Characterization of the Genetic Landscape of Feline Oral Squamous Cell Carcinoma

A Thesis presented to

the Faculty of the Graduate School

at the University of Missouri – Columbia

In Partial Fulfilment

Of the Requirements for the Degree

Master of Science

By

Alana R. Rodney

Dr. Wes Warren, Thesis Advisor

May 2022
The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

CHARACTERIZATION OF THE GENETIC LANDSCAPE OF FELINE ORAL SQUAMOUS CELL CARCINOMA

Presented by Alana R. Rodney
A candidate for the degree MASTER OF SCIENCE
We hereby certify that in our opinion it is worthy of acceptance

Wes Warren, PhD – Thesis Advisor

Leslie A. Lyons, PhD

Owen Skinner, BVSc, DECVS, DACVS-SA, MRCVS

Chris Elsik, PhD
Acknowledgements

First, I would like to thank Dr. Warren for his great mentorship over the past several years. Learning from you has been a great honor, and I am excited to take these skills to complete a PhD this fall. I would also like to thank my committee, Dr. Lyons whose mentorship, support, and encouragement has been invaluable, especially during the last months of preparation. I would also like to thank Dr. Owen Skinner for his help through the writing process and consistent constructive and helpful feedback. I very much appreciated Dr. Elsik’s calm cool and collected teaching style, as script writing and bioinformatic can be very stressful, and it was great to have a soft reassuring pace to land during frustrating times. Finally, I want to thank my lab mates, friends, and family, because without their support, I would not be where I am today.
**TABLE OF CONTENTS**

ACKNOWLEDGEMENTS........................................................................................................ii

TABLE OF CONTENTS........................................................................................................iii

LIST OF TABLES..................................................................................................................v

LIST OF FIGURES.................................................................................................................vi

LIST OF ABBREVIATIONS.....................................................................................................vii

ABSTRACT.............................................................................................................................viii

CHAPTERS

1. INTRODUCTION..............................................................................................................1

2. LITERATURE REVIEW....................................................................................................7
   2.1 Feline Cancer Landscape.........................................................................................7
   2.2 Lymphoma...............................................................................................................7
   2.3 Soft Tissue Sarcoma...............................................................................................9
   2.4 Feline Mammary Cancer.......................................................................................12
   2.5 Oral Squamous Cell Carcinoma.............................................................................14
   2.6 Pulmonary Carcinoma............................................................................................16
   2.7 Conclusion..............................................................................................................17

3. EXOME DESIGN............................................................................................................19
   3.1 Introduction .............................................................................................................19
   3.2 Methods..................................................................................................................21
LIST OF TABLES

Table

3.1 Description and Diseases of 41 cats for WES evaluation ....................... 27

3.2 Summary of Metrics Across both Cohorts .......................................... 30

4.1 Cohort Characteristics ........................................................................... 43

4.2 Variants in Common with Known HNSCC Causes variants ....................... 51
LIST OF FIGURES

FIGURE

1.1 Summary of Molecular Markers................................. 5
3.1 Proportion of Bases Covered with Exome Probe............... 31
4.1 Pathology of Oral Squamous Carcinoma.......................... 43
4.2 Variants called with GATK-Mutect2.......................... 47
4.3 Number of Non-Synonymous Mutations.......................... 48
4.4 Location of p53 Mutations........................................ 49
LIST OF ABBREVIATIONS

WGS: Whole Genome Sequencing
WES: Whole Exome Sequencing
FeLV: Feline Leukemia Virus
FIV: Feline Immunotherapy Virus
STS: Soft Tissue Sarcoma
FeSV: Feline Sarcoma Virus
FMC: Feline Mammary Cancer
ISS: Injection Site Sarcoma spelling?
FOSCC: Feline Oral Squamous Cell Carcinoma
HNSCC: Human Head and Neck Squamous Cell Carcinoma
PKD: Polycystic Kidney Disease
ADPKD: Autosomal Dominant Polycystic Kidney Disease
TMB: Tumor Mutational Burden
ABSTRACT

Over 94 million domestic cats are susceptible to cancers and other common and rare diseases. While cancer treatment in cats increasingly mirrors that available in humans, treatment failures are more frequent. There are no FDA or USDA approved cancer drugs for cats and a paucity of cancer treatments beyond surgery, radiotherapy, and cytotoxic chemotherapy indicate an urgent need to define the molecular properties of aggressive feline cancers including common cancers such as soft tissue sarcoma, mammary carcinoma, lymphoma, and feline oral squamous cell carcinoma. Whole exome sequencing (WES) is a proven strategy to study these disease-causing variants in humans and other organisms. To examine the effectiveness of WES in the study of feline cancer and Mendelian diseases, whole exome sequencing was conducted on 41 cats with known and unknown genetic diseases and traits, of which ten cats had matching whole genome sequence (WGS) data available, used to validate WES performance. Within the 41 cats, we identified 31 previously known causal variants and discovered new gene candidate variants, including novel missense variance for polycystic kidney disease and atrichia in the Peterbald cat. The WES data was sufficient to identify novel gene candidate alleles for diseases and traits in a feline model. Feline oral squamous cell carcinoma (FOSCC) is a cancer of the squamous cell lining in the oral cavity and represents up to 80% of all oral cancers in cats, with a low one-year survival rate of <10%. The cancer pathology associated with feline oral squamous cell carcinoma is similar to human head and neck squamous cell carcinoma (HNSCC), which accounts for 90% of oral cancers in humans. FOSCC may present as a potential model to study HNSCC due to spontaneous formation, similar genetic landscape, pathology, and survival rates. We have generated single nucleotide variant calls using GATK-Mutect2 on six cats with FOSCC and have fully annotated and identified driver genes in common with HNSCC. Due to low sample size, a
larger cohort is needed to draw stronger association of disease-causing traits. Our results show some overlap in the genetic landscape of both cancers, with five samples having mutations in p53, a common mutation in HNSCC, and two samples having four genes in common with HNSCC each. Several samples with mutations in p53 and mutations in genes implicated in HNSCC suggests that the domestic cat could be a viable model for HNSCC.
CHAPTER 1

INTRODUCTION

1. Introduction

Today, at least 94 million cats reside in US households that account for 7.4 billion USD in veterinary services\(^1^2\). While companion animals, hereafter referred to as cats and dogs, share environments, disease predispositions, risk factors and occurrence rates are distinct in differing breeds, let alone species\(^3\). From a cancer perspective, there is a rather different profile of disease in cats than dogs. A major complication to finding the origins of substantial differences are companion animals are less likely to show symptoms until later stages, and without the ability to articulate symptoms, leading to challenges in identifying the earlier stages of disease. Cats and dogs both have a shorter average life span than humans, and a smaller window to study disease progression presents an opportunity for identification of potential therapies that may be successful in humans\(^4\). Dogs present with cancers of various types, including a higher incidence than cats but no confirmed explanations why\(^3^5\). For example, osteosarcoma is found at a rate of 72 to 126 cases reported per 10,000 dogs depending on the breed, compared to 4.9 cases per 100,000 cats\(^6^7\). The varying history between cat and dog domestication, with dogs having lower rates of heterozygosity due to strict breeding practices, may offer some genetic or epigenetic explanations for these large differences in cancer rates. Dog and cat domestication differ greatly, with dog domestication events estimated around 15,000 to 30,000 years ago depending on the study\(^8\), compared with cat domestication where available evidence estimated the domestication period up to 9,000 years ago\(^9\). Genetic data shows dog domestication began in middle eastern
and east Asian grey wolves, with two periods of intense selection\textsuperscript{10}. The first selection period began approximately 10,000 years ago, where dogs were selected for docility and better interactions with humans leading to ancient breeds, like the chow chow and New Guinea singing dog that are significantly divergent from modern breeds\textsuperscript{10}. Modern dog breeds are a result of a second intense selection period in the Victorian era, 1830 - 1900, where dogs began to be bred under the direction of breed clubs\textsuperscript{10}. These practices resulted in a wide range of variations with 195 AKC registered breeds, and over 350 breeds known worldwide\textsuperscript{11}. Breeds vary in size, skeletal and cranial proportion, and characteristics including sight, smell, and tendencies for herding, swimming, aggression and laziness\textsuperscript{8}. This wide divergence of dog breed phenotypic diversity is largely due to the discrete fixation of DNA variants that have a large effect in individual lineages, which are then crossed into other groups, followed by the selection of this trait in the F2 generation\textsuperscript{8,10}. This intense artificial selection causes the haplotype where the variants are fixed to rise in frequency that in combination with breeding of related individuals, leads to reduced levels of genome-wide heterozygosity. In dogs, this inbreeding is likely an important contributor to higher rates of specific diseases in some breeds versus others. For example, a study evaluating degenerative myopathy in Collies found a heterozygous carrier rate of 27.6\% for a mutation in SOD1, which was implicated in disease progression in the small (less than 100) populations of collies in Japan. This study highlights the risk of inbreeding on the occurrence of degenerative myopathy\textsuperscript{12}. The genetic origins of these higher breed-specific incidence rates, however, is understudied and often unknown.

The genetic history of the domestic cat is dramatically different to the dog. Cats were suggested to cohabit with humans in Cyprus around 9,000 years ago. Domestication is thought to be mutualistic between humans and cats, where cats were first adopted to control rodents eating the
grains stored within agricultural villages near the Mediterranean fertile crescent. Today, domestic cats are considered a subspecies of wildcats due to their morphological and genetic similarities, and they are often visibly indistinguishable from wild cats, especially those cats with a tabby coat. Cats have fewer than 40 true breeds appearing in only the last 150 years, compared to over 400 recognized dog breeds, and most domestic cats are categorized as random bred, leading to lower rates of inbreeding in the general cat population. Further differing from the dog, coat color, length, and texture trait aesthetics are the main drivers of breed formation. While the establishment of cat breeds began with random-bred cats throughout the world, allowing the outcrossing of many breeds, where in dogs mating of close relatives leads to an increased risk of genetic disorders and diseases. It is all these genetic features of feline genome evolution, i.e., increased rates of heterozygosity and a less stringent genetic bottleneck, that has led to hypotheses that domestic cats can offer a contrasting model of why cat cancer types differ substantially from the dog in many aspects.

Tumors in companion animals are staged based on tumor specific protocols and are classified from stages I to IV as in humans. Staging informs the extent of cancer locally, distally and in the draining lymph nodes to determine the course of treatment. Risk factors for many cancers in companion animals include exposure to tobacco smoke, tinned tuna, flea collars, infection of the feline leukemia virus, and reoccurring inflammation at sites of trauma due to injections and microchipping. Many cancers in cats are very aggressive and treatments available include surgery, radiation, chemotherapy, and ultimately palliative care. While various treatments can be effective with early intervention, new therapies are desperately needed to improve prognosis.

Companion animals and humans suffer from many of the same ailments, with over 70 genes shown in cats to contain single and multiple DNA variants in common with humans contributing
to disease causation such as cardiomyopathy and polycystic kidney disease. Genomic medicine is currently being practiced in human health where one sequences the genome, coding regions, or select genes of the patient to identify variants that cause or contribute to the morbidity of a disease in order to improve treatment options. This process is rapidly being implemented in cancer therapeutic decisions with druggable gene mutations as the outcome but not in companion animals. Recent whole genome sequencing (WGS) and whole exome sequencing (WES) cat studies have shown many more causative or associated disease variants await discovery. By uniquely identifying a homozygous missense variant in NPC1, a gene responsible for causing Niemann-Pick disease type C1, a cat with an undiagnosed neurological disorder can now inform the veterinarians treatment strategy. In non-disease cases, for instance breeds with rare phenotypes like dwarfism, WGS was able to identify a novel structural variant in the UGDH gene associated with cats displaying dwarfism. With actionable genetic information available, veterinarians can be aware of disease risk and proactively treat patients to prevent severe symptoms. However, the practice of genomic medicine in veterinary clinics is very limited today. In cancer, this is due to the scarcity of WGS or WES studies that ascertain somatic mutations linked to tumor progression. Our knowledge of variants associated with domestic cat cancers should substantially increase, thus, genomic medicine will be feasible for cats in the near future with cancer or unknown ailments. A survey of the literature follows for the most common types of cat cancer with an emphasis on what is known for genetic causality. From this point forward, a discussion of the most common cat cancer types with contrasts to dog, where appropriate, is presented. While feline cancer is rare, with one study citing cats having an incidence rate of 63 per 100,000 cats per year compared to dog with about 143 per 100,000 dogs per year both compared to humans, which have a 1 in 3 chance in developing cancer. Cancers
such as lymphoma, fibrosarcoma and mammary cancers are commonly seen while tumors of the brain, lung, and liver are much less frequent, but all often result in poor prognosis due to their late stage of detection. The possible risk factors, prognosis, and when known molecular causes for these feline cancers, including lymphoma, fibrosarcoma, oral squamous cell carcinoma, and pulmonary neoplasia are briefly reviewed (Figure 1.1).

**Figure 1.1** Summary of the most common types of cancer in cats with candidate driver genes of each cancer.
CHAPTER 2

LITERATURE REVIEW

2.1 Feline Cancer Molecular Landscape

While cat cancers are less frequent compared to dogs, there are several that are often seen in cats. The most common by far is lymphoma followed by soft tissue sarcoma, which includes injection site sarcoma. Mammary cancer accounts for 12% of all cats’ cancers, followed by oral squamous cell carcinoma which is the most common oral cancer diagnosed in cats, with pulmonary carcinoma diagnosed less often, but often resulting in metastasis. While several reports have advocated for the overlooked potential of studying the only other spontaneous naturally occurring model of cancer outside human, the dog lost in this discussion are the interesting genetic features of domestic cat cancer even with its lower rate of incidence. Improved genome references, better gene annotation, and greater WGS sampling of domestic cats for variant ascertainment, have established resources for more transcriptomic and genomic investigations of cat cancer. Addressing this, in turn, will facilitate the finding of shared and novel molecular drivers of cat cancer compared to the human and dog.

2.2 Lymphoma.

A malignancy of lymphocytes, various versions of lymphoma are the most common neoplasms seen in cats. Lymphoma is typically considered a system-wide disease, as these cells travel through the entire body through the lymphatic system. In the cat, lymphoma can arise at any site but generally presentations are alimentary/gastrointestinal, peripheral nodal, and extra nodal, and most frequently is seen in the lymph nodes, spleen and bone marrow with B, T, or NK cells forming the neoplastic population. Feline leukemia virus (FeLV) is the most common cause of
lymphoma in cats, however after the introduction of the FeLV vaccine in the 1980’s in combination with the loss of infected animals, there has been a decline in FeLV-associated lymphoma. The median age of cats diagnosed with lymphoma has increased to 12 years from 3 to 5 years of age as a result\(^{37}\). There is an association between feline immunodeficiency virus (FIV) infection and increased incidence of lymphoma, resulting in a five-fold increase of developing feline lymphoma\(^{38}\), suggesting that FIV plays an indirect role in tumorigenesis causing chronic dysregulation of the immune system and activation of oncogenic pathways. Siamese and Oriental breeds are reported to be at higher risk, with one study reporting a 1.5:1 male to female ratio and another study finding no association with neutering status\(^{37,34}\).

Clinically, identification of lymphoma includes a complete blood count (CBC) with differential cell and platelet count, serum biochemistry profile, urinalysis and a retroviral screen\(^{35}\). Tumors are graded and then staged to evaluate the extent of the disease. Tumors that present as low grade have strong remission rates with conservative treatment protocols including oral chlorambucil and prednisolone resulting in average survival rates of 1.5 to 3 years\(^{39}\). Most feline lymphomas present as medium and high-grade tumors at any anatomical site that require more aggressive multiagent combination therapies. CHOP therapies, consisting of a combination of cyclophosphamide (C), hydroxydaunorubicin (H), oncovin (O) and prednisone (P)\(^{35,20}\), are most successful in treating medium and high-grade lymphoma. CHOP is a modified treatment protocol based on human protocols, and only have complete remission rates of 50 to 60%, with an even lower survival rate of less than one year. Radiation treatment has been demonstrated as an alternative effective therapy because lymphatic cells are sensitive to radiation\(^{34}\). Lastly, surgery is typically not an adequate treatment for this disease but is most often used to achieve a diagnosis or relieve gastrointestinal obstruction but has no clear effect on survival\(^{21,40}\).
With the use of genomic medicine approaches, human patient survival times for blood borne cancers have increased and that has already accelerated investigations of all types of lymphoma maladies among companion animals, albeit at a much slower pace in cats. Interestingly, we can perhaps also study heritable risk since there is a lymphoma predisposition observed in the oriental cat breeds. Changes in oncogenic pathways, epigenetic changes and signal transduction alterations found in human are suspected to be at work in the domestic cat but again unproven to date. One cat study saw increased Bcl-2 expression which controls cellular proliferation and cell cycle apoptosis. In this study, lymphoma cell lines had higher levels of Bcl-2 compared to normal peripheral mononuclear blood cells which had lower levels of Bcl-2 expression, suggesting that Bcl-2 expression may be useful in the differential diagnosis of feline tumors.

Since very few studies explore the underlying driver genes of feline lymphoma or their accompanying transcriptomes cat lymphoma outcomes haven’t changed much. In the past, gene microarray techniques were utilized to broadly understand the role of chromosomal aberrations and gene expression changes associated with lymphoma but no treatment changes were implemented as a result. The adoption of fast developing genomic methodologies promises a brighter therapeutic outcome for feline lymphoma.

**2.3 Soft Tissue Sarcoma**

Soft tissue sarcomas (STS) are locally aggressive cancers that arise from mesenchymal tissues and can arise in connective tissues including the muscle, adipose neurovascular, fascial, and fibrous tissues. This type of cancer is relatively common with an incidence rate of 17 per 100,000 cats. Soft tissue sarcoma can arise at any anatomical location, but most commonly occurs in subcutaneous tissue and is the second most prevalent skin tumor in cats. In cats, STS
has three known subtypes: spontaneous formation of fibrosarcoma, feline sarcoma virus and injection site sarcoma (ISS). Fibrosarcomas can be rapidly growing and present as multiple cutaneous or subcutaneous nodules. These nodules are normally locally invasive and can metastasize to the lungs and other sites. The feline sarcoma virus (FeSV) is a recombinant version of the FeLV, where the recombination of the FeLV genome with cellular oncogenes of the infected cat causes uncontrolled cell proliferation. Through this recombination, FeSV can acquire one of the multiple oncogenes including \( FES, FMS, \) or \( FGR \). Oncogenes associated with FeSV were discovered by transforming fibroblasts and producing a fibrosarcoma that is multicentric in younger cats. Older studies show that only about 2% of the fibrosarcomas found in cats are virally induced and as cases of FeLV have decreased so have cases of FeSV.

The most common histological type of fibrosarcoma is ISS. While ISS is included as a type of soft tissue sarcoma, they are histologically heterogeneous and have distinctive biological behavior, are more aggressive, and appear in younger cats when compared to non-injection site fibrosarcoma. ISS is often thought to be the result of the proliferation of fibroblasts and myofibroblasts at the sites of chronic inflammation caused by a sustained immune response induced by the injected material. It is also characterized as the proliferation of atypical spindle cells and a variable amount of multinucleated giant cells, which are not generally seen in non-injection site feline fibrosarcomas. In this cancer, cats develop a subcutaneous inflammatory response at the site of injections and the tumor generally develops within the first 3 years of vaccination. This inflammation has been associated with the use of vaccines that contain aluminum-based adjuvants, and while the increasing use of non-adjuvant vaccines has caused a reduction of ISS, it remains a concern in feline veterinary medicine.
associated sarcomas have been reported to have elevated mitotic rates, with higher-grade
classification, less differentiated, and contained cells with more variable size and shape
compared to non-vaccine site sarcomas. Other causes of ISS include microchipping, the use of
long-acting injectable antibiotics, and injectable glucocorticoids. ISS typically presents as a
subcutaneous mass at the site of injection. Cats can also present with decreased appetite, and
difficulty breathing if the cancer has metastasized, most frequently to the lungs. Staging is an
important step in ISS treatment plans since this tumor type is highly aggressive and treatment
generally includes tumor resection depending on the stage and grade of the tumor. Of note,
there are marked institutional variations here with some centers that routinely involve radiation
treatment in treatment while others perform radical excision as a locally curative intent therapy
without RT (as long as margins are acceptable).

The study of molecular causes associated with ISS have been limited to a few genes and
typically rely on the understanding of the analogous injection site sarcoma cancer type in human
for the most likely driver gene candidates. A cohort of ISS cats was found to harbor homozygous
deletions in PTEN, amplification of KIT, and an overall DNA copy number imbalance that was
correlated to a more aggressive tumor behavior. Another study found that increased expression
of MMP-9, MMP-2, and TIMP-2 in all ISS tumors investigated was suggestive of tumor
aggressiveness. With the aggressive nature and often unknown specific causes of ISS, early
detection and surgery are still the first-line defenses against this cancer type with druggable
genes not on the clinical horizon.

Excitement and perhaps too much anticipation is now placed on immunotherapy opportunities.
Yet, one study showed that viruses expressing interleukin-2 are not very effective, reducing the
tumor reoccurrence to 28% compared to 52% at 12 months. Combinatorial strategies used to
treat human head and neck cancers (HNSCC) and melanomas, such as targeted type I interferon pathway inhibitors and oncolytic viruses could offer better tumor control in feline sarcomas, with one feline study showing reoccurrence rates dropping from 61% to 28%. Inflammation at injection sites causes mutations in many oncogenic pathways, potentially leading to injection site sarcomas; however, improved vaccine technology, early detection, and better adjuvant therapy, seem to be reducing ISS incidence.

2.4 Feline Mammary Cancer.

Feline mammary cancer (FMC) is the third most diagnosed feline neoplasm. Approximately 80 - 96% of FMCs are malignant with multiple tumors being common at diagnosis. FMC carries an incidence rate of 24.4 cats per 100,000 female cats, and accounts for 12% of all tumors in cats regardless of sex. A major risk factor for FMC is age, with risk of diagnosis becoming significantly increased at 7 to 9 years, and mean age of diagnosis being 10 to 12 years in age. Breed is another major risk factor with studies suggesting short-haired cats have a higher risk. Siamese cats are at twice the risk of developing FMC and are diagnosed at a younger age.

Another major risk factor is hormonal exposure, with gonadally intact cats having a 7-fold higher risk of developing FMC than spayed cats; a similar effect is observed in dogs. Cats exposed to regular progesterone have an increased risk of FMC, so ovariohysterectomy, a tissue source of this steroid, is protective against mammary tumor development. It is a highly aggressive cancer, with locally infiltrative and metastatic features, as the primary tumor metastasizes frequently to regional lymph nodes and lungs. Fortunately during routine veterinary visits FMC based on the palpation of firm nodular masses on the mammary gland. Thoracic radiographs, abdominal ultrasound, and fine-needle aspiration are then often used to assess the extent of FMC. Just as in human breast cancer, most feline tumors are classified as
moderately or poorly differentiated\textsuperscript{63}. In today's practice, many veterinarians have adopted the Elston and Ellis histological grading system that is considered the “gold standard” grading system for human breast cancer\textsuperscript{64}. Feline mammary histological grading is based on cellular differentiation and degree of tubule formation; nuclear pleomorphism; and mitotic frequency\textsuperscript{63}. Many factors affect the prognosis of FMC including grade, mitotic count, and disease stage. Tumors smaller than 2 cm have seen a survival a mean of 54 months, tumors 2-3 cm average of 24 months and 3-6cm have an average survival of only 6 months\textsuperscript{60,63}. Once again, the first line of FMC treatment is surgical excision, with studies showing bilateral mastectomy improves survival time\textsuperscript{65}. FMC is most likely to be hormone receptor-negative and patients presenting with triple-negative mammary carcinoma, an aggressive form of cancer lacking secretion of progesterone, human epidermal growth factor and progesterone receptors\textsuperscript{66}, had the lowest survival rates\textsuperscript{67,68}. Multiple genes with specific somatic mutations have been associated with FMC. Elevated expression of the oncogene \textit{HER2}, which stimulates downstream activation of the AKT pathway\textsuperscript{69} that promotes growth factor-mediated cell growth, proliferation, migration, and survival is highly correlated with malignancy and tumor differentiation\textsuperscript{69,70}. Loss of the \textit{PTEN} gene, a negative regulator of the AKT pathway, has been shown in 76\% of FMC cases and in higher-grade tumors predicting a poor prognosis\textsuperscript{70}. Recurrent deletions among B2 and E3 chromosomes detected in higher-grade FMC tumors exhibit suggested these genomic imbalances contribute to poor outcome\textsuperscript{71,72}. Copy number gain of \textit{EBR2} was greater than 3-fold in high-grade compared to intermediate-grade mammary gland cat tumors that alludes to the important role this gene plays as part of this chromosome instability correlation with a worse outcome\textsuperscript{2,71}. There is evidence that molecular markers cyclin A, \textit{p53}, \textit{RON}, \textit{VEGF}, and \textit{COX} can be predictors of outcome for
Studies have shown VEGFR-3 highly to moderately expressed in all FMC. All these studies demonstrate much more FMC sample sequencing will be needed to resolve the somatic mutations with the highest occurrence by gene.

As stated earlier for other cancer types immunotherapy is also being explored for FMC treatment. Recent studies suggest immunotherapy strategies targeting programmed cell death 1 (PD-L1) and programmed cell death ligand 1 (PD-L1) genes, FMC may be a new promising line of treatment. PD-L1 has been implicated in a long list of human cancers, including breast, lung, and metastatic melanoma, and promotes immune suppression and tumor escape.

Nascimento et al discovered elevated levels of PD-1 and PD-L1 in the serum of cats with FMC, and with a newly FDA approved monoclonal antibody that blocks PD-L1 binding better therapeutic outcomes are hopefully possible for FMC.

2.5 Oral Squamous Cell Carcinoma.

Feline oral squamous cell carcinoma (FOSCC) is a cancer of the squamous cell lining in the oral and oropharyngeal cavity of the cat and may spread into deeper tissues. These tumors are commonly found in the gingiva, tongue, and sublingual region. FOSCC is the fourth most common cancer in cats and represents 70-80% of all oral cancers and shares many risk factors and molecular markers to human head and neck squamous cell carcinoma (HNSCC). These similarities may present an opportunity to test evolving therapies in HNSCC but also offer comparative oncological models of their molecular circuits. These tumors are usually locally invasive, however can invade the local bone tissue. FOSCC uncommonly metastasize to distant locations but one study found 35% of cats had metastasis to local lymph nodes. FOSCC is typically a deadly disease for cats with median survival post-diagnosis only reaching a few months, and a one-year survival rate of less than 10%. Cats presenting with FOSCC are an
average age of 12.5 - 13 years, with no sex or breed disposition\textsuperscript{76,80}. Signs of this disease include oral pain, difficulty swallowing, excessive salivation, anorexia, and loss of teeth\textsuperscript{20}. In cats, risks associated with FOSCC include exposure to tobacco smoke as well as flea collars and feeding some canned foods\textsuperscript{81}. Human papillomavirus, a known cause of HNSCC, has been associated with more negative outcomes in humans; however, current studies show FOSCC is more closely related to HPV-negative HNSCC\textsuperscript{82}. Molecular markers thus far include abnormal $p53$, and $\text{CK2}$ expression as well as genes implicated in angiogenesis, e.g. $\text{LOX}$ and $\text{COX}\textsuperscript{83}$. $\text{EGFR}$ overexpression is seen in FOSCC and humans, and has been associated with proliferation and migrations of squamous carcinoma cells to other sites\textsuperscript{84,85}. Recent studies show that $\text{CD147}$, a surface cluster protein found in HNSCC tumors is associated with poor prognosis\textsuperscript{80}. Since FOSCC $\text{CD147}$ associated expression has been observed perhaps its manipulation in the tumor is a therapeutic target\textsuperscript{76}. Treatment of FOSCC is first surgical removal from the oral cavity, if feasible. In contrast to some other cat cancer types, FOSCC surgery is often followed by high recurrence rates with estimates ranging from 15.4 to 38\%\textsuperscript{86-88}. Other standards of treatment following surgery include, radiation, and chemotherapy; but again poor outcomes are the result\textsuperscript{89,90}. Starting with FOSCC cell lines, treatment with actinomycin D, dinaciclib, flavopiridol, and methotrexate shows promise yet this drug mixture has not been tested in clinical settings\textsuperscript{77}. While current treatments offer unacceptable FOSCC control, new potential strategies that include $\text{EGFR}$ inhibitors, $\text{CK2}$ inhibitors, $\text{COX/LOC}$ inhibitors, or methods to reverse hypoxia\textsuperscript{83} can hopefully avoid euthanasia that remains the most common veterinarian recommendation in FOSCC cases.
2.6 Pulmonary Carcinoma.

Primary pulmonary carcinoma is rare a type of lung cancer in the cat and most commonly is derived from the epithelium of the airways or the alveolar parenchyma. This lung cancer is found in about 2.2 per 100,000 cats, almost all tumors are malignant and the most common type of primary lung tumors are adenocarcinomas (60 - 80%), with sarcomas, squamous cell carcinomas, and adenomas less common. Primary lung neoplasia is less common than metastatic tumors that spread to the lung from other primary cancers, most commonly from the breast. Pulmonary carcinoma can also present as feline lung-digit syndrome, an atypical pattern of metastasis where the primary lung tumor cells travel to the digits of the cat, particularly weight bearing regions. This presentation is associated with a poor prognosis.

Metastasis to the digits may be due, in part, to high digital blood flow to compensate for heat loss. Risk factors of pulmonary carcinoma include exposure to secondhand smoke and as in other cat cancers increased age but with no higher prevalence among breeds or sex. A recent study has concluded that in areas with high radon exposure, companion animals have a 2-fold higher incidence of primary pulmonary neoplasia. Signs of this cancer include coughing, weight loss, lethargy, lameness, and respiratory changes. Computed tomography and thoracic radiographs are commonly used for staging of these tumors prior to removal. Multiple driver genes have been implicated in human lung cancer studies, but none thus far have been investigated in cats. Oncogenes such as KRAS, EGFR, BRAF, PIK3CA, and MET have each been shown to harbor somatic mutations in as many as 33% of lung tumors. P53 or STK11 mutations in combination with KRAS mutations have been linked with low survival rates and have been found to cause early-onset cancer in mice. EGFR is implicated in 20% of lung cancer in humans and is responsible for the early stages of epidermoid carcinoma development.
Tyrosine kinase inhibitors, such as gefitinib and erlotinib have been shown to improve survival rates of human patients with susceptible \textit{EGFR} mutations\textsuperscript{100}. Treatment of this cancer includes surgical resection as the first line of defense in companion animals as well. Partial or complete lobectomies are generally performed, where a cuff of normal tissue is also removed to ensure a wide margin is obtained\textsuperscript{91}. Chemotherapy is a standard approach in human lung cancer, however very few trials have been conducted to evaluate its efficiency in cats\textsuperscript{91}. In humans, patients often develop chemoresistance and thereafter have poor survival rates. A rare feline study of chemotherapy resistance found expression of proteins that are associated with human multi-drug chemotherapy resistance in primary lung tumors, P-glycoprotein \textit{ABCB1}, multi-drug resistant protein, and lung resistance-related protein, all of which regulate the export of chemotherapeutic drugs outside of the cell and could present future targets\textsuperscript{101}. While the underlying feline genetic mechanism of pulmonary carcinomas, i.e., driver genes with high recurrence, is unknown, further investigation into the predicted similar somatic mutational landscape from the vast human lung cancer sequencing cases for example \textit{EGFR}, may present as treatment options for cats outside surgery and chemotherapy.

\subsection{2.7 Conclusions}

Cats provide humankind with crucial companionship and should receive the best veterinary care. Nearly 6 million cats will be diagnosed with cancer each year\textsuperscript{3}, and with very little known about the genetic landscape of cancer, many pets will face euthanasia. Adopting the use of genomic medicine approaches being developed for humans may provide an improved standard of care for the cat with cancer. A critical gap in knowledge that can be filled with larger sampling of cat cancers of all types using WGS or WES methods will find somatic or germline variants implicated in cancers. This new approach to therapy may allow for more accurate and earlier
diagnosis and precise therapy leading to an improved prognosis rate in several diseases, and, importantly, deadly cancers. With many overlapping molecular and environmental causes of cancer in cats and humans, several types of cancer present opportunities for comparative therapeutics, such as mammary and oral squamous cell carcinoma. While current treatment is limited to surgery, chemotherapy, and radiation, new treatment approaches are on the horizon.
CHAPTER 3
A Domestic Cat Whole Exome Sequencing Resource for Trait Discovery

3.1 Introduction
Genomic medicine promises new avenues of disease treatment in veterinary medicine. However, the appropriate resources are not yet readily available for robust implementation in clinical practice. One resource which has been successfully applied to the diagnosis of rare diseases in humans is whole exome sequencing (WES) analysis, a cost-effective method for identifying potentially impactful DNA variants in the coding regions of genes. DNA base changes in the exome can alter amino acids in proteins or disrupt their overall structure, so focusing on these regions offers a more direct and biologically interpretable approach to searching for putative disease variants. In comparison, whole genome sequencing (WGS) captures DNA variants spanning the entire genome. However, as the vast majority of the identified variants are within non-coding regions, much of the variation is difficult to interpret. WES also allows for deeper coverage of target sequences due to lower cost compared to WGS. The present study seeks to develop and validate the use of WES as a viable approach for determining novel disease variants in cats.

Over the last decade, a surge of studies using next generation sequencing (NGS), in particular WES, has led to many novel discoveries of candidate disease-causing variants across species. WES is recognized as an efficient means for genome resequencing and is the primary NGS approach used to help diagnose human patients with rare genetic diseases. By selectively sequencing all protein-coding regions to a deeper depth than WGS, WES is a dependable method for finding biallelic exonic variants causative of Mendelian inherited diseases that rarely appear in healthy populations. In humans, WES is commonly used to
find genetic causes in a wide range of diseases, even complex neurological conditions such as autism spectrum disorder\textsuperscript{105}. Its widespread use has led to the discovery of therapeutic targets for drug development and genetic markers for innovative clinical applications\textsuperscript{106}. Tumor WES has been especially successful by cost-effectively providing somatic variant information about a patient’s normal and tumor exomes, supporting the identification of recurrent somatic mutations among known oncogenes that may suggest a mechanism of action and targets for potential drug therapies\textsuperscript{107}. The significant depth of exome coverage is integral to overcoming diluted somatic variant allele frequencies (VAF) due to tumor clonality and purity issues.

Exome sequencing has also proven successful in non-human species. Mouse WES studies have found strong candidate alleles for models of orofacial clefting, urogenital dysmorphology, and autoimmune hepatitis\textsuperscript{108}. In companion animals, the development of dog WES has demonstrated that causative allele discovery for common diseases has great potential\textsuperscript{109}. Some examples in dogs include the discovery of a two-base pair deletion in \textit{SGCD} causing muscular dystrophy, and a splice site variant in \textit{INPP5E} which is associated with cystic renal dysplasia\textsuperscript{110}. As there are many isolated breeds of domestic dogs, this species is an important genetic resource for cancer studies, for which WES demonstrated dogs have similar oncogene variant patterns to humans\textsuperscript{111}. However, many oncogene variants are not equivalent to a WES analysis of human, and canine bladder cancers identified novel mutations in \textit{FAM133B}, \textit{RAB3GAP2}, and \textit{ANKRD52} that are unique to canine bladder cancer, emphasizing the need to understand the biological differences in origin\textsuperscript{112}.

Similar to canines, domestic cats have long been recognized for their potential in modeling human diseases, such as retinal blindness\textsuperscript{113,114}. Approximately 150 variants in domestic cats are associated with over 100 genetic traits or diseases, many mimicking human
disease phenotypes. As feline genomic resources continue to advance, more diseases caused by single base variants are being discovered, such as two novel forms of blindness in Persian and Bengal cats. However, a feline WES resource has not been described to date for the discovery of novel disease gene candidates. Here we describe the first feline exome resource, a WES analysis of 41 cats, and its use in the discovery of known and novel variants associated with feline phenotypes, healthy and diseased. A comparison of WES and WGS methods was also completed to understand the efficiency, depth of coverage, and sequence specificity, for variant calling from each approach.

3.2 Methods

Exome Design. The annotated exons from the Felis_catus_9.0 reference genome assembly were used as the basis to design the exome capture probes, incorporating the NCBI RefSeq release 92 annotation, containing 19,590 refGene names completed by Roche Sequencing Solutions. The coding sequences (CDS) for the primary chromosomes were extracted and consolidated into a non-overlapping set of features, and repetitive probes were removed totaling 35,724,716 bases divided over 201,683 regions. Of those bases, only 395,115 bp are not covered directly or indirectly. GO functions for removed genes were olfactory genes or unidentifiable. Since Y chromosome genes are not represented in the Felis_catus_9.0 reference, a set of coding sequence features from the Felis catus Y chromosome genomic sequence (NCBI accession KP081775) was used. The cat exome panel was designed by Roche Sequencing Solutions (Madison, USA). A capture probe dataset was constructed for the full cat genome by tiling variable-length probes, ranging from 50 - 100 bases in length, at a five-base step across all sequences. Each capture probe was evaluated for repetitiveness by constructing a 15-mer histogram from the full genome sequence and then calculating the average 15-mer count across each probe, a sliding
window size of 15 bases across the length of each probe. Any probe with an average 15-mer count greater than 100 was considered to be repetitive and excluded from further characterization. Non-repetitive probes were then scored for uniqueness by aligning each capture probe to the full cat genome using SSAHA v3119. A close match to the genome was defined as a match length of 30 bases, allowing up to five insertions/deletions/substitutions. Capture probes were selected for each coding sequence feature by scoring one to four probes in a 20-base window, based on repetitiveness, uniqueness, melting temperature, and sequence composition, and then choosing the best capture probe in that window. The start of the 20 base windows was then moved 40 bases downstream and the process repeated. Selected probes were allowed to start up to 30 bases before the 5’ start of each feature and overhang the 3’ end by 30 bp. A maximum of five close matches in the genome was allowed when selecting the capture probes.

**Samples and DNA Isolation.** Cat DNA samples for WES were donated by owners and archived in the Lyons Feline Genetics Laboratory at the University of Missouri, College of Veterinary Medicine in accordance with the University of Missouri Institutional Animal Care and Use Committee protocol study protocols 9056, 9178, and 9642. DNA was isolated from 41 whole blood or tissue cat samples using standard organic methods120 and verified for quantity and quality by DNA fluorescence assay (Qubit, Thermo Fisher) and ethidium bromide staining after 0.7% agarose gel electrophoresis. Ten cats with existing whole genome sequence (WGS) data were initially tested, followed by 31 novel cats for additional screening.

**Sequencing.** All WGS cat data used in this study was obtained from Buckley et al, and library preparation was completed by Washington University.121 Genomic DNA (250 ng) was fragmented on the Covaris LE220 instrument targeting 250 bp inserts. Automated dual indexed libraries were constructed with the KAPA HTP library prep kit (Roche) on the NGS platform.
The libraries were PCR-amplified with KAPA HiFi for 8 cycles. The final libraries were purified with a 1.0x AMPureXP bead cleanup and quantitated on the Caliper GX instrument (Perkin Elmer) and were pooled pre-capture generating a total 5µg library pool. Each library pool was hybridized with a custom NimbleGen probe set (Roche), targeting 35.7 Mb. The libraries were hybridized for 16 - 18 hours at 65°C followed by washing to remove non-specific hybridized library fragments. Enriched library fragments were eluted following isolation with streptavidin-coated magnetic beads and amplified with KAPA HiFi Polymerase prior to sequencing. PCR cycle optimization was performed to prevent over-amplification of the libraries. The concentration of each captured library pool was determined via qPCR utilizing the KAPA library Quantification Kit (Roche) to produce appropriate cluster counts prior to sequencing. The Illumina NovaSeq6000 instrument was used to generate paired-end 2 x 150 bp length sequences to yield an average of 14 Gb of data per 35.7 Mb target exome, producing ~80x exome sequencing depth of coverage. Exome sequencing data are available at the Sequence Read Archive under accession number PRJNA627536.

**Variant Discovery.** The following tools/packages were applied to WGS and WES samples in accordance with variant processing as previously described\(^\text{121, 71}\): BWA-MEM version 0.7.17\(^\text{122}\), Picard tools version 2.1.1 (http://broadinstitute.github.io/picard/), Samtools version 1.9\(^\text{123}\), and Genome Analysis toolkit version 3.8\(^\text{124,125,126}\) by Rueben Buckley. Code used for the variant calling workflow can be found at https://github.com/mu-feline-genome/batch_GATK_workflow. For WES processing, GATK tools were restricted to exons annotated in Ensembl release 99 with an additional 100 bp of flanking sequence\(^\text{127}\). Following processing, samples were genotyped in three separate cohorts. The first cohort consisted of all 41 WES samples. The second and third cohorts were ten matched WES and WGS samples. Variants in all three cohorts were tagged
using the same variant filtering criteria. For SNVs, the filtering criteria were QD < 2.0, FS > 60.0, SOR > 3.0, ReadPosRankSum < -8.0, MQ < 40.0, and MQRankSum < -12.5. For indels, the filtering criteria were QD < 2.0, FS > 200.0, SOR > 10.0, and ReadPosRankSum < -20.0. Although five Y chromosome genes were included in the exome probe set, these genes had not been added to the aligning reference. For WGS/WES comparison, matched WES/WGS samples were annotated using variant effect predictor (VEP). Variants from both cohorts were independently tagged as to whether they were biallelic, SNVs, or passed filtering criteria. Before analysis, variants flanking the exome primary target regions +/- 2bp were removed. Variant processing and comparisons were performed in the R statistical environment using the vcf R package. Common variants between both platforms were determined as those at the same position with the same reference and alternate alleles. Exclusive variants were determined as those where the position and/or the alleles were specific to a particular platform.

**Disease and trait variant detection.** Variants for all 41 cats were evaluated using VarSeq software (GoldenHelix, Inc.). SNVs were annotated as having high, moderate, or low impacts on gene function. High impact variations were those that were a protein-truncating variant caused by stop gain or loss and splice-site acceptor or donor mutations. Moderate impacts include missense mutations or in-frame insertions, and lastly, low impact variants are characterized by synonymous base changes, splice region variants, or intron variants. Known variants for diseases and traits were evaluated in each cat.

**Polycystic Kidney Disease.** A pointed cat of the Siberian breed (a.k.a. Neva Masquerade, a pointed Siberian) was diagnosed with polycystic kidney disease based on signs of renal disease (polydipsia, polyuria) and ultrasonography (Table 3.1, cat 37). DNA was submitted using buccal swabs and a whole blood sample to two different commercial testing laboratories in which both
confirmed the absence of the currently known autosomal dominant polycystic kidney disease in polycystin-1 (PKD1). The dam and a sibling were also reported as having PKD by ultrasonography but were not available for genetic analyses.

Cystinuria. A three-month-old European shorthair kitten from the isle of Korfu, Greece, was presented to the AniCura Small Animal Hospital, Bielefeld, FRG, for heavy straining during urination, and the owner reported the kitten would fall over from time to time (Table 1, cat 17). The kitten had been pretreated with two injections of cephalaxine and dexamethasone for suspected cystitis, however, difficulty in urination worsened. Upon hospital admission, the kitten was in good general condition. Abdominal palpation revealed an enlarged urinary bladder. Abdominal X-ray showed over 30 radiolucent urinary stones up to a diameter of half of the width of the last rib. Urinary bladder stones and some urethral stones were removed via cystolithotomy and retrograde flushing of the urethra. Urinary stones were submitted for infraspectroscopic stone analysis. Stone analysis revealed pure cystine stones and a diagnosis of cystinuria was made. Urinary stones reoccurred at six months of age, but the kitten was otherwise healthy.

3.3 Results

Phenotype cohort. WES was performed on 41 individual cats, representing a variety of different diseases and traits, some with known disease alleles (Table 3.1). The 41 cats can be further divided into two separate cohorts: the first is the initial ten cats that had nine known variants for various diseases and aesthetic traits, e.g., coat colors and fur types. These 10 cats also had matched WGS data, which was used to assess the efficacy of WES. The second cohort of 31 represents genetically uncharacterized cats. These cats represented 11 different breeds and include 14 random-bred cats. Groups of cats with similar genetic backgrounds were used to evaluate causes for mediastinal lymphoma, a seizure disorder, eyelid colobomas,
hypothyroidism, hypovitaminosis D, blue eyes of Ojos Azules breed, and curly hair coat of the Tennessee Rex. Five cats were reported with cardiac diseases, including hypertrophic cardiomyopathy (HCM). At least seven neurological disorders are represented in the study population, generally representing novel presentations in random-bred cats. Overall, the 41 cats had approximately 31 different unknown disease presentations.
### Table 3.1 Description and diseases of 41 cats for WES evaluation.

<table>
<thead>
<tr>
<th>No.</th>
<th>Id.</th>
<th>Breed</th>
<th>Sex</th>
<th>Disease / Trait</th>
<th>Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19725</td>
<td>Lykoi</td>
<td>F</td>
<td>Lykoi</td>
<td>HR</td>
</tr>
<tr>
<td>2</td>
<td>13230</td>
<td>Mixed Breed</td>
<td>F</td>
<td>Bengal PRA / Bobbed tail</td>
<td>KIF3B / HES7</td>
</tr>
<tr>
<td>3</td>
<td>14056</td>
<td>Mixed Breed</td>
<td>M</td>
<td>Persian PRA / Long</td>
<td>APL1 / FGF5</td>
</tr>
<tr>
<td>4</td>
<td>17994</td>
<td>Mixed Breed</td>
<td>F</td>
<td>Hydrocephalus</td>
<td>GDF7</td>
</tr>
<tr>
<td>5</td>
<td>19067</td>
<td>Munchkin</td>
<td>F</td>
<td>Dwarfism / Dominant White</td>
<td>UGDH / KIT</td>
</tr>
<tr>
<td>6</td>
<td>5012</td>
<td>Oriental</td>
<td>M</td>
<td>Lymphoma</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>20382</td>
<td>Peterbald</td>
<td>M</td>
<td>Hairless</td>
<td>LPAR6*</td>
</tr>
<tr>
<td>8</td>
<td>11615</td>
<td>Random Bred</td>
<td>M</td>
<td>Dominant White</td>
<td>KIT</td>
</tr>
<tr>
<td>9</td>
<td>18528</td>
<td>Random Bred</td>
<td>M</td>
<td>Spotting</td>
<td>KIT</td>
</tr>
<tr>
<td>10</td>
<td>20424</td>
<td>Siberian</td>
<td>F</td>
<td>Long / Cardiac disease</td>
<td>FGF5 / Candidate</td>
</tr>
<tr>
<td>11</td>
<td>22550</td>
<td>Bengal</td>
<td>F</td>
<td>Polyneuropathy</td>
<td>Unknown</td>
</tr>
<tr>
<td>12</td>
<td>20957</td>
<td>Devon Rex</td>
<td>U</td>
<td>Papilloma virus</td>
<td>Unknown</td>
</tr>
<tr>
<td>13</td>
<td>22752</td>
<td>Devon Rex</td>
<td>M</td>
<td>Neurological disorder</td>
<td>Unknown</td>
</tr>
<tr>
<td>14–15</td>
<td></td>
<td>Ojos Azules</td>
<td>1M:1F</td>
<td>Ojos Azules</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>20964</td>
<td>Oriental</td>
<td>F</td>
<td>Cardiac disease</td>
<td>Unknown</td>
</tr>
<tr>
<td>17</td>
<td>22728</td>
<td>Random bred</td>
<td>F</td>
<td>Cystinuria</td>
<td>SLC3A1*</td>
</tr>
<tr>
<td>18</td>
<td>20617</td>
<td>Random Bred</td>
<td>M</td>
<td>Neuronal ceroid lipofuscinosis</td>
<td>CLN6*</td>
</tr>
<tr>
<td>19</td>
<td>20948</td>
<td>Random Bred</td>
<td>M</td>
<td>Cinnamic acid urica</td>
<td>Unknown</td>
</tr>
<tr>
<td>20</td>
<td>21153</td>
<td>Random Bred</td>
<td>M</td>
<td>Ambulatory paraparesis</td>
<td>Unknown</td>
</tr>
<tr>
<td>21</td>
<td>22287</td>
<td>Random Bred</td>
<td>F</td>
<td>Myotonia congenita</td>
<td>Unknown</td>
</tr>
<tr>
<td>22</td>
<td>22397</td>
<td>Random Bred</td>
<td>M</td>
<td>Neurological disorder</td>
<td>Unknown</td>
</tr>
<tr>
<td>23</td>
<td>22505</td>
<td>Random Bred</td>
<td>M</td>
<td>Cardiac disease</td>
<td>Unknown</td>
</tr>
<tr>
<td>24</td>
<td>22623</td>
<td>Random Bred</td>
<td>U</td>
<td>Pycnodysostosis</td>
<td>Candidate</td>
</tr>
<tr>
<td>25</td>
<td>22740</td>
<td>Random Bred</td>
<td>F</td>
<td>Epidemolysis bullosa</td>
<td>Unknown</td>
</tr>
<tr>
<td>26–27</td>
<td></td>
<td>Random Bred</td>
<td>1F:1M</td>
<td>Eyelid coloboma</td>
<td>Unknown</td>
</tr>
<tr>
<td>28</td>
<td>22751</td>
<td>Random Bred</td>
<td>M</td>
<td>Ehlers-Danlos</td>
<td>Unknown</td>
</tr>
<tr>
<td>29–30</td>
<td></td>
<td>Random Bred</td>
<td>2F</td>
<td>Hypothyroidism</td>
<td>Candidate</td>
</tr>
<tr>
<td>31–32</td>
<td></td>
<td>Savannah</td>
<td>2M</td>
<td>Hypovitaminosis D</td>
<td>Unknown</td>
</tr>
<tr>
<td>33</td>
<td>21984</td>
<td>Scottish Fold</td>
<td>F</td>
<td>Cardiacosis D</td>
<td>Candidate</td>
</tr>
<tr>
<td>34–35</td>
<td></td>
<td>Selkirk Rex</td>
<td>1F:1U</td>
<td>Seizures</td>
<td>Unknown</td>
</tr>
<tr>
<td>36</td>
<td>20953</td>
<td>Siamese</td>
<td>F</td>
<td>Cardiac disease</td>
<td>Candidate</td>
</tr>
<tr>
<td>37</td>
<td>22622</td>
<td>Siberian</td>
<td>U</td>
<td>PKD</td>
<td>PKD2*</td>
</tr>
<tr>
<td>38</td>
<td>22711</td>
<td>Singapura</td>
<td>F</td>
<td>Hypovitaminosis D</td>
<td>Candidate</td>
</tr>
<tr>
<td>39–40</td>
<td></td>
<td>Tennessee Rex</td>
<td>1F:1M</td>
<td>Rexoid hair coat</td>
<td>Unknown</td>
</tr>
<tr>
<td>41</td>
<td>6623</td>
<td>Oriental</td>
<td>M</td>
<td>Lymphoma</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
713 | 41 breeds | 19F:18M:4U | ~31 diseases & traits |

A complete description of diseases and traits for entire cohort. Candidate genes are potential genes that been identified with less evidence of a causal mutations. U: Unknown sex; F: Female; M: Male *Mutations as tentative causal variants for diseases presented
Sequence coverage and specificity. To assess the performance of this feline exome resource, deep coverage WES data was produced for ten cats with WGS data for comparison. After mapping to Felis_catus_9.0, base quality trimming, and PCR duplicate removal, the average percentage of reads uniquely mapped was 82% (Table 3.2). The average sequencing depth was 267x with a range of 76x to 458x (Supplementary Table 1). Assessing the depth of coverage, of the 201,683 exonic targets, 98.1% aligned with coverage of >20x. An average of 6.98% of the total reads aligned outside of the targeted regions of the genome (Supplementary Table 2). For the uncharacterized 31 cat exomes, the sequencing depth was adjusted to typical human WES studies; for this group of cats, we estimated the average depth of coverage to be 80x. 96.41% of exonic targets aligned with a coverage of >20x, ranging from 91-98%. An average of 10.41% of total reads aligned off-target is slightly higher when compared to the first 10 higher-coverage cats that can be attributed to lower sequencing depth in the larger cohort. As expected, overall, there is a reduction in mapping at lower depth of coverage; for example, at 40x, 93.5% of targeted bases were covered (Figure 1), conversely, 99% are covered at 2x.

Known variant validation. To further analyze the effectiveness of WES for variant detection, we examined each sample for the presence of known trait-causing variants. The Felis_catus_9.0 Ensembl release 99 gene annotation was used with a selection of exons with +/- 30 bp to match exome capture design and variants were browsed using the VarSeq software (GoldenHelix, Inc). The majority of the previously published 115 trait causing variants in the domestic cat that have been documented as causal for diseases and traits affect either the coding regions or a splice donor/acceptor site. Of these known variants, 44 were identified in our WES cohort. All variants for coat colors and diseases expected to be present in the ten cats were identified,
including the alleles in the loci for Agouti (ASI-allele a\textsuperscript{133}), Brown (TYRP1 – b\textsuperscript{134}), Color (TYR – cs\textsuperscript{135}),

Table 3.2 Summary of Metrics across both Cohorts

<table>
<thead>
<tr>
<th></th>
<th>Average-First 10</th>
<th>Range-First 10</th>
<th>Average-Cohort of 31</th>
<th>Range-Cohort of 31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Coverage</td>
<td>267x</td>
<td>76-485x</td>
<td>80x</td>
<td>60-108x</td>
</tr>
<tr>
<td>% of Bases Covered</td>
<td>99.1%</td>
<td>92.3-100%</td>
<td>96.4%</td>
<td>91-98%</td>
</tr>
<tr>
<td>% Reads Aligned</td>
<td>99.9%</td>
<td>99.9-100%</td>
<td>82%</td>
<td>75-85%</td>
</tr>
</tbody>
</table>
Figure 3.1 The proportion of bases covered with the exome capture probes. The initial 10 samples are colored in red, with the X axis showing the depth of coverage, which is how many times a nucleotide base is covered starting at a depth of 10x and increasing to 50x.
Dense ($MLPH - d^{136}$), Longhair ($FGF5 - l^{137}$), Lykoi ($HR - hrTN, hrVA^{138}$), Bengal progressive retinal degeneration ($KIF3B^{139}$) and Persian progressive retinal degeneration ($AIPL1^{115}$), hydrocephalus ($GDF7^{140}$), and others. The cats also had variants known to affect cat blood type as well$^{141,142}$. In accordance with the limitations of our feline exome capture design, neither known structural nor intronic variants were detected. When analyzing discordant reads in a WGS dwarf sample, a deletion and rearrangement indicating a structural variant (SV) was visible in the $UDGH$ gene$^{115}$, but no read discordance was found in the WES analysis (Supplementary Figure 1). In addition, the $KIT$ intron one SV for White and Spotting were not identified$^{143}$. Therefore, the WES approach will fail to identify many complex SVs, an important limitation to consider for future feline trait discovery efforts.

**Novel candidate variant discovery.** Novel DNA variants were explored as putatively causal for diseases and traits in 33 cats. A novel frameshift mutation in polycystin 2 ($PKD2^{144}$), a gene associated with polycystic kidney disease (PKD) was predicted to disrupt protein function in a Siberian cat shown by ultrasound to have PKD. This mutation, a single-base deletion, causes a truncated protein (p.Lys737Asnfs*2). This variant was heterozygous in the affected cat and unique to the exome data and was not identified in the 195-cat cohort of the 99 Lives variant dataset$^{121}$. This variant was also identified in both grandparents on the dam’s side of the pedigree, although kidney ultrasound was not available. However, analysis of other Siberian cats with PKD diagnosed by ultrasound failed to identify the c.2211delG variant in $PKD2$, suggesting that this could be a private variant and that other disease-causing PKD variants are yet to be discovered in this breed.

A variant in the lysophosphatidic acid receptor 6 ($LPAR6$) gene associated with the autosomal recessive rexoid (Marsella wave) coat of the Cornish rex breed was detected in a
Peterbald cat, which is a hairless breed\textsuperscript{145}. However, the hairless trait is considered autosomal dominant by cat breeders. The annotation predicts a c.249\textsuperscript{del}G causing a p.Phe84\textsuperscript{Leufs*10}; therefore, this Peterbald cat likely is compound heterozygous for two mutations juxtaposed in LPAR6. This variant was heterozygous in the affected cat, unique to the exome data and not identified in the 99 Lives variant dataset.

A known feline disease variant was also re-identified\textsuperscript{121}. A solute carrier family 3-member 1 (\textit{SLC3A1}) variant was homozygous in a Greek cat presenting with cystinuria. The c.1342\textsuperscript{C>T} variant, causing a p.Arg448Trp at position A3:66539609 has been previously documented to be associated with this condition\textsuperscript{146}. No other cat in the exome dataset had this variant. Many of the variants associated with cat blood group B and its extended haplotype were detected in 11 cats, suggesting five cats as type B, one was confirmed\textsuperscript{141}. Variants were detected in \textit{APOBEC3}, which is associated with feline immunodeficiency virus (FIV) infection in cats, and three cats had the allelic combination producing the IRAVP amino acid haplotype that is associated with FIV resistance\textsuperscript{147}. Novel findings included two cats that were heterozygous for a porphyria variant in UROS (c.140\textsuperscript{C>T}, c.331\textsuperscript{G>A})\textsuperscript{148,149}, one cat which was homozygous for FXII deficiency variant (FXII_1631G>C)\textsuperscript{148}, and had died as a kitten, and one cat which was heterozygous for a copper metabolism deficiency in \textit{ATP7B}\textsuperscript{150}. Additional variants for neuronal ceroid lipofuscinosis, pycnodysostosis, Ehlers-Danlos syndrome, hypothyroidism, and hypovitaminosis D, and several individual-specific variants for hypertrophic cardiomyopathy are under further investigation (Table 3.1).

3.4 Discussion

In humans, WES has flourished over the past few years and is becoming more common in the practice of genomic medicine, especially newborn screening\textsuperscript{151}. Well-annotated genomes and
extensive resources, such as for human and mouse, have led to the development of various exome capture products ranging from those with a very limited focus, e.g., oncogene panels, to more extensive designs including 5’ and 3’ untranslated regions, predicted regulatory elements, and non-coding RNAs. For other mammals, exome capture designs have ranged from 44.6 Mb in pigs\textsuperscript{152} to 146.8 Mb in rats\textsuperscript{153,154}, illustrating the variation in experimental objectives. This is not currently the case for veterinary medicine due to several factors: a dog or cat owner's unwillingness to incur the costs, lower accuracy of available genome references\textsuperscript{155}, and the uncertainty of treatment options driven by sequence variant data. In companion animals, only the domestic dog has exome capture probes available, which span 53 to 152 Mb with an overlap of 34.5 Mb between the capture designs\textsuperscript{156,157}. In this study, a feline exome resource was developed by designing capture probes against the annotated Felis_catus_9.0 genome assembly, a highly contiguous assembly that enabled efficient probe design\textsuperscript{155}. The targeted 35.7 Mb accounts for the exons and 30bp of flanking sequences to minimize the loss of detectable splice donor and acceptor variants.

Success in disease variant identification in any species using WES is dependent on multiple factors, including mode of inheritance, sequencing depth, and efficient probe design that covers the regions of interest with high specificity, minimizing the number of off-target reads. Sequence coverage of $\geq 20x$ is generally regarded as the standard to efficiently detect heterozygous variants\textsuperscript{158}. At this threshold, an acceptable average target coverage of 96.4\% was obtained in our study. In our first WES experiment of 10 cats, we achieve maximum exonic coverage of 99\% with a mean depth of 267x at aligned bases. However, we have found this high-depth approach is not necessary or cost-efficient for the discovery of feline associated disease variants. The first domestic dog exome design\textsuperscript{156}, which covered 52.8 Mb distributed over 203,059 regions, had a
range of 87-90% mapped reads at a 102x mean sequencing depth. An updated canine design\textsuperscript{156} had 93.5% of the targeted bases (<53 Mb) covered to at least 1X depth of coverage, while in our feline exome design, the on-target reads were nearly 100% at 10x sequencing depth. Whilst absolute dog and cat exome comparisons are difficult due to the differences in annotation, genome assembly accuracy, and design techniques, both of these resources reveal acceptable performance.

The intended application of the cat WES was twofold: the identification of heritable, Mendelian diseases and traits, and somatic mutations in cancer. In this study, the focus was the former and included the assessment of the efficiency of the feline exome design for SNV discovery against ten matched WGS samples. The matched WGS and WES cats had an average of 30x and 267x depth of coverage, respectively, with the vast majority of SNVs and indels in overlapping regions being detected by both platforms. Altogether, these findings suggest the use of this feline exome probe set was extremely consistent with variant discovery from WGS, where 99.4% were uncovered in WGS while only 1.5% were absent from the WES cats. Consistent with large cohort human studies, indel discovery was less consistent (92.5% overlap) with 12.2% of WGS indels absent from WES data owing to the well-known short-read misalignment problem in regions with indels of varying size. Differences in the number of common variants between platforms is due to differential filtering, as common variants were identified prior to when filtering was performed. The percentage of exclusive variants per platform also varied according to variant impact, with high impact variants representing the largest percentage of exclusive variants for their impact class. Since high impact mutations are generally rare due to their impact on normal gene function, their enrichment within platform exclusive variant sets is
expected. In the same manner, as low impact variants have no impact on gene function, they are
less likely to be identified as platform exclusive within their variant class.

Previously characterized and unknown germline or somatic variants of clinical
significance, the former often not identifiable without the parents, were investigated to confirm if
each were identical or unique to genes associated with each disease or phenotype in prior studies.

Known variants were first confirmed to validate the accuracy of the cat exome design for the
following aesthetic traits: Agouti, Brown, Dense, Gloves, Dilution, Extension, Long, Lykoi, and
hairless coat types. In addition, disease variants were found in genes earlier shown to be
candidate alleles in hydrocephalus, hypertrophic cardiomyopathy, and progressive retinal
atrophy. These results importantly validate our design is capable of detecting variants with
prior trait association. Nonetheless, a primary study objective was to find new potential causal
variants in our small mixed disease and trait cohort of 31 domestic cats. This cohort was
searched to find novel candidate variants for three diseases and traits; feline autosomal dominant
polycystic kidney disease (ADPKD), atrichia, hypotricha. ADPKD is a common inherited
autosomal dominant disease affecting about 6% of the world's cats and is characterized by
fluid-filled cysts that form in the bilateral kidneys that often leads to renal failure. Many of the
features of feline ADPKD are similar to human ADPKD and recent studies demonstrated the
utility of the cat model. The c. 10063C>A mutation in exon 29 of PKD1 was the only
known causative allele for feline ADPKD, however, for human ADPKD, variants are found
throughout PKD1. A variant in polycystin 2 (PKD2), c.2211delG at position B1:134992553,
causes a p.Lys737Asnfs*2 and was identified in a Siberian cat from Europe, indicating
additional alleles may be segregating for ADPKD in cats.
Domestic cats have various forms of atrichia and hypotrichia, which even though each is characterized by baldness or loss of hair coat, are not considered diseased cats since breeders have selected upon these observed traits to develop new breeds. Only two breeds are recognized as completely hairless, the Sphynx and Donskoy. Donskoy cats are a breed of Russian cats in which loss of hair is determined by a semi-dominant allele\textsuperscript{163}. Peterbald cats were bred in Russia in 1994 as a product of a Donskoy and an Oriental Shorthair cross, and are often born with no hair, or lose their hair over time\textsuperscript{164}. Cornish Rex, a hypotrichia breed, that is characterized by a curly coat, is caused by a homozygous deletion mutation in \textit{LPAR6}\textsuperscript{165}. The Peterbald cat had an \textit{LPAR6} 4 base pair deletion that is in juxtaposition to a compound heterozygote for the Cornish rex deletion variant. Both variants result in premature stop codons a few amino acids downstream of the variant site. Other disease-associated variants were re-identified, such as cystinuria variants, in which the cat was homozygous and affected. Determination of allele frequencies through the 99 lives project\textsuperscript{155} improved the identification of cats that were heterozygous for variants associated with recessive diseases, such as, porphyria\textsuperscript{148}, Factor XII deficiency\textsuperscript{166}, and copper metabolism\textsuperscript{150}. The inclusion of 99 Lives WGS data was central to establishing the likelihood of variants being causal for diseases and further cross-species explorations of variant frequencies promises to better define variants of uncertain significance\textsuperscript{167}.

Clinical use of sequence variant information in companion animals is in the very early stages, which hampers the ability of veterinarians to rapidly diagnose some diseases without standard or unclear phenotypic determinants. In the future, it could be used to adapt treatments to the specific animal and disease type\textsuperscript{168}. Many diagnosed rare diseases have a poor prognosis, with some less than 90 days; thus, cost-effective sequencing approaches may help discover alternate and more effective treatments. The Undiagnosed Diseases Program of the National
Institutes of Health routinely uses WES for this purpose of finding treatments where none exist, suggesting veterinary medicine could benefit in the same manner. We confirm here, as other studies have shown, that WES is cost-effective, data process-efficient (by requiring less computing time), and easier to use than WGS for inferring a variant’s biological relevance. As in the dog, a first step is offered toward the use of feline WES for robust disease variant detection, including the validation of previously identified causal alleles and the discovery of novel candidate variants that we suggest are of interest for further experimental scrutiny. We have developed domestic cat-specific WES, and importantly, based on our findings, validated its use for the evaluation of potential disease variants for the future practice of feline genomic medicine.
CHAPTER 4

A Genetic Profile of Feline Oral Squamous Cell Carcinoma

4.1 Introduction

Feline oral squamous cell carcinoma (FOSCC) is the fourth most common cancer, and the most commonly found oral tumor in cats\textsuperscript{172}, with a one-year survival rate of less than 10\%\textsuperscript{76}. This cancer affects squamous cell lining in the oral and oropharyngeal cavity; gingiva, tongue, and sublingual regions\textsuperscript{172} and rarely metastasizes to distant locations; however, local bone invasion and lymph nodes can be affected. FOSCC is often diagnosed at stages too late to intervene with symptoms including oral pain successfully clinically, difficulty swallowing, loss of teeth, and anorexia. Early studies have shown that the use of flea collars, feeding with mostly canned foods were found to increase developing FOSCC 5-fold compared to those who did not have canned foods or tuna\textsuperscript{20}. FOSCC has a limited number of treatment options, with the best typically being surgical resection of the tumor from the oral cavity, if feasible, followed by radiation and/or chemotherapy if concerns arise about risk of recurrence or metastasis based on histopathology. However, due to the high reoccurrence rate, these therapies often have poor outcomes and often the recommendation is euthanasia may end up being the best option\textsuperscript{77,173}. FOSCC presents with similar risk factors and molecular mechanisms as human head and neck cancer (HNSCC) and may present an opportunity as a model for therapy\textsuperscript{174}.

HNSCC is the sixth most common cancer found among humans worldwide, with 550,000 new cases per year and also has a low 5-year survival rate of less than 50\% with 275,000 deaths per year\textsuperscript{175,176}. Like FOSCC, if HNSCC is diagnosed in early stages survival rates are much higher at 82\%, compared to 26\% if the tumor is distantly metastasized\textsuperscript{76,177,178}. Known risk
factors for HNSCC include exposure to tobacco smoke, alcohol, and infection with HPV\textsuperscript{177-179}. Little is understood about molecular mechanism similarities in cats and humans with only candidate gene approaches in the cat. Both humans and cats show the perturbed function of p53, causing issues in a cell’s metabolism, cycle arrest, and apoptosis\textsuperscript{180}. FOSCC and HNSCC are both similar in disease progression and biologic behavior with both cancers being locally invasive with metastasis to regional lymph nodes\textsuperscript{174}. Overexpression of EGFR is found in 69 to 100% of FOSCC and 90% of HHNSCC, causing cell cycle progression, uncontrolled proliferation, and invasion through the activation of intracellular tyrosine kinase and is associated with poor prognosis in human\textsuperscript{84,85,174,181}. Similarities such as low survival rates, similar genetic, and morphological profiles in feline may offer a viable model to study HNSCC.

Naturally occurring animal models of cancer are becoming more integral to a better understanding of tumor evolution and progression such as HNSCC compared to rodent models\textsuperscript{174,182}. Murine models, for example, lack important factors that contribute to spontaneous tumor formation and follow-up adaptive immune responses. Understanding the mutually exclusive genetic environment of FOSCC will allow us to determine how informative a comparative model model is to study which molecular candidates are involved in tumors formation in both FOSCC and HNSCC. With many uncertainties in the genetics of FOSCC, we aimed to characterize the mutational and transcriptional profile of FOSCC thus gaining insight into its molecular pathogenies when comparing to HNSCC. To accomplish this, WES and RNAseq was performed. FOSCC tumor tissue and matching blood samples were used for WES, and RNA-seq was generated on FOSCC tumor tissue and oral cavity samples from healthy cats. Following these experiments, we searched among candidate genes for monotherapy matches that
offer a retrospective view of missed mutation druggable mutations for treating feline patients with FOSCC.

4.2 Methods

Clinical samples. Cryopreserved tissues corresponding to 6 FOSCC tumor samples and 6 matching whole blood samples, as well as 3 normal oral mucosal samples collected from healthy animals, were used (Table 4.1). Tumor samples had been collected during standard-of-care surgical procedures and stored by the Cornell Veterinary Biobank until retrieved for analysis. Sample collection was performed in accordance with a protocol (#2005-0151) approved by Cornell University’s Institutional Animal Care and Use Committee. Accordingly, informed consent to authorize the use of tissue samples and clinical data for research purpose was obtained from cat owners prior to sample collection, and undue harm was never inflicted to client-owned cats for the purposes of this study; all methods were performed in accordance with the relevant guidelines and regulations. The diagnosis of FOSCC was validated using routine hematoxylin and eosin-stained samples archived by the Anatomic Pathology Section at Cornell University’s College of Veterinary Medicine by a board-certified veterinary pathologist (ADM); tumors were diagnosed following previously described criteria183 while blinded to molecular assays.

FOSCC histology. Histological assessment of tissues was done using routine hematoxylin and eosin-stained samples archived by the Anatomic Pathology Section at Cornell University’s College of Veterinary Medicine by a board-certified veterinary pathologist (ADM); tumors were diagnosed following previously described criteria183 while blinded to molecular assays (Figure 4.1).
### Table 4.1 Cohort Characteristics

<table>
<thead>
<tr>
<th>Sample</th>
<th>Location Collected</th>
<th>Sex</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>23263</td>
<td>Tongue</td>
<td>MC</td>
<td>Adult</td>
</tr>
<tr>
<td>9895</td>
<td>Mandible/Gingiva</td>
<td>FS</td>
<td>Adult</td>
</tr>
<tr>
<td>26903</td>
<td>Mandible</td>
<td>unknown</td>
<td>Adult</td>
</tr>
<tr>
<td>7741</td>
<td>Oral Cavity</td>
<td>MC</td>
<td>Adult</td>
</tr>
<tr>
<td>24147</td>
<td>Oral Cavity/Gingiva</td>
<td>MC</td>
<td>Adult</td>
</tr>
<tr>
<td>605591</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Adult</td>
</tr>
</tbody>
</table>

MC = male castrated

FS = female spayed
**Figure 4.1 Pathology of Oral Squamous Carcinoma**

A. Underlying a moderately hyperplastic gingival epithelium are ribbons, cords, and trabeculae of neoplastic squamous epithelial cells.  
B. Neoplastic squamous epithelial cells surround and produce brightly eosinophilic keratin (arrows).  
C. Neoplastic squamous epithelial cells that are enmeshed in abundant scirrhous response (asterisk) are associated with marked bony invasion and remodeling (arrow).  
D. Neoplastic squamous epithelial cells with dyskeratosis (arrow) invade the dentin layer of a tooth (asterisk). Images provided by......

**Image info**

A. 79666; 40x mag  
B. 16236; 100x mag  
C. 76610; 200x mag  
D. 89317; 200x mag
Somatic Variant Calling, Annotation, and Filtering. Raw sequence reads were mapped to Felis_Catus_9.0 reference using Burrows-Wheeler Aigner (BWA) v0.7.17. Sam files were sorted and converted to bam and merged using Piccard tools v2.18.9. PCR and optical duplicates were marked using Piccard tools v.2.19.9. These files were then processed through the Genome Analysis Toolkit (GATK) v.4.0.1 for base quality rescore calibration. Mutect2 was used to identify somatic single nucleotide variants (SNVs) and insertions and deletions (indels) and filtered through the standard MuTect2 filters, such as \textit{t\_lod\_fstar}, filters out variants with insufficient evidence of presence in tumor sample and panel\_of\_normals, which filters out variants present in at least two samples in the panel of normal (Supplemental Figure 4.1). VCF files were further filtered for missing data and minor allele frequencies less than .1\% using VCFtools. SNVs were then annotated using Variant effect Predictor (VEP) v. 101.0. TMB was calculated using the number of non-synonymous SNVs, somatic mutations altering the amino acid sequence, found per megabase in the coding regions. We used an exome size of 35MB that was calculated based on the size of coding regions in the cat genome.

Measuring tumor mutational burden (TMB). TMB has been reported to help classify whether a cancer type is more amenable to immunotherapy due to the presentation of multiple antigen targets. For each VCF file, Felis_Catus_9.0 was set as the reference and Ensembl (release 102) genes was used as the definition. TMB was calculated using the number of non-synonymous SNVs, somatic mutations altering the amino acid sequence, found per megabase in the coding regions. We used an exome size of 35MB that was calculated based on the size of coding regions in the cat genome.

Driver Database Version 3/OncoKB: Driver Database version 3 was accessed online on 3/15/21(http://driverdb.tms.cmu.edu.tw/cancer). Cancer driver genes database was selected, and then the HNSCC data set was selected that included 263 genes. In addition, OncoKB was
accessed on the same day, and 10 HNSCC druggable genes were obtained. All gene sets were used for further investigation of their presence in the FOSCC samples.

4.3 Results

**FOSCC cohort characteristics.** In this study, we analyzed six FOSCC samples collected by the Cornell Veterinary Biobank between 6/19/14 and 9/19/19. Only 6 samples were available at the time of analysis. Samples included three castrated males, one spayed female, and two samples of unknown sex (Table 4.1). Samples were collected from various locations in the head and neck region, including the mandible, gingiva, and tongue. Due to varied anatomical location and sex, we were unable to account for differences in sex or spatial context for tumor genetic changes.

**SNV annotation.** In the six samples, we found 1,057 synonymous and nonsynonymous variants with a mean of 176 variants per FOSCC (Figure 4.2 and 4.3). After variant visualization using IGV, we found a false call rate of 5% across all 6 samples for 20 genes (Supplemental Figure 4.2). Among all somatic SNVs, 56 and 731 were nonsense and missense, respectively. Only one gene, *TP53*, a commonly mutated gene in many cancer types, including head and neck cancer\(^\text{180}\) showed multiple occurrences of SNVs with 83% showing somatic mutations as differing positions (Figure 4.4). Four samples had mutations in *TP53*, with three being missense and one a frameshift.

TMB has emerged as a biomarker for human patient stratification toward immunotherapy. However, the prognostic value of TMB across cancer types is uncertain. Using our small cohort we generated a preliminary TMB estimate of FOSCC to compare to earlier estimates in HNSCC. Using non-synonymous somatic mutations, we calculated the TMB for each feline tumor (Supplemental Table 4.1). The mean TMB for all six samples was 3.7 with a range of 1.4 - 8.5.
We were not able to determine if survival was associated with TMB score due to all cats being euthanized soon after the time of tumor resection.
Figure 4.2 All variants called, including synonymous variants, using GATK-Mutect2.
Figure 4.3 Number of Non-Synonymous somatic variants per sample. Non-Synonymous somatic variants called across all samples. Average of 176 variants over six samples including 56 nonsense high impact mutations and 731 missense mutations.
Figure 4.4 Location of the four TP53 mutations found in five of our FOSCC samples.
**FOSCC comparison with HNSCC.** We compared the somatic mutations of FOSCC to genes shown have highly recurrent mutations in HNSCC to assess gene model similarities. Genes harboring FOSCC non-synonymous SNVs were matched to the same HNSCC genes in the Driver Database 3 and OncoKB databases. The TP53 gene has the most frequent somatic mutations in HNSCC, and was also the most recurrent in FOSCC, with four of the samples having a missense mutation, and sample 24147 containing a compound heterozygous mutation, specifically a missense and splice region variant (Figure 4.4). A missense mutation in KAT2B was present in two samples, with several other missense mutations present in only one gene per sample (Table 4.2). Samples 9895 and 23263 had the most genes implicated in HNSCC with four genes in common each (Table 4.2). This data indicates some evidence for overlap in mutational background between HNSCC and FOSCC but is very preliminary at this stage.

**4.4 Discussion**

In this first study of the somatic mutations and gene expression variation present in FOSCC, we describe the similarities and differences when comparing to HNSCC. FOSCC and HNSCC have similarly low survival rates and morphological profiles that suggests a FOSCC model may offer some comparative insight into HNSCC therapeutic strategies or mechanism of action. The scope of FOSCC somatic mutations, however, is not known which we partially address by exploring comparable molecular features between species, i.e., somatic mutation and gene expression patterns. We found several genes (Table 4.2) overlapping with HHNSCC, for example the surprising overlap with TP53, the most recurrent in our FOSCC samples. TP53 is the most mutated in all cancer types, but also has the highest prevalence of driver mutations in HNSCC (41%).
<table>
<thead>
<tr>
<th>Sample</th>
<th>Gene</th>
<th>Annotation</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>7741</td>
<td>TP53</td>
<td>Frameshift/Splice region variant</td>
<td>E1:2541686</td>
</tr>
<tr>
<td>9895</td>
<td>TP53</td>
<td>Missense</td>
<td>E1:17726778</td>
</tr>
<tr>
<td></td>
<td>MED12L</td>
<td>Missense</td>
<td>C2:115279831</td>
</tr>
<tr>
<td></td>
<td>PXYLP1</td>
<td>Missense</td>
<td>C2:124530451</td>
</tr>
<tr>
<td>26903</td>
<td>TP53</td>
<td>In-Frame Deletion</td>
<td>E1:2541698</td>
</tr>
<tr>
<td></td>
<td>KAT2B</td>
<td>Missense</td>
<td>C2:139167628</td>
</tr>
<tr>
<td>23263</td>
<td>KAT2B</td>
<td>Missense</td>
<td>C2:139217714</td>
</tr>
<tr>
<td></td>
<td>ARID1A</td>
<td>Missense</td>
<td>C1:20180520</td>
</tr>
<tr>
<td></td>
<td>KMT2D</td>
<td>Missense</td>
<td>B4: 78023698</td>
</tr>
<tr>
<td></td>
<td>TP53</td>
<td>Frameshift</td>
<td>E1:2541322</td>
</tr>
</tbody>
</table>

**Table 4.2** Variant annotations for variants found in FOSCC in common with known HNSCC.
Given *TP53* is a tumor suppression gene, and certain somatic mutations are associated with lower survival outcomes in patients with HNSCC\(^\text{189}\). With *TP53* missense variants in three of the samples, an insertion in one, and a compound heterozygous mutation in another it is reasonable to predict these cats experienced rapid tumor progression yet missing clinical follow up prevented us from drawing this conclusion *(Figure 4.4)*. Coincidently, *TP53* mutations have been implicated in other FOSCC studies, with mutations in this gene found in 24 - 69% of cancers\(^\text{190-192}\). In HNSCC, *TP53* mutations are found in 70% of all cases, with variation in *TP53* being a predictive marker for immunotherapy in those with metastatic HNSCC\(^\text{180}\).

As estimates of the TMB have been used to predict positive patient response to immune checkpoint inhibitor therapy in some cancer types, e.g. non-small cell lung cancer and melanoma\(^\text{193,194}\), we sought to compare FOSCC to HNSCC for this metric. Recent studies on HNSCC have found that mutations in *TP53* are associated with high TMB and low overall survival rates, and high TMB patients responded well to immunotherapy\(^\text{195,196}\). In our FOSCC samples the average TMB was 3.7 with a range of 1.4 - 8.5. Cancer studies that calculated HNSCC TMB as high (>5.0) or low (<5.0), show higher TMB is associated with poor prognosis\(^\text{197,198}\). Our first TMB estimates in FOSCC fall within the observed HNSCC range suggesting some benefits could be gained for immunotherapy outcomes. More FOSCC sample sequencing to obtain better estimates of TMB is needed as well as the future availability of immune checkpoint inhibitors that could collectively be used to substantially improve outcomes for this very lethal cancer.

Other FOSCC recurrent or single gene mutations of interest were *KAT2B, ARID1A, MED12L, HOXB3*, and *PXYLP1* each with interesting features for comparative inference *(Table 3)*. Two samples had a missense variant in *KAT2B* *(Table 3)*. *KAT2B* is a gene that codes for an enzyme that functions as a histone acetyltransferase (HATs) and mutations in this gene have
been implicated in many diseases including cancers\textsuperscript{199}; however, it has not been evaluated in FOSCC. \textit{KAT2B} is part of a family of lysine transferases that are responsible for the acetylation of genes that targets a broad range of proteins and can function as tumor suppressors and oncogenes\textsuperscript{200}. \textit{KAT2B} is also responsible for inhibiting cell cycle progression and counteracting mitogenic activity. HNSCC cell line studies have shown universal loss of \textit{KAT2B}\textsuperscript{175}, and a study using HNSCC tumors found significantly lower expressions of \textit{KAT2B} compared to the normal tissue\textsuperscript{201}. \textit{KMT2D} was identified in only one sample but has similar epigenetic properties to \textit{KAT2B}. Studies completed by The Cancer Genome Atlas have shown mutations often occurring in \textit{KMT2D}, keeping chromatin in an open state, thereby promoting gene expression\textsuperscript{202,203}. Mutations in both \textit{KAT2B} and \textit{KMT2D} possibly induce epigenetic changes in HHNSCC and FOSCC which alternative to immunotherapy could and may open an avenue to study epigenetic drug control of both human and feline oral squamous cell carcinoma\textsuperscript{200,203}. 

A missense mutation in \textit{ARID1A} was also found in one of our samples (Table 4.2). \textit{ARID1A} is a gene that is often found to be deleted in many human cancers; however, to our knowledge the effect of this gene in feline cancers is unknown. \textit{ARID1A} functions as a tumor suppressor and tumor stemness repressor by disrupting the perturbed function of p53 or PTEN pathways\textsuperscript{204,205}. The upregulation of miR-31 is known to have oncogenic properties in human head and neck cancer and studies show that elevated expression of this miRNA causes reduced expression of \textit{ARID1A}, and patients with low expression of \textit{ARID1A} are found to have the worst survival rates\textsuperscript{205,206}. 

Several other genes were found to overlap in HHNSCC and FOSCC. \textit{MED12L}, \textit{HOXB3}, and \textit{PXYP1} were also identified as single nucleotide variants in one sample each (Table 4.4); however, these genes are understudied in both HHNSCC and FOSCC. Mediator Complex
Subunit 12L (MED12L) works by activating the kinase activity of CDK8 which regulates the growth and division of cells\textsuperscript{207}. Studies report significant differentiation of the expression of MED12L in many cancers including head and neck cancer; however, an altered MED12 complex is altered in 3.05% of HNSCC patients\textsuperscript{208,209}. HOX genes regulate a wide range of cell activity including proliferation and migration. HNSCC studies have shown am overall elevation in all HOX genes, including HOXB3\textsuperscript{210-212}. There was no overlap in actionable genes from OncoKB related to head and neck cancer. We believe this is due to the low sample size, and further studies need to be conducted.

In FOSCC, we have identified the somatic mutations landscape by exome sequencing. For the best-known cancer driver gene, TP53, we observe mutations that despite their presence in FOSCC, are not the same variants as those observed in human head and neck cancer. But several other genes also overlap between the two types of cancer in our small cohort suggesting the use of similar genes that initiate tumorigenesis and perhaps future comparative models of treatment. This small exploratory study demonstrated the ability to call variants unique to feline oral squamous cell carcinoma tumors, identified common genes between human and feline oral squamous carcinoma. This study presents a starting point to study FOSCC in a larger cohort of feline oral squamous carcinoma patients and draw more similarities between human and feline oral cancer. With the further development of this technology, we may be able to diagnose this cancer at earlier stages using genomic methods, as well as possibly develop immunotherapy treatments for both cats and humans.
References


Morris, D. *Cat Watching*. (Ebury PR, 1985).


Michael R. Green, J. S. *Molecular Cloning: A Laboratory Manual Fourth.*


Drögemüller, C., Rüfenacht, S., Wichert, B. & Leeb, T. Mutations within the FGF5 gene are associated with hair length in cats. *Animal genetics* 38, 218-221, doi:10.1111/j.1365-2052.2007.01590.x (2007).


Supplementary Table 3.1

<table>
<thead>
<tr>
<th>Cat</th>
<th>TOTAL READS</th>
<th>Unique Reads</th>
<th>Duplicate Reads</th>
<th>% Unique Reads</th>
<th>Mapped Unique Reads</th>
<th>MEAN COVERAGE</th>
<th>MEDIAN COVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>215022748</td>
<td>176614831</td>
<td>38407917</td>
<td>82.14</td>
<td>176570410</td>
<td>99.97</td>
<td>373</td>
</tr>
<tr>
<td>2</td>
<td>145242280</td>
<td>118605692</td>
<td>26636588</td>
<td>81.66</td>
<td>118567772</td>
<td>99.97</td>
<td>199</td>
</tr>
<tr>
<td>3</td>
<td>83764798</td>
<td>68089930</td>
<td>15674868</td>
<td>81.29</td>
<td>68065004</td>
<td>99.29</td>
<td>211</td>
</tr>
<tr>
<td>4</td>
<td>164681954</td>
<td>134926578</td>
<td>29755376</td>
<td>81.93</td>
<td>134884666</td>
<td>99.97</td>
<td>232</td>
</tr>
<tr>
<td>5</td>
<td>192670988</td>
<td>158132174</td>
<td>34538815</td>
<td>82.07</td>
<td>158090364</td>
<td>99.97</td>
<td>331</td>
</tr>
<tr>
<td>6</td>
<td>66980388</td>
<td>53730990</td>
<td>13249398</td>
<td>80.22</td>
<td>53700410</td>
<td>99.94</td>
<td>88</td>
</tr>
<tr>
<td>7</td>
<td>237374507</td>
<td>195097488</td>
<td>42277019</td>
<td>82.19</td>
<td>195050455</td>
<td>99.98</td>
<td>416</td>
</tr>
<tr>
<td>8</td>
<td>55650338</td>
<td>45654534</td>
<td>9995804</td>
<td>82.04</td>
<td>45633063</td>
<td>99.95</td>
<td>76</td>
</tr>
<tr>
<td>9</td>
<td>170319229</td>
<td>139649516</td>
<td>30669713</td>
<td>81.99</td>
<td>139610319</td>
<td>99.97</td>
<td>289</td>
</tr>
<tr>
<td>10</td>
<td>259726266</td>
<td>213580145</td>
<td>46146121</td>
<td>82.23</td>
<td>213530500</td>
<td>99.98</td>
<td>458</td>
</tr>
</tbody>
</table>

Mean 159143350 130408188 28735162 81.94 130370296 99.97 267 195

68
<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>57911238</td>
<td>48615406</td>
<td>9295832</td>
<td>83.95</td>
<td>99.84</td>
<td>78.8</td>
<td>71</td>
</tr>
<tr>
<td>22</td>
<td>82008038</td>
<td>64579700</td>
<td>17428338</td>
<td>78.75</td>
<td>99.85</td>
<td>108.2</td>
<td>97</td>
</tr>
<tr>
<td>23</td>
<td>92835910</td>
<td>72587612</td>
<td>20248298</td>
<td>78.19</td>
<td>99.88</td>
<td>103.5</td>
<td>94</td>
</tr>
<tr>
<td>24</td>
<td>63399300</td>
<td>52257320</td>
<td>11141980</td>
<td>82.43</td>
<td>99.85</td>
<td>85.1</td>
<td>77</td>
</tr>
<tr>
<td>25</td>
<td>53514212</td>
<td>44731352</td>
<td>8782860</td>
<td>83.59</td>
<td>99.85</td>
<td>75.1</td>
<td>68</td>
</tr>
<tr>
<td>26</td>
<td>46245536</td>
<td>38645300</td>
<td>7600236</td>
<td>83.57</td>
<td>99.85</td>
<td>60.0</td>
<td>54</td>
</tr>
<tr>
<td>27</td>
<td>66297558</td>
<td>55692274</td>
<td>10605284</td>
<td>84.00</td>
<td>99.86</td>
<td>88.7</td>
<td>80</td>
</tr>
<tr>
<td>28</td>
<td>63195078</td>
<td>52759900</td>
<td>10435178</td>
<td>83.49</td>
<td>99.86</td>
<td>83.6</td>
<td>76</td>
</tr>
<tr>
<td>29</td>
<td>54045256</td>
<td>44167266</td>
<td>9877990</td>
<td>81.72</td>
<td>99.77</td>
<td>70.1</td>
<td>64</td>
</tr>
<tr>
<td>30</td>
<td>49086732</td>
<td>40778946</td>
<td>8307786</td>
<td>83.08</td>
<td>99.78</td>
<td>66.5</td>
<td>60</td>
</tr>
<tr>
<td>31</td>
<td>46079700</td>
<td>37352402</td>
<td>8715568</td>
<td>81.08</td>
<td>99.83</td>
<td>60.1</td>
<td>54</td>
</tr>
<tr>
<td>32</td>
<td>61427656</td>
<td>47373818</td>
<td>14053838</td>
<td>77.12</td>
<td>99.85</td>
<td>77.2</td>
<td>70</td>
</tr>
<tr>
<td>33</td>
<td>72513806</td>
<td>56798782</td>
<td>15715024</td>
<td>78.33</td>
<td>99.89</td>
<td>90.7</td>
<td>83</td>
</tr>
<tr>
<td>34</td>
<td>46864618</td>
<td>39837652</td>
<td>7026966</td>
<td>85.01</td>
<td>99.88</td>
<td>63.7</td>
<td>57</td>
</tr>
<tr>
<td>35</td>
<td>57105380</td>
<td>45115204</td>
<td>11990176</td>
<td>79.00</td>
<td>99.91</td>
<td>62.2</td>
<td>57</td>
</tr>
<tr>
<td>36</td>
<td>49270872</td>
<td>41010968</td>
<td>8259904</td>
<td>83.24</td>
<td>99.85</td>
<td>63.0</td>
<td>57</td>
</tr>
<tr>
<td>37</td>
<td>57055464</td>
<td>47567866</td>
<td>9487598</td>
<td>83.37</td>
<td>99.84</td>
<td>75.9</td>
<td>69</td>
</tr>
<tr>
<td>38</td>
<td>72860408</td>
<td>60362784</td>
<td>12497624</td>
<td>82.85</td>
<td>99.86</td>
<td>97.2</td>
<td>88</td>
</tr>
<tr>
<td>39</td>
<td>65352134</td>
<td>52724620</td>
<td>12627514</td>
<td>80.68</td>
<td>99.82</td>
<td>83.8</td>
<td>76</td>
</tr>
<tr>
<td>40</td>
<td>61091136</td>
<td>49625330</td>
<td>11465806</td>
<td>81.23</td>
<td>99.85</td>
<td>77.7</td>
<td>69</td>
</tr>
<tr>
<td>41</td>
<td>53287642</td>
<td>41298322</td>
<td>11989320</td>
<td>77.50</td>
<td>99.85</td>
<td>65.9</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>61853485</td>
<td>50145194</td>
<td>11708291</td>
<td>81.07</td>
<td>99.85</td>
<td>80</td>
<td>72</td>
</tr>
<tr>
<td>Cat</td>
<td>Median Coverage</td>
<td>% target bp covered &gt;10x</td>
<td>% target bp covered &gt;20x</td>
<td>% target bp covered &gt;30x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>----------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>264</td>
<td>99</td>
<td>100</td>
<td>103</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>181</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>103</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>237</td>
<td>99</td>
<td>100</td>
<td>101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>79</td>
<td>99</td>
<td>97</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>292</td>
<td>100</td>
<td>101</td>
<td>104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>69</td>
<td>99</td>
<td>97</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>209</td>
<td>99</td>
<td>99</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>320</td>
<td>100</td>
<td>101</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Mean | 195 | 99 | 99 | 99 |

<table>
<thead>
<tr>
<th>Cat</th>
<th>Median Coverage</th>
<th>% target bp covered &gt;10x</th>
<th>% target bp covered &gt;20x</th>
<th>% target bp covered &gt;30x</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>54</td>
<td>98</td>
<td>95</td>
<td>86</td>
</tr>
<tr>
<td>26</td>
<td>54</td>
<td>98</td>
<td>95</td>
<td>87</td>
</tr>
<tr>
<td>41</td>
<td>57</td>
<td>98</td>
<td>93</td>
<td>83</td>
</tr>
<tr>
<td>34</td>
<td>57</td>
<td>98</td>
<td>96</td>
<td>88</td>
</tr>
<tr>
<td>35</td>
<td>57</td>
<td>98</td>
<td>96</td>
<td>89</td>
</tr>
<tr>
<td>20</td>
<td>57</td>
<td>98</td>
<td>96</td>
<td>89</td>
</tr>
<tr>
<td>36</td>
<td>57</td>
<td>98</td>
<td>96</td>
<td>89</td>
</tr>
<tr>
<td>30</td>
<td>60</td>
<td>98</td>
<td>96</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>61</td>
<td>98</td>
<td>96</td>
<td>90</td>
</tr>
<tr>
<td>29</td>
<td>64</td>
<td>98</td>
<td>96</td>
<td>91</td>
</tr>
<tr>
<td>11</td>
<td>65</td>
<td>99</td>
<td>96</td>
<td>90</td>
</tr>
<tr>
<td>25</td>
<td>68</td>
<td>98</td>
<td>97</td>
<td>92</td>
</tr>
<tr>
<td>40</td>
<td>69</td>
<td>99</td>
<td>96</td>
<td>91</td>
</tr>
<tr>
<td>37</td>
<td>69</td>
<td>98</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>32</td>
<td>70</td>
<td>99</td>
<td>97</td>
<td>93</td>
</tr>
<tr>
<td>21</td>
<td>71</td>
<td>96</td>
<td>91</td>
<td>85</td>
</tr>
<tr>
<td>13</td>
<td>72</td>
<td>99</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>16</td>
<td>73</td>
<td>99</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>28</td>
<td>76</td>
<td>99</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>39</td>
<td>76</td>
<td>99</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>24</td>
<td>77</td>
<td>99</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>15</td>
<td>80</td>
<td>99</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>27</td>
<td>80</td>
<td>99</td>
<td>97</td>
<td>94</td>
</tr>
<tr>
<td>14</td>
<td>82</td>
<td>99</td>
<td>98</td>
<td>95</td>
</tr>
<tr>
<td>33</td>
<td>83</td>
<td>99</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td>17</td>
<td>84</td>
<td>99</td>
<td>98</td>
<td>95</td>
</tr>
<tr>
<td>18</td>
<td>85</td>
<td>97</td>
<td>94</td>
<td>91</td>
</tr>
<tr>
<td>38</td>
<td>88</td>
<td>99</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td>23</td>
<td>94</td>
<td>99</td>
<td>98</td>
<td>97</td>
</tr>
<tr>
<td>19</td>
<td>95</td>
<td>99</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td>22</td>
<td>97</td>
<td>99</td>
<td>98</td>
<td>96</td>
</tr>
<tr>
<td>Mean</td>
<td><strong>72.00</strong></td>
<td><strong>98.45</strong></td>
<td><strong>96.41</strong></td>
<td><strong>91.84</strong></td>
</tr>
</tbody>
</table>

1681

1682

1683

1684

1685
Supplementary Figure 3.1

This is a visual of the reads that overlap the dwarfism structural variant in the dwarfism cat sample. The first row is the WES reads showing no evidence of a structural variant in the UDGH gene. Row 2 is the WGS reads showing a deletion and rearrangement (green and red). Thus, showing that WES does not adequately cover structural variants.
Supplemental Figure 4.1 Workflow of GATK-Mutect2 Pipeline
Tumor Realigned BAM File
Create tumor quality score recalibration table - GATK
Tumor Recalibration Table
Apply recalibration - GATK
Tumor Recalibrated BAM File
Panel of Normals
Somatic SNV and indel calling - Mutect2 GATK
VCF with Candidate Somatic Variants
Normal Realigned BAM File
Create normal quality score recalibration table - GATK
Normal Recalibration Table
Apply recalibration - GATK
Normal Recalibrated BAM File
Panel of Normals
Somatic SNV and indel calling - Mutect2 GATK
VCF with Candidate Somatic Variants
Supplementary Figure 4.2. GATK-Mutect2 called a SNV at position 167,962,120 but was determined to be a false call because there was only one variant called out of all the reads.
Supplementary Table 4.1. Tumor mutational burden calculated for all samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>TMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>7741</td>
<td>8.5</td>
</tr>
<tr>
<td>9895</td>
<td>2.6</td>
</tr>
<tr>
<td>26903</td>
<td>1.5</td>
</tr>
<tr>
<td>60551</td>
<td>1.4</td>
</tr>
<tr>
<td>23263</td>
<td>6</td>
</tr>
<tr>
<td>24147</td>
<td>2.5</td>
</tr>
</tbody>
</table>