATGGAGTACCTGAGGTTGAGCTGCTAGGT |
| POT1 | 2 | 2 | chr14 | 10988286 | 10988406 | GATATGCCACTTGAATTAATAAAAGAAGTGAATGGAGTACCTGAG GTTGAGCTGCTAGGTGAATTAAACAAAGTAGAGTTCAGGTCTCTA CAACCAGGTTGTGATTATAGCAAGAAATAT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| POT1 | 2 | 3 | chr14 | 10988346 | 10988466 | GAATTAAACAAAGTAGAGTTCAGGTCTCTACAACCAGGTTGTGAT TATAGCAAGAAATATATCCAGGTAAGAATTTTCATAACTCCGAATG CTTTTTTTTTCTTCCAAAAGAAATGCTCA |
| POT1 | 3 | 1 | chr14 | 10997222 | 10997342 | TTGCTTTTAAAATAGTTATTCCAATTTTAACTTATCTCTTGGTGTTG TGAGAAGCTCTTTTGGGTTTTCTGTATAACCATTTTTCTCTCATTTT AGGGAAAACTTTCACTTATTTCTTCT |
| POT1 | 3 | 2 | chr14 | 10997282 | 10997402 | TGGGTTTTCTGTATAACCATTTTTCTCTCATTTTAGGGAAAACTTT CACTTATTTCTTCTTTAACTTATGGTGGAAACAGAACAGCTGTTTT GAAGATTGCACTACAAGGTATACAATAT |
| POT1 | 3 | 3 | chr14 | 10997342 | 10997462 | TTAACTTATGGTGGAAACAGAACAGCTGTTTTGAAGATTGCACTA CAAGGTATACAATATAATAAGACCATTTCTGCTTCACCTTTTGCTT ATAATCTACAGATGGATAACTTAATATGT |
| POT1 | 4 | 1 | chr14 | 11000574 | 11000694 | TTGTGTGTCTTAAACTGAATCTAATTAGTGTTTTTCTGTATGTGTG TTTTTTCTCCTAGAACAAGAAAATTCAAATTCAGACACCAACACAG TCAATGTACTATTTTTTGGAAAATTAGC |
| POT1 | 4 | 2 | chr14 | 11000634 | 11000754 | ACAAGAAAATTCAAATTCAGACACCAACACAGTCAATGTACTATTT TTTGGAAAATTAGCAAAGGACTGTGCAAAAGTTTTCTATGAAGGG GTAAATTACTTTCATGTTACTTTGTTTTA |
| POT1 | 4 | 3 | chr14 | 11000694 | 11000814 | AAAGGACTGTGCAAAAGTTTTCTATGAAGGGGTAAATTACTTTCAT GTTACTTTGTTTTAGGAATATATTTAAAATTATAGATGCTATATTCA TCCTATAGATCTTTTTTTGTTTGGTTT |
| POT1 | 5 | 1 | chr14 | 11011501 | 11011621 | TTATATACATTCAGAATTACAAATATTTTGATGATTTAAAACCACTT TTTCCTATGTAGGACACCATTGCAGTGGCTGGATTTGCTGTCTTG TCAGCATCTTCATCTACAAGCAGAGATG |
| POT1 | 5 | 2 | chr14 | 11011561 | 11011681 | ACACCATTGCAGTGGCTGGATTTGCTGTCTTGTCAGCATCTTCAT CTACAAGCAGAGATGGTAAACATTGTTTATTTTTGAAAGCATTTGA GGACTTCAAATCAACTGTTTATGTAAGTA |
| POT1 | 5 | 3 | chr14 | 11011621 | 11011741 | GTAAACATTGTTTATTTTTGAAAGCATTTGAGGACTTCAAATCAAC TGTTTATGTAAGTATGTAACTTTTAGTCCAATCATATGGTTTTAATG AGTTAAGGGGATGGAATAGTATTTCTG |
| POT1 | 6 | 1 | chr14 | 11012367 | 11012487 | TCAAATTGTTTAGGTGCAATTAAACGTCTGTAATCACAGTGGAGG TTGTTTATATACAAGTTAATGCTCTTTATTTGTTTTCAGGTTTATGT TAAATCTGTAAGAGATTTTTCTGCAGAA |
| POT1 | 6 | 2 | chr14 | 11012427 | 11012547 | TTAATGCTCTTTATTTGTTTTCAGGTTTATGTTAAATCTGTAAGAGA TTTTTCTGCAGAATCAATGTCTTTGGTAAGATGATACTCAATTTTA ATTGGAAATATGCACACACGCGCACAC |
| POT1 | 6 | 3 | chr14 | 11012487 | 11012607 | TCAATGTCTTTGGTAAGATGATACTCAATTTTAATTGGAAATATGC ACACACGCGCACACGTAAAGATACACACATGTATGTACATACATA TATATTTGATCCGCTTTATAGACTCATTT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| POT1 | 7 | 1 | chr14 | 11014524 | 11014644 | ACATTATTCTAAAATGTAAATACTTAGTATTGATATATTCACTGGCT TCTCTTTCTTAGGTTCCAACAACACACTATATATATACACCCCTGA ATCAACTTAAGGGTGGCACAGTCGTCA |
| POT1 | 7 | 2 | chr14 | 11014584 | 11014704 | TTCCAACAACACACTATATATATACACCCCTGAATCAACTTAAGGG TGGCACAGTCGTCAATGTCTATGGTGTTGTGAAGTTCTTTAAGCC TCCATATCTAAGCAAAGGAACTGGTAGGT |
| POT1 | 7 | 3 | chr14 | 11014644 | 11014764 | ATGTCTATGGTGTTGTGAAGTTCTTTAAGCCTCCATATCTAAGCAA AGGAACTGGTAGGTATTAAAATTGGTAAAGATTTTGATGGGCTGG AACTGATTTATTGGTTTCAGAAACATACT |
| POT1 | 8 | 1 | chr14 | 11030199 | 11030319 | TCCAGATTATTGCTCAGTTGTAACAATTGTGGACCAGACAAATGT AAAGCTAACCTGCATGCTCTTTAGTGGAAATTATGAAGCCCTTCC AATAATTTATAAAAATGGAGATATTGTTCG |
| POT1 | 8 | 2 | chr14 | 11030319 | 11030439 | CTTTCACAGGCTGAAGGTATAGTTCATACTATCTAGAACTTATCTG TTTGAAGAGAACTCTATTTTTCACTAATACTAGTTAGCACTAATTAT GTGGTTACTGTGGTAGAACATGACATC |
| POT1 | 9 | 1 | chr14 | 11033511 | 11033631 | TTTCTGCTCTGCCCAACTTTCTTAGATCCAAGTATATAAAAATGAA ACTCAAGGTATCACCAGCTCTGGCTTTGCATCTTTGACATTTGAG GGGACTTTGGGAGCTCCTGTCATACCTCG |
| POT1 | 9 | 2 | chr14 | 11033631 | 11033751 | TACTTCAAGCAAGTATTTTAACTTCACTGCTGAGGACCAGAAAAT GGTAGAAGCTTTACGTGTTTGGGCATCTACTCATATTTCACCTTCT TCAACATTAGTAAAATTGTGTGATGTTCA |
| POT1 | 9 | 3 | chr14 | 11033751 | 11033871 | ACCATTGCAGTATTTTGACCTGACTTGTCAGCTCTTGGGCAAAGC AGAAGTGGATGGAGCATCATTTCTTTTAAAGGTAGGTATTTTAAAT GCTTTAAGTTGTAAGTTAACACCACCTAC |
| POT1 | 10 | 1 | chr14 | 11037754 | 11037874 | TCTACTCTGTTATGGCAGGTATGGGATGGCACCAGGACCCCGTT CCCATCATGGAGAGTCTTAATACAAGACCTTGTACTTGAAGGAGA TTTAAGTCATATCCATCGGCTGCAGAATCTG |
| POT1 | 10 | 2 | chr14 | 11037874 | 11037994 | ACAGTGGACATTTTAGTCTATGATAATCATGTTCAAGTGGCAAAAT CTCTAAAGGTGATGTTGGTAGCAAATGTTTGCTCTTTTTTTTTTTTT TTTTTTTATTTTATTTATGATAGGCAC |
| POT1 | 11 | 1 | chr14 | 11043257 | 11043377 | TAAATAATAAATAATGTATTTTTAGGTTGGAAGCTTTCTTAGAATTT ATAGCCTCCATACCAAACTTCAGTCACTGAATTCAGAGAACCAGG CAACGTTGTTAGCTCTAGAGTTTCATCT |
| POT1 | 11 | 2 | chr14 | 11043377 | 11043497 | CCATGGGGGTACCAGTTACGGCCGGGGAATCAGAGTCTTGCCA GAGAGTAACTCTGATGTGGATCAACTGAAAAAGTAAGCATATCAT ATGCCTATTTAGAAATTTCAAATAAGGTATTA |
| POT1 | 12 | 1 | chr14 | 11044464 | 11044584 | ACTAGTCATATTGACTATTTTATGCAATTTGATGTGAAATTATGTTC TTTCAATTACAGGGTTTTAGAATCTGCAGATTTGACAGCCAGTCA GTGTTCCGATGCTATATGTCATTCAGAA |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| POT1 | 12 | 2 | chr14 | 11044524 | 11044644 | GTTTTAGAATCTGCAGATTTGACAGCCAGTCAGTGTTCCGATGCT ATATGTCATTCAGAACAGGAGGACAGCCTTCGGACTCTTTCGAGT AAGGACAGTGTTTAATGTTAACGGTTTAAT |
| POT1 | 12 | 3 | chr14 | 11044584 | 11044704 | CAGGAGGACAGCCTTCGGACTCTTTCGAGTAAGGACAGTGTTTA ATGTTAACGGTTTAATCTTTGCTAAGTAAAATGGCTTTATCAGTAT TCTCTTTTTTAATACTTTCCTATCTCTGCC |
| POT1 | 13 | 1 | chr14 | 11048884 | 11049004 | AGTTCCTGTCAGCAATTTTTCAGTGCTACATGGAATTTTTAACGGA TTTTAAAATTGTTTAATTCTTTTTCCCTTCTCCTTCGTGTAGGCTCT GGATCAGTATCACTGTATGAAATAGAA |
| POT1 | 13 | 2 | chr14 | 11048944 | 11049064 | AATTCTTTTTCCCTTCTCCTTCGTGTAGGCTCTGGATCAGTATCAC TGTATGAAATAGAAAGATGTCAGCAACTATCAGCTACAAGTAAGA TCAGTTCATTTTTAAAATATGAGTGTGAT |
| POT1 | 13 | 3 | chr14 | 11049004 | 11049124 | AGATGTCAGCAACTATCAGCTACAAGTAAGATCAGTTCATTTTTAA AATATGAGTGTGATAGTGTGCATAATAAAATTGTCTTACCAACTAA TTTGTTCCTATCTAATATTGAGAGCAGA |
| POT1 | 14 | 1 | chr14 | 11051255 | 11051375 | TCTAAAAATTCTTTTAAGTACTTACAGATCATCAGTATTTGGAGAA AACCCCAATATGTACTATTTTGAAACAAAAAGCTCCTCAGCAGTAT CGTATCCGAGCCAAATTAAGGTCATATA |
| POT1 | 14 | 2 | chr14 | 11051375 | 11051495 | AGCCCAGAAGACTCTTTCAGTCTGTTAAACTCCATTGTCCTAAAT GTCATGCACTGTAAGTATTTTCTGTAATAAAACAAAAGTTTTTATTT CATAATATTTATGCTTTCTTATATATCA |
| POT1 | 15 | 1 | chr14 | 11052916 | 11053036 | AGTTTATACCAATTGGTTTGTATTTCGTCATTGATGATTTTAGGCA AGAAGTTCCACATGAGGGTGATATGGATATAATTTTACAAGAGGG TGCAAGTAAGACCCCAGATACCAAACTAC |
| POT1 | 15 | 2 | chr14 | 11053036 | 11053156 | AAAATACATCATTATATGATTCAAAAATGTGGACCACTACAGATCA AGGAGAACGAAATGTGGCTGTTCATTTTGTAAAAAATAATGGTATT CTTCCACTTTCAAGTGACTGTCTGATTT |
| POT1 | 15 | 3 | chr14 | 11053156 | 11053276 | TGATAGAAGGTATTTAAATTACTGTTTTATTAGAATACAGCTTTTGT CTATGTCTTTCTTTGATTAATGATCATTAAATATATTTTAGGGCAAC CTAGGACAATTCGATAAAATGGAATA |
| POT1 | 16 | 1 | chr14 | 11056487 | 11056607 | GTTAATAGGAGGTACACTCAGTGAAATCTGCAAACTCGCAGACAA GTTTCATAGTGTAATTCCTGTGAGATCTGGCCATGAAGACCTGGA AGTCCTTGATCTTTCAGCACCATTCCTTAT |
| POT1 | 16 | 2 | chr14 | 11056607 | 11056727 | ACAAGGGAAAATTCATCACTATGGGTATGCTTTCAAATCTGAGAA TTTTGAGAATTAAAAACTGCATAGAAATAACTTTTATTACTTTATCC CACTAGTTGCTGTATATTCAATGGATAC |
| POT1 | 17 | 1 | chr14 | 11059564 | 11059684 | ATAATAAATGAGTTATAATAATTATTTTTAAATGCCTGTGGTTTTTT TCTATCATTAATTTCAGATGTAAACAATGTTCTACTTTGAGACCAA TTCAAAATCTAAATTCCCTGGTTGATA |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| POT1 | 17 | 2 | chr14 | 11059624 | 11059744 | TCAGATGTAAACAATGTTCTACTTTGAGACCAATTCAAAATCTAAA TTCCCTGGTTGATAAAACACAATGGATTCCATCTTCTGTGGCTGA AGGTGAGCTAAATTTGTGTGTCGTGCAAC |
| POT1 | 17 | 3 | chr14 | 11059684 | 11059804 | AAACACAATGGATTCCATCTTCTGTGGCTGAAGGTGAGCTAAATT TGTGTGTCGTGCAACTGTGACCTCACTGATTATATTTCAGTGTTT GTGGGTATAATAAAACCGGAGCAAAGACTG |
| POT1 | 18 | 1 | chr14 | 11061680 | 11061800 | CATAAGACAATTTTTACTTCACTTTGATTATAAGGAATAGAAGATT CACATAGTTTGATCTTTTTTTCCCTTCACTAGTACTGGGTATTGTA CCCCTTCAACATGTGTTTGTAATGACAT |
| POT1 | 18 | 2 | chr14 | 11061740 | 11061860 | TTTTTTTCCCTTCACTAGTACTGGGTATTGTACCCCTTCAACATGT GTTTGTAATGACATTTACTCTTGATGATGGAACAGGAGTATTGGA AGCTTATCTCATGGATTCTGTAAGTAACA |
| POT1 | 18 | 3 | chr14 | 11061800 | 11061920 | TTACTCTTGATGATGGAACAGGAGTATTGGAAGCTTATCTCATGG ATTCTGTAAGTAACAGAGTTGGTATAGGTATTTATAACAAAAAAAA AATTTTTTTAAATTCTAATGTATGTATTC |
| POT1 | 19 | 1 | chr14 | 11063716 | 11063836 | TCAAATAAATAAAATCTTCAAAAAAATAAAAAAGATTAATTTCCTTT TTTCCTCTAAATTTTTTCAAGGACAAATTCTTCCAGATCCCAGCAT CAGAAATCCTCATCAATGATGACCTTC |
| POT1 | 19 | 2 | chr14 | 11063776 | 11063896 | TTTTCAAGGACAAATTCTTCCAGATCCCAGCATCAGAAATCCTCAT CAATGATGACCTTCAGCAGAGTATGGATTCGATCATGGATATGTT TTGTCCTTCAGGAATAAAAATTGGTAGGC |
| POT1 | 19 | 3 | chr14 | 11063836 | 11063956 | AGCAGAGTATGGATTCGATCATGGATATGTTTTGTCCTTCAGGAA TAAAAATTGGTAGGCAATAATATTTTTAACATTCATTCTTTTTTTGA AACAGTACTTGACATGTATCTAATATGT |
| POT1 | 20 | 1 | chr14 | 11066241 | 11066361 | TTTCAGAGATACTTGTTTAACGTTAACAGTTAAATAACTTTTTTTTT TCTCATTTGTAGATGCATATCCATGGTTGGAATGCTTCATCAAGTC GTATAATGTCACAAGTGGAACAGAGCA |
| POT1 | 20 | 2 | chr14 | 11066301 | 11066421 | TGCATATCCATGGTTGGAATGCTTCATCAAGTCGTATAATGTCAC AAGTGGAACAGAGCAGCAAATTTGCTATCAGATTTTTGACACCAC AGTTGCAGAAGATATAATCTAATATTGTCA |
| POT1 | 20 | 3 | chr14 | 11066361 | 11066481 | GCAAATTTGCTATCAGATTTTTGACACCACAGTTGCAGAAGATATA ATCTAATATTGTCATTTAGCTTAGCATATGTAAAATATTATAATTTT AGAAAACATAATTTTCTGTGAGATTGT |
| KRAS | 1 | 1 | chr27 | 22261655 | 22261775 | AAAGGTGTTGATAGAGTGGGTTATACTAACCTCATGTTCTAATTTA TATTAACCACATGTTCTAATTGAGTTACATTTTCATTATTTTTATTA TAAGGCCTGCTGAAAATGACTGAATAT |
| KRAS | 1 | 2 | chr27 | 22261715 | 22261835 | TCTAATTGAGTTACATTTTCATTATTTTTATTATAAGGCCTGCTGAA AATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAG GCAAGAGTGCCTTGACGATACAGCTAATT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| KRAS | 1 | 3 | chr27 | 22261775 | 22261895 | AAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAAGAGTGCCTT GACGATACAGCTAATTCAGAATCACTTTGTGGATGAATATGATCC TACAATAGAGGTAAATTTTGTTTTAATATTA |
| FBXW7 | 12 | 1 | chr15 | 50189250 | 50189370 | GCAGGGAGTATATCGTCTACACATTTGGACAAATTTGTCTTTTCT GCTCTTCACTTCATGTCCACATCAAAGTCCAGCACCAGCAGCTTG GTTTCCTCAGTCCCATTCCGACTCCCAACT |
| FBXW7 | 12 | 2 | chr15 | 50189370 | 50189490 | GCACACACCAGCTTTGTGTTTGAGGCTCTGATCCGCCACACGAC TCCCCCACTCCCCCCACTCTCCAATGTGACTAGGTTTCGAATAAA TTCACCCGTTTTCAAGTCCCATAGTTTTACA |
| FBXW7 | 12 | 3 | chr15 | 50189490 | 50189610 | GTTCCATCATCTGAGCTGGTAATTACAAAGTTCTTGTTGAACTGTA AACAGGTCACAGCACTCTGATGCTTGTTAGGACCTAGGCAAAAC CAAAGGAATTCAATTACTGGTTAGAAATCA |
| FBXW7 | 11 | 1 | chr15 | 50190660 | 50190780 | CAAGACTTACCTTGCAATGTTTGTAAACACTGTCCTGTTTTGATAT CCCAGATTTTAACTGTAGAATCTGCATTCCCAGAGACAAGAATAT TGTCTTTGAGTTCCATTCCACTTGTTAAC |
| FBXW7 | 11 | 2 | chr15 | 50190780 | 50190900 | GACTGGTGTCCTGTTAACGTGTGAATGCAATTCCCTGTCTCCACA TCCCACACTCGGATTGATGTATCAAGAGATCCACTCACCACATGG ATGCCATCAAACTGCAACAGACACAGCATA |
| FBXW7 | 10 | 1 | chr15 | 50192364 | 50192484 | GGATCCCTTACCTGTAATGAATAGACTCTATTAGTATGCCCCTGC AACGTGTGTAGACAGGTCTCGGTCTCCGGATCCCACACCTTCAC CATAAAATCATATGCTCCACTAACAACCCTC |
| FBXW7 | 10 | 2 | chr15 | 50192484 | 50192604 | CTGCCATCATACTGAACACAGCGGACTGCTGCTACATGACCCAT CAAAACATGTAAACACTGGCCTGTCTCAATATCCCAAACCCTAAG AGTGGCATCTCGAGAACCGCTAACAACTCTG |
| FBXW7 | 9 | 1 | chr15 | 50194539 | 50194659 | AATTCCCAACCATTACAGGATTTTCCCTTACCTTTTTTCATGGAGA TGCATACAGCGGACAGTGGAAGTATGCCCATATAAGGTATGTATA CATTCCCCAGTCTCTGCATTCCACACTTT |
| FBXW7 | 9 | 2 | chr15 | 50194659 | 50194779 | TAGAGTCCGATCTGTAGATCCACTAATGATGATATTGTCTCTCATC TGTGATGACCATACTCCACCTGTGTGTCCCACTAATGTTCTCAGA CACTGGAAAAACACTTAAGATGATTACTT |
| FBXW7 | 8 | 1 | chr15 | 50195903 | 50196023 | TTCCCTGATTGATCCATGCAAAAACATAAACAAAATGGGAAGTAT AGAAAGAAGTATGCTTACTTTGCCTGTGACTGCTGACCAAACTTT TAAAGTGTTGTCATCAGAACCACTAACTAT |
| FBXW7 | 8 | 2 | chr15 | 50195963 | 50196083 | TACTTTGCCTGTGACTGCTGACCAAACTTTTAAAGTGTTGTCATCA GAACCACTAACTATTCGGTTACCACAAAACTGTAAGCATGTGATC ACATGATCATCATGTCCTTTCAGAACCTA |
| FBXW7 | 8 | 3 | chr15 | 50196023 | 50196143 | TCGGTTACCACAAAACTGTAAGCATGTGATCACATGATCATCATG TCCTTTCAGAACCTACAAGATAAATGTACAGATTAGAAATATGTTA TTAAATTATATTTCTTAAAATACTGGTAA |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| FBXW7 | 7 | 1 | chr15 | 50197239 | 50197359 | TCAGTTACCTTAGGAGATTTGAGTTCTCCTCGCCTCCAGTTAGTA TCAATTCTATGCTGTCTGATGTATGCACTTTTCCATGGACTGTGTA TGAAACCTGGTTTTATTACTTTTCTTCTC |
| FBXW7 | 7 | 2 | chr15 | 50197359 | 50197479 | TTGATGTGCAATGGTTCATCAATACCTGAAATTTTAAAGAGTCACC AATAGTAAAAATTACACATTTTATAATATATCTCATTTTCCCCTATG CTCAAAACCACTGAAATACAGGACATT |
| FBXW7 | 6 | 1 | chr15 | 50199853 | 50199973 | CAGAATCATTTTGCTTTTTCAAAGTATAAAAATTTTCTCTTTGTGAT TTTGAACCTTACCCTCTTCTTTGCATTTCTCTCTCCAGAGAAGGTT GTCTTCAGCCAAAATTCTCCAGTAGCG |
| FBXW7 | 6 | 2 | chr15 | 50199913 | 50200033 | CTCTTCTTTGCATTTCTCTCTCCAGAGAAGGTTGTCTTCAGCCAAA ATTCTCCAGTAGCGACACGTCTGAGCCGCTTGTAGCAGGTCTTT GGGTTCCAGGAATGAAAGCACATAAAGTGC |
| FBXW7 | 6 | 3 | chr15 | 50199973 | 50200093 | ACACGTCTGAGCCGCTTGTAGCAGGTCTTTGGGTTCCAGGAATG AAAGCACATAAAGTGCCAGCTATGAAAAGGAAAAAAGCATAGTAT ACTCATATTTTAAATATATACTTATTTAACA |
| FBXW7 | 5 | 1 | chr15 | 50204890 | 50205010 | GGATGAAAAATATAATTGATAATCCTTACCTCTTTAGGGAGCAAG GAAATGAAGTCTCGTTGAAACTGGGGTTCTATCACTTGCATCATA TGTTTTACTTGGGTTGGTTCACAACTATCA |
| FBXW7 | 5 | 2 | chr15 | 50205010 | 50205130 | ATAAGCTCATCTAAAGCAAGCAATTTCTCTGGTCCACTCCAGCTC TATGAAAAAGGAAGTAGAAATTTTTAGTATTTTAACAGATGCTCAA ATTACAAACTCATAAAGGAAAAACCACAT |
| FBXW7 | 4 | 1 | chr15 | 50214599 | 50214719 | TTCAAGTACTGTGTTTAGACATGTCACCTGAAACATTTTCAGCCAT TCCTGGAGGCCTGTAGGTGGTTGGACAGATGTAATTCGGCGTCG TTGCTGCCCTTGGCCATTGGCTGCTCTCAG |
| FBXW7 | 4 | 2 | chr15 | 50214719 | 50214839 | GTCTCCAAAAGTTGTTGGTGTTGCTGAACATGGTACAAGCCCAGT GGTACTACAAAAAAAAACAAAAAAAAACAAAAAAACAAAACAAAAA AACAAACCCACAAGAGTGGAAACAGAGAT |
| FBXW7 | 3 | 1 | chr15 | 50217939 | 50218059 | TTTGGAATCATCACTTCTTTAAAGAAACATAAGAATAATAAGAGAA TTAATTATTATACTCAGTATTACCTTGTGTATTCTGAGACTTTGCAT GGTTTCTTTCCCAAAGAAAAAGAGCGG |
| FBXW7 | 3 | 2 | chr15 | 50217999 | 50218119 | CAGTATTACCTTGTGTATTCTGAGACTTTGCATGGTTTCTTTCCCA AAGAAAAAGAGCGGACCTCAGAACCATGGTCCAATTTTCTTTTCA TCTATAAGGTAAAGCAAATAAGATAGGTT |
| FBXW7 | 3 | 3 | chr15 | 50218059 | 50218179 | ACCTCAGAACCATGGTCCAATTTTCTTTTCATCTATAAGGTAAAGC AAATAAGATAGGTTTAAAAAAAAAAGTCACCTGTCCCAACTACTAA ATCTTAACATATGCATTGGCAATACTGA |
| FBXW7 | 2 | 1 | chr15 | 50281778 | 50281898 | TTTGTTGTCTTTGTATAGAATGGGGAGGAGAGTTGGTGAATAGGC AGGTCCACGATACTAGTGGAGTTTGTGACACTGTTACTATGTGTA TGGTCATCTTCCCTGCTACTGTCATCAGAT |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| FBXW7 | 2 | 2 | chr15 | 50281898 | 50282018 | TGATCAAAATCATCACTCTCCTGGTCCATTTCCTCTTCTTCTTCAT CCTCCTCATCTTGTTCACCAGCATGTTCTTCATCTTCTTCTTGCTC TTCTTGGTTTCCTGAGGAGTCCTCATCT |
| FBXW7 | 2 | 3 | chr15 | 50282018 | 50282138 | ACCGAAATGAATCGATTATTGTTTTCTTCCAAATGTCCTTGCTGGG AATCATTTTGGTCTCCAGGTCTAGGTTCTGCTCCGACGACTTCAC CATTCCTTGCAGTGTGCTCCTCCTCTTGT |
| FBXW7 | 2 | 4 | chr15 | 50282138 | 50282258 | TGTCTGAGCTGTTGCTGCTGTTGCTGCTGCTGCTGTTCCTCCTCT ACCACACGATTCATCTGTTCCTCATCTGCCTGGCTTGAGGAAGG GTTACCTCTCAGAGAGCCTCCAGTTCGTCGT |
| FBXW7 | 2 | 5 | chr15 | 50282258 | 50282378 | CGTTTGCTGCCCACAGAGAGCAGTTCCTGATTCATTTCCAAAAGC CAGCTTGCTACTTCTTGGACCTTGGCTAGACTATCAGAAGATGCA AAGGATTTCACATTCTGTAGGGAAAAATGA |
| EZH2 | 3 | 1 | chr16 | 1953097 | 1953217 | AGTATGTTTAGTTCCAATCGTCAGAAAATTTTGGAAAGAACGGAA ATCTTAAACCAAGAATGGAAACAGCGAAGGATACAGCCTGTGCA CATCCTGACTTCTGTGAGCTCATTGCGCGGG |
| EZH2 | 3 | 2 | chr16 | 1953157 | 1953277 | TGGAAACAGCGAAGGATACAGCCTGTGCACATCCTGACTTCTGT GAGCTCATTGCGCGGGACCAGGGAGGTTGGTTAACTCACGCACT GTAGTAATTATCGGTGCATTCTGGTGTTTGCT |
| EZH2 | 3 | 3 | chr16 | 1953217 | 1953337 | ACCAGGGAGGTTGGTTAACTCACGCACTGTAGTAATTATCGGTG CATTCTGGTGTTTGCTCTTGCAATGTTGGTTGGTTATTGATAGGA AGTCATTCACCTCAGGATGTCCGTCTTTTTA |
| EZH2 | 4 | 1 | chr16 | 1961697 | 1961817 | GTGTGTTAATTTCTTTCCCTTCCCTGCAGGCCATTCACCTGCTCTT CCCTCTACGATTTTCCATAACACAGTTTTTACTTGGAACCAGCCTT CTGCCAAGAGTCTCAGTTTGGTTGTGTA |
| EZH2 | 4 | 2 | chr16 | 1961817 | 1961937 | CTCCTACAACTACTATTTTTGGCTTGACTTCCCTCTCCAGCTCCCA GTGGTCCAGCCACACAGTCGCCTATTGTACCAGCTGGTTGATGT AAGTCTATCCATCCTCAGAGGAACCCGGTT |
| EZH2 | 5 | 1 | chr16 | 1965306 | 1965426 | TACCACTTGAGGTTGTCCTTTTATTCTTATGAGTCAATTTTATTTGT TTCTGTTCTAGTGTTCAGTGACCAGTGACTTGGATTTTCCAACAC AAGTCATCCCATTAAAGACTCTGAATGC |
| EZH2 | 5 | 2 | chr16 | 1965426 | 1965546 | GGTTGCTTCAGTACCCATAATGTATTCTTGGTCTCCCCTTCAGCA GAATTTTATGGTATGTACATCTGTGATCTGCTTACTTTATTATCAC TTATTTTAAATAAAACACACTTTTCTTGT |
| EZH2 | 6 | 1 | chr16 | 1968623 | 1968743 | GATTTAAAAAATATTTTGATACTTCAGATTAACTTGCTTTTAAAATA TGTTTATTTTTAGGTGGAAGATGAAACTGTTTTACATAACATTCCG TATATGGGGGATGAAGTTTTAGATCAG |
| EZH2 | 6 | 2 | chr16 | 1968683 | 1968803 | GTGGAAGATGAAACTGTTTTACATAACATTCCGTATATGGGGGAT GAAGTTTTAGATCAGGATGGGACTTTCATTGAAGAACTAATAAAA AATTATGATGGAAAAGTACATGGGGATAGA |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| EZH2 | 6 | 3 | chr16 | 1968743 | 1968863 | GATGGGACTTTCATTGAAGAACTAATAAAAAATTATGATGGAAAA GTACATGGGGATAGAGGTGAGCCCTGTGTTTATTCTCTTCAAAAA AGGGGCCAAAGAATTGAAAGATATTTTGAA |
| EZH2 | 7 | 1 | chr16 | 1969602 | 1969722 | TGAATTTCAGAATGTGGGTTTATAAATGATGAAATTTTTGTGGAGT TGGTGAATGCTCTTGGTCAATACAATGATGATGATGATGATGATG ATGGAGATGATCCTGATGAAAGAGAAGAA |
| EZH2 | 7 | 2 | chr16 | 1969722 | 1969842 | AAACAGAAAGATCTAGAGGAGAACAGAGATGGTATGTCTGTTTAT CAGGTTTGTGTGTAGCCTTGACAGCATTTTTCAAAAACAAAAAATC ACAAACTGTTATTAATTCCTTCTTTCTCT |
| EZH2 | 8 | 1 | chr16 | 1970655 | 1970775 | GTGGTTGCATGGAGAAGGAAGGTTTGAAAACTTGTTTGCTTTCAT CTTTTTTCCCTTTCATATTTCCCCATTTAGATAAAGAAAGCCGCCC ACCTCGGAAATTTCCTTCTGATAAAATTT |
| EZH2 | 8 | 2 | chr16 | 1970715 | 1970835 | TATTTCCCCATTTAGATAAAGAAAGCCGCCCACCTCGGAAATTTC CTTCTGATAAAATTTTTGAAGCCATTTCCTCAATGTTTCCAGATAA GGGCACAGCAGAAGAACTAAAGGAAAAGT |
| EZH2 | 8 | 3 | chr16 | 1970775 | 1970895 | TTGAAGCCATTTCCTCAATGTTTCCAGATAAGGGCACAGCAGAAG AACTAAAGGAAAAGTAAGAGTTGTTCTTTTGAGGCAATCTTGCCT TGTATGGTATGTTAGCTCAGCATTGCGGTT |
| EZH2 | 9 | 1 | chr16 | 1971322 | 1971442 | CGTTTTTCTTGGTCAATGTCAGGTATAAAGAGCTCACCGAGCAGC AGCTCCCAGGTGCGCTTCCTCCTGAATGTACCCCGAACATCGAT GGACCAAATGCTAAATCTGTTCAGAGGGAGC |
| EZH2 | 9 | 2 | chr16 | 1971442 | 1971562 | AAAGCTTGCATTCATTTCATACGCTTTTCTGTAGGCGATGTTTTAA ATATGACTGCTTCCTACATCGTAAGTGCAGTTATTGTACGTTTTCT TTTTCTATCTTTGTTGTGGAATTATGTC |
| EZH2 | 10 | 1 | chr16 | 1979980 | 1980100 | GGCTTTTTCCACTGAAACTGTGGAGATGAAAAAGTAAATATAATTA TACATGGATGTTTTCTGTTTATTTTGCAGCATTTCATGCAACACCC AATACTTACAAGCGGAAGAACACAGAAA |
| EZH2 | 10 | 2 | chr16 | 1980040 | 1980160 | CTGTTTATTTTGCAGCATTTCATGCAACACCCAATACTTACAAGCG GAAGAACACAGAAACCGCTCTAGACAACAAACCTTGTGGACCAC AGTGTTACCAGCACTTGGTAAGCTCGGGTG |
| EZH2 | 10 | 3 | chr16 | 1980100 | 1980220 | CCGCTCTAGACAACAAACCTTGTGGACCACAGTGTTACCAGCACT TGGTAAGCTCGGGTGCTGCGTGATCCAGAGGGCATTTGAGGAGT TTTGTGGTGAACTGTGGCGAGGTCCATGAAG |
| EZH2 | 11 | 1 | chr16 | 1981573 | 1981693 | GAGGGAGCAAAGGAGTTTGCGGCTGCCCTCACTGCCGAGCGCA TCAAGACCCCGCCCAAGCGGCCCGGCGGCCGCAGGAGAGGCC GGCTTCCCAACAATAGCAGCCGGCCCAGCACCCCC |
| EZH2 | 11 | 2 | chr16 | 1981693 | 1981813 | ACCATTAATGTGCTGGAGTCCAAGGACACGGACAGTGACAGGGA AGCGGGGACTGAGACGGGCGGGGAGAACAACGATAAGGAGGAG GAGGAGAAGAAGGACGAGACCTCCAGCTCCTCG |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| EZH2 | 12 | 1 | chr16 | 1982496 | 1982616 | CCAGTGCTTCACTTTGGTTTTCCAGAAGCAAATTCTCGGTGTCAA ACGCCAATAAAGATGAAGCCAAACATTGAGCCTCCCGAGAACGT GGAGTGGAGCGGCGCTGAGGCCTCCATGTTC |
| EZH2 | 12 | 2 | chr16 | 1982616 | 1982736 | CGGGTGCTCATCGGTACCTACTATGACAATTTCTGTGCCATTGCC AGGCTCATCGGGACCAAGACTTGTCGGCAGGTGAGCATGTCTGA CACCGCAGTTTGAGAGCTTGAGATGGGGGAT |
| EZH2 | 13 | 1 | chr16 | 1982985 | 1983105 | CAGCAGGGGCGGGTCTGGGGCGCTTTGCCGGAGCTGCCAGTGC TCACAGCCCTGTGACCACACCTCTTGTTAGGTGTATGAGTTCCGT GTCAAAGAATCGAGTATCATTGCTCCAGCCCC |
| EZH2 | 13 | 2 | chr16 | 1983045 | 1983165 | CACCTCTTGTTAGGTGTATGAGTTCCGTGTCAAAGAATCGAGTAT CATTGCTCCAGCCCCGGCCGAGGACGTGGACACCCCCCCAAGG AAGAAGAAGAGGAAACACCGGTAAGGCTGGGT |
| EZH2 | 13 | 3 | chr16 | 1983105 | 1983225 | GGCCGAGGACGTGGACACCCCCCCAAGGAAGAAGAAGAGGAAA CACCGGTAAGGCTGGGTGGGTGGCTTCTGCTGTGGGTTGGATG CCAGAGTCATGTGGCCGCACACGCTCGCTCGCCA |
| EZH2 | 14 | 1 | chr16 | 1984305 | 1984425 | CGCAGGCAGCTTCTGTTGGGTTTGTGCCGTGACCCGACTCCTTA GCCTCACAAGGATGGCTTCTGTTTTGTGTCTTGCAGGTTGTGGG CGGCACACTGCAGAAAGATACAGCTGAAAAAG |
| EZH2 | 14 | 2 | chr16 | 1984365 | 1984485 | TTCTGTTTTGTGTCTTGCAGGTTGTGGGCGGCACACTGCAGAAA GATACAGCTGAAAAAGGGTTAGCATCCTCCGTTCATAATTAACTT AAAAATGCCTGCTCATTTGTTTGAGAATAGT |
| EZH2 | 14 | 3 | chr16 | 1984425 | 1984545 | GGTTAGCATCCTCCGTTCATAATTAACTTAAAAATGCCTGCTCATT TGTTTGAGAATAGTAACTGGATTAGCGTTTGGGATGTTCTGATGT AATTGACTTCAGTACCCGAAGCCCATTTA |
| EZH2 | 15 | 1 | chr16 | 1984739 | 1984859 | ATTCGCGGGCGTCGTGTTCCCTCGCCTCTGGATTGACCCGCGCT GCTCTCCTTCCGCAGACGGCTCCTCTAACCACGTTTACAACTATC AACCCTGTGACCATCCACGGCAGCCTTGTGA |
| EZH2 | 15 | 2 | chr16 | 1984799 | 1984919 | CGGCTCCTCTAACCACGTTTACAACTATCAACCCTGTGACCATCC ACGGCAGCCTTGTGACAGTTCGTGCCCTTGTGTGATAGCACAAA ATTTTTGTGAGAAGTTTTGTCAATGTAGCTC |
| EZH2 | 15 | 3 | chr16 | 1984859 | 1984979 | CAGTTCGTGCCCTTGTGTGATAGCACAAAATTTTTGTGAGAAGTT TTGTCAATGTAGCTCAGAGTGTAAGTATTTCGTTTTGATGACTTGT GTGAGTTCCTGTTTTATTTATTTATTTAT |
| EZH2 | 16 | 1 | chr16 | 1985613 | 1985733 | CCCGCCTGAGGACCACCCTGTCCCTGAAGGTCAAAACCGCTTTC CTGGGTGCCGCTGCAAAGCGCAGTGCAACACGAAGCAGTGCCC GTGCTACCTGGCCGTGCGCGAGTGCGACCCCGA |
| EZH2 | 16 | 2 | chr16 | 1985733 | 1985853 | CCTGTGCCTCACGTGCGGCGCCGCCGACCACTGGGACAGCAAG AACGTGTCCTGCAAGAACTGCAGCATCCAGCGCGGCTCCAAGAA GGTACGCACCGCCCCACTGGCTGGGGGCGGGGC |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| EZH2 | 17 | 1 | chr16 | 1987210 | 1987330 | CTGGGCTGTCACGTAGCTAACTCTCCGTCTTTGCTCTCCTTCGGT GATGCGATTGTGTGCGTTTCCTTCTCTAGCACTTGCTGCTGGCAC CGTCTGACGTGGCAGGCTGGGGGATTTTCA |
| EZH2 | 17 | 2 | chr16 | 1987270 | 1987390 | GTTTCCTTCTCTAGCACTTGCTGCTGGCACCGTCTGACGTGGCA GGCTGGGGGATTTTCATCAAAGACCCTGTGCAGAAAAATGAATTC ATCTCGGAGTACTGTGGAGAGGTAAGGCAGG |
| EZH2 | 17 | 3 | chr16 | 1987330 | 1987450 | TCAAAGACCCTGTGCAGAAAAATGAATTCATCTCGGAGTACTGTG GAGAGGTAAGGCAGGGGGACCCAGCCCCGGGCGGCGCCGCAC GTACGAACCACATGGGTTCGGGGGTGTCGTGTG |
| EZH2 | 18 | 1 | chr16 | 1987880 | 1988000 | AGCAACACCAGCCGTAGGGACCATTCGTTGGGGTTCCTCATGGG GGTCTGAGGTTGTCCTGATCACCTGATCCAGGTCTTGTCCTCTGC CTATCTTTTTAGATTATTTCTCAAGATGAAG |
| EZH2 | 18 | 2 | chr16 | 1987940 | 1988060 | GATCACCTGATCCAGGTCTTGTCCTCTGCCTATCTTTTTAGATTAT TTCTCAAGATGAAGCAGACAGAAGAGGGAAAGTCTACGATAAATA CATGTGCAGCTTTCTCTTCAACTTGAACA |
| EZH2 | 18 | 3 | chr16 | 1988000 | 1988120 | CAGACAGAAGAGGGAAAGTCTACGATAAATACATGTGCAGCTTTC TCTTCAACTTGAACAATGGTACGTTTGACTTGGGGGCCCCCGGG ACGCATCGAGGAGCCCCTTCTCCTTATGAGC |
| EZH2 | 19 | 1 | chr16 | 1988610 | 1988730 | ATTGTCAGAACTTTGTGTTTACAGACCTAGGACTGAACCTATTCA CTGGACTATACTTACTTTTTCTTCTTTTTAGATTTTGTGGTGGATG CAACCCGCAAGGGTAACAAAATTCGTTTT |
| EZH2 | 19 | 2 | chr16 | 1988670 | 1988790 | TTTTTCTTCTTTTTAGATTTTGTGGTGGATGCAACCCGCAAGGGTA ACAAAATTCGTTTTGCAAATCACTCAGTAAATCCAAACTGCTATGC AAAAGGTAGGTGCCGATGAAGTCCGAAT |
| EZH2 | 19 | 3 | chr16 | 1988730 | 1988850 | GCAAATCACTCAGTAAATCCAAACTGCTATGCAAAAGGTAGGTGC CGATGAAGTCCGAATCCCCAACAGAACTGCTTGCAGACCAGGGA GGGGGGACTGGAGCAGCCTGGGACTGACGTG |
| EZH2 | 20 | 1 | chr16 | 1988840 | 1988960 | GACTGACGTGGACTTTGTGCGTGAGTCAGGTCTTCCCCGTGGCC TGTCTTCATCCAACTCACTGATACTGTGTCTCTCTCTCCAGTTATG ATGGTTAACGGCGATCACAGGATAGGGATT |
| EZH2 | 20 | 2 | chr16 | 1988900 | 1989020 | ACTGATACTGTGTCTCTCTCTCCAGTTATGATGGTTAACGGCGAT CACAGGATAGGGATTTTTGCTAAGAGAGCCATCCAGACCGGTGA AGAGCTGTTTTTTGATTACAGGTTGGTAACA |
| EZH2 | 20 | 3 | chr16 | 1988960 | 1989080 | TTTGCTAAGAGAGCCATCCAGACCGGTGAAGAGCTGTTTTTTGAT TACAGGTTGGTAACAGCATATTTCTGTGTGACCTGAGGGCTTCAT CTTCCTGGACTCTGGGCTGTGGACAAAGGC |
| EZH2 | 21 | 1 | chr16 | 1991574 | 1991694 | GTTTTAACACACACGCTGAGCGGAGCGGCTTGGTTGTGTGGGTC TCCCCCTGTGCTGGATGCTCACTCATTCCTTTTTCCTGCTGTTTC AGATACAGCCAGGCTGATGCCCTGAAGTACG |

| Gene | Exon | Probe | Chromosome | Start | Stop | Seq |
|---|---|---|---|---|---|---|
| EZH2 | 21 | 2 | chr16 | 1991634 | 1991754 | GCTCACTCATTCCTTTTTCCTGCTGTTTCAGATACAGCCAGGCTG ATGCCCTGAAGTACGTGGGCATCGAAAGAGAGATGGAAATCCCT TGACGTCTGCTACCTCTTCCTCCCCCTCCTC |
| EZH2 | 21 | 3 | chr16 | 1991694 | 1991814 | TGGGCATCGAAAGAGAGATGGAAATCCCTTGACGTCTGCTACCT CTTCCTCCCCCTCCTCCTTTCTGGAACAGCTGCCTTAGCTTCAGG AACCTCGAGTACTGTGGGCAATTTAGAAAAA |

## Table 2. Dog Information

| Sample | Breed | Age | Age Bracket | Sex | Coding Mutations | Non-coding Mutations | Coding Depth | NTs |
|---|---|---|---|---|---|---|---|---|
| AKC4 | Alaskan Malamute | 2.5 | Young | Male | 27 | 31 | 10691.9012 | 186648520 |
| AKC3 | Alaskan Malamute | 3.74 | Young | Male | 28 | 33 | 12724.0464 | 222250918 |
| AKC72 | Alaskan Malamute | 5.69 | Adolescent | Female | 39 | 34 | 11233.5814 | 196329303 |
| AKC24 | Alaskan Malamute | 6.5 | Adolescent | NA | 29 | 28 | 8352.40844 | 145824699 |
| AKC104 | Alaskan Malamute | 6.56 | Adolescent | Male | 38 | 21 | 11758.5547 | 205339641 |
| AKC2 | Alaskan Malamute | 7.75 | Adolescent | Male | 56 | 48 | 11588.7495 | 202385922 |
| AKC102 | Alaskan Malamute | 9.88 | Older | Male | 54 | 34 | 12329.9825 | 215318485 |
| AKC16 | Alaskan Malamute | 10.55 | Older | Female | 26 | 20 | 7806.35484 | 136267730 |
| AKC1 | Alaskan Malamute | 11.41 | Older | Male | 36 | 27 | 7388.95866 | 129040774 |
| AKC101 | Alaskan Malamute | 14.05 | Older | Female | 43 | 28 | 9769.25271 | 170522306 |
| AKC99 | Boxer | 0.88 | Young | Female | 17 | 23 | 10680.1924 | 186657722 |
| AKC34 | Boxer | 1.19 | Young | Female | 18 | 18 | 10927.2654 | 190975818 |
| AKC98 | Boxer | 1.21 | Young | Male | 19 | 15 | 7576.62265 | 132416634 |
| AKC33 | Boxer | 1.36 | Young | Male | 30 | 32 | 15252.5251 | 266568382 |
| AKC109 | Boxer | 1.89 | Young | Male | 25 | 29 | 13154.2797 | 229897347 |
| AKC40 | Boxer | 2.41 | Young | Male | 72 | 63 | 12569.9449 | 219684927 |
| AKC110 | Boxer | 2.54 | Young | Female | 54 | 45 | 12451.2997 | 217611365 |
| AKC93 | Boxer | 2.64 | Young | Female | 71 | 39 | 12200.6404 | 213230592 |
| AKC100 | Boxer | 3.89 | Young | Male | 22 | 28 | 11649.8404 | 203604260 |
| AKC26 | Boxer | 4.15 | Adolescent | Female | 25 | 16 | 6751.97746 | 118004310 |
| AKC29 | Boxer | 4.63 | Adolescent | Female | 45 | 21 | 13051.0699 | 228093549 |
| AKC28 | Boxer | 4.77 | Adolescent | Female | 46 | 43 | 14062.449 | 245769421 |
| AKC65 | Boxer | 4.94 | Adolescent | Female | 22 | 22 | 11016.2412 | 192530847 |
| AKC113 | Boxer | 5.12 | Adolescent | Female | 38 | 37 | 12693.1097 | 221837478 |
| AKC30 | Boxer | 6.33 | Adolescent | Female | 99 | 117 | 12175.6031 | 212793016 |
| AKC31 | Boxer | 6.91 | Adolescent | Male | 66 | 32 | 14950.2884 | 261286191 |

| Sample | Breed | Age | Age Bracket | Sex | Coding Mutations | Non-coding Mutations | Coding Depth | NTs |
|---|---|---|---|---|---|---|---|---|
| AKC27 | Boxer | 8.71 | Older | Male | 60 | 31 | 10550.6063 | 184392946 |
| AKC64 | Boxer | 9.1 | Older | Male | 17 | 14 | 8804.39921 | 153874485 |
| AKC111 | Boxer | 9.86 | Older | Female | 35 | 39 | 11976.4948 | 209313200 |
| AKC108 | Boxer | 10.3 | Older | Male | 41 | 35 | 9922.01644 | 173208641 |
| AKC112 | Boxer | 10.45 | Older | Male | 45 | 24 | 13009.4302 | 227365811 |
| AKC44 | Dalmatian | 1.66 | Young | Female | 36 | 31 | 12666.0422 | 221364420 |
| AKC45 | Dalmatian | 2.08 | Young | Female | 78 | 60 | 12661.2288 | 221280296 |
| AKC8 | Dalmatian | 3.33 | Young | Male | 13 | 21 | 10241.1745 | 178985007 |
| AKC6 | Dalmatian | 3.66 | Young | Female | 21 | 15 | 10359.783 | 181057927 |
| AKC67 | Dalmatian | 3.67 | Young | Female | 13 | 11 | 9103.2497 | 158915430 |
| AKC42 | Dalmatian | 4.16 | Adolescent | Male | 26 | 31 | 11392.4537 | 199105914 |
| AKC43 | Dalmatian | 4.25 | Adolescent | Female | 27 | 19 | 9667.21761 | 168905626 |
| AKC41 | Dalmatian | 7.08 | Adolescent | Female | 54 | 29 | 13181.5076 | 230373208 |
| AKC68 | Dalmatian | 7.17 | Adolescent | Female | 25 | 13 | 7198.89416 | 125757482 |
| AKC66 | Dalmatian | 9.83 | Older | Female | 21 | 15 | 8981.02907 | 156961445 |
| AKC46 | Dalmatian | 10.21 | Older | Female | 82 | 39 | 11334.7989 | 197928259 |
| AKC7 | Dalmatian | 11.16 | Older | Male | 22 | 28 | 7245.01144 | 126621065 |
| AKC47 | Dalmatian | 11.54 | Older | Female | 23 | 22 | 10140.1311 | 177219072 |
| AKC5 | Dalmatian | 14.99 | Older | Female | 22 | 21 | 6757.13981 | 118020204 |
| AKC50 | German Shephard | 1.32 | Young | Male | 50 | 35 | 12510.2727 | 218466893 |
| AKC12 | German Shephard | 1.95 | Young | Female | 34 | 19 | 10319.5398 | 180282361 |
| AKC60 | German Shephard | 2.03 | Young | Male | 33 | 24 | 12668.8862 | 221236760 |
| AKC59 | German Shephard | 2.5 | Young | Female | 20 | 16 | 13142.9268 | 229528074 |
| AKC11 | German Shephard | 2.53 | Young | Female | 24 | 20 | 9854.50407 | 171961096 |
| AKC62 | German Shephard | 3.03 | Young | Male | 35 | 32 | 13596.0361 | 237441174 |
| AKC63 | German Shephard | 3.2 | Young | Female | 27 | 25 | 11807.6491 | 206196976 |
| AKC49 | German Shephard | 4.95 | Adolescent | Female | 30 | 17 | 9696.72338 | 169275700 |

| Sample | Breed | Age | Age Bracket | Sex | Coding Mutations | Non-coding Mutations | Coding Depth | NTs |
|---|---|---|---|---|---|---|---|---|
| AKC61 | German Shephard | 5.45 | Adolescent | Female | 38 | 27 | 12920.865 | 225637065 |
| AKC92 | German Shephard | 7 | Adolescent | Female | 19 | 27 | 12450.4304 | 217421866 |
| AKC48 | German Shephard | 9.36 | Older | Male | 43 | 27 | 11463.3316 | 200172696 |
| AKC51 | German Shephard | 9.56 | Older | Female | 42 | 31 | 12195.2594 | 212965815 |
| AKC73 | German Shephard | 12.02 | Older | Female | 38 | 26 | 8587.915 | 149936408 |
| AKC106 | Golden Retriever | 0.5 | Young | Male | 42 | 27 | 14840.4552 | 259366635 |
| AKC69 | Golden Retriever | 1.67 | Young | Female | 9 | 20 | 10935.1863 | 190950224 |
| AKC85 | Golden Retriever | 2.18 | Young | Female | 17 | 19 | 10814.4311 | 189003813 |
| AKC14 | Golden Retriever | 2.19 | Young | Female | 16 | 13 | 6156.93999 | 107112285 |
| AKC80 | Golden Retriever | 2.44 | Young | Female | 10 | 12 | 10438.7389 | 182093362 |
| AKC105 | Golden Retriever | 2.5 | Young | Female | 35 | 33 | 13994.3406 | 244397164 |
| AKC70 | Golden Retriever | 3.02 | Young | Female | 26 | 22 | 10588.8655 | 184892180 |
| AKC71 | Golden Retriever | 5.24 | Adolescent | Female | 10 | 11 | 7651.18579 | 133513192 |
| AKC116 | Golden Retriever | 5.26 | Adolescent | Male | 45 | 37 | 13396.2428 | 234126135 |
| AKC86 | Golden Retriever | 5.26 | Adolescent | Male | 28 | 33 | 11344.8802 | 197888745 |
| AKC84 | Golden Retriever | 5.77 | Adolescent | Male | 31 | 24 | 12818.4669 | 223861706 |
| AKC75 | Golden Retriever | 7.96 | Adolescent | Male | 45 | 32 | 11551.0876 | 201693540 |
| AKC76 | Golden Retriever | 8.04 | Older | Female | 38 | 21 | 10339.3145 | 180514092 |
| AKC81 | Golden Retriever | 8.34 | Older | Female | 33 | 17 | 11015.5646 | 192243633 |
| AKC74 | Golden Retriever | 9.46 | Older | Female | 31 | 27 | 10283.0751 | 179563057 |
| AKC114 | Golden Retriever | 10.76 | Older | Male | 30 | 15 | 11717.0306 | 204614505 |
| AKC77 | Golden Retriever | 10.87 | Older | Male | 27 | 16 | 9448.12314 | 164964230 |
| AKC107 | Golden Retriever | 11.83 | Older | Male | 62 | 36 | 12188.0397 | 213010370 |
| AKC115 | Golden Retriever | 14.17 | Older | Male | 36 | 27 | 11385.1563 | 198978377 |
| AKC90 | Labrador Retriever | 1.46 | Young | Female | 19 | 13 | 12814.1956 | 223953697 |
| AKC52 | Labrador Retriever | 1.59 | Young | Male | 35 | 17 | 12030.1317 | 210250611 |
| AKC94 | Labrador Retriever | 2.71 | Young | Male | 33 | 19 | 13107.8865 | 229086533 |

| Sample | Breed | Age | Age Bracket | Sex | Coding Mutations | Non-coding Mutations | Coding Depth | NTs |
|---|---|---|---|---|---|---|---|---|
| AKC82 | Labrador Retriever | 3.71 | Young | Male | 35 | 19 | 9638.05758 | 167711840 |
| AKC83 | Labrador Retriever | 4.04 | Adolescent | Female | 22 | 13 | 7908.04717 | 137963791 |
| AKC53 | Labrador Retriever | 5.13 | Adolescent | Female | 26 | 31 | 11481.8453 | 200668211 |
| AKC54 | Labrador Retriever | 7.63 | Adolescent | Female | 42 | 31 | 12019.1689 | 210059015 |
| AKC91 | Labrador Retriever | 8 | Adolescent | Male | 30 | 26 | 11664.8282 | 203714560 |
| AKC87 | Labrador Retriever | 8.37 | Older | Male | 31 | 20 | 9295.4068 | 162325689 |
| AKC88 | Labrador Retriever | 10.37 | Older | Female | 48 | 26 | 13210.4334 | 230720220 |
| AKC55 | Labrador Retriever | 11.29 | Older | Male | 30 | 28 | 10319.8069 | 180359266 |
| AKC89 | Labrador Retriever | 11.95 | Older | Male | 65 | 38 | 14383.1702 | 251374665 |
| AKC39 | Samoyed | 1.36 | Young | Male | 24 | 19 | 11040.5243 | 192789635 |
| AKC9 | Samoyed | 2.52 | Young | Female | 45 | 23 | 11281.876 | 197015400 |
| AKC79 | Samoyed | 2.63 | Young | Male | 8 | 13 | 7824.50839 | 136608092 |
| AKC97 | Samoyed | 3.87 | Young | Male | 31 | 24 | 13676.561 | 238847461 |
| AKC96 | Samoyed | 6.37 | Adolescent | Male | 32 | 34 | 12548.302 | 219143546 |
| AKC36 | Samoyed | 6.98 | Adolescent | Male | 24 | 17 | 7869.24285 | 137389111 |
| AKC38 | Samoyed | 7.73 | Adolescent | Female | 27 | 21 | 9579.0646 | 167250468 |
| AKC37 | Samoyed | 10.89 | Older | Male | 28 | 16 | 10167.5814 | 177495469 |
| AKC10 | Samoyed | 11.07 | Older | Male | 43 | 21 | 10047.3486 | 175426707 |
| AKC78 | Samoyed | 13.36 | Older | Female | 43 | 38 | 11222.9841 | 195964526 |
| AKC35 | Samoyed | 13.8 | Older | Female | 28 | 22 | 7652.44275 | 133588693 |

## Table 3. R Package Citations

| Package | Version | Citation |
|---|---|---|
| base | 4.3.0 | R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. |
| ggpubr | 0.6.0 | Kassambara A (2023). _ggpubr: 'ggplot2' Based Publication Ready Plots_. R package version 0.6.0, <https://CRAN.R-project.org/package=ggpubr>. |
| cowplot | 1.1.1 | Wilke C (2020). _cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'_. R package version 1.1.1, <https://CRAN.R-project.org/package=cowplot>. |
| ggbeeswarm | 0.7.2 | Clarke E, Sherrill-Mix S, Dawson C (2023). _ggbeeswarm: Categorical Scatter (Violin Point) Plots_. R package version 0.7.2, <https://CRAN.R-project.org/package=ggbeeswarm>. |
| tidyquant | 1.0.7 | Dancho M, Vaughan D (2023). _tidyquant: Tidy Quantitative Financial Analysis_. R package version 1.0.7, <https://CRAN.R-project.org/package=tidyquant>. |
| scales | 1.2.1 | Wickham H, Seidel D (2022). _scales: Scale Functions for Visualization_. R package version 1.2.1, <https://CRAN.R-project.org/package=scales>. |
| ggpmisc | 0.5.2 | Aphalo P (2022). _ggpmisc: Miscellaneous Extensions to 'ggplot2'_. R package version 0.5.2, <https://CRAN.R-project.org/package=ggpmisc>. |
| reshape2 | 1.4.4 | Wickham H (2007). ,ÄúReshaping Data with the reshape Package.,Äù _Journal of Statistical Software_, *21*(12), 1-20. <http://www.jstatsoft.org/v21/i12/>. |
| ggtext | 0.1.2 | Wilke C, Wiernik B (2022). _ggtext: Improved Text Rendering Support for 'ggplot2'_. R package version 0.1.2, <https://CRAN.R-project.org/package=ggtext>. |
| janitor | 2.2.0 | Firke S (2023). _janitor: Simple Tools for Examining and Cleaning Dirty Data_. R package version 2.2.0, <https://CRAN.R-project.org/package=janitor>. |
| sigfit | 2.2 | Gori K, Baez-Ortega A (2018). ,Äúsigfit: flexible Bayesian inference of mutational signatures.,Äù _bioRxiv_, 372896. doi:10.1101/372896 <https://doi.org/10.1101/372896>. |
| gsubfn | 0.7 | Grothendieck G (2018). _gsubfn: Utilities for Strings and Function Arguments_. R package version 0.7, <https://CRAN.R-project.org/package=gsubfn>. |

| Package | Version | Citation |
|---|---|---|
| broom | 1.0.4 | Robinson D, Hayes A, Couch S (2023). _broom: Convert Statistical Objects into Tidy Tibbles_. R package version 1.0.4, <https://CRAN.R-project.org/package=broom>. |
| conflicted | 1.2.0 | Wickham H (2023). _conflicted: An Alternative Conflict Resolution Strategy_. R package version 1.2.0, <https://CRAN.R-project.org/package=conflicted>. |
| cli | 3.6.1 | Cs√°rdi G (2023). _cli: Helpers for Developing Command Line Interfaces_. R package version 3.6.1, <https://CRAN.R-project.org/package=cli>. |
| dbplyr | 2.3.2 | Wickham H, Girlich M, Ruiz E (2023). _dbplyr: A 'dplyr' Back End for Databases_. R package version 2.3.2, <https://CRAN.R-project.org/package=dbplyr>. |
| dplyr | 1.1.2 | Wickham H, Fran√ßois R, Henry L, M√ºller K, Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.2, <https://CRAN.R-project.org/package=dplyr>. |
| dtplyr | 1.3.1 | Wickham H, Girlich M, Fairbanks M, Dickerson R (2023). _dtplyr: Data Table Back-End for 'dplyr'_. R package version 1.3.1, <https://CRAN.R-project.org/package=dtplyr>. |
| forcats | 1.0.0 | Wickham H (2023). _forcats: Tools for Working with Categorical Variables (Factors)_. R package version 1.0.0, <https://CRAN.R-project.org/package=forcats>. |
| ggplot2 | 3.4.2 | Wickham H (2016). _ggplot2: Elegant Graphics for Data Analysis_. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>. |
| googledrive | 2.1.0 | D'Agostino McGowan L, Bryan J (2023). _googledrive: An Interface to Google Drive_. R package version 2.1.0, <https://CRAN.R-project.org/package=googledrive>. |
| googlesheets4 | 1.1.0 | Bryan J (2023). _googlesheets4: Access Google Sheets using the Sheets API V4_. R package version 1.1.0, <https://CRAN.R-project.org/package=googlesheets4>. |
| haven | 2.5.2 | Wickham H, Miller E, Smith D (2023). _haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files_. R package version 2.5.2, <https://CRAN.R-project.org/package=haven>. |
| hms | 1.1.3 | M√ºller K (2023). _hms: Pretty Time of Day_. R package version 1.1.3, <https://CRAN.R-project.org/package=hms>. |
| httr | 1.4.6 | Wickham H (2023). _httr: Tools for Working with URLs and HTTP_. R package version 1.4.6, <https://CRAN.R-project.org/package=httr>. |

| Package | Version | Citation |
|---|---|---|
| jsonlite | 1.8.4 | Ooms J (2014). ,ÄúThe jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects.,Äù _arXiv:1403.2805 [stat.CO]_. <https://arxiv.org/abs/1403.2805>. |
| lubridate | 1.9.2 | Grolemund G, Wickham H (2011). ,ÄúDates and Times Made Easy with lubridate.,Äù _Journal of Statistical Software_, *40*(3), 1-25. <https://www.jstatsoft.org/v40/i03/>. |
| magrittr | 2.0.3 | Bache S, Wickham H (2022). _magrittr: A Forward-Pipe Operator for R_. R package version 2.0.3, <https://CRAN.R-project.org/package=magrittr>. |
| modelr | 0.1.11 | Wickham H (2023). _modelr: Modelling Functions that Work with the Pipe_. R package version 0.1.11, <https://CRAN.R-project.org/package=modelr>. |
| pillar | 1.9.0 | M√ºller K, Wickham H (2023). _pillar: Coloured Formatting for Columns_. R package version 1.9.0, <https://CRAN.R-project.org/package=pillar>. |
| purrr | 1.0.1 | Wickham H, Henry L (2023). _purrr: Functional Programming Tools_. R package version 1.0.1, <https://CRAN.R-project.org/package=purrr>. |
| ragg | 1.2.5 | Pedersen T, Shemanarev M (2023). _ragg: Graphic Devices Based on AGG_. R package version 1.2.5, <https://CRAN.R-project.org/package=ragg>. |
| readr | 2.1.4 | Wickham H, Hester J, Bryan J (2023). _readr: Read Rectangular Text Data_. R package version 2.1.4, <https://CRAN.R-project.org/package=readr>. |
| readxl | 1.4.2 | Wickham H, Bryan J (2023). _readxl: Read Excel Files_. R package version 1.4.2, <https://CRAN.R-project.org/package=readxl>. |
| reprex | 2.0.2 | Bryan J, Hester J, Robinson D, Wickham H, Dervieux C (2022). _reprex: Prepare Reproducible Example Code via the Clipboard_. R package version 2.0.2, <https://CRAN.R-project.org/package=reprex>. |
| rlang | 1.1.1 | Henry L, Wickham H (2023). _rlang: Functions for Base Types and Core R and 'Tidyverse' Features_. R package version 1.1.1, <https://CRAN.R-project.org/package=rlang>. |
| rstudioapi | 0.14 | Ushey K, Allaire J, Wickham H, Ritchie G (2022). _rstudioapi: Safely Access the RStudio API_. R package version 0.14, <https://CRAN.R-project.org/package=rstudioapi>. |
| rvest | 1.0.3 | Wickham H (2022). _rvest: Easily Harvest (Scrape) Web Pages_. R package version 1.0.3, <https://CRAN.R-project.org/package=rvest>. |

| Package | Version | Citation |
|---|---|---|
| stringr | 1.5.0 | Wickham H (2022). _stringr: Simple, Consistent Wrappers for Common String Operations_. R package version 1.5.0, <https://CRAN.R-project.org/package=stringr>. |
| tibble | 3.2.1 | M√ºller K, Wickham H (2023). _tibble: Simple Data Frames_. R package version 3.2.1, <https://CRAN.R-project.org/package=tibble>. |
| tidyr | 1.3.0 | Wickham H, Vaughan D, Girlich M (2023). _tidyr: Tidy Messy Data_. R package version 1.3.0, <https://CRAN.R-project.org/package=tidyr>. |
| xml2 | 1.3.4 | Wickham H, Hester J, Ooms J (2023). _xml2: Parse XML_. R package version 1.3.4, <https://CRAN.R-project.org/package=xml2>. |
| tidyverse | 2.0.0 | Wickham H, Averick M, Bryan J, Chang W, McGowan LD, Fran√ßois R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, M√ºller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). ‚ÄúWelcome to the tidyverse.‚Äù _Journal of Open Source Software_, *4*(43), 1686. doi:10.21105/joss.01686 <https://doi.org/10.21105/joss.01686>. |

# BIBLIOGRAPHY


Adnan, K., Akbar, R., Khor, S. W., & Ali, A. B. A. (2020). Role and Challenges of Unstructured Big Data in Healthcare. In N. Sharma, A. Chakrabarti, & V. E. Balas, *Data Management, Analytics and Innovation* Singapore.

Ahn, E. H., Hirohata, K., Kohrn, B. F., Fox, E. J., Chang, C. C., & Loeb, L. A. (2015). Detection of Ultra-Rare Mitochondrial Mutations in Breast Stem Cells by Duplex Sequencing. *PLoS One*, *10*(8), e0136216. https://doi.org/10.1371/journal.pone.0136216

Akiyama, M., Kyoizumi, S., Hirai, Y., Kusunoki, Y., Iwamoto, K. S., & Nakamura, N. (1995). Mutation frequency in human blood cells increases with age. *Mutat Res*, *338*(1-6), 141-149. https://doi.org/10.1016/0921-8734(95)00019-3

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nat Genet*, *47*(12), 1402-1407. https://doi.org/10.1038/ng.3441

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Ng, A. W. T., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., Getz, G., Rozen, S. G., & Stratton, M. R. (2018). The Repertoire of Mutational Signatures in Human Cancer. *bioRvix*.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., . . . PedBrain, I. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415-421. https://doi.org/10.1038/nature12477

Alexandrov, L. B., & Stratton, M. R. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev*, *24*, 52-60. https://doi.org/10.1016/j.gde.2013.11.014

APPA . *2019–2020 APPA National Pet Owners Survey Statistics: Pet Ownership and Annual Expenses.* American Pet Products Association; Stamford, CT, USA: 2020.

American Veterinary Medical Association. *AVMA Pet Ownership and Demographics Sourcebook, 2017–2018 ed.* American Veterinary Medical Association; Schaumburg, IL, USA: 2018.

Baer, C., Miyamoto, M., & Denver, D. (2007). Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature reviews. Genetics*, *8*(8). https://doi.org/10.1038/nrg2158

Banfield Pet Hospital state of pet health. [(accessed on 26 March 2021)]; Available online: https://www.banfield.com/Home/pet-health/State-of-pet-health

Bartlett, P. C. a. (2010). Disease surveillance and referral bias in the veterinary medical database. *Preventive Veterinary Medicine*, *94*(3-4), 264--271 , pmid = 20129684. https://doi.org/10.1016/J.PREVETMED.2010.01.007

Bonnett, B. N., Egenvall, A., Hedhammar, A., & Olson, P. (2005). Mortality in over 350,000 insured Swedish dogs from 1995-2000: I. Breed-, gender-, age- and

cause-specific rates. *Acta Vet Scand*, *46*(3), 105-120. https://doi.org/10.1186/1751-0147-46-105

Boyko, A. R. (2011). The domestic dog: man's best friend in the genomic era. *Genome Biol*, *12*(2), 216. https://doi.org/10.1186/gb-2011-12-2-216

Boyko, A. R., Quignon, P., Li, L., Schoenebeck, J. J., Degenhardt, J. D., Lohmueller, K. E., Zhao, K., Brisbin, A., Parker, H. G., vonHoldt, B. M., Cargill, M., Auton, A., Reynolds, A., Elkahloun, A. G., Castelhano, M., Mosher, D. S., Sutter, N. B., Johnson, G. S., Novembre, J., . . . Ostrander, E. A. (2010). A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol*, *8*(8), e1000451. https://doi.org/10.1371/journal.pbio.1000451

Cadieu, E., Neff, M. W., Quignon, P., Walsh, K., Chase, K., Parker, H. G., Vonholdt, B. M., Rhue, A., Boyko, A., Byers, A., Wong, A., Mosher, D. S., Elkahloun, A. G., Spady, T. C., Andre, C., Lark, K. G., Cargill, M., Bustamante, C. D., Wayne, R. K., & Ostrander, E. A. (2009). Coat variation in the domestic dog is governed by variants in three genes. *Science*, *326*(5949), 150-153. https://doi.org/10.1126/science.1177808

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T. H. H., Sanders, M. A., Lawson, A. R. J., Harvey, L. M. R., Bhosle, S., Jones, D., Alcantara, R. E., Butler, T. M., Hooks, Y., Roberts, K., Anderson, E., Lunn, S., Flach, E., Spiro, S., Januszczak, I., . . . Martincorena, I. (2022). Somatic mutation rates scale with lifespan across mammals. *Nature*, *604*(7906), 517-524. https://doi.org/10.1038/s41586-022-04618-z

Campisi, J. (2008). Aging and cancer cell biology, 2008 [https://doi.org/10.1111/j.1474-9726.2008.00383.x]. *Aging Cell*, *7*(3), 281-284. https://doi.org/https://doi.org/10.1111/j.1474-9726.2008.00383.x

Cervantes, R. B., Stringer, J. R., Shao, C., Tischfield, J. A., & Stambrook, P. J. (2002). Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proc Natl Acad Sci U S A*, *99*(6), 3586-3590. https://doi.org/10.1073/pnas.062527199

Chawla, N. V., & Davis, D. A. (2013). Bringing big data to personalized healthcare: a patient-centered framework. *Journal of general internal medicine*, *28 Suppl 3*(Suppl 3). https://doi.org/10.1007/s11606-013-2455-8

Clements, D. N., Handel, I. G., Rose, E., Querry, D., Pugh, C. A., Ollier, W. E. R., Morgan, K. L., Kennedy, L. J., Sampson, J., Summers, K. M., & de Bronsvoort, B. M. C. (2013). Dogslife: A web-based longitudinal study of Labrador Retriever health in the UK. *BMC Veterinary Research*, *9*. https://doi.org/10.1186/1746-6148-9-13

Creevy, K. E., Akey, J. M., Kaeberlein, M., Promislow, D. E. L., & Dog Aging Project, C. (2022). An open science study of ageing in companion dogs. *Nature*, *602*(7895), 51-57. https://doi.org/10.1038/s41586-021-04282-9

Da Costa, R. E. P., Kinsman, R. H., Owczarczak-Garstecka, S. C., Casey, R. A., Tasker, S., Knowles, T. G., Woodward, J. L., Lord, M. S., & Murray, J. K. (2022). Age of sexual maturity and factors associated with neutering dogs in the UK and the Republic of Ireland. *Vet Rec*, *191*(6), e1265. https://doi.org/10.1002/vetr.1265

Davis, B. W., & Ostrander, E. A. (2014). Domestic dogs and cancer research: a breed-based genomics approach. *ILAR J*, *55*(1), 59-68. https://doi.org/10.1093/ilar/ilu017

Dobson, J. M., Samuel, S., Milstein, H., Rogers, K., & Wood, J. L. N. (2002). Canine neoplasia in the UK: Estimates of incidence rates from a population of insured dogs. *Journal of Small Animal Practice*, *43*(6), 240--246 , pmid = 12074288. https://doi.org/10.1111/J.1748-5827.2002.TB00066.X

Egenvall, A., Bonnett, B. N., Olson, P., & Hedhammar, k. (1998). Validation of computerized Swedish dog and cat insurance data against veterinary practice records. *Preventive Veterinary Medicine*, *36*(1), 51--65 , pmid = 9677627. https://doi.org/10.1016/S0167-5877(98)00073-7

Egenvall, A., Ndtvedt, A., Penell, J., Gunnarsson, L., & Bonnett, B. N. (2009). Insurance data for research in companion animals: benefits and limitations. *Acta veterinaria Scandinavica*, *51*, 42 , pmid = 19874612. https://doi.org/10.1186/1751-0147-51-42

Elvers, I., Turner-Maier, J., Swofford, R., Koltookian, M., Johnson, J., Stewart, C., Zhang, C. Z., Schumacher, S. E., Beroukhim, R., Rosenberg, M., Thomas, R., Mauceli, E., Getz, G., Palma, F. D., Modiano, J. F., Breen, M., Lindblad-Toh, K., & Alföldi, J. (2015). Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res*, *25*(11), 1634-1645. https://doi.org/10.1101/gr.194449.115

Fleming, J. M., Creevy, K. E., & Promislow, D. E. (2011). Mortality in north american dogs from 1984 to 2004: an investigation into age-, size-, and breed-related causes of death. *J Vet Intern Med*, *25*(2), 187-198. https://doi.org/10.1111/j.1939-1676.2011.0695.x

Fontana, L., Partridge, L., & Longo, V. D. (2010). Extending healthy life span--from yeast to humans. *Science*, *328*(5976), 321-326. https://doi.org/10.1126/science.1172539

Frantz, L. A. F., Bradley, D. G., Larson, G., & Orlando, L. (2020). Animal domestication in the era of ancient genomics. *Nature Reviews Genetics*, *21*(8), 449-460. https://doi.org/10.1038/s41576-020-0225-0

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., . . . Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. *PLoS Genet*, *10*(1), e1004016. https://doi.org/10.1371/journal.pgen.1004016

Galibert, F., Quignon, P., Hitte, C., & Andre, C. (2011). Toward understanding dog evolutionary and domestication history. *C R Biol*, *334*(3), 190-196. https://doi.org/10.1016/j.crvi.2010.12.011

Genovese, G., Kahler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landen, M., Hoglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., . . . McCarroll, S. A. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*, *371*(26), 2477-2487. https://doi.org/10.1056/NEJMoa1409405

Glickman, L. T., Moore, G. E., Glickman, N. W., Caldanaro, R. J., Aucoin, D., & Lewis, H. B. (2006). Purdue University-Banfield National Companion Animal Surveillance Program for emerging and zoonotic diseases. *Vector-Borne and Zoonotic Diseases*, *6*(1), 14--23 , pmid = 16584323. https://doi.org/10.1089/VBZ.2006.6.14

Goto, M. (1997). Hierarchical deterioration of body systems in Werner's syndrome: implications for normal ageing. *Mech Ageing Dev*, *98*(3), 239-254. https://doi.org/10.1016/s0047-6374(97)00111-5

Grist, S. A., McCarron, M., Kutlaca, A., Turner, D. R., & Morley, A. A. (1992). In vivo human somatic mutation: frequency and spectrum with age. *Mutat Res*, *266*(2), 189-196. https://doi.org/10.1016/0027-5107(92)90186-6

Grossglauser, M., & Saner, H. (2014). Data-driven healthcare: from patterns to actions. *European journal of preventive cardiology*, *21*(2 Suppl). https://doi.org/10.1177/2047487314552755

Gupta, N. K., & Rohil, M. K. (2020). Big Data Security Challenges and Preventive Solutions. In N. Sharma, A. Chakrabarti, & V. E. Balas, *Data Management, Analytics and Innovation* Singapore.

Guy, M. K., Page, R. L., Jensen, W. A., & Olson, P. N. a. (2015). The golden retriever lifetime study: Establishing an observational cohort study with translational relevance for human health. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1673 , pmid = 26056371). https://doi.org/10.1098/RSTB.2014.0230

Hoffman, J. M., Creevy, K. E., Franks, A., O'Neill, D. G., & Promislow, D. E. L. (2018). The companion dog as a model for human aging and mortality. *Aging Cell*, *17*(3), e12737. https://doi.org/10.1111/acel.12737

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. A., Ladenvall, C., Getz, G., . . . Ebert, B. L. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*, *371*(26), 2488-2498. https://doi.org/10.1056/NEJMoa1408617

Jimenez, A. G. (2016). Physiological underpinnings in life-history trade-offs in man's most popular selection experiment: the dog. *J Comp Physiol B*, *186*(7), 813-827. https://doi.org/10.1007/s00360-016-1002-4

Kakiuchi, N., & Ogawa, S. (2021). Clonal expansion in non-cancer tissues. *Nat Rev Cancer*, *21*(4), 239-256. https://doi.org/10.1038/s41568-021-00335-3

Karlsson, E. K., Baranowska, I., Wade, C. M., Salmon Hillbertz, N. H., Zody, M. C., Anderson, N., Biagi, T. M., Patterson, N., Pielberg, G. R., Kulbokas, E. J., 3rd, Comstock, K. E., Keller, E. T., Mesirov, J. P., von Euler, H., Kampe, O., Hedhammar, A., Lander, E. S., Andersson, G., Andersson, L., & Lindblad-Toh, K. (2007). Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet*, *39*(11), 1321-1328. https://doi.org/10.1038/ng.2007.10

Karlsson, E. K., Sigurdsson, S., Ivansson, E., Thomas, R., Elvers, I., Wright, J., Howald, C., Tonomura, N., Perloski, M., Swofford, R., Biagi, T., Fryc, S., Anderson, N., Courtay-Cahen, C., Youell, L., Ricketts, S. L., Mandlebaum, S., Rivera, P., von Euler, H., . . . Lindblad-Toh, K. (2013). Genome-wide analyses implicate 33 loci

in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biol*, *14*(12), R132. https://doi.org/10.1186/gb-2013-14-12-r132

Karyadi, D. M., Karlins, E., Decker, B., vonHoldt, B. M., Carpintero-Ramirez, G., Parker, H. G., Wayne, R. K., & Ostrander, E. A. (2013). A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. *PLoS Genet*, *9*(3), e1003409. https://doi.org/10.1371/journal.pgen.1003409

Kennedy, S. R., Loeb, L. A., & Herr, A. J. (2012). Somatic mutations in aging, cancer and neurodegeneration. *Mech Ageing Dev*, *133*(4), 118-126. https://doi.org/10.1016/j.mad.2011.10.009

Kennedy, S. R., Salk, J. J., Schmitt, M. W., & Loeb, L. A. (2013). Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLoS Genet*, *9*(9), e1003794. https://doi.org/10.1371/journal.pgen.1003794

Kennedy, S. R., Schmitt, M. W., Fox, E. J., Kohrn, B. F., Salk, J. J., Ahn, E. H., Prindle, M. J., Kuong, K. J., Shen, J. C., Risques, R. A., & Loeb, L. A. (2014). Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc*, *9*(11), 2586-2606. https://doi.org/10.1038/nprot.2014.170

Kenyon, C. J. (2010). The genetics of ageing. *Nature*, *464*(7288), 504-512. https://doi.org/10.1038/nature08980

Kim, R. N., Kim, D. S., Choi, S. H., Yoon, B. H., Kang, A., Nam, S. H., Kim, D. W., Kim, J. J., Ha, J. H., Toyoda, A., Fujiyama, A., Kim, A., Kim, M. Y., Park, K. H., Lee, K. S., & Park, H. S. (2012). Genome analysis of the domestic dog (Korean Jindo) by massively parallel sequencing. *DNA Res*, *19*(3), 275-287. https://doi.org/10.1093/dnares/dss011

Kirkwood, T. B. (1977). Evolution of ageing. *Nature*, *270*(5635), 301-304. https://doi.org/10.1038/270301a0

Koh, G., Degasperi, A., Zou, X., Momen, S., & Nik-Zainal, S. (2021). Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer*, *21*(10), 619-637. https://doi.org/10.1038/s41568-021-00377-7

Kol, A., Arzi, B., Athanasiou, K. A., Farmer, D. L., Nolta, J. A., Rebhun, R. B., Chen, X., Griffiths, L. G., Verstraete, F. J. M., Murphy, C. J., & Borjesson, D. L. (2015). Companion animals: Translational scientist's new best friends. *Science Translational Medicine*, *7*(308 , pmid = 26446953). https://doi.org/10.1126/SCITRANSLMED.AAA9116

Kraus, C., Pavard, S., & Promislow, D. E. (2013). The size-life span trade-off decomposed: why large dogs die young. *Am Nat*, *181*(4), 492-505. https://doi.org/10.1086/669665

Kraus, C., Snyder-Mackler, N., & Promislow, D. E. L. (2023). How size and genetic diversity shape lifespan across breeds of purebred dogs. *Geroscience*, *45*(2), 627-643. https://doi.org/10.1007/s11357-022-00653-w

Krimmel, J. D., Schmitt, M. W., Harrell, M. I., Agnew, K. J., Kennedy, S. R., Emond, M. J., Loeb, L. A., Swisher, E. M., & Risques, R. A. (2016). Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic TP53 mutations in noncancerous tissues. *Proc Natl Acad Sci U S A*, *113*(21), 6005-6010. https://doi.org/10.1073/pnas.1601311113

Kucab, J. E., Zou, X., Morganella, S., Joel, M., Nanda, A. S., Nagy, E., Gomez, C., Degasperi, A., Harris, R., Jackson, S. P., Arlt, V. M., Phillips, D. H., & Nik-Zainal, S. (2019). A Compendium of Mutational Signatures of Environmental Agents. *Cell*, *177*(4), 821-836 e816. https://doi.org/10.1016/j.cell.2019.03.001

Lewis, T. W., Wiles, B. M., Llewellyn-Zaidi, A. M., Evans, K. M., & O'Neill, D. G. (2018). Longevity and mortality in Kennel Club registered dog breeds in the UK in 2014. *Canine Genetics and Epidemiology*, *5*(1), 10. https://doi.org/10.1186/s40575-018-0066-8

Li, Y., Wang, G. D., Wang, M. S., Irwin, D. M., Wu, D. D., & Zhang, Y. P. (2014). Domestication of the dog from the wolf was promoted by enhanced excitatory synaptic plasticity: a hypothesis. *Genome Biol Evol*, *6*(11), 3115-3121. https://doi.org/10.1093/gbe/evu245

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., Zody, M. C., Mauceli, E., Xie, X., Breen, M., Wayne, R. K., Ostrander, E. A., Ponting, C. P., Galibert, F., Smith, D. R., DeJong, P. J., . . . Lander, E. S. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, *438*(7069), 803-819. https://doi.org/10.1038/nature04338

Loeb, L. A. (2016). Human Cancers Express a Mutator Phenotype: Hypothesis, Origin, and Consequences. *Cancer Res*, *76*(8), 2057-2059. https://doi.org/10.1158/0008-5472.CAN-16-0794

Lynch, M. (2010). Evolution of the mutation rate. *Trends Genet*, *26*(8), 345-352. https://doi.org/10.1016/j.tig.2010.05.003

Lynch, M., Ackerman, M. S., Gout, J. F., Long, H., Sung, W., Thomas, W. K., & Foster, P. L. (2016). Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet*, *17*(11), 704-714. https://doi.org/10.1038/nrg.2016.104

Marsden, C. D., Ortega-Del Vecchyo, D., O'Brien, D. P., Taylor, J. F., Ramirez, O., Vilà, C., Marques-Bonet, T., Schnabel, R. D., Wayne, R. K., & Lohmueller, K. E. (2016). Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc Natl Acad Sci U S A*, *113*(1), 152-157. https://doi.org/10.1073/pnas.1512501113

McEwen, S. A., & Collignon, P. J. (2018). Antimicrobial Resistance: a One Health Perspective. *Microbiology Spectrum*, *6*(2 , pmid = 29600770). https://doi.org/10.1128/MICROBIOLSPEC.ARBA-0009-2017

McGreevy, P., Thomson, P., Dhand, N. K., Raubenheimer, D., Masters, S., Mansfield, C. S., Baldwin, T., Magalhaes, R. J. S., Rand, J., Hill, P., Peaston, A., Gilkerson, J., Combs, M., Raidal, S., Irwin, P., Irons, P., Squires, R., Brodbelt, D., & Hammond, J. (2017). Vetcompass Australia: A national big data collection system for veterinary science. *Animals*, *7*(10). https://doi.org/10.3390/ANI7100074

Michell, A. R. (1999). Longevity of British breeds of dog and its relationships with sex, size, cardiovascular variables and disease. *Vet Rec*, *145*(22), 625-629. https://doi.org/10.1136/vr.145.22.625

Modiano, J. F., Breen, M., Burnett, R. C., Parker, H. G., Inusah, S., Thomas, R., Avery, P. R., Lindblad-Toh, K., Ostrander, E. A., Cutter, G. C., & Avery, A. C. (2005). Distinct B-cell and T-cell lymphoproliferative disease prevalence among dog

breeds indicates heritable risk. *Cancer Res*, *65*(13), 5654-5661. https://doi.org/10.1158/0008-5472.CAN-04-4613

North American Pet Health Insurance Association (NAPHIA) [(accessed on 26 March 2021)]; Available online: https://naphia.org/industry-data/section-2-total-pets-insured/

O'Neill, D. G., Church, D. B., McGreevy, P. D., Thomson, P. C., & Brodbelt, D. C. (2014). Approaches to canine health surveillance. *Canine Genetics and Epidemiology*, *1*(1). https://doi.org/10.1186/2052-6687-1-2

Orgel, L. E. (1963). The maintenance of the accuracy of protein synthesis and its relevance to ageing. *Proc Natl Acad Sci U S A*, *49*(4), 517-521. https://doi.org/10.1073/pnas.49.4.517

Owczarek-Lipska, M., Jagannathan, V., Drogemuller, C., Lutz, S., Glanemann, B., Leeb, T., & Kook, P. H. (2013). A frameshift mutation in the cubilin gene (CUBN) in Border Collies with Imerslund-Grasbeck syndrome (selective cobalamin malabsorption). *PLoS One*, *8*(4), e61144. https://doi.org/10.1371/journal.pone.0061144

Parker, H. G. (2012). Genomic analyses of modern dog breeds. *Mamm Genome*, *23*(1-2), 19-27. https://doi.org/10.1007/s00335-011-9387-6

Parker, H. G., Kim, L. V., Sutter, N. B., Carlson, S., Lorentzen, T. D., Malek, T. B., Johnson, G. S., DeFrance, H. B., Ostrander, E. A., & Kruglyak, L. (2004). Genetic structure of the purebred domestic dog. *Science*, *304*(5674), 1160-1164. https://doi.org/10.1126/science.1097406

Patterson, D. F. (2000). Companion animal medicine in the age of medical genetics. *J Vet Intern Med*, *14*(1), 1-9. https://www.ncbi.nlm.nih.gov/pubmed/10668810

Paynter, A.N.; Dunbar, M.D.; Creevy, K.E.; Ruple, A. Veterinary Big Data: When Data Goes to the Dogs. *Animals* 2021, *11*, 1872. https://doi.org/10.3390/ani11071872

Perri, A. R., Feuerborn, T. R., Frantz, L. A. F., Larson, G., Malhi, R. S., Meltzer, D. J., & Witt, K. E. (2021). Dog domestication and the dual dispersal of people and dogs into the Americas. *Proceedings of the National Academy of Sciences*, *118*(6), e2010083118. https://doi.org/doi:10.1073/pnas.2010083118

Plant, J. D., Lund, E. M., & Yang, M. (2011). A case-control study of the risk factors for canine juvenile-onset generalized demodicosis in the USA. *Veterinary Dermatology*, *22*(1), 95--99 , pmid = 20707860. https://doi.org/10.1111/J.1365-3164.2010.00922.X

Plassais, J., Kim, J., Davis, B. W., Karyadi, D. M., Hogan, A. N., Harris, A. C., Decker, B., Parker, H. G., & Ostrander, E. A. (2019). Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat Commun*, *10*(1), 1489. https://doi.org/10.1038/s41467-019-09373-w

Priester, W. A., & McKay, F. W. (1980). The occurrence of tumors in domestic animals. *Natl Cancer Inst Monogr*(54), 1-210. https://www.ncbi.nlm.nih.gov/pubmed/7254313

Proschowsky, H. F., Rugbjerg, H., & Ersboll, A. K. (2003). Mortality of purebred and mixed-breed dogs in Denmark. *Prev Vet Med*, *58*(1-2), 63-74. https://doi.org/10.1016/s0167-5877(03)00010-2

Pugh, C. A., de Bronsvoort, B. M. C., Handel, I. G., Summers, K. M., & Clements, D. N. (2015). Dogslife: A cohort study of Labrador Retrievers in the UK. *Preventive Veterinary Medicine*, *122*(4), 426--435 , pmid = 26189582. https://doi.org/10.1016/J.PREVETMED.2015.06.020 , file = :E\:/Brendan F Kohrn/Documents/Mendeley Desktop/Pugh et al. - 2015 - Dogslife A cohort study of Labrador Retrievers in the UK.pdf:pdf

Radford, A., Tierney, Coyne, K. P., Gaskell, R. M., Noble, P. J., Dawson, S., Setzkorn, C., Jones, P. H., Buchan, I. E., Newton, J. R., & Bryan, J. G. E. (2010). Surveillance: Developing a network for small animal disease surveillance. *Veterinary Record*, *167*(13), 472--474 , pmid = 20871079. https://doi.org/10.1136/VR.C5180

Raghavan, M., Glickman, N., Moore, G., Caldanaro, R., Lewis, H., & Glickman, L. (2007). Prevalence of and risk factors for canine tick infestation in the United States, 2002-2004. *Vector-Borne and Zoonotic Diseases*, *7*(1), 65--75 , pmid = 17417959. https://doi.org/10.1089/VBZ.2006.0570

Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L. B., Hood, L., & Galas, D. J. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*, *328*(5978), 636-639. https://doi.org/10.1126/science.1186802

Rowell, J. L., McCarthy, D. O., & Alvarez, C. E. (2011). Dog models of naturally occurring cancer. *Trends Mol Med*, *17*(7), 380-388. https://doi.org/10.1016/j.molmed.2011.02.004

Ruple, A., Jones, M., Simpson, M., & Page, R. (2021). The Golden Retriever Lifetime Study: Assessing factors associated with owner compliance after the first year of enrollment. *Journal of Veterinary Internal Medicine*, *35*(1), 142--149 , pmid = 33191623. https://doi.org/10.1111/JVIM.15921

Ruple, A., MacLean, E., Snyder-Mackler, N., Creevy, K. E., & Promislow, D. (2022). Dog Models of Aging. *Annu Rev Anim Biosci*, *10*, 419-439. https://doi.org/10.1146/annurev-animal-051021-080937

Salk, J. J., Loubet-Senear, K., Maritschnegg, E., Valentine, C. C., Williams, L. N., Higgins, J. E., Horvat, R., Vanderstichele, A., Nachmanson, D., Baker, K. T., Emond, M. J., Loter, E., Tretiakova, M., Soussi, T., Loeb, L. A., Zeillinger, R., Speiser, P., & Risques, R. A. (2019). Ultra-Sensitive TP53 Sequencing for Cancer Detection Reveals Progressive Clonal Selection in Normal Tissue over a Century of Human Lifespan. *Cell Rep*, *28*(1), 132-144.e133. https://doi.org/10.1016/j.celrep.2019.05.109

Salk, J. J., Maritschnegg, E., Speiser, P., Zeillinger, R., Risques, R., & Loeb, L. A. (2017). Abstract 3041: Duplex Sequencing detects cancer-associated mutations arising during normal aging: Clonal evolution over a century of human lifetime. *Cancer Research*, *77*(13_Supplement), 3041-3041. https://doi.org/10.1158/1538-7445.Am2017-3041

Salk, J. J., Schmitt, M. W., & Loeb, L. A. (2018). Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet*, *19*(5), 269-285. https://doi.org/10.1038/nrg.2017.117

Schiffman, J. D., & Breen, M. (2015). Comparative oncology: what dogs and other species can teach us about humans with cancer. *Philos Trans R Soc Lond B Biol Sci*, *370*(1673). https://doi.org/10.1098/rstb.2014.0231

Schmidt, C. W. (2019). Environmental Factors in Successful Aging: The Potential Impact of Air Pollution. *Environ Health Perspect*, *127*(10), 102001. https://doi.org/10.1289/EHP4579

Schmitt, M. W., Fox, E. J., Prindle, M. J., Reid-Bayliss, K. S., True, L. D., Radich, J. P., & Loeb, L. A. (2015). Sequencing small genomic targets with high efficiency and extreme accuracy. *Nat Methods*, *12*(5), 423-425. https://doi.org/10.1038/nmeth.3351

Schmitt, M. W., Kennedy, S. R., Salk, J. J., Fox, E. J., Hiatt, J. B., & Loeb, L. A. (2012). Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A*, *109*(36), 14508-14513. https://doi.org/10.1073/pnas.1208715109

Sirilla, J., Copelan, E. A., Elder, P. J., & Davis, R. (2006). Creation of a research database to comply with the HIPPA Privacy Act. *Biology of Blood and Marrow Transplantation*, *12*(2), 149. https://doi.org/10.1016/J.BBMT.2005.11.459 , file = :E\:/Brendan F Kohrn/Documents/Mendeley Desktop/Sirilla et al. - 2006 - Creation of a research database to comply with the HIPPA Privacy Act.pdf:pdf

Small Animal Veterinary Surveillance Network (SAVSNET) [(accessed on 1 June 2021)]; Available online: https://www.liverpool.ac.uk/savsnet/using-savsnet-data-for-research/

Starkey, M. P., Scase, T. J., Mellersh, C. S., & Murphy, S. (2005). Dogs really are man's best friend - Canine genomics has applications in veterinary and human medicine! *Briefings in Functional Genomics and Proteomics*, *4*(2), 112--128 , pmid = 16102268. https://doi.org/10.1093/BFGP/4.2.112

Steensma, D. P., Bejar, R., Jaiswal, S., Lindsley, R. C., Sekeres, M. A., Hasserjian, R. P., & Ebert, B. L. (2015). Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*, *126*(1), 9-16. https://doi.org/10.1182/blood-2015-03-631747

Sun, Z., Strang, K. D., & Pambel, F. (2020). Privacy and security in the big data paradigm. *Journal of Computer Information Systems*, *60*(2), 146-155. https://doi.org/10.1080/08874417.2017.1418631

Sutter, N. B., Bustamante, C. D., Chase, K., Gray, M. M., Zhao, K., Zhu, L., Padhukasahasram, B., Karlins, E., Davis, S., Jones, P. G., Quignon, P., Johnson, G. S., Parker, H. G., Fretwell, N., Mosher, D. S., Lawler, D. F., Satyaraj, E., Nordborg, M., Lark, K. G., . . . Ostrander, E. A. (2007). A single IGF1 allele is a major determinant of small size in dogs. *Science*, *316*(5821), 112-115. https://doi.org/10.1126/science.1137045

Tarone, L., Barutello, G., Iussich, S., Giacobino, D., Quaglino, E., Buracco, P., Cavallo, F., & Riccardo, F. (2019). Naturally occurring cancers in pet dogs as pre-clinical models for cancer immunotherapy. *Cancer Immunol Immunother*. https://doi.org/10.1007/s00262-019-02360-6

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., . . .

Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*, *47*(D1), D941-D947. https://doi.org/10.1093/nar/gky1015

Tonomura, N., Elvers, I., Thomas, R., Megquier, K., Turner-Maier, J., Howald, C., Sarver, A. L., Swofford, R., Frantz, A. M., Ito, D., Mauceli, E., Arendt, M., Noh, H. J., Koltookian, M., Biagi, T., Fryc, S., Williams, C., Avery, A. C., Kim, J. H., . . . Lindblad-Toh, K. (2015). Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers. *PLoS Genet*, *11*(2), e1004922. https://doi.org/10.1371/journal.pgen.1004922

Trainor, K. J., Wigmore, D. J., Chrysostomou, A., Dempsey, J. L., Seshadri, R., & Morley, A. A. (1984). Mutation frequency in human lymphocytes increases with age. *Mech Ageing Dev*, *27*(1), 83-86. https://doi.org/10.1016/0047-6374(84)90084-8

Turan, Z. G., Parvizi, P., Donertas, H. M., Tung, J., Khaitovich, P., & Somel, M. (2019). Molecular footprint of Medawar's mutation accumulation process in mammalian aging. *Aging Cell*, *18*(4), e12965. https://doi.org/10.1111/acel.12965

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Rosengren Pielberg, G., Sigurdsson, S., Fall, T., Seppala, E. H., Hansen, M. S., Lawley, C. T., Karlsson, E. K., Consortium, L., Bannasch, D., Vila, C., Lohi, H., Galibert, F., Fredholm, M., Haggstrom, J., Hedhammar, A., . . . Webster, M. T. (2011). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet*, *7*(10), e1002316. https://doi.org/10.1371/journal.pgen.1002316

The Veterinary Companion Animal Surveillance System Overview. [(accessed on 27 March 2021)]; Available online: https://www.rvc.ac.uk/vetcompass/about/overview

The Veterinary Medical Databases. [(accessed on 27 March 2021)]; Available online: https://vmdb.org/

Veterinary Information Network. [(accessed on 26 March 2021)]; Available online: https://news.vin.com/default.aspx?pid=210&Id=9305364&useobjecttypeid=10&fromVINNEWSASPX=1

Vijg, J., & Dong, X. (2020). Pathogenic Mechanisms of Somatic Mutation and Genome Mosaicism in Aging. *Cell*, *182*(1), 12-23. https://doi.org/10.1016/j.cell.2020.06.024

Vijg, J., & Suh, Y. (2013). Genome instability and aging. *Annu Rev Physiol*, *75*, 645-668. https://doi.org/10.1146/annurev-physiol-030212-183715

Vonholdt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., Degenhardt, J. D., Boyko, A. R., Earl, D. A., Auton, A., Reynolds, A., Bryc, K., Brisbin, A., Knowles, J. C., Mosher, D. S., Spady, T. C., Elkahloun, A., Geffen, E., Pilot, M., . . . Wayne, R. K. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*, *464*(7290), 898-902. https://doi.org/10.1038/nature08837

Wayne, R. K., & vonHoldt, B. M. (2012). Evolutionary genomics of dog domestication. *Mamm Genome*, *23*(1-2), 3-18. https://doi.org/10.1007/s00335-011-9386-7

Wilcox, B., & Walkowicz, C. (1995). Atlas of Dog Breeds of the World.

Wiles, B. M., Llewellyn-Zaidi, A. M., Evans, K. M., O'Neill, D. G., & Lewis, T. W. (2017). Large-scale survey to estimate the prevalence of disorders for 192 Kennel Club

registered breeds. *Canine Genet Epidemiol*, *4*, 8. https://doi.org/10.1186/s40575-017-0047-3

Williams, G. C. (1957). PLEIOTROPY, NATURAL-SELECTION, AND THE EVOLUTION OF SENESCENCE. *Evolution*, *11*(4), 398-411. https://doi.org/10.1111/j.1558-5646.1957.tb02911.x

Woolley, C. S. C., Handel, I. G., Bronsvoort, B. M., Schoenebeck, J. J., & Clements, D. N. (2020). Is it time to stop sweeping data cleaning under the carpet? A novel algorithm for outlier management in growth data. *PLoS One*, *15*(1 , pmid = 31978151). https://doi.org/10.1371/JOURNAL.PONE.0228154

Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., Ozenberger, B. A., Welch, J. S., Link, D. C., Walter, M. J., Mardis, E. R., Dipersio, J. F., Chen, F., Wilson, R. K., Ley, T. J., & Ding, L. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*, *20*(12), 1472-1478. https://doi.org/10.1038/nm.3733

Yordy, J., Kraus, C., Hayward, J. J., White, M. E., Shannon, L. M., Creevy, K. E., Promislow, D. E. L., & Boyko, A. R. (2020). Body size, inbreeding, and lifespan in domestic dogs. *Conserv Genet*, *21*(1), 137-148. https://doi.org/10.1007/s10592-019-01240-x

Zandvliet, M. (2016). Canine lymphoma: a review. *Vet Q*, *36*(2), 76-104. https://doi.org/10.1080/01652176.2016.1152633

# VITA

Ashley Paynter as born and raised in New York City. Growing up, Ashley always loved animals, plants, the rainforest, and asking questions about the world. In high school, Ashley participated in a research internship for students at the American Museum of Natural History, which was where her love for research began. Ashley attended Binghamton University in 2012, where she studied Lyme disease and completed a Bachelor of Science in Biology. Ashley also continued to spend summers working at the American Museum of Natural History where she studied the genetics of a contagious cancer in bivalves.

In 2017, Ashley started a doctoral degree at the University of Washington, in the department of Biology. She soon joined the lab of Daniel Promislow, who leads the Dog Aging Project. Ashley spent her years in graduate school studying comparative oncology, cancer genetics, and evolutionary genetics. In 2020, Ashley started her own podcast called Decolonizing Science.