THE USE OF MASSIVELY PARALLEL SEQUENCING TO STUDY THE VIROME,
EPIGENOME AND GENOME OF CANINE AND FELINE CANCERS

A thesis
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
Of the requirements for the degree
Doctorate of Philosophy

By
SHIRLEY CHU
Dr. Jeffrey Bryan, Thesis Supervisor
DECEMBER 2019

© Copyright by Shirley Chu 2019

All Rights Reserved

The undersigned, appointed by the dean of the Graduate School, have examined the
dissertation entitled

THE USE OF MASSIVELY PARALLEL SEQUENCING TO STUDY THE VIROME,
EPIGENOME AND GENOME OF CANINE AND FELINE CANCERS

Presented by Shirley Chu,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Jeffrey Bryan

Assistant Professor Angela McCleary-Wheeler

Associate Professor James Amos-Landgraf

Dean Carolyn Henry

ACKNOWLEDGEMENTS

I would like to acknowledge Jeffrey Bryan, who has been my scientific and clinical advisor during my internship, residency and PhD. I am lucky to have him as a mentor because he truly cares about his patients, clients, colleagues and mentees. He inspires his mentees with his accomplishments, positive attitude, and unwavering support. Because he believed in me I had the opportunity to develop an interest in genomics and Golden Retrievers. Genomics and Golden Retrievers are now my main research interests and I hope to dedicate my career advancing cancer genomics in veterinary medicine and improving the lives of Golden Retrievers through early detection, more effective treatment and preventative measures in cancer.

My committee members Dr. Kristen Taylor, Dr. Gerald Arthur, Dr. Carolyn Henry, Dr. James Amos-Landgraf and Dr. Angela McCleary-Wheeler have provided the much needed encouragement and support during this process. Drs. MacDonald-Dickinson, Elliot, Saam and Sabhlok were the medical oncologists who shared their passion for treating cancer patients with me during my DVM training. I could not have asked for a better oncology family, Drs. Menard, Robinson, Rippey, Norquest, Reich, Couto, Flesner, Donnelly, Ehling, Maitz, Willcox, Bechtel, Selting, Rindt and Kumar; Anastasia, Deb, Jeanna, Mindy, Kaitlin, Jeff, Brianna, Claire, Kelsey, Mary and Gaye. Thank you Diane Brown for giving me the opportunity to ask and answer my own research question. I am honored to have had the support of the Canine Health Foundation. Thank you PEO for supporting me and the thousands of female

researchers. To all the clients that I have come across, thank you for trusting me and reminding me why I need to improve the lives of our companions.

Lastly, I would like to thank my family. I am grateful for all the sacrifices that were made to give me the opportunity to keep reaching higher. Thank you OLG for introducing me to bioinformatics, providing limitless and selfless support and understanding. I dedicate this dissertation to SC.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF FIGURES.....	vi
LIST OF TABLES.....	viii
ABSTRACT.....	ix
INTRODUCTION.....	1
CHAPTER 1. A VIROME SEQUENCING APPROACH TO FELINE ORAL SQUAMOUS CELL CARCINOMA TO EVALUATE CAUSATIVE FACTORS.....	6
1.1 Abstract.....	6
1.2 Introduction.....	9
1.3 Materials and Methods.....	15
1.4 Results.....	21
1.5 Discussion.....	29
1.6 Conclusions.....	43
CHAPTER 2. GENOME WIDE EXPLORATION OF THE METHYLOME IN DIFFUSE LARGE B CELL LYMPHOMA IN GOLDEN RETRIEVERS REVEALS A STABLE HYPERMETHYLOME.....	45
2.1 Introduction.....	47
2.2 Materials and Methods.....	54
2.3 Results and Discussion.....	62
2.4 Conclusions.....	92

CHAPTER 3. UNRAVELING THE CHAOTIC GENOMIC LANDSCAPE OF CANINE OSTEOSARCOMA WITH CURRENT SEQUENCING TECHNOLOGIES AND BIOINFORMATICS APPROACHES.....	94
3.1 Abstract.....	94
3.2 Introduction.....	97
3.3 Materials and Methods.....	99
3.4 Results.....	104
3.5 Discussion.....	126
3.6 Conclusions.....	138
CONCLUSIONS AND FUTURE DIRECTIONS.....	143
BIBLIOGRAPHY	149
VITA.....	185

LIST OF FIGURES

Figure 1.1 Summary of the viral groups and families that were identified in this ViroCap study.

Figure 1.2. Summary of the viruses identified by ViroCap from each patient in the FOSCC, positive and negative control groups.

Figure 1.3. Read alignment from the FOSCC sample OSCC8 to FPV-3 (GenBank NC_021472.1)

Figure 2.1. Size distribution of differentially methylated regions (DMRs) and CG islands (CGIs).

Figure 2.2. Subgroups of canine DLBCL can be defined by promoter-TSS methylation signature. Heatmap of the DMRs from MIRA-seq of the 23 DLBCL and 7 normal B cell samples. The heatmap shows unsupervised hierarchical clustering of the 1654 hypermethylated promoter-TSS DMRs that were identified with a $q < 0.001$ (Benjamin-Hochberg correction). Log2 of the rpkm values of each DMR was used.

Figure 2.3. Normal T and B cells can be defined by DNA methylation signatures. Heatmap of the DMRs from MIRA-seq of the 8 normal T cell samples and 7 normal B cell samples. The heatmap shows supervised hierarchical clustering of the 167 promoter-TSS DMRs that were hypermethylated in T cells and the 90 DMRs that were hypermethylated in B cells ($q < 0.001$, Benjamin-Hochberg correction).

Figure 2.4. Hypermethylated DMRs follow the distribution of CGIs and are located around the transcription start site (TSS). For the hypermethylated DMRs, the median, mean and range is 217 bp, 882 bp, -535,391 bp to 560,181 bp respectively from TSS. For the hypomethylated DMRs, the median, mean and range is 1067 bp, 1614 bp, -28 bp to 556,107 bp respectively from TSS. For the CGIs, the median, mean and range is 230 bp, 3087 bp, -535,693 bp to 560,156 bp respectively from TSS.

Figure 2.5. Hypermethylated DMRs are located in CGIs. Distribution of B vs. DLBCL DMRs in relation to the nearest CGI. CGI's in UCSC were defined as GC content $\geq 50\%$, length > 200 bp, ratio of observed CG to expected CG > 0.6 .

Figure 2.6. Comparison of the hypermethylated genes identified with previous whole methylation canine studies. The Infinium study used the human HM850 microarray in buffy coats from dogs with high-grade B-cell lymphoma. The DREAM study used restriction endonuclease digestion in the CLBL-1 cell line. The microarray study used a canine microarray in whole lymph nodes of dogs with DLBCL.

Figure 2.7. Comparison of the hypomethylated genes identified with previous whole methylation canine studies. The Infinium study used the human HM850 microarray

in buffy coats from dogs with high-grade B-cell lymphoma. The microarray study used a canine microarray in whole lymph nodes of dogs with DLBCL.

Figure 3.1. Filtering algorithm for single nucleotide variants (SNVs)

Figure 3.2. 88.6% of the mutations in *TP53* (ENSCAFT00000026465.3) in canine OSA occurred in the DNA binding domain.

Figure 3.3. Strong correlation between SNVs seen in the primary compared to the metastatic lesion in the Sheepdog.

Figure 3.4. Absence of kataegis in canine OSA. Rainfall plot showing the intermutation distance versus genomic position for validated SNVs.

Figure 3.5. Genes were more commonly mutated by SVs versus SNVs in canine OSA. Venn diagram depicting the proportion of (a) SNVs and (b) SVs that were shared between dogs and between the primary and metastatic lesion

Figure 3.6. Plots depicting SNVs, CNAs, LOH and SVs in coding regions.

Figure 3.7. Chromosomal segments were not heterozygous in the germline. The selected chromosome, chr14, was also less affected by chromosomal lesions than other chromosomes (normal in the Sheepdog).

Figure 3.8. Summary of loss of heterozygosity (LOH) for both patients.

Figure 3.9. Segmental copy number neutral loss of heterozygosity (CN-LOH) in the middle of chr5 in both dogs.

Figure 3.10. Chr26 was the most affected by structural variants (SVs). (a) Representative CNA (top panel), somatic LOH (middle panel) and germline LOH (bottom panel) plots from the primary lesion in the Lab for chr26. (b) Close up view of CNAs for *PTEN* in the Sheepdog primary lesion (top panel), Sheepdog metastatic lesion (middle panel) and Lab primary lesion (bottom panel).

Figure 3.11. A focal deletion was noted in the *CDKN2A/B* locus, chr11:41,223,260-41,246,135 overlapped in both dogs.

Figure 3.12. Few SVs resulted in expressed gene fusions. (a) *GRK3-HPS4* was caused by an inversion in the Lab primary and metastatic lesion. (b) *PICALM-DLG2* was caused by a deletion in the Sheepdog primary and metastatic lesion.

LIST OF TABLES

Table 1.1. Patient characteristics and read information for each sample that was sequenced by ViroCap

Table 2.1. Clinical information of the Golden Retrievers that were sequenced by MIRA-seq

Table 2.2. Raw reads and QC statistics for Golden Retrievers with DLBCL and normal control dogs that were sequenced with MIRA-seq

Table 2.3. Location of the hypermethylated DMRs in DLBCL compared to normal B cells as annotated by Homer

Table 3.1. Signalment and diagnostic information for the dogs that were sequenced in the canine OSA study

Table 3.2. Treatment and outcome for the dogs that were sequenced in the canine OSA study

Table 3.3. Sequencing coverage and tumor purity for each sample

ABSTRACT

The remarkable advancements in sequencing technologies have allowed the entire mutational landscape of hundreds of different types of human cancers to be defined. This knowledge gives patients two new categories of treatment options, small molecule inhibitors and targeted immunotherapy. Veterinary oncology has been slow to enter the field of genome-wide studies, largely due to costs. Five years ago the first whole genome of a canine cancer patient was published. Because of NIH support and the steady decline in sequencing costs, the number of published canine tumor sequences has exploded this past year as these projects are being concluded.

This dissertation contributes to this growing knowledgebase. The scope was diverse and included exploration in metagenomics, epigenomics, genomics and transcriptomics in canine and feline cancers. This research affirmed previous notions that papillomavirus is not a common cause of feline oral squamous cell carcinoma (FOSCC) and added to the understanding that it is not caused by any other DNA or integrated dsRNA virus. In canine DLBCL (cDLBCL) we provided more evidence that the subgroups seen in human DLBCL (hDLBCL) cannot be faithfully reproduced in the methylome of cDLBCL, widespread promoter hypermethylation was present in cDLBCL, likely at a greater scale than DNA mutations, and neurogenic/neuroendocrine, *HOX* and Wnt pathways were epigenetically targeted. In canine osteosarcoma (OSA) we built upon the genome model by adding loss of heterozygosity data and illustration of metastatic lesions. We also provided support to target the *CDK4*, Wnt and mTOR pathways.

Finally we highlight unexpected discoveries such as the discovery of Epstein Barr virus in FOSCC, stability of the methylome in Golden Retriever DLBCL, epimutations in the *TBX* pathway in cDLBCL, dysregulation of human melanoma pathways in canine OSA, absence of dysregulation of *HER2* in canine OSA, absence of genomic evidence of a telomere maintenance mechanism, recurrent shattering of chr26 in canine OSA and *HSP90AB1*, *MITF* and *NOTCH2*, as potentially actionable genes.

These data will provide a significant resource for the community to generate new hypotheses and answer key questions about the molecular drivers of canine and feline malignancies

INTRODUCTION

Cancer is characterized by mutations. Mutations in oncogenes and tumor suppressor genes make up the classic Hallmarks of Cancer (1). Mutations can be DNA mutations or epimutations which comprise of changes in DNA methylation, histones and non-protein coding RNAs. Newer research shows that epimutations, deregulation of cellular genetics and stromal and immune cells in the local tumor microenvironment can promote and enable carcinogenesis but the end result is invariably DNA mutations (2). Defining the mutational landscape of tumors has thus been the focus of research in human cancers since the normal human genome was sequenced in 2003 (3). A catalogue of human tumor sequences have been amassed with multi-institutional efforts such as The Cancer Genome Atlas (TCGA) from the US National Institutes of Health, 100,000 Genomes Project from the UK National Institute for Health Research and National Health Service and the 25K Project from the International Cancer Genome Consortium. The results from these concerted projects are made immediately available to the public, in raw formats that allow hypothesis generation and further analysis with user-friendly databases that are accessible to researchers and clinicians who do not have bioinformatics training. The human oncology sequencing experience is a great example that can be followed and improved upon for veterinary oncology.

In comparison the normal canine and feline genomes were first sequenced in 2012 and 2014 respectively. The first whole genome sequencing (WGS) project in a canine cancer was conducted from two dogs with transmissible venereal tumor (4).

Later that year, WGS was performed on three dogs with canine mammary carcinomas (5). In this 2014 study whole exome sequencing (WES) and RNA-seq were also performed. Since then, 7 melanoma, 27 osteosarcoma and 77 glioma (preprint) tumors have been sequenced with a variety of methods which all include >30x WGS coverage (6-8). The majority of these efforts were funded by an NIH P30 Cancer Center Support mechanism. Continued financial support for studies with translational value will lead to an exponential increase in canine cancer sequencing efforts. Canine cancers with the highest translational value include rare cancers in humans that are much more common in certain breeds of dogs such as aggressive T-cell lymphoma (Boxers), osteosarcoma (large/giant breed dogs), hemangiosarcoma (Golden Retrievers), histiocytic sarcoma (Bernese Mountain Dogs), *BRAF*^{wt} melanoma, muscle invasive transitional cell carcinoma (Scottish Terriers), and glioma (Boxers). In addition, cancers that are common in both species such as non-Hodgkin lymphoma and mammary carcinoma (intact female dogs) serve as relevant translational comparators. It is thus a very exciting time in veterinary cancer genetics, as we have just had a peek at the naked genome of cancer in dogs.

The first feline tumor has yet to be sequenced by any means. The reasons behind this are likely that the feline genome reference is not as high quality as the canine build and contains many gaps and unaligned reads. The ability to performing clinical trials in a feline patient may limit the translational potential of a feline cancer model. Due to the small body size and less social nature of most cats repeated blood and tumor collections for pharmacodynamic and pharmacokinetic,

and phase II and III studies may not be possible. The first challenge will be overcome in the near future with the re-sequencing and assembly of the feline genome. Regardless of these challenges, cats are the best naturally occurring model for many human cancers, including Human Papillomavirus negative (HPV) head and neck squamous cell carcinoma (HNSCC), immunodeficiency virus associated aggressive lymphoma, papillomavirus associated Bowenoid *in-situ* carcinoma, and squamous cell carcinoma.

This dissertation encompasses three next generation sequencing projects in veterinary cancers. The first utilizes a targeted enrichment hybridization technique, ViroCap, to sequence all DNA containing viruses in feline oral squamous cell carcinoma (FOSCC) and normal mucosal samples (9). The purpose of this study was to find an association between a virus and FOSCC if FOSCC is virally induced. The secondary purpose of this study was to describe the viral microbiome of a feline patient for the first time. The third purpose of the study was to validate ViroCap, a technique created to evaluate human samples, as a tool for viral metagenomics in companion animals. The second project utilizes an enrichment pull down method to evaluate whole genome DNA methylation in diffuse large B cell lymphoma (DLBCL). The purposes of this study were to characterize the methylome (hyper and hypomethylated genes, chromosomal distribution, distribution in regards to genomic features) in DLBCL in Golden Retrievers. Once the methylome was defined the following hypotheses were tested (1) the methylation pattern in DLBCL would be distinct from normal cells; (2) subgroups of DLBCL could be defined by the methylome; (3) hypermethylated regions compared to normal B cells would be

preferentially located in the promoter-TSS (transcriptions start site); (4) hypermethylated regions would be in genes that are known cancer associated genes, specifically tumor suppressor genes (TSG); (5) hypermethylated regions would parallel genes seen in human DLBCL (hDLBCL). Cancers are known to have a heterogenous DNA mutational landscape. The advantages that epigenetic profiling has over DNA mutation profiling is that DNA methylation changes occur before DNA mutations and DNA methylation changes are potentially reversible. Changes that occur early in the founding clone are likely to have a high prevalence in the tumor population despite tumor heterogeneity from divergence that occurs over time. A highly prevalent change is an ideal trait for a biomarker. Identification of early changes may also provide clues about the pathogenic changes that drive carcinogenesis. The third study utilizes a comprehensive sequencing approach with WGS, WES and RNA-seq on primary and pulmonary metastatic lesions in canine osteosarcoma. The purposes of this study were to (1) add to the growing genomic database in canine OSA; (2) adapt a comprehensive human cancer informatics pipeline for the canine species; (3) compare the primary and metastatic lesions; (4) evaluate the genetic similarities between pediatric and canine OSA and thus the utility of dogs as a translational model; and (5) identify actionable driver pathways. This is the first study to perform WGS on a metastatic lesion and perform loss of heterozygosity (LOH) analysis in canine osteosarcoma. This study serves as the most comprehensive genomic analysis of a canine patient with osteosarcoma.

At the start of my PhD there was only one study that performed WGS on a canine cancer, mammary carcinoma (5). The field of veterinary cancer genomics

has grown exponentially since this time. Despite the growing number of studies the numbers are still low compared to human oncology. New driver mutations, tumor subtypes, markers of response, diagnostic biomarkers and causative factors are still being identified in human and canine studies. This dissertation serves as a significant contribution to this growing knowledgebase.

CHAPTER 1:

A virome sequencing approach to feline oral squamous cell carcinoma to evaluate causative factors

Shirley Chu^{a*}, Todd N. Wylie^{b,c}, Kristine M. Wylie^{b,c}, Gayle C. Johnson^e, Zachary L. Skidmore^b, Obi L. Griffith^{b,d}, Jeffrey N. Bryan^a

^aDepartment of Veterinary Medicine and Surgery, University of Missouri, 900 E. Campus Drive, Columbia, MO 65211, USA

^bMcDonnell Genome Institute, Washington University, 4444 Forest Park Ave, St. Louis, MO, 63108, USA

^cDepartment of Pediatrics, Washington University, 660 S. Euclid Ave, St. Louis, MO, 63110, USA

^dDepartment of Medicine, Washington University, 660 S. Euclid Ave, St. Louis, MO, 63110, USA

^eDepartment of Veterinary Pathobiology, University of Missouri, 800 E. Campus Loop, Columbia, MO 65211, USA

*Correspondence: Shirley Chu chus@missouri.edu

Abstract

Feline oral squamous cell carcinoma (FOSCC) may be the best naturally-occurring model of human head and neck squamous cell carcinoma (HNSCC), but there is sparse evidence to support that these two diseases are molecularly similar or that

they share causative factors. HNSCC can be broadly divided into human papillomavirus (HPV)-negative cancers and HPV-positive cancers where HPV is the causative agent. There has been conflicting evidence about the presence of papillomavirus in FOSCC samples. Previous studies have used both species-specific and species-nonspecific PCR primers that may be insensitive to the detection of papillomaviruses and other viruses that may be divergent from known sequences.

ViroCap is a targeted capture and next generation sequencing tool for a comprehensive set of vertebrate viral genomes. ViroCap was designed to identify all known vertebrate DNA and RNA viruses. Reported here is a metagenomic study for DNA viruses in 20 FOSCC, 9 normal feline oral mucosal, and 8 suspected papillomavirus positive control samples. We tested the hypothesis that viruses would be enriched in FOSCC compared to normal oral mucosa.

The virome of the FOSCC and normal feline oral mucosa consisted of feline foamy virus (35% and 22%), feline torque teno virus (5% and 0%), alphaherpesvirus (5% and 0%), FIV (0% and 11%), Epstein-Barr virus (5% and 0%) and feline papillomavirus (5% and 0% respectively). *Felis catus* papillomavirus-3 was found in 1/20 FOSCC samples. A virus was not associated consistently with FOSCC. FOSCC appears most closely related to HPV-negative HNSCC. Future research on FOSCC should focus on identifying genetic and environmental causes.

Keywords: feline oral squamous cell carcinoma, papillomavirus, sequencing, virus, virome, metagenomic

Introduction

Feline oral squamous cell carcinoma (FOSCC) is the 4th most common feline neoplasm and there is no effective treatment for this disease. In one study of 31 cats treated with an accelerated radiation therapy and concomitant IV carboplatin protocol, the median survival time was 163 days (10). Survival may be longer for cats with a complete response to treatment; mandibular, tonsillar, cheek, and rostral locations; and smaller tumors. Regardless, almost all cats with FOSCC are euthanized due to local disease progression.

Lip and oral cavity cancer was the 15th most frequently diagnosed cancer in the world in 2015 (11). Despite surgery, radiation therapy and chemotherapy, 50% of human patients die of loco-regional recurrence within 5 years (12). Causes include tobacco use, alcohol consumption and human papillomavirus (HPV) infection. Human head and neck squamous cell carcinomas (HNSCC) can be divided into HPV-positive and HPV-negative cancers. HPV-positive HNSCC tend to affect younger, more educated, and higher income patients than HPV-negative HNSCC, and have a better prognosis either due to less aggressive biologic behavior, patient, or treatment factors. Human papillomavirus has been implicated in 25-60% of HNSCC, especially in oropharyngeal (base of the tongue, tonsils, walls of the pharynx and soft palate) squamous cell carcinoma (SCC). Neoplastic transformation can be caused by viral integration and/or episomal expression of viral oncogenes. Papilloma subtype is known to be important since HPV-16 has been identified in

94.7% of HPV-positive oral and oropharyngeal cancers (13) (14). Both the Gardasil® (HPV 6/11/16/18) and Cervarix® (HPV16/18) vaccines that are labeled for the prevention of HPV-associated cancers of the cervical, vulvar, vaginal and anal mucosa may protect against HNSCC. Data regarding this indication are currently lacking.

FOSCC has commonly been proposed as the best model for HNSCC since FOSCC occurs naturally more commonly in cats than any other domesticated species and exhibits similar biologic behavior, resistance to treatment, and outcome to the human disease (15). While it is likely that FOSCC is the best naturally-occurring model for HNSCC, underlying genetic, epigenetic, and viral factors have not been extensively correlated to the known features of the subtypes of HNSCC.

Understanding the cause of FOSCC could benefit feline patients as well as human patients.

The papillomavirus family is divided into 36 known genera with the largest being the alpha-PVs, which includes subtypes HPV-16 and 18 that have the highest malignant transformation potential (14). Papillomaviruses are non-enveloped circular double-stranded DNA (dsDNA) viruses with 5 or 6 early (E1, E2, E4, +/-E5, E6, E7) genes and two late genes (L1 and L2). The early genes encode proteins for viral replication and virulence, and the late genes encode capsid proteins. All papillomaviruses contain a highly conserved L1 gene and thus family-specific consensus PCR primers can be used to amplify a region in this gene.

There are currently ~150 identified HPV species and only 5 *Felis catus* PV (FPV) species known to infect domestic cats (16). Papillomavirus infection has been implicated in many feline neoplasms, including Bowenoid in situ carcinomas (BISC), cutaneous SCC, and less commonly FOSCC (Supplementary Table 1). Feline papillomavirus DNA, especially FPV-2, and induced changes in cell regulation (increased p16 identified by IHC staining) have been detected in the majority of BISCs (17-19). Papillomavirus DNA and increased p16 have also been detected in 75% of UV-protected cutaneous SCC, and thus PV is likely a causative agent in feline BISCs and UV-protected cutaneous SCC (17, 20).

There have been conflicting results in seeking the presence of papillomavirus DNA in FOSCC. Most studies have not found strong evidence to support a viral etiology (21). The results of the studies are summarized in Supplementary Table 2. The most recently published study evaluated 8 formalin-fixed, paraffin embedded (FFPE) samples with the PV L1 and E1 consensus primers, MY09/11 and CP4/5 respectively, and FPV-2, 3, 4 type specific L1 primers. The investigators did not find PV in any of the oral lesions (22). In summary, to date papillomavirus has only been found in 6 FOSCC samples in peer-reviewed journals and the association between papillomavirus and FOSCC thus remains weak.

At the 2015 Veterinary Cancer Society (VCS) conference an Austrian study using subtype specific primers (primers used were not published in the abstract or

described in the presentation) for FPV-1, 2 and 3 in 12 FOSCC lesions detected FPV-2 DNA in 2 FOSCC lesions and FPV-3 in 1 FOSCC lesion. The FAP59/64 consensus primers detected PV amplicons in all FOSCC lesions and these amplicons were unlike the known subtypes. The use of negative controls was not reported (23). This study has yet to be published. The reason for these starkly conflicting results is unknown, but may be due to differences in sample preparation, PCR methodologies, geographic etiologic differences, or different subject pools. Since the vast majority of PV studies in cats, except the VCS study, have been done by one group from New Zealand, previous studies may have included a biased and homogenous population. Papillomavirus in FOSCC has only previously been investigated in the US with samples from 2 other institutions, University of Tennessee (24, 25) and The Ohio State University (26). This current study will evaluate PV status in a new population of American cats.

Next generation sequencing (NGS) can substantially increase the sensitivity and specificity of virus detection as it is not limited by primer specificity or the selection of specific assays in a PCR reaction. Amplicon-based survey sequencing approaches, such as 16S rRNA gene sequencing, have been utilized to study bacterial diversity, but a similar method cannot be used for viruses due to the lack of universally conserved genes. Targeted broad viral characterization and discovery methods have included microarray-based methods such as the Virochip® and PathoChip®, which detect viral and microbial pathogens respectively (27, 28). The

disadvantages of microarrays are that they target smaller genomic regions and do not provide sequence information.

ViroCap, a hybridization-based capture and NGS approach, has the ability to enrich nucleic acids from all DNA and RNA viruses in samples from vertebrate hosts (excluding endogenous retroviruses). Thirty-four families are included. Using k-mer and clustering analysis tools, 1 billion base pairs (bp) of viral genomes were condensed into 200 million bp for targeted sequence capture. Compared to PCR, ViroCap can detect viruses that are divergent from reference genome sequences (e.g. anellovirus family), and it has the ability to generate complete or nearly complete genome sequences (breadth of coverage (BoC)) because probes are tiled across the full length of the genomes. A higher BoC allows the opportunity to distinguish amongst closely related subtypes or even amongst strains. ViroCap has been utilized on human vaginal swabs (29), whole blood, plasma, cerebrospinal fluid, nasopharyngeal swabs, tracheal aspirates, skin swabs, and stool (9, 30) and in a Coronavirus outbreak in Canada Geese (31). To the author's knowledge, the current study would be the first companion animal study to utilize a targeted capture and NGS strategy to study the virome. ViroCap is the most definitive method used to date to characterize the virome of the oral mucosa of cats and to find a viral cause of FOSCC. The aim of the analysis is to provide evidence that FOSCC will be either a good model for HPV-negative or positive HNSCC. If the results provide evidence of PV-induced FOSCC, further studies would be warranted

to establish cause and effect with the goal of developing a prophylactic vaccine to prevent this grave disease.

The proposed study is a metagenomic study of the virome in normal feline oral mucosa and FOSCC samples. Using a sensitive and specific test of targeted capture and NGS we will first define the virome of normal feline oral mucosa and then look for viral associations with FOSCC. The identification of viruses in FOSCC that are not commonly found in normal oral mucosa or are enriched in FOSCC would support a possible viral cause of FOSCC. We hypothesized that viral genomic material will be found in many mucosal samples as part of the normal flora, but that a particular genus of virus will not be overrepresented in cats with FOSCC. A viral cause will be found in FOSCC and FOSCC should be studied as a model for HPV-positive HNSCC.

Materials and Methods

Patients

Formalin-fixed, paraffin embedded samples from 20 cats diagnosed with FOSCC by a board-certified diagnostic pathologist in 2012-2013 were obtained from the University of Missouri Veterinary Medical Diagnostic Laboratory (MU VMDL) (samples OSCC1-OSCC20 in Table 1). Banked FFPE samples from 8 tumors with that were considered to be caused by PV on histopathology were obtained from the MU VMDL (samples PV1-PV8 in Table 1). Presumed PV-negative controls consisted of 9 fresh frozen (FF) tumor negative oral mucosal biopsy samples (combined tongue and gingival mucosa) from adult cats obtained from the University of Missouri Veterinary Health Center and Central Missouri Humane Society (samples N1-N9 in Table 1). None of the samples in this study have been used in any previous study. Fresh frozen samples were stored at -80°C until DNA extraction was performed.

DNA Isolation

DNA was isolated from FFPE samples with the QIAamp DNA FFPE Tissue Kit (Qiagen; 56404) according to the manufacturer's protocol. DNA extractions were repeated until 1 ug of DNA was obtained. DNA was isolated from fresh frozen samples with the DNeasy Blood and Tissue Kit (Qiagen; 69506). DNA was quantified with the Qubit dsDNA HS Assay kit (Invitrogen; Q32851) on the Qubit 3.0 fluorometer (Invitrogen; Q33216). DNA quality was assessed with the NanoDrop (Thermo Fisher Scientific; 2000).

Library preparation

Genomic DNA was submitted to the McDonnell Genome Institute at Washington University for automated library preparation. Automated dual indexed libraries were constructed with 100-250ng of genomic DNA utilizing the KAPA HTP Library Kit (KAPA Biosystems; KK8234) on the SciClone NGS instrument (Perkin Elmer; CLS145321) targeting 250bp inserts.

Capture

FFPE libraries from the 28 FOCCC and positive controls were pooled and captured separately from the pool of 9 fresh frozen negative control libraries. Each library pool was hybridized with a custom NimbleGen targeted probe set (Roche), ViroCap (9) targeting a pan-virome space, with additional Integrated DNA Technologies (IDT) Lockdown probes targeting (HPV) human papilloma viruses (32, 33). The IDT probes were added at a molar concentration equivalent to the NimbleGen probe concentration, thus representing the HPV regions in a similar quantity as the ViroCap regions. The libraries were hybridized for 72 hours at 47°C followed by stringent washing. Enriched ssDNA library fragments were amplified with KAPA HiFi HotStart polymerase (KAPA Biosystems; KK2600) prior to sequencing. The concentration of each captured library pool was accurately determined through qPCR according to the manufacturer's protocol (KAPA Biosystems) to produce cluster counts appropriate for the Illumina HiSeq2500 1T platform. The two library capture pools were normalized and pooled prior to data generation.

Sequencing

Captured reads were sequenced on one lane of the HiSeq2500 1T as 2 x 126 bp reads generating on average 1.3Gb of data per sample.

Sequence Analysis

FASTQ files were used as input and adaptor trimming was done with fqtrim. Low-complexity reads were removed with VSEARCH – fastx_mask. Remaining reads were mapped to felis_catus_8.0/felCat8 with BWA-MEM. Unmapped reads were mapped to a comprehensive viral nucleotide database ($\sim 1.5 \times 10^6$ viral sequences from NCBI GenBank and RefSeq) with BWA-MEM. Unmapped reads also were mapped to a translated viral amino acid database (6-frame) with DIAMOND (diamond blastx). Resulting mapped reads to the viral nucleotide and translated viral databases were merged, and merged reads were aligned to NCBI nt + viral RefSeq with BWA-MEM. Viral reads unmapped by NCBI nt were mapped using DIAMOND to NCBI nr to increase detection of more divergent viruses. This pipeline was repeated for the second read in the pair. Taxonomy reporting was then performed on mapped reads. Bacteriophage were not further analyzed. This workflow is similar to the one that was used in a study of the vaginal virome in preterm birth (29).

Samples were further analyzed in-depth individually. Accessions from identified viral taxa (e.g. order, family or genus) were obtained from the NCBI Taxonomy

Browser. Paired reads from each identified taxon were extracted from the sample, competitively aligned with BWA-MEM to the identified reference genome of the viral taxon and host species of the tissue sample, duplicate reads were removed, and read counts for each identified accession were tallied. The complete genomic sequence of the species/strain that was most represented was downloaded from GenBank. The extracted paired reads were aligned with BWA-MEM to the genomic sequence to identify single nucleotide polymorphisms (SNPs), identify physical form of the virus (e.g. circularized or integrated), and to quantitate genomic coverage. Aligned reads were visualized in IGV. Average genomic coverage (depth of coverage) was estimated from base representation by the extracted, deduplicated and aligned reads for each base of the genome. Breadth of coverage was estimated in SAMtools and was defined as the percent of bases of the reference viral genome that was covered by sequence reads at a level of 5x or higher.

The detection of false positives (e.g. by index swapping or contamination from sample handling) were minimized with the definition of strict coverage criteria. Positive samples were defined as any sample that had $\geq 0.5x$ viral genome coverage after reads were mapped and deduplicated with Picard. Exceptions to this rule included samples that were manually reviewed to confirm presence of a virus (e.g. HPV in N6). Presence of a virus was confirmed if reads were distributed across the viral genome.

Papillomavirus validation

Conventional PCR

The degenerate consensus FAP59/FAP64 primers were used. Details of the PCR conditions can be found in Supplementary Table 3. Each sample was amplified in duplicate and a no template control was included in each assay. The assay was repeated if the duplicates did not yield the same results. The sample was determined to be positive if the results of the duplicates were concordant and positive.

Immunohistochemistry

The positive control samples were evaluated for PV via IHC prior to nucleic acid sequencing. Biopsy specimens were fixed by immersion in neutral buffered 10% formalin. Specimens were sampled for examination and dehydrated, then embedded in paraffin by standard methods. Five µm sections were prepared and stained by hematoxylin and eosin (HE). Additional 5 µm thick specimens were prepared and placed on glass slides, then deparaffinized and rehydrated by immersion in xylene and graded alcohols to distilled water. Staining (Biocare Medical; IPS0001) was done in a Biocare IntelliPATH FLX autostainer. Sections were then treated with 3% hydrogen peroxide for 15 minutes to block endogenous peroxidase activity and washed in buffer. Non-specific reagent binding was blocked by treating with Sniper (Biocare Medical; BS966) for 10 minutes with a second buffer wash. Slides were treated with mouse monoclonal anti-papillomavirus antibody which is a mixture of the BPV-1/1H8 against SDS disrupted BPV-1 and CAMVIR-1 against L1 of HPV-16

(Abcam; ab2417) at a dilution of 1:600 for 30 minutes, rinsed with buffer, and treated with anti-mouse HRP EnVision+ (Agilent DAKO, K4001) for 30 minutes. After rinsing with distilled water Romulin Red chromogen (Biocare Medical; RAEC810) was applied and CAT hematoxylin (Biocare Medical; CATHE) applied as a counter stain. After a final rinse, tissues were dehydrated with graded alcohols and xylene and coverslipped. For each specimen a second slide was stained with an irrelevant antibody to demonstrate the specificity of staining and a positive control was positively and negatively stained for each run.

Results

Patients

Twenty FFPE FOSCC, 9 FF normal oral mucosal, 8 clinically suspected PV-positive control samples were included in the viral capture. The median age of the cats with FOSCC was 15 years (range 11-19 years), Table 1. The median age of the negative control cohort was likely younger but the age of most of the cats was unknown. Of the 20 FOSCC samples, 12 (60%) were from the gingiva, 5 (25%) were from the tongue and the location of 3 (15%) samples was unknown. Of the 9 negative control cats, the cats were euthanized for leukemia or stage 5 lymphoma, squamous cell carcinoma of the nasal planum, pulmonary angiomatosis, valvular endocarditis, trauma, chronic renal disease, FIP, FIV and FeLV, Table 1.

Sequencing

The average raw reads sequenced per sample was 13,255,356 (range 3,732,410-49,712,932). The average raw reads per fresh frozen normal mucosal sample was 26,789,154 (range 19,796,205- 49,712,932) and per FFPE tumor sample was 8,905,206 (range 3,732,410-24,972,777), Table 1. Between 19.7 and 56.4% of raw reads belonged to PV in samples that contained PV (sample N6 not included).

Virome of the oral mucosa of cats with and without FOSCC

The virome of the FOSCC and normal feline oral mucosa consisted of retroviruses, anelloviruses, adenoviruses, herpes viruses and papillomaviruses (Figures 1 and 2).

More specifically, feline papillomavirus was found in 5% and 0% ($p=1.00$, Fisher's exact test, 2-tailed), feline foamy virus in 35% and 22% ($p=0.67$), feline torque teno virus in 5% and 0% ($p=1.00$), alphaherpesvirus in 5% and 0% ($p=1.00$), FIV in 0% and 11% ($p=0.09$), Epstein-Barr virus in 5% and 0% ($p=1.00$) respectively in FOSCC and normal oral mucosa. The prevalence was not statistically significant for any virus and thus a virus was not associated consistently with FOSCC.

Papillomavirus

The results of conventional PCR, IHC and ViroCap were concordant for all samples except PV2 and PV5 (Supplementary Table 4). PV2 was a papilloma on the skin of a dog and contained CPV-6 by ViroCap. PV2 was also positive on IHC but negative on PCR. The FAP59/64 primers have never been shown to amplify CPV-6 and this may explain PCR failure. PV5 was a sarcoid from the eyelid of a horse and contained BPV-1 on ViroCap and PCR but failed to be detected with IHC. This can be expected since BPV-1 virion production is frequently aborted and may not be supported in equids (34) and if present would be confined to the epidermis (35). The diffuse melanin granules in the epidermis of PV5 or prolonged formalin fixation may have interfered with L1 IHC interpretation.

Papillomavirus was detected in 5% (1/20) of the FOSCC samples and 11% (1/9) of the normal mucosal samples. The FOSCC gingival sample, OSCC8, had a high viral load, >2.1 million reads, with feline PV (FPV-3), likely indicative of a true infection. The normal mucosal sample, N6, had a very low PV load, 56 reads, with a human PV

(HPV-28) and likely represented contamination (Supplementary Table 5, Supplementary Figure 1).

Papillomavirus was not detected by any method for the following positive control samples, PV1, PV6, PV7 and PV8, Supplementary Table 4. These samples were from feline lesions that were diagnosed as sarcoids (fibropapillomas) and oral papillomas via histopathology with hematoxylin and eosin staining. Based on these results, it appears that at least a portion of histologically diagnosed feline sarcoids and oral papillomas are not virally induced.

Five novel variants of PV were discovered with ViroCap (Supplementary Table 6, Figure 3 and Supplementary Figures 2-5). GenBank accessions for the new variants are KY802017, KY825186, KY825187, KY825188, and KY886226. All SNPs in the L1 gene in the novel variants were synonymous and thus the L1 protein sequences were the same as the reference genomes (Supplementary Table 7). KY886226 represents a more accurate genome for BPV-1 than GenBank accession NC_001522.1 since it includes the nucleotide that is missing from the reference at position 1205 and encodes a longer E2 protein that is more consistent with the verified E2 protein in UniProt (ID P03122). The full list of variants can be found in Supplementary Table 7.

For the 5 novel PV variants, read overhangs consistent with integration sites were not observed. The reads that aligned to the ends of the reference genome contained

overhangs that mapped to the other end of the genome. This is consistent with a circularized virus. Thus in all samples where papillomavirus infection was identified, the virus existed as an episome and was not integrated into the host genome.

Foamy virus

Foamy virus was found in 30% (10/33) of all feline samples, 35% (7/20) of the FOSCC samples and 22% (2/9) of the normal oral mucosal samples (Supplementary Table 8). The difficulties of mapping short reads to repeat regions (e.g. long terminal repeat/LTR) prevented prediction of the whole genome sequence and location of genomic integration sites.

Many SNPs were detected throughout the foamy virus genomes when compared to the closest relative. At least two distinct foamy viral species were detected, 40% (4/10) of the foamy virus positive cats had a foamy virus envelop (*ENV*) gene that was most closely related to the Coleman strain of the foamy virus (GenBank AB052797.1), and 60% of foamy virus positive cats had an *ENV* gene that was most closely related to the Puma strains of foamy virus (GenBank KC292054.1), Supplementary Figures 6 and 7.

Anelloviruses or torque teno viruses

Anellovirus was detected in 24% of cats (8/33; 23% in non-FOSCC patients and 25% in FOSCC patients) in the current study but the coverage was <0.5x in all but 2

of the samples and thus insufficient to confidently conclude the presence of anellovirus. The prevalence in our study was thus only 6% (2/33) of cats, OSCC1 and OSCC8, Supplementary Table 9. Low read counts were seen in both samples and there were gaps in coverage.

Herpesviruses

Only 6% (2/33) of cats, OSCC19 and OSCC21, had evidence of felid herpes virus 1 (FHV-1), which is an alphaherpesvirus that causes upper respiratory tract infections in cats, Supplementary Table 10. The BoC was low at 2.8% and 19.3%. Feline Calicivirus (FCV) likely was present but was not detected because this is an ssRNA that is not integrated into the host genome.

Epstein Barr Virus (EBV; Human herpesvirus-4; HHV-4), a gammaherpesvirus, was detected in 3% (1/33) of all cats. The infected cat, OSCC8, had FOSCC but was not FIV or FeLV positive. The DoC of EBV was approximately 2x (Supplementary Table 5). The BoC was 7.1%.

Orthoretrovirinae subfamily of retroviruses (includes FeLV and FIV)

The orthoretrovirinae are single stranded RNA reverse transcriptase (ssRNA-RT) viruses that would only be detected in this study if integrated into the host genome. FIV belongs to the lentivirus genus. FIV was detected with ViroCap in 11% (2/9) of cats with normal oral mucosa. The FIV-positive cats were the FIV SNAP positive cat, N3, and a stray cat N9 where FIV testing was not done. Low levels, 0.8x and 0.3x

genomic coverage, of integrated FIV were detected in N3 and N9 respectively, Supplementary Table 11 and Supplementary Figure 8. Feline leukemia virus (FeLV) belongs to the gammaretrovirus genus. Exogenous FeLV (exFeLV) was not detected in the cat that was FeLV positive on SNAP test, N2. To rule out the possibility that exFeLV reads were excluded by because of similarity to endogenous FeLV (enFeLV) in the *Felis catus* genome, all sequenced reads in cat N2 that aligned to the *Felis catus* or exFeLV (GenBank NC_001940.1) genomes were identified and re-analyzed. Cat N2, the cat that was FeLV positive on SNAP test, had 24,457 unique reads that aligned with uneven distribution to exFeLV (GenBank NC_001940.1), Supplementary Table 12 and Supplementary Figure 9a. The pattern of high DoC, 332.7x, and the lower BoC, 34.4%, were consistent with enFeLV. A similar pattern was seen in samples N1, N3, N4, N5, N6, N8 and N9 and thus exFeLV infection was not suspected in these cats. Cats, N1, N3, N4, N6, N8, accordingly were negative on SNAP FeLV testing at the time of tissue collection and were not clinically suspected to be FeLV/FIV positive. FeLV/FIV testing was not done on samples N5, N7 or N9. In addition, the region of the U3 long terminal repeat (LTR) that contains the CCAAT enhancer (nucleotide position 271 bp) and TATA boxes (nucleotide positions 177 and 312 bp) found in exFeLV did not contain aligned reads in any patient except N7 and there were gaps in coverage in *ENV* and the 5' end of the gap-pro-pol polyprotein (Supplementary Figure 9b). In contrast, cat N7 had a high BoC, 95.3%, and had coverage of the U3 LTR and was thus suspected to be infected with exFeLV. Cat N7 was not FeLV/FIV tested but was diagnosed with feline infectious peritonitis (FIP) on necropsy. FIP is caused by infection with a mutated feline coronavirus, a

ssRNA virus, and immunosuppression by FeLV has been associated with increased risk of FIP (36).

Non-felid viruses

Siadenovirus was detected in sample PV3, 33% (1/3) of all canine samples. The DoC and BoC were low, 0.3x and 0% respectively, Supplementary Figure 10. BLAST of the reads that mapped to the hexon and fiber genes aligned solely to the turkey hemorrhagic enteritis viruses. A canine siadenovirus has not been reported. A turkey hemorrhagic enteritis like virus was thus present in PV3.

Gallid herpesvirus 3 (GaHV-3) was present in 2.7% (1/37) of all samples, OSCC13, Supplementary Table 5. BLAST of representative aligned reads were a match only to the Marek's disease viruses (MDV-3 and MDV-2) that infect chickens. This sample also contained another avian virus, avian leucosis virus (ALV). To determine if these avian viruses were present as a contaminant from chicken tissues, all of the sequenced reads from OSCC13 were aligned to the *Gallus gallus* genome (Galgal4). OSCC13 had 98,146 reads (deduplicated with a mapping quality of ≥ 60) that mapped to Galga4 with a BoC of 0.011%. In comparison, representative FOSCC FFPE samples that did not contain avian viruses, OSCC12-14, had 9144-16,699 reads with a BoC of 0.001-0.002%. GaHV-3 and ALV were thus likely contaminants in sample OSCC13.

A Jaagsiekte retrovirus (ovine pulmonary carcinoma virus) like virus was present in 10.8% (4/37) of all samples, OSCC4, OSCC10 OSCC19 and OSCC21 (Supplementary Table 5). The pattern of SNPs was identical in all samples and thus the virus in all samples came from the same source (Supplementary Figure 11). The viral load was highest in OSCC10. OSCC10 could have cross-contaminated the other samples or all of these samples could have been contaminated by the same sheep tissue source.

Sequence data for the 37 samples were deposited into the Sequence Read Archive (SRA) as under BioProject PRJNA553834.

Discussion

This is the first study to use a comprehensive capture approach to characterize the virome of the oral mucosa of domestic cats. ViroCap not only enabled the detection of PVs but also the complete genomic characterization of 5 new PV variants and physical form of the virus in client-owned cats (Supplementary Table 6). ViroCap sequencing revealed that the oral mucosae from healthy cats and those with FOSCC infrequently contain DNA viruses and that infection with presumably benign foamy viruses was the most common.

Papillomavirus is an unlikely cause of FOSCC

It is currently unknown if high-risk PVs that have a tropism for the oral mucosa exist in cats. High-risk PV genotypes in HNSCC and cervical SCC include, HPV-16, 18, 31, 33, 35, 45, 51, 52, 56, 58, and 59 (37, 38). HPV-16 is the cause of at least 87% of HPV-positive oral and oropharyngeal SCC in people (13, 14, 37). High-risk HPVs belong to the alphapapillomavirus genus. Alphapapillomaviruses by definition only affect primates. To demonstrate that PV causes FOSCC, ideally all 3 of the following criteria would be met: (1) PV DNA in the lesion, (2) expression of oncogenes such as *E6* and *E7*, (3) disruption of tumor suppressor gene, *TP53* and *RB1*, function by viral oncoproteins.

The first criterion was specifically evaluated in this study and was not met.

Papillomavirus was only detected with ViroCap in 5% of all FOSCC samples, a lower

prevalence than in the normal mucosal control samples. These results are consistent with all published literature that includes >1 FOSCC patient, where PV infection was detected in only 0-6.9% of patients (Supplementary Table 2) (22, 25, 26, 39, 40). These results are also consistent with absence of PV detected in OSCC of other felids, e.g. lynx (41) and most reports in the domestic dog (42-45). However, reports exist that identify PV in a relatively low proportion of canine OSCC, 15.8% to 18% (46, 47) but higher than that identified in cats.

It is possible that PV was not detected but causes FOSCC via a 'hit and run' mechanism, where E6 and E7 contribute to tumor initiation, the virus is lost in overt neoplasms, and it is not required to maintain the transformed phenotype. Evidence of this theory has been presented from *in vitro* and *in vivo* rodent models of cutaneous SCC. The hit and run hypothesis has been theorized for feline BISC (18) and demonstrated in a Cre-loxP-based transgenic mouse model of HPV-38 and UV-induced cutaneous SCC (48, 49) and murine fibroblasts that were treated with TPA (12-O-tetradecanoylphorbol-13-acetate) or DMBA (7,12-dimethylbenz[a]-anthracene) and BPV-4 (50, 51). The hit and run hypothesis is unlikely in FOSCC due to the rarity of detecting FPV in FOSCC and in the oral mucosa of cats, lack of *in vivo* evidence that FPV causes premalignant/early changes in the oral mucosa, and the fact that PV is usually detectable with sensitive methods such as NGS in HNSCC lesions that have a genetic signature consistent with PV induced lesions (38).

The only PV virus that was isolated from a FOSSC lesion was FPV-3. FPV-3 is not as commonly isolated as FPV-2 in cats, Supplementary Table 1. FPV-3 infection is commonly pathogenic and was found in cutaneous SCC, nasal planum SCCs and BISC lesions (20, 22, 52-54). FPV-3 could have been a by-stander or incidental infection and not the cause of the carcinoma. The authors plan to further investigate PV as a cause of FOSSC in patient OSCC8 by demonstrating evidence of criteria 2 and 3 through analysis of expression of oncogenes and interaction with host tumor suppressor genes in a future study. Increased expression of p16, E6 and E7 and decreased expression of pRb and p53 would be expected in high-risk PV induced lesions but are less likely in FOSSC based off of results from previous studies. Decreased p53 and pRB expression and increased p16 expression was found in only 2.3% of FOSSC samples and all samples with high p16 expression were negative for PV (26). High-risk PVs encode an E6 protein that binds to p53 sufficiently to induce ubiquitin proteolysis (37). E7 similarly induces pRb degradation. However, E7 of FPV-3 is not expected to behave like in high-risk PVs. The critical cysteine amino acid in the LXCXE motif on E7, that is conserved in HPV-16, FPV-1 and FPV-2, which is responsible for pRb binding was not present in FPV-3, Supplementary Figure 12. FPV-3 either does not bind to pRb and cause its degradation or binds pRb in a yet to be discovered novel mechanism that is different than that of high-risk HPVs.

There is thus currently no evidence that PV is associated with changes in oncoproteins or tumor suppressor proteins in FOSSC analogous to high-risk HPVs. Three other characteristics of FOSSC decrease the likelihood that FOSSC is a good

model for high-risk PV induced SCC in people: the physical form of PV in the host, patient age and tumor location. Another characteristic of high-risk HPVs is integration into the host genome. Initially after infection HPV is maintained as an episome. Integration into the host genome is not necessary to cause HNSCC but is common, estimated to occur in 39-71% of patients with HPV-positive HNSCC (37, 38, 55), and can aid in oncogenesis. Viral transcripts from integrated PV are more stable and expressed at higher levels than episomal virus (37). Integration breakpoints preferentially occurred in or around genic regions in the host genome (38, 55). Breakpoints in the virus genome preferentially occurred in the early genes E1 and E2, disrupting their expression, which leads to the deregulation of negative feedback control of E6 and E7. Integration sites may also result in local gene amplification, structural rearrangements, mRNA transcript abundance and splicing, epigenetic changes (e.g. regulatory elements, histone and methylation) that alter gene expression and contribute to genetic instability (38, 55). In the current study, discordant read pairs (one end of the paired end read mapped to the host genome and the other mapped to the virus), or split reads or sudden drops in coverage were not seen and thus PV did not appear to be integrated into the feline genome. PVs in PV2, PV3, PV4, PV5 and OSCC8 were thus apparently maintained as a full-length episome.

Secondly, FOSCC is almost exclusively a cancer of geriatric patients. This is in contrast to PV-positive HNSCC being most prevalent in younger patients compared to carcinogen induced HNSCC. Thirdly, HPV-positive tumors are the most common

in the oropharynx (especially the tonsils and less often the soft palate, and the base of the tongue) and are rare in the oral cavity or non-oropharyngeal sites (such as the gingiva, hard palate, buccal mucosa, lip, mobile tongue, floor of the mouth, larynx and hypopharynx) (56). The base of the tongue in human medicine is defined as the tonsillar tissue caudal to the circumvallate papilla (57). The base of the tongue in veterinary medicine is often misrepresented as underneath the tongue where tonsillar tissue does not exist. The tonsils are especially at risk due to a discontinuous basement membrane in the palatine, lingual, tubal and adenoid tonsils, known as Waldeyer's tonsillar ring, that allow HPV to access the basal keratinocytes (37, 57). FOSCC most commonly occurs in the oral cavity proper and is rare in the oropharynx (58). The only PV-positive FOSCC lesion in this study, OSCC8, was located in the gingiva and thus non-oropharyngeal. There were no tonsillar FOSCC lesions included in this study, due to the rarity of this lesion at the MU VMTH. None of the PV-positive FOSCC lesions from previous studies were classified as oropharyngeal, Supplementary Table 2. To the author's knowledge PV status of oropharyngeal FOSCC has been evaluated in only 9 cats to date (25, 39, 40) and all lesions were PV-negative. Tonsillar FOSCC lesions have been associated with a better prognosis, than FOSCC lesions located in the oral cavity, when treated with an accelerated radiation therapy and carboplatin protocol (10). If there is a feline correlate with HPV-positive OPSCC, tonsillar FOSCC would likely be the best candidate and thus if future studies investigating PV status in FOSCC are done, tonsillar FOSCCs should be included.

Non-viral causes of FOSCC may include age, genetics and exposure to carcinogens which can affect the oral cavity via ingestion or grooming. The only epidemiologic study to date on causes of FOSCC identified 4 possible carcinogenic agents: tobacco smoke, canned diets, canned tuna in the diet, and flea collars. These agents were associated with relative risk ratios of 2.3 (not significant statistically), 3.6, 4.7 and 5.3 respectively (59). This study also identified flea shampoo usage with a relative risk ratio of 0.1, which is suggestive for carcinogen exposure from grooming as a cause of FOSCC. Carcinogenic agents in HNSCC have been more thoroughly investigated. Strong associations with HNSCC have been found for exposure of the oral mucous membranes to alcohol (relative risk of 2.1-21.1 for light to heavy drinkers), tobacco (relative risk in women 12.96 and 5.45 in men), betel or areca nut (relative risk 2.56-10.98) and combined betel nut with tobacco exposure (relative risk 7.74-24.84) (60-62). Many of the environmental carcinogen exposure in cats and in humans are expected to be shared but alcohol and betel nut exposure would be expected to be negligible in cats. Potential carcinogenic causes of FOSCC would be likely to share a similar oncogenic mechanism, namely oxidation. Future studies on the etiology of FOSCC should focus on carcinogen exposure. Identifying a PV-negative mutational landscape, e.g. with *CDKN2A* and *TP53* mutations, in FOSCC with NGS would also support a carcinogenic cause (55) but this awaits genome-wide, large-scale sequencing of feline tumors.

Other viruses

Foamy virus was the most prevalent virus in normal and FOSCC oral mucosa

Feline foamy or syncytial viruses are ssRNA-RT of the spumaretrovirus subfamily. Feline foamy virus does not appear to be transmitted via friendly social interaction but once a cat is infected a persistent active infection ensues and virus is continuously shed from the oral mucosa after a 15-day incubation period (63). A prevalence of 36% (27% in feral cats and 43% in previously owned cats) was seen in homeless cats from South Australia (64). The prevalence seen in this study was 30% and thus similar to the previous study. Foamy viruses have been used in gene therapy due to the preferential integration into transcriptionally inactive regions and resulting apparent lack in pathogenicity. Complementary DNA conversion was not done on isolated genomic material and thus RNA viruses, such as foamy virus, would have been detected only if the virus was converted to DNA via reverse transcriptase and integrated into the host genome. Integration sites could only be speculated in 4 cases, PV7, OSCC10, OSCC16 and OSCC4 and all locations were in intergenic regions.

Torque teno virus (anellovirus) was the second most prevalent virus but was detected at very low coverage

Anelloviruses are small, non-enveloped circularized ssDNA viruses that have a dsDNA intermediate and have a worldwide distribution. Anelloviruses have been detected in cat feces, saliva and serum (65, 66). Anelloviruses are not known to be

pathogenic in cats and pathogenicity in other species is controversial with the exception of chicken anemia virus (CAV) and porcine circovirus, which is closely related to anelloviruses (67). In humans, anellovirus loads have been used to assess immunocompetency after organ transplants and anthropogenic pollution (68). Prevalence in feline serum of 12.5-43% has been seen worldwide (65, 69, 70). Confirmed and suspected prevalence in all cats in this study were, 5% and 24% respectively. Low coverage with a virome approach likely still indicates presence of anellovirus since anelloviruses are known for genomic diversity and viral loads can vary drastically, especially in response to immunosuppression (71). To confirm the suspected higher prevalence a targeted PCR approach can be done.

Feline herpesvirus was uncommonly detected in cats without signs of respiratory illness

Upper respiratory tract infections in cats are commonly associated with the detection of one or more of the following pathogens: FHV-1, FCV, *Chlamydia felis*, *Bordetella bronchiseptica*, and *Mycoplasma felis* (Hoskins et al, 1998, Sykes et al, 1999, Foley et al, 2002, Chandler et al, 2004, Bannasch, Foley, 2005, Helps et al, 2005, Ruch-Gallie et al, 2008 and Hartmann et al, 2010). Prevalence of the FHV-1 glycoprotein B (gB) gene with qPCR in clinical and non-clinical animals in short-term shelters was highest at 59%/41%, in long-term sanctuaries 19%/8%, in foster care programs 23%/6%, and trap-neuter-return programs 17%/5% respectively. These results may be falsely high since the majority (120/150) of cats in short-term shelters were vaccinated with a modified live vaccine within 21 days of sample

collection (72). Other studies evaluating FHV-1 with PCR in conjunctival and gingival samples of clinical and non-clinical cats reported that only 1-6% cats were positive for FHV-1 (73). In the current study, a similarly low prevalence of 6% was found. None of the normal cats had clinical signs of a URI and it is likely that none or a few of the cats with FOSCC had signs of a URI. Both cats that were positive for herpesvirus were owned cats, OSCC19 and OSCC21, and it was not noted in their biopsy submission that they had signs of a URI.

Epstein Barr Virus was detected in a cat with FOSCC and concurrent FPV-3 infection

Epstein Barr virus or an EBV-like virus was detected at low levels in OSCC8. Human gammaherpesviruses, such as EBV, was deemed an unlikely source of contamination since EBV is shed via human saliva, plasma and blood and personal protective equipment that is routinely worn during surgery or necropsy should have minimized contamination with human bodily fluids. The MU VMDL also does not process human samples and thus cross contamination was unlikely. The oral mucosal sampling route used in this study would have been appropriate for detection of an active EBV infection since EBV replicates in the B-lymphocytes and epithelial cells in the oropharynx and is transmitted via oral secretions. EBV or an EBV-like virus has been detected previously in dogs with serology, PCR and electron microscopy (74-76). EBV or an EBV-like virus has been detected previously in cats with serology (74). The EBV indirect immunofluorescence assay showed cross-reactivity to canine alphaherpesvirus in 10% (1/10) of canine positive control

samples (74) and has been shown to cross-react with simian gammaherpesvirus (77). The first and only *Felis catus* gammaherpesvirus, FcaGHV-1, was described in 2014 (78) and the prevalence of FcaGHV-1 DNA in FIV-negative pet cats from around the world have ranged from 9.6-19.1% and an increased odds for infection was found for male, FIV-positive, and older cats (79). FcaGHV-1 could have led to false positive serologic tests in cats. The complete genome of the first and only feline gammaherpesvirus to date was released on GenBank in 2018 and the ViroCap reagent was produced in 2014 and thus was not included in the version of ViroCap that was included in this study (80). Alignment of the reads to the FcaGHV-1 type strain (GenBank accession NC_028099.1) did not show any reads and thus FcaGHV-1 could not explain the EBV DNA found in OSCC8. EBV or an EBV-like virus was thus present in OSCC8. EBV has been linked to human HNSCC. The presence of circulating EBV DNA in the blood pre and post treatment and antibody titers have been shown to be independent negative prognostic indicators in nasopharyngeal carcinomas (56). EBV infection has also been shown to increase the risk of HOSCC (81). Interestingly, OSCC8 was the only FOSCC sample that was PV-positive. This is the first study to find EBV DNA in a cat and further screening of FOSCC patients for EBV DNA would be recommended to explore if EBV has a role in a subset of FOSCC patients.

Exogenous FeLV was not detected in the cat that was ELISA FeLV positive

Exogenous FeLV was not detected with ViroCap from the cat, N2, which was diagnosed with an active FeLV infection with the IDEXX FeLV/FIV ELISA SNAP test.

The SNAP test detects the presence of soluble viral capsid protein p27 in the blood. The SNAP test has been reported to have a sensitivity and specificity of 91.3-100% and 92.3-100% respectively and especially in cat populations with low FeLV prevalence false positives are possible (82, 83). The SNAP FeLV positive cat could have been a false positive since a confirmatory test such as PCR, virus isolation or immunofluorescence assay was not done and the screening test was not repeated. False positives may occur due to anti-mouse IgG antibodies, found in $\leq 0.5\%$ of domestic cats, that bridge the murine-derived monoclonal antibodies and solid phase antibodies without the presence of p27 (84). The cause of the false positive results were reported to the manufacturers in a subsequent study and the problem was rectified in $\frac{1}{4}$ kits when the samples were retested with the improved kits (85). The improved test kits contain reagents in the conjugate that block feline anti-mouse antibodies from binding to the mouse antibodies (83). The false positivity rate caused by feline anti-mouse IgG antibodies with the test kits available today is unknown but is likely lower than $\leq 0.5\%$ due to technologic improvements.

Alternatively, if N2 was truly antigen positive, negative proviral DNA is possible if the cat had a focal/atypical infection, later developed a regressive FeLV infection after initial antigen testing or ViroCap was technically unable to detect exFeLV in the conditions of this study. With a focal or regressive infection the SNAP test would have been only transiently positive (or negative) and proviral DNA would only have been detected at low levels from selected tissues (36, 86). Feline oral mucosa may not have contained sufficient exFeLV for detection by ViroCap. This study would have only been able to detect integrated exFeLV and thus a sample with a higher

number of lymphocytes (e.g. blood, lymph node aspirate, bone marrow) or salivary gland would have been a more appropriate sample for testing proviral FeLV status. PCR for exFeLV proviral DNA or RNA would also be expected to be more sensitive than ViroCap (87). It has been previously reported that differentiating endogenous from exogenous FeLV with whole genomic approaches has been difficult due to the presence of enFeLV in all domestic cats, the marked genomic similarity between enFeLV and exFeLV and within host FeLV genomic heterogeneity (88, 89).

ViroCap confirmed the presence of FIV in a known FIV infected cat

On the other hand, ViroCap demonstrated that it was able to detect other types of retroviruses. FIV provirus was detected at low levels in 2 cats, N3 and N9, Supplementary Table 11. ViroCap results were consistent with the FIV SNAP test (detects antibodies to the p24 core antigen) for N3. N9 was a stray cat that presented for a large erosive lesion on the nasal planum that was suspected to be a squamous cell carcinoma. The FeLV/FIV status of N9 was not known and thus the ViroCap result cannot be confirmed. Similarly with FeLV, increased read counts and thus sensitivity for FIV provirus detection would likely have been obtained if a lymphocyte rich sample type was tested. Alternatively, fresh frozen oral mucosal tissue would be a more suitable sample type since both of these viruses are shed in the saliva during active infection but cDNA conversion of genomic material would need to be done for RNA detection, which was not done in this study.

Possible contamination with avian ovine, and human viruses

One sample, OSCC13, contained two chicken viruses, GaHV-3 and ALV. Gallid herpesvirus type 2 (GaHV-2) or Marek's disease virus serotype 1 (MDV-1) causes a rapid onset T-cell lymphoma and demyelinating disease in chickens, Marek's disease. It is a disease that is still present in the US at an estimated rate of 0.001% (based on broiler condemnation rates) in 2011 (90) and diagnosed on histopathology samples submitted to the MU VMTH (personal communication). GaHV-3 also known as MDV-2, is not oncogenic in chickens, and is used in the bivalent vaccine formulation. It is likely that this sample was contaminated with materials from chickens during sample preparation (e.g. carcass storage, necropsy table, scroll preparation). Similarly, PV3 contained an adenovirus possibly from contamination from turkeys and OSCC4, OSCC10, OSCC19 and OSCC21 contained retrovirus possibly from contamination from sheep. HPV-28 was detected at a low level in sample N6. HPV-28 causes benign flat plane warts in humans and was likely present in N6 due to contamination from tissue/sample handling (91). The microtome and instruments were cleaned with 95% ethanol before scrolls were prepared for each sample. Forensic studies have shown that wiping down surfaces with ethanol are not sufficient to remove contamination (92). Cleaning with a dilute (0.9-6%) hypochlorite solution or commercial decontamination solution (e.g. DNA ZAP TM) instead are effective methods and essentially removed all contaminating DNA (93). In future studies, where sensitive genomic work is involved modified cleaning techniques will be requested.

Limitations

A limitation of the current study is that RNA viruses that had not integrated into the host genome would not have been detected. Oncogenic viruses such as single-stranded RNA (ssRNA) viruses that belong to the Flaviviridae family (e.g. hepatitis C virus) would not have been detected; although an oncogenic ssRNA virus is not currently known to infect cats. The ViroCap reagent does include probes for RNA viruses and thus has the capability to identify RNA viruses but would require conversion of genomic RNA into cDNA, which was not done in this study, since PV, a DNA virus, was the main target. A future metagenomic study that includes capture of RNA viruses would be required to completely rule out a viral cause of FOSCC, but RNA viruses would be considered a highly unlikely cause of FOSCC since a non-integrated RNA virus has not caused oral squamous cell carcinoma in any species to date.

Only 20 FOSCC samples were tested in this study. In comparison, the largest study investigating the presence of PVs in FOSCC included 43 cats (26). A larger study that includes FOSCCs in the tonsillar regions, could be pursued to more definitively resolve the role of PV in FOSCC. Papillomavirus was found at very high viral loads in the positive control samples. The spike-in of PV probes in addition to the standard ViroCap reagent was likely not necessary for the positive controls. In addition, an increased number of viral reads from test samples would have been obtained if the final library concentrations of the PV positive controls were decreased in the pool

prior to capture. As a result a PV spike-in would only be recommended in future ViroCap studies if PV was the virus of greatest interest and at least a 100 times dilution of each PV positive control would be recommended. Higher sensitivity for PV detection was a trade-off for detection of other viruses that were present at lower viral loads and for including more samples per lane of sequencing.

Another limitation is that treatment and prognostic data were not available for the cats with FOSCC. In humans with HNSCC and cats with cutaneous SCC, PV infection has been associated with a better prognosis. Unless there is a funded clinical trial, owner and patient commitment required for treatment, significant side effects of radiation therapy treatment, poor long-term prognosis and cost associated with treatment will likely preclude the ability to make conclusions about prognosis for PV-positive FOSCC patients and treatment outcomes.

Conclusion

FOSCC is not a good model for PV-induced squamous cell carcinoma in humans. If PVs have a role in FOSCC it is at most a supplementary or uncommon role and elucidation of other factors, e.g. environmental factors, would make a bigger impact on the role of FOSCC as a naturally occurring HNSCC model and clinical outcome of FOSCC.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Declaration of Interest

None.

Funding Information

This work was supported by a University of Missouri Phi Zeta Veterinary Honor Society resident grant.

Acknowledgements

The authors would like to acknowledge Maren Fleer at the University of Missouri for technical assistance.

CHAPTER 2:

Genome wide exploration of the methylome in Diffuse Large B Cell Lymphoma in Golden Retrievers reveals a stable hypermethylome

Shirley Chu^a, Anne Avery^d, Janna Yoshimoto^d, Tasha Miller^e, Heather Wilson-Robles^e,
Obi L. Griffith^{b,c}, Jeffrey N. Bryan^a

^aDepartment of Veterinary Medicine and Surgery, University of Missouri, 900 E.
Campus Drive, Columbia, MO, 65211, USA

^bDepartment of Medicine, Washington University, 660 S. Euclid Ave, St. Louis, MO,
63110, USA

^cMcDonnell Genome Institute, Washington University, 4444 Forest Park Ave, St.
Louis, MO, 63108, USA

^dDepartment of Microbiology, Immunology and Pathology, Colorado State
University, 200 West Lake Street, Fort Collins, CO 80523, USA

^eDepartment of Small Animal Clinical Sciences, Texas A&M University, 408 Raymond
Stotzer Parkway, College Station, TX 77845

Keywords: DLBCL, methylome, MIRA, sequencing, Golden Retriever

Abbreviations:

cDLBCL: canine diffuse large B-cell lymphoma

hDLBCL: human diffuse large B-cell lymphoma

COO: cell of origin

ABC: activated B cells

GCB: germinal center B cells

PBMC: peripheral blood mononuclear cells

GEP: gene expression profiling

NHL: non-Hodgkin lymphoma

CpG: cytosine-phospho-guanosine dinucleotide

GR: Golden Retriever

Introduction

Diffuse Large B Cell Lymphoma (DLBCL) is the most common subtype of lymphoma in humans and in dogs, accounting for 21% and 52% of all lymphomas respectively (94). Canine DLBCL (cDLBCL) shows remarkable similarities to human DLBCL (hDLBCL); both are treated with and respond to CHOP-like chemotherapy protocols, share a common cytologic and histologic appearance and share common dysregulated pathways e.g. NF- κ B (95-98). The importance of the canine model for hDLBCL has been proven with the completion of clinical trials in companion dogs that have provided the necessary data to bring drugs into human clinical trials, including a topoisomerase 1 inhibitor (LMP744) (99), an autologous vaccine with heat shock protein peptide complexes with hydroxylapatite (100), and PI3K δ inhibitor (RV1001) (101).

Genomic sequencing is a crucial tool to allow us to understand changes that lead to cellular transformation, biological behavior of tumors, creation of non-invasive diagnostics, and tumor classification for prognostication. As the cost of sequencing decreases, large scale genomic studies have become feasible in veterinary oncology. The first large scale genomic studies in cDLBCL have focused on gene expression profiles (GEP). Gene expression profiling in hDLBCL has identified 3 main subtypes of hDLBCL that differ in prognosis and mirror the cell of origin, germinal center B cell-like (GCB), activated B cells-like (ABC), and unclassified or primary mediastinal (102). DLBCL is thought to arise from B-lymphocytes that have encountered antigen, become activated by T-helper cells and migrated into the lymph node

follicle. GCB cells are centroblasts in the secondary or mature follicle (germinal center) that are undergoing continual somatic hypermutation (SHM) of their B-cell receptor (BCR). GCBs then become activated B cells (ABC) or post-GC cells when SHM is complete and the cells exit the germinal center to become plasma cells or memory cells. Key genes drive B cell development and thus the different stages of B cell development can be identified by gene expression. For example, GCB cells express *CD10*, *BCL6* and *EZH2*, whereas ABCs are known for constitutive activation of the NF- κ B pathway and classically express *IRF4/MUM1*.

It is less clear that the same 3 subgroups exist in cDLBCL. RNA expression microarray data was only able to subgroup cDLBCL into the ABC and GCB subgroups when clustered with hDLBCL microarray data using the 25-gene panel used in hDLBCL (103). The resulting subgroups also did not differ in prognosis. On a second experiment using a panel of dog specific genes that were identified to separate the groups with supervised clustering, a prognostic difference was only then found. In support of an ABC-like subgroup cDLBCL patients in the ABC group had higher expression of genes in the NF- κ B pathway (e.g. *NFKB1*, *NFKB1B*, *RELA*) as would be expected in hDLBCL. Conclusions from this data were that a set of genes, that are different than in hDLBCL, can be used to identify dogs that will have a poorer prognosis and can be concluded to have an ABC-like phenotype. An alternative hypothesis is that the 3 subgroups do not exist in cDLBCL and cDLBCL may be more homogenous than hDLBCL and more closely resemble ABC than GCB hDLBCL (96).

More recent 'omic studies in hDLBCL and cDLBCL have evaluated the methylome or genome-wide DNA methylation. DNA methylation is reversible but it is perhaps the most stable epigenetic change. In normal cells DNA methylation is crucial for normal cell development, differentiation, imprinting and X-chromosome inactivation. As such epigenetic programming is precisely controlled in normal B-cell development. On the other hand, aberrant DNA methylation is recognized as a consistent characteristic in cancers across different cell types and species (104). The functional result of DNA methylation can vary depending on the cell type, cell environment, gene, and location within the gene or regulatory region. The most consistent consequence of DNA methylation is methylation at CpG-rich promoters, which is often associated with transcriptional silencing. In cancer, two main patterns of methylation are seen: (1) global DNA hypomethylation which leads to genomic instability and (2) focal hypermethylation at regulatory regions near the promoter of tumor suppressor genes that can contribute to the other hallmarks of cancer. One example is the hypermethylation of the promoter region of adenomatous polyposis coli (*APC*), which can lead to continued self-renewal and metastasis in early onset colorectal cancer. Another example is the hypermethylation of *CDKN2A* which can lead to continued growth signals in many malignancies including breast, head and neck cancers, gliomas, and melanomas (105). Focal hypermethylation has been utilized for subtyping hDLBCL. Supervised hierarchical clustering of hDLBCL with the HELP (*HpaII* tiny fragment enrichment by ligation-mediated PCR) promoter microarray showed a 91% predictive accuracy

in classification of patients into the ABC and GCB subgroups but this was not tested on a new validation cohort of patients. (106). It can be concluded from this study that ABC and GCB subgroups are more clearly separated on the transcriptome but the methylome can also be used to highlight the differences between the subgroups.

Genome wide DNA methylation in cDLBCL has been evaluated in only three studies as far as the authors are aware. The first study is a custom canine promoter microarray study in the whole lymph nodes from 37 dogs (various breeds) with DLBCL. Conclusions from this study were that polycomb repressive complexes in stem cells were notably affected by hypermethylation. Functional pathways affected included homeobox, bone morphogenic protein, and Wnt pathways. Hypermethylated sequences were enriched for promoter, 5'-UTR, and exonic regions (Bonferroni-adjusted p-values < 0.006), whereas hypomethylated sequences were enriched for exonic, 3'-UTR, and downstream regions (107).

The second study used sequential restriction enzyme digestion in the CLBL-1 cDLBCL cell line and thus was not restricted to the CGI biased microarray in the Ferraresso study (108). This study concluded that thousands of CpG sites in CLBL-1 gained methylation at CGI sites and lost methylation at normally methylated non-CGI sites. Over-represented hypermethylated sites included the promoter CGI sites in *TWIST2*, *TLX3* and genes with characteristics of 'Homeobox' and 'DNA-binding' (DAVID).

The most recent genome wide study, evaluated global methylation in peripheral leukocytes (buffy coat) of dogs with different types of aggressive and indolent lymphomas by HPLC and immunocytochemistry (109). Global hypomethylation was found in dogs with lymphoma. A subset of samples were evaluated with a human promoter microarray and the biological process (BP) pathways that were enriched included development processes, regulation of cell development and response to external stimulus.

In all of the previous cDLBCL studies the differentially methylated genes in cDLBCL and hDLBCL have not been described to be similar, either because the analysis was not done or the similarities were not present. Instead similarities were described at the level of the pathways. Pathways that were common in hDLBCL and cDLBCL included the homeobox and developmental pathways. In general, although similarities in dysregulated genes at the genetic and epigenetic level can be shared, the genetic landscape is marked with differences (95, 96, 103). Despite the genetic differences when looking at the individual gene level, the clinical similarities and driver pathway similarities have shown the value of dogs as a model for hDLBCL.

Methylated-CpG island recovery assay and next generation sequencing (MIRA-seq), relies on the high affinity of the MBD2b-MBD3L1 complex for CpG-methylated DNA. MIRA-seq is an affinity enrichment-based method to identify highly methylated areas using a whole genome based approach. Compared to the gold standard of whole genome bisulfite sequencing (BS-seq) for methylation analysis, MIRA-seq is

more sensitive for CpG-rich DNA compared to BS-seq. Since cancer methylomes are known to have stark differences in areas of high-CpG-density such as CGIs, this potential bias of MIRA-seq, can be an advantage for interrogating cancer methylomes. MIRA-seq was used in the current study for the following advantages over microarrays: the number of evaluable loci are not limited by the space restrictions on the chip, and better resolution of methylated CpGs is possible.

Due to financial limitations veterinary genomic studies are comparatively small compared to human genomic studies. Small sample sizes limit the ability to identify subgroups. Restricting the background genomic variability from different dog breeds is a method that can be used to overcome sample size limitations. To obtain as homogenous group of dogs with naturally occurring DLBCL as possible, the study group in this study was limited to Golden Retrievers from the US. The Golden Retriever breed was selected due to the relatively low genetic variation in this breed and the high risk of developing lymphoma, estimated to be a 11.5% lifetime risk (diagnosis confirmed by veterinarian) (110) and an overall risk of 3.12 when compared to mixed breed dogs (111). The study was restricted to Golden Retrievers in the US since an increased risk of lymphoma has not been documented for any other country in the EU, except for Golden Retrievers in the UK (112).

The current study differs from the previous cDLBCL methylation studies in that genetic variability due to breed differences was minimized by only including Golden Retrievers and genetic variability from contamination with other cell types (e.g. T

cells, other white blood cells and connective tissue cells) was minimized by only including normal control samples with immunophenotypically sorted cells. This study is also the only study in cDLBCL tumor samples that is not limited to the promoter regions.

The objective of this study was to characterize the methylome (hyper and hypomethylated genes, chromosomal distribution, distribution in regards to genomic features) in DLBCL in Golden Retrievers. Once the methylome was defined the following hypotheses were tested 1) the methylation pattern in DLBCL would be distinct from normal cells; 2) subgroups of DLBCL could be defined by the methylome; 3) hypermethylated regions compared to normal B cells would be preferentially located in the promoter-TSS (transcriptions start site); 4) hypermethylated regions would be in genes that are known cancer associated genes, specifically tumor suppressor genes (TSG); 5) hypermethylated regions would parallel genes seen in human DLBCL (hDLBCL).

Materials and Methods

Patients

Samples were submitted from the Veterinary Cancer Center (CT), Southwest Veterinary Oncology (AZ), the Oncology Service (VA), New England Veterinary Oncology Group (MA), Front Range Oncology Group (CO), Oakland Veterinary Referral Services (MI), Georgia Veterinary Specialists (GA), Sugar Land Veterinary Specialists (TX), Animal Cancer Care Clinic (FL), Colorado State University Veterinary Teaching Hospital (CO), University of Missouri Veterinary Health Center (MO), and Texas A&M (TX). All dogs with lymphoma that were included in the study were client owned purebred Golden Retrievers. All dogs had at least stage 3 lymphoma with diffuse peripheral lymphadenopathy. A board-certified clinical pathologist confirmed diagnosis of medium to large cell lymphoma with frequent mitoses on cytologic interpretation. Full staging diagnostics with abdominal ultrasound, thoracic radiographs and bone marrow aspirate were not performed on all dogs. Client consent was obtained prior to inclusion in this study. Lymph node samples were either biopsy or fine needle aspirate samples that were collected as part of the patient's diagnostic plan. Dogs were excluded if they were ever treated with chemotherapy for lymphoma prior to sample collection.

The MIRA-seq lymphoma affected group included lymph node samples from 23 Golden Retrievers with treatment naïve DLBCL (fine needle aspirates). The MIRA-seq normal control dogs included CD21+ cells from 7 dogs (5 from lymph nodes and 2 from blood), CD4+ T cells from 4 dogs (all from lymph nodes), CD8+ T cells from 3

dogs (all from lymph node) and CD5+ cells from 2 dogs (all from blood). A clinical summary of the dogs included in the study can be found in Table 1.

Table 1: Clinical information of the Golden Retrievers that were sequenced by MIRA-seq

Sample	Breed	Sex	Age	Tissue Source	Treatment	Survival (days)	Cause of death
DLBCL1	GR	MI	7	LN FNA	CHOP	234	LSA
DLBCL2	GR	FS	10	LN FNA	Doxorubicin	206	LSA
DLBCL3	GR	MC	11	LN FNA	N/A	N/A	N/A
DLBCL4	GR	MC	6	LN FNA	Prednisone	34	LSA
DLBCL5	GR	MC	6	LN FNA	COP/CHOP	387	LSA
DLBCL6	GR	FS	14	LN FNA	Doxorubicin	305	LSA
DLBCL7	GR	FS	6	LN FNA	Prednisone	N/A	N/A
DLBCL8	GR	MC	8	LN FNA	Doxorubicin	>107	LTFU
DLBCL9	GR	FS	10	LN FNA	Prednisone	>1	LTFU
DLBCL10	GR	FS	6	LN FNA	CHOP	126	LSA
DLBCL11	GR	MC	8	LN FNA	Doxorubicin	219	LSA
DLBCL12	GR	FS	8	LN FNA	CCNU CHOP	289	LSA
DLBCL13	GR	FS	9	LN FNA	CHOP	129	LSA
DLBCL14	GR	MC	5	LN FNA	Prednisone	N/A	LTFU
DLBCL15	GR	FS	9	LN FNA	N/A	N/A	N/A
DLBCL16	GR	FS	11	LN FNA	Doxorubicin	>512	Renal failure
DLBCL17	GR	MC	14	LN FNA	N/A	N/A	N/A
DLBCL18	GR	MC	9	LN FNA	N/A	N/A	N/A
DLBCL19	GR	FS	10	LN FNA	Doxorubicin	>92	LTFU
DLBCL20	GR	FS	13	LN FNA	N/A	N/A	N/A
DLBCL21	GR	MC	13	LN FNA	N/A	N/A	N/A
DLBCL22	GR	FS	10	LN FNA	N/A	N/A	N/A
DLBCL23	GR	MC	8	LN FNA	CHOP	341	LSA
CD21_1	BG	FS	Young adult	Lymph node biopsy; flow sorted	N/A	N/A	Purpose bred dog
CD21_2	BG	FS	Young adult	Lymph node biopsy; flow sorted	N/A	N/A	Purpose bred dog
CD21_3	GR	FS	10	PBMC; flow sorted	N/A	4947	LarPar
CD21_4	LR	MC	10	Lymph node biopsy; flow sorted	Amputation and carboplatin	160	OSA
CD21_5	MBD	FS	7	Lymph node biopsy; flow sorted	N/A	2574	Necrotizing pancreatitis
CD21_6	GR	MC	6	PBMC; flow sorted	N/A	>3804	Alive
CD21_7	BH	MI	7	LN; flow sorted	N/A	2684	Suspect GDV

CD4_1	BG		Young adult	Lymph node biopsy; flow sorted	N/A	N/A	Purpose bred dog
		FS					
CD4_2	LR		10	Lymph node biopsy; flow sorted	Amputation and carboplatin	160	OSA
		MC					
CD4_3	MBD		7	Lymph node biopsy; flow sorted	N/A	2574	Necrotizing pancreatitis
		FS					
CD4_4	BH	MI	7	Lymph node biopsy; flow sorted	N/A	2684	Suspect GDV
CD8_1	MBD		7	Lymph node biopsy; flow sorted	N/A	2574	Necrotizing pancreatitis
		FS					
CD8_2	LR		10	Lymph node biopsy; flow sorted	Amputation and carboplatin	160	OSA
		MC					
CD8_3	BH	MI	7	Lymph node biopsy; flow sorted	N/A	2684	Suspect GDV
CD5_1	GR		10	PBMC; flow sorted	N/A	4947	LarPar
		FS					
CD5_2	GR	MC	6	PBMC; flow sorted	N/A	>3804	Alive

GR: Golden Retriever; BG: Beagle; MBD: mixed breed dog; FS: female spayed; MC: male castrated; LN FNA: lymph node fine needle aspirate; PBMC: peripheral blood mononuclear cells; TIC: tumor initiating cell; TCC: transitional cell carcinoma; LarPar: laryngeal paralysis; GDV: gastric dilation and volvulus

Flow cytometry

Submitting clinics obtained lymph node aspirates by aspirating material from the node into a solution of saline and 10% serum from the patient or another healthy dog. This suspension was pelleted and resuspended in 1 ml of lysis buffer (0.15 M NH₄CL, 1 M KH₂PO₄, 0.1 mM Na₂EDTA, 1 M HCL at a pH of 7.2-7.4) for 5 minutes to remove red blood cell contamination. Samples were then pelleted again and resuspended in between 150 ul and 1000 ul of flow cytometry buffer. A 96-well plate

was used in which 25 µl of cell suspension was added to individual wells plus 25 µl of a cocktail of antibodies, Supplementary Table 1 (not provided for the dissertation). Samples were incubated for 15 minutes at room temperature and then washed twice in flow cytometry buffer. Samples then were resuspended in flow cytometry buffer with 10 µg/ml of propidium iodide for dead cell exclusion, and analyzed within 1 hour. Flow cytometry was performed at Colorado State University. Samples were analyzed using a 3 laser Coulter Gallios. All data analysis was carried out using Kaluza software (Beckman Coulter). Only dogs with a CD21 lymphocytosis on lymph node aspirates were included in the study and CD34+ dogs were excluded.

Magnetic column based manual cell separation

If flow sorting was unavailable, the Miltenyi MACS system was used for separation of B cells from canine PBMCs and homogenized lymph nodes. The manufacturer's protocol was followed. Specifically, cells were run through a pre-separation filter (#130-041-407, Miltenyi), washed in PBS with 1% BSA, incubated with 1/100 dilution of mouse anti-canine CD21 (#MCA1781R, Bio-Rad), washed with PBS with 1% BSA, incubated with goat anti-mouse IgG microbeads (130-048-402, Miltenyi), washed with MACS buffer (1:20 dilution of 130-091-376 with #130-091-222, Miltenyi), eluted in MS column (#130-041-201, Miltenyi). An aliquot of the eluent was incubated with mouse anti-canine CD21 PE (#MCA1781PE, Bio-Rad) and rat anti-canine CD5 (#MCA1037A647, Bio-Rad) as per the manufacturer's protocol and counted on the BD FACScan for purity estimates.

PBMC Processing

All normal dogs had a CBC with blood smear evaluation by a board-certified clinical pathologist to confirm the absence of atypical lymphocytes or a lymphocytosis.

Blood samples were collected in mononuclear cell preparation (CPT) sodium citrate tubes (#362761, BD Biosciences) and processed according to the manufacturer's protocol.

DNA library preparation and sequencing

Genomic DNA was extracted with the DNeasy blood and tissue kit (Qiagen Sciences, Germantown, MD, USA). Spectrophotometric and fluorometric DNA quantification was done with the NanoDrop and Qubit platforms respectively (NanoDrop products, Wilmington, DE, USA and Life Technologies, Grand Island, NY, USA). One ug of genomic DNA was sonicated with the Covaris M220 focused ultrasonicator to ~350 bp fragments (Covaris Inc, Woburn, MA, USA). Clean up was accomplished with 1.6x AMPure XP beads (Beckman Coulter Inc, Brea, CA, USA). Library preparation was performed with the NEBNext Ultra DNA Library Prep Kit for Illumina (#E7370, New England BioLabs Inc, Ipswich, MA, USA). The MethylCollector Ultra kit was used for the MIRA (#55005, Active Motif, Carlsbad, CA, USA). PCR enrichment and indexing was done with the NEBNext Multiplex Oligos for Illumina kit (#E7335, New England BioLabs Inc, Ipswich, MA, USA). PCR was performed on the C1000 Thermal Cycler (Bio-Rad, Hercules, CA, USA) with the following conditions, 98°C for 30s, 11 cycles of 98°C for 10s, 65°C for 30s, 72°C for 30s and a final extension of 72°C for 5 minutes.

The conditions were revised to 98°C for 30s, 4 cycles of 98°C for 10s, 65°C for 75s, and a final extension of 72°C for 5 minutes, when the PCR buffer was changed to the Q5 Hot Start HF PCR Master Mix. Libraries were cleaned twice with 1x Agencourt AMPure XP beads (A63881, Beckman Coulter, Indianapolis, IN, USA). Qubit quantification and fragment analysis was done on the resulting library (Advanced Analytical Technologies, Ames, IA, USA). DNA libraries were pooled to 3-4 samples per lane on the HiSeq 2500 sequencer and sequenced as 100 bp single-end reads (Illumina, San Diego, CA, USA). The last samples (LSA culture and PBMC samples) were pooled 8 samples per lane on the NextSeq sequencer and sequenced as 75 bp single-end reads due to replacement of the HiSeq with the NextSeq at the University of Missouri DNA core.

Quality editing and control

Reads were trimmed 15 bp at the 5' end and filtered for quality (> 30% N calls or > 10% of the sequence with Phred quality < 20 were omitted) using Cutadapt v1.6 (113). Trimmomatic v0.32 was used to further select high quality reads (114). Reads were dropped if the read length <20. Reads were trimmed if the average quality per base < 20 (4-base sliding window). The ends of the reads were trimmed if the quality of the base was <10. Base quality was measured with the simplified log-likelihood approach. FastQC v0.11.2 was used to obtain summary QC statistics for the remaining reads.

Alignment

Remaining reads were then aligned to the *Canis familiaris* reference CanFam3.1 with BWA-MEM v0.7.10 with default parameters (115).

MEDIPS Analysis of MIRA-seq data

Data was analyzed with the Bioconductor MEDIPS package, version 1.34.0 (116).

The following parameters were specified: unique reads, reads were extended 300 nucleotides, and each chromosome was divided into 100 nucleotide fragments for binning purposes. Only the autosomes were analyzed. Differential methylation was evaluated with the following parameters, Benjamini-Hochberg correction method, q-value of 0.001, >2 fold change, EdgeR, ≥ 10 reads per 100 bp window.

DMRs were annotated to Ensembl canfam3.1 version 96 (last modified on 3/13/19, 2:04:00 PM) with HOMER v4.8. Mitochondrial, scaffold and sex chromosomal DNA were excluded. The promoter was defined as 1000 bp upstream and 100 bp downstream of the TSS (default in HOMER).

DMRs were also manually annotated as in a promoter by two additional methods.

Firstly, if the start or end of the DMR was within 500 bp upstream or 200 bp downstream of the TSS, the DMR was labeled as belonging to the promoter-TSS region. Secondly, if the DMR was in the region outlined with previous CAGE studies (Cap Analysis of Gene Expression) (117, 118), the DMR was labeled as belonging to the promoter-TSS region. The Eukaryotic Promoter Database (EPD) has 7545

experimentally validated promoters, Cf_EPDnew_001 (119). This database represents the CAGE data that was further analyzed (Ensembl version 92) and verified to only contain promoters that are associated with protein coding genes, on full chromosomes and annotated in UCSC with known transcripts.

Pathway analysis

Pathway analysis was done with open access software, David v6.8 and ConsensusPathDB v34. Overrepresentation analysis based on the hypergeometric test was done with ConsensusPathDB.

Data visualization

Heatmaps were done with gplots in R. Hierarchical clustering was done with the Euclidian average method. The log2 of the reads per kilobase of transcript, per million mapped reads (rpkm) values were used.

Results and Discussion

1. MIRA-seq Quality Control

Saturation analysis resulted in true correlation and estimated correlation 84.7-97.1% and 90.6-98.5% respectively for all samples, Table 2. An estimated saturation of $\geq 90\%$ indicates that there were a sufficient number of mapped reads for that sample to generate reproducible coverage of the entire canine genome and thus coverage was sufficient to compare samples to each other (120). RelH (based on the relative frequency of CpGs within the binned regions) and GoGe (based on the observed/expected ratio of CpGs within the binned regions) were 2.4-3.8 and 1.7-2.2 respectively. The GoGe is usually 1.6 to 2.0 in enriched samples and 0.8 to 1.0 for input samples (120) and thus confirms that the libraries in this study were adequately enriched. After read editing the final libraries contained 22,721,628-55,389,227 reads with an average of 35,946,025 reads per sample.

Table 2: Raw reads and QC statistics for Golden Retrievers with DLBCL and normal control dogs that were sequenced with MIRA-seq

Sample	Gro up	#reads (MIRA)	#reads analyzed (unique)	SR %	EstS R%	Rel ER %	O/E ER %	^a No n- spec ific	^b No cov erag e	^c Cove rage dept h
DLBCL1	B	51,359,995	47,609,175	93.8	96.9	2.9	1.9	7.4	1.8	60.7
DLBCL2	B	41,399,042	40,082,968	89.2	94.4	2.6	1.7	7.9	2.2	59.8
DLBCL3	B	58,373,584	54,040,941	95.1	97.5	3.4	1.9	3.9	2.8	64.6
DLBCL4	B	41,738,784	39,581,211	91.4	95.5	2.8	1.8	5.3	3.5	57.1
DLBCL5	B	26,935,182	26,022,969	89.3	94.5	3.0	1.8	5.6	7.7	42.3
DLBCL6	B	31,008,012	29,525,053	95.2	90.6	3.0	1.9	4.7	6.4	47.7
DLBCL7	B	47,157,428	32,940,236	95.0	97.4	3.7	2.0	3.7	9.3	48.7
DLBCL8	B	37,920,175	33,576,547	90.7	95.2	2.9	1.8	6.7	4.9	52.0
DLBCL9	A	43,321,850	41,653,868	90.6	95.4	2.7	1.7	8.8	2.3	56.8
DLBCL10	B	43,534,419	41,453,661	91.8	95.8	2.8	1.8	8.1	2.4	58.9
DLBCL11	B	49,690,644	44,515,151	97.1	98.5	3.8	2.1	4.5	6.4	47.5
DLBCL12	A	47,493,046	44,245,570	93.4	96.7	2.9	1.9	7.1	3.7	53.7
DLBCL13	B	39,332,981	37,097,831	94.2	97.1	3.2	2.0	6.5	5.3	46.1

DLBCL14	B	39,147,192	36,242,187	94.8	97.4	3.4	2.1	5.0	9.3	44.8
DLBCL15	B	33,439,549	31,313,655	95.1	97.5	3.7	2.1	3.9	10.7	41.8
DLBCL16	A	38,252,308	36,120,041	93.8	96.8	3.2	2.0	6.0	6.6	47.2
DLBCL17	A	34,993,436	33,681,982	90.1	94.9	2.7	1.9	8.5	10.3	46.6
DLBCL18	A	38,036,341	35,809,462	94.7	97.3	3.2	2.0	6.7	8.2	43.4
DLBCL19	A	31,899,654	29,648,174	95.6	97.8	3.6	2.2	6.0	11.3	36.0
DLBCL20	A	34,044,753	32,737,763	91.5	95.6	2.8	1.9	8.4	8.6	42.2
DLBCL21	A	33,611,449	31,773,791	93.8	96.9	3.1	2.0	7.4	8.5	40.1
DLBCL22	A	37,033,917	35,274,909	93.3	96.6	2.7	1.9	9.6	5.5	41.5
DLBCL23	A	34,183,497	32,232,529	94.9	97.4	3.3	2.1	7.6	9.3	37.4
CD21_1	N/A	33,845,383	31,984,743	95.0	97.5	3.2	2.1	7.4	10.9	36.1
CD21_2	N/A	36,878,359	34,915,675	94.4	97.2	3.1	2.0	7.2	9.0	40.5
CD21_3	N/A	37,165,903	34,505,939	95.2	97.6	3.3	2.0	7.0	6.8	41.6
CD21_4	N/A	31,066,504	29,250,908	95.6	97.8	3.6	2.1	5.1	16.4	36.2
CD21_5	N/A	33,522,807	32,396,448	92.1	96.0	2.9	1.9	7.4	8.6	41.2
CD21_6	N/A	23,425,701	22,721,628	90.8	95.3	3.0	2.0	7.9	13.8	31.4
CD21_7	N/A	34,660,459	33,772,102	87.5	93.6	2.5	1.7	10.7	8.2	46.4
CD4_1	N/A	37,484,427	34,640,638	96.2	98.0	3.5	2.1	6.8	12.9	37.7
CD4_2	N/A	36,326,844	33,919,969	95.0	97.5	3.2	2.0	6.8	7.1	40.6
CD4_3	N/A	58,344,528	55,389,227	94.6	97.3	2.8	1.8	8.3	6.8	60.8
CD4_4	N/A	40,227,668	38,221,981	88.7	94.2	2.4	1.7	7.6	8.6	51.7
CD8_1	N/A	47,942,251	45,741,691	94.1	97.0	2.9	1.9	6.6	7.5	52.6
CD8_2	N/A	31,386,418	29,872,097	92.9	96.4	3.0	1.9	7.7	6.7	38.9
CD8_3	N/A	37,196,347	35,932,345	87.0	93.2	2.4	1.7	7.9	9.5	50.3
CD5_1	N/A	37,973,552	36,042,632	91.7	95.8	2.6	1.8	10.5	5.1	45.5
CD5_2	N/A	25,985,281	25,407,292	84.7	92.1	2.4	1.8	11.4	9.1	32.4

SR: saturation analysis; ER: enrichment analysis.

^a% reads that do not contain GC

^b% CG's with 0% coverage

^c% CG's with >5x coverage

2. Differential Methylation Analysis

Normal B cells compared to DLBCL

MIRA-seq identified more DMRs than previous methylation studies done with other methods.

The following groups were tested for differential methylation, normal B cells from peripheral blood and lymph node (n=7) versus DLBCL (n=23), Table 1.

This resulted in 13,982 DMRs that were hypermethylated in DLBCL compared to normal B cells and 4467 DMRs that were hypomethylated in DLBCL compared to normal B cells, Supplemental Table 2 (not provided in this dissertation).

Considering only DMRs in genes with an Ensembl gene ID, this corresponds to 8468 hypermethylated and 2127 hypomethylated DMRs respectively. These DMRs represent 4932 hypermethylated and 1461 hypomethylated unique genes. The human microarray study in high-grade B cell lymphoma found 525 genes with differential methylation ($p < 0.05$) (109). The restriction digestion study found 8759 hypermethylated and hypomethylated DMRs in the CLBL-1 cell line (CGI and non-CGI) (108). The canine microarray study found 1011 hypermethylated and 183 hypomethylated CpG regions (107). The MIRA-seq method thus provided the most comprehensive methylome compared to previous studies. MIRA-seq has been previously shown to be more sensitive than other methods such as MeDIP-seq, reduced representation bisulfite sequencing (RRBS) and the Infinium microarray (121).

The mean and median size of the hypermethylated DMRs were 424 bp and 298 bp (range 98-5398 bp) and of the hypomethylated DMRs were 275 bp and 198 bp (range 98-2898 bp) respectively, Figure 1. There was a peak in both hypermethylated and hypomethylated DMRs at ~100 bp which corresponds to the length of the fragment that was sequenced. There was a second much smaller peak ~150-200 bp which likely corresponds to the peak in CGI size.

The 5 most significantly methylated genes when the rpkkm values were compared between the groups were, chr29: 40704302-40704700 (intergenic), promoter-TSS of *PPP1R14C*, transcription termination site (TTS) of lincRNA *ENSCAFG00000037434*, intron 5/16 of *CSMD2*, and intron 4/5 of *ENSCAFG00000029547* or intron 31/34 of *COL6A1*. All of the above DMRs were either beside a gap in the Ensembl canine build or had low rpkkm values except for *PPP1R14C*. *PPP1R14C* or KEPI (kinase-enhanced phosphatase inhibitor) is an inhibitor of protein phosphatase 1 (*PP1*) (122). *PP1* is normally involved in signaling pathways for neuronal activity, protein synthesis, metabolism, muscle contraction, and cell division. Inhibition of *PP1* has been shown to contribute to chemoresistance in leukemic cells (123). Evaluation of RNA levels for *PPP1R14C* will be done with previous genome wide RNA-seq data to determine if promoter hypermethylation was correlated with gene silencing of *PPP1R14C*. If decreased expression of *PPP1R14C* is noted further studies to correlate this gene with chemosensitivity in canine lymphoma are warranted.

Figure 1: Size distribution of differentially methylated regions (DMRs) and CG islands (CGIs).

Most of the known lymphoma associated genes that had DMRs had both hypermethylated and hypomethylated DMRs, suggesting recurrent DNA methylation epigenetic changes to key genes

Five percent (210/ 3943) of the unique hypermethylated genes were represented in the COSMIC Cancer Gene database v89. Almost half (44.2%) of the genes were identified as TSGs. Notably, genes with a role in lymphoma, hDLBCL or NHL included *ARHGEF10L*, *CIITA*, *CSF3R*, *ETV6*, *FAS*, *FAT4*, *LEF1*, *MYC*, *SYK*, *CARD11*, *CD79B*, *EP300*, *IKZF1*, *PRDM1*, *TP63*, *BCL2*, *BCL7A*, *CCND2*, *FGFR1*, *NFKB2*, *PAX5* and *PIM1*. *BCL2*, *CCND2*, *FAS*, *NFKB2* and *PIM1* are also NF- κ B targets (124). *CARD11* is also recurrently mutated in ABC hDLBCL and is an important positive regulator of

NF-kb (124). Also 5.7% (64/1119) of the unique hypomethylated genes were represented in the COSMIC Cancer Gene database v89. Again almost half (46.9%) of the genes were identified as TSGs. Notably, genes with a role in lymphoma, hDLBCL or NHL included *FAS*, *FAT4*, *IKZF1*, *PRDM1*, *ARHGEF10L* and *FGFR10P*.

All of the genes in the hypomethylated set were also included in the hypermethylated set of COSMIC genes, *FAS*, *FAT4*, *IKZF1*, *PRDM1* and *ARHGEF10L*, with the exception of *FGFR10P*. In comparison, only 9.3% of the genes with DMRs had both hypermethylated and hypomethylated DMRs. There were 2 hypomethylated and 1 hypermethylated DMRs in *FAS*. There are 2 canine transcripts of *FAS* and the hypermethylated DMR was in the promoter-TSS of *ENSCAFT00000024890* and a hypomethylated DMR was in the promoter-TSS of *ENSCAFT00000024887*. The other hypomethylated DMR was located in intron 2/9 or 1/7. We can speculate that these methylation differences contribute to regulation of alternative transcripts. For *FAT4* there were 3 DMRs, 2 hypermethylated and 1 hypomethylated DMR. The hypermethylated DMRs were located in the promoter-TSS region. Recurrent mutations have been found in *FAT4* in hDLBCL, mantle cell lymphoma, splenic marginal zone lymphoma, natural killer/T cell lymphoma, large granular lymphoma (125-129). *FAT4* is a protocadherin in the Hippo signaling pathway and acts as a TSG. The hypomethylated DMR was located on exon 14/17 and had a low rpkm value even in the normal B cells and thus this DMR may not be significant. For *IKZF1* there were 4 hypomethylated DMRs and 1 hypermethylated DMR. All of these DMRs were in

introns. For *PRDM1* there was 1 hypermethylated and 1 hypomethylated DMR. Both DMRs were in introns. *PRDM1*, a zinc finger protein, encodes BLIMP1 (B lymphocyte-induced maturation protein 1) that is classically inactivated in ABCs compared to GCBs (130). Evaluation of the protein levels of BLIMP1 would be crucial to determine the significance of the DMRs in *PRDM1* given the relatively low fold change, 2.43-2.90, in both of these DMRs and the intronic location.

Pathways that were overrepresented in the hypermethylome included HOX, WNT and TBX pathways. Pathways that were overrepresented in both the hypermethylome and hypomethylome included neurologic and PI3K-AKT pathways.

KEGG pathways that were hypermethylated included neurologic and cardiac pathways and the PI3K-Akt and Wnt/Hippo signaling pathways, Supplementary Table 3 (not provided in this dissertation). GO terms that were hypermethylated again included neurologic, cardiac and Wnt pathways. Protein domains that were hypermethylated included homeobox, tyrosine kinase (e.g. *EGF*), Wnt, G-protein coupled receptor, immunoglobulin, voltage-gated calcium channels, transcription factor and neurotransmitter-gated ion channel domains.

The KEGG pathway that was hypomethylated was also the PI3K-Akt signaling pathway, Supplementary Table 4 (not provided in this dissertation). GO terms that were hypomethylated again included neurologic pathways, more specifically

pathways involved with microtubules (e.g. semaphorin, plexin, neuropilin, kinesin genes) and neurotransmitter activity. The specific neurologic GO terms that were enriched in the hypermethylated genes were different than for the hypomethylated genes. Protein domains that were hypomethylated included tyrosine kinase (e.g. EGF, ephrin), tyrosine phosphatase, semaphorin, FERM, G-protein coupled receptor, immunoglobulin and cadherin domains, Supplementary Table 5 (not provided in this dissertation).

The *HOX* and *NKX* genes and the homeobox domains were notably absent in the hypomethylated genes while 75.8% (25/33) and 92.3% (12/13) of the known *HOX* and *NKX* genes respectively in the dog (Ensembl) were hypermethylated, Supplementary Tables 6-8 (not provided in this dissertation). The *HOX* genes are transcription factors that are responsible for specifying body position in the developing embryo. They have been recurrently implicated in cancer, most commonly as an oncogene, but also as a TSG (131). As a TSG the *HOX* genes have been implicated in breast and lung carcinomas. In acute leukemias the *HOX* genes have been commonly implicated as an oncogene (131) but also less frequently as a TSG (132). Hypermethylation of the *HOX* genes was seen in clusters C, D and E in hDLBCL. Interestingly the *HOX* genes were not differentially methylated when all of the hDLBCL samples were compared to normal (133) and other B cell lymphomas such as follicular (134) and mantle cell lymphoma (135). A potential mechanism of *HOX* gene hypermethylation may be *EZH2* overexpression. This mechanism has been shown in hDLBCL (136) and *EZH2* overexpression has been shown in cDLBCL

(137). The functional consequence of hypermethylation of the *HOX* genes in lymphoma is largely unknown and should be further studied.

The Tbox (*TBX*) transcription factors and *EOMES* were also completely absent in the hypomethylated genes. The *TBX* genes are evolutionarily conserved transcription factors that function in organ development (differentiation, proliferation, tissue integrity and epithelial to mesenchymal transition). They can function as TSG or oncogenes. *TBX3* was silenced by hypermethylation in human glioblastoma, bladder and gastric cancer (138). *EOMES*, activates CD8+ cells and functions as a TSG by inducing anti-tumor immunity. *TBX5* acts as a TSG in colonic carcinoma. Of the known canine *TBX* genes, 58.8% (10/17) were hypermethylated and none were hypomethylated. As far as the author is aware, the *TBX* genes have been found to be hypermethylated in some hDLBCLs, clusters E and F (133) but have not been previously implicated as a key pathway in hDLBCL or cDLBCL.

The Wnt pathway regulates stemness and development and is commonly dysregulated in cancer (139). Seventy eight percent (15/19) of the known canine Wnt genes had hypermethylated DMRs and only 15.8% (3/19) had a hypomethylated DMR. Interestingly, the negative regulators of the canonical Wnt pathway, the secreted frizzled-related proteins (*SFRPs*), dickkopf proteins (*DKKs*) and *WIFI* were also hypermethylated and none were hypomethylated. Multiple components of the Wnt pathway with seemingly opposite functions were hypermethylated and thus the consequence of this hypermethylation cannot be

speculated on methylation data alone. Likely multiple levels of methylation-induced dysregulation occur in the Wnt pathway in cDLBCL.

Mutations in the *TBX*, *WNT* and *HOX* genes were not seen in a comprehensive analysis of the genome of 40 hDLBCL cases and thus dysregulation of these genes appears to be epigenetic (127).

cDLBCL has a distinct methylation signature from normal B and T cells and can be separated from normal cells based on the hypermethylated DMRs in the promoter-TSS regions alone

Two subgroups were seen in cDLBCL

All of the normal samples separated from the cDLBCL samples with supervised hierarchical clustering of the 1654 promoter-TSS DMRs, Figure 2. Two main subgroups were noted in the cDLBCL samples, group A and B, Table 2. The groups were almost at an equal frequency, 43.5% in group A and 56.5% in group B. The DMRs in subgroup B were more highly methylated than group A and thus group A more closely resembled normal than group B. Comparison of the MEDIPS QC metrics between the 2 subgroups, yielded significant differences in non-specific read pull-down $p=0.0031$ (% of reads that do not contain CG) and depth of CG coverage (% CG's with $>5x$ coverage) $p=0.0281$, Table 2. There were insignificant differences between the read count $p=0.17$, MEDIPS analyzed reads $p=0.34$, % true saturation $p=0.90$, % estimated saturation $p=0.47$, % relative enrichment $p=0.34$, %

observed/expected enrichment $p=0.44$, % CG's with 0% coverage $p=0.22$, Table 2.

The libraries in group A had more contaminating non-CG reads, a lower depth of CG coverage and the DMRs were less methylated than group B. This could be due to technical variability in producing or sequencing the libraries (batch effect) or to real differences in the methylomes of the group A and B samples. Technical variability could arise from reduced binding affinity of the methyl binding domain (MBD) proteins for methylated CG sites in the group A samples. Reduced binding affinity could have been due to a different lot of the MBD enzyme or magnetic beads purchased, age and condition of the enzyme and magnetic beads used, carry over of residual supernatant with unbound reads, and variability in the concentration of the salt in the washes.

Figure 2: Subgroups of canine DLBCL can be defined by promoter-TSS methylation signature. Heatmap of the DMRs from MIRA-seq of the 23 DLBCL and 7 normal B cell samples. The heatmap shows unsupervised hierarchical clustering of the 1654 hypermethylated promoter-TSS DMRs that were identified with a $q < 0.001$ (Benjamin-Hochberg correction). Log2 of the rpkm values of each DMR was used.

The 2 subgroups are likely due to a batch effect, since all of the libraries in group A were produced and sequenced at a later date compared to those in group B, with the exception of DLBCL9 and DLBCL12. The libraries in group A were produced with the NEBNext DNA Library Prep Master Mix Set that contained the Q5 Hot Start HiFi

PCR Master Mix that replaced the Phusion High-Fidelity 2X PCR Master mix in earlier kits. As a result the PCR amplification cycles also increased to produce libraries with similar concentrations. It is possible that the Q5 Hot Start polymerase and increased PCR amplification cycles led to preferential amplification of GC poor reads. Depletion of reads with GC contents <11% and >56% was seen with PCR (140). Modifications of the cycle conditions, that included a lower ramp speed, extending the initial denaturation step from 30 s to 3 mins and the denaturation step at each cycle from 10 s to 80 s, adding 2M betaine and using the Phusion polymerase preserved reads with a high-GC content. The Phusion polymerase resulted in better representation of high-GC loci than the AccuPrime Taq HiFi (140). Like the Q5 polymerase, the AccuPrime polymerase is also a hot start polymerase. Future studies should minimize batch effects by producing and sequencing libraries with the same lot of reagents and using increased input DNA to avoid PCR amplification before sequencing.

To confirm that the 2 subgroups were due to a batch effect from library preparation or sequencing, unsupervised hierarchical clustering was done on an equivalent number of DMRs, to the promoter-TSS DMR analysis, on a random set of hypermethylated or hypomethylated DMRs. The clustering was faithfully reproduced except for sample DLBCL9 that now correctly clustered with the libraries that were produced later, further evidence of a batch effect. We hypothesized that the difference in methylation between cDLBCL subgroups in Golden Retrievers is small and the batch effect seen in this study was larger than the

difference in methylation between cDLBCL subgroups (if subgroups are present).

Pathway analysis of the DMRs that separate group A from B samples is underway to determine if there is a plausible biological versus technical explanation for the difference.

Normal T cells compared to normal B cells

T cells and B cells have distinct methylation signatures that correlate with T-cell receptor (TCR) and B-cell receptor (BCR) signaling function

The following groups were tested for differential methylation, normal T cells (CD5+ PBMC, CD4+ lymph node and CD8+ lymph node) and normal CD21+ B cells (PBMC and lymph node). Supervised hierarchical clustering with all, genic only and promoter-TSS DMRs all resulted in perfect separation of T and B cells, Figure 3.

Pathway analysis with genic DMRs with over-representation analysis in ConsensusPathDB (minimum overlap with input list=2, $p < 0.01$) of DMRs that were hypermethylated in T cells compared to B cells had the following top 5 enriched KEGG pathways, BCR signaling pathway, phospholipase D signaling pathway, Fc gamma R-mediated phagocytosis, axon guidance and Type II diabetes mellitus, Supplementary Table 9 (not provided in this dissertation). The top cluster in DAVID contained the BCR signaling pathway, Fc epsilon RI signaling pathway and NK cell mediated cytotoxicity pathway with a cluster enrichment score of 2.73, Supplementary Table 10 (not provided in this dissertation). Hypermethylated genes included phenotypic markers used in IHC and flow cytometry, e.g. *BCR*, *CD19*,

CD22, *CD79B*, and *IGHM*. Conversely, pathway analysis of DMRs that were hypermethylated in B cells compared to T cells enriched for TCR, pathways in cancer, *EGFR1*, TH17 cell differentiation, Th1 and Th2 cell differentiation and IL2, Supplementary Table 11 (not provided in this dissertation). Hypermethylated genes included *CD5*, a phenotypic marker for T-cells used in IHC and flow cytometry. The results were also confirmed in DAVID, Supplementary Table 12 (not provided in this dissertation). Pathways in cancer and the TCR signaling pathway were amongst the top 5 enriched KEGG pathways in DAVID. DNA methylation differences in T and B cells thus can be defined by methylation signatures at the genic level alone.

3. Distribution of DMRs

Chromosome and gene feature distribution

DMRs were preferentially located in CGIs and the first exon

The DMRs were enriched in the 5'UTR, promoter-TSS, exon, 3'UTR and TTS and were least likely to be located in the intron and intergenic space, Tables 3 and 4. The hypermethylated DLBCL DMRs (all genic locations included) were distributed around the TSS, Figure 4, more specifically slightly downstream of the TSS. The median distance to the TSS for the hypermethylated DMRs was seen at 217 bp downstream of the TSS. The median, average and range for the length of the first exon (all canine genes from Ensembl) in *Canis familiaris* was 174 bp, 311 bp, 0-8007

bp. The distance to the TSS of the first exon is 82 bp (median), 145 bp (average), - 885 bp to 3899 bp (range) from the TSS. Thus the DMRs were mostly located in the first exon.

Figure 3: Normal T and B cells can be defined by DNA methylation signatures. Heatmap of the DMRs from MIRA-seq of the 8 normal T cell samples and 7 normal B cell samples. The heatmap shows supervised hierarchical clustering of the 167 promoter-TSS DMRs that were hypermethylated in T cells and the 90 DMRs that were hypermethylated in B cells ($q < 0.001$, Benjamin-Hochberg correction).

Table 3: Location of the hypermethylated DMRs in DLBCL compared to normal B cells as annotated by Homer

Annotation	Number of peaks	Total size (bp)	Log2 enrichment
5UTR	233	1,727,330	4.41
Promoter	1640	36,921,368	2.81
Exon	1077	33,686,211	2.33
TTS	559	35,401,096	1.32
3UTR	162	11,789,802	1.11
Intron	4782	861,185,853	-0.19
Intergenic	5529	1,221,715,378	-0.49

Table 4: Location of the hypomethylated DMRs in DLBCL compared to normal B cells as annotated by Homer

Annotation	Number of peaks	Total size (bp)	Log2 enrichment
5UTR	16	1,727,330	2.19
Exon	197	33,686,211	1.53
Promoter	150	36,921,368	1.00
3UTR	45	11,789,802	0.91
TTS	108	35,401,096	0.59
Intergenic	2340	1,221,715,378	-0.083
Intron	1611	861,185,853	-0.12

Figure 4: Hypermethylated DMRs follow the distribution of CGIs and are located around the transcription start site (TSS). For the hypermethylated DMRs, the median, mean and range is 217 bp, 882 bp, -535,391 bp to 560,181 respectively from TSS. For the hypomethylated DMRs, the median, mean and range is 1067 bp, 1614 bp, -28 bp to 556,107 respectively from TSS. For the CGIs, the median, mean and range is 230 bp, 3087 bp, -535,693 bp to 560,156 respectively from TSS.

This pattern is associated with the distribution of CGIs in the canine genome. There was a secondary preference of hypermethylated DMRs to be located slightly upstream of the TSS, Figure 4. The association of hypermethylated DMR distribution with CGI distribution is again shown when the distance to the nearest CGI of each DMR was plotted, Figure 5. In regards to the hypermethylated DLBCL

DMRs, 48.9% of the DMRs were in an island (CGI), 20.7% were in the shores (within 2000 bp of the CGI) and 30.3% were in the open sea. In comparison CGI's and shores represent only 1.69% and 8.27% respectively of the genomic space.

Figure 5: Hypermethylated DMRs are located in CGIs. Distribution of B vs. DLBCL DMRs in relation to the nearest CGI. CGI's in UCSC were defined as GC content $\geq 50\%$, length >200 bp, ratio of observed CG to expected CG >0.6 .

Hypomethylated DMRs retain this preference but are more likely to be distributed throughout the genome than hypermethylated DMRs

On the other hand, the hypomethylated DMRs are more evenly distributed throughout the genome than the hypermethylated DMRs but retain a preference for downstream of the TSS with a median distance of 1067 bp from the TSS, Figure 4 and Tables 3 and 4. In regards to the hypomethylated DLBCL DMRs, there remains a preferential distribution at the CGI but the DMRs are more likely to be dispersed throughout the genome as 15.1% of the DMRs were in the island, 25.2% were in the shores and 59.7% were in the open sea.

4. Comparison to other studies

Canine DNA methylation studies

A methylation signature, consisting of genes hypermethylated in cDLBCL, can be used to define cDLBCL indicating that the cDLBCL hypermethylation is stable and conserved between dog breeds and geography.

Upon intersection of DMR locations in this study compared to the 1,011 hypermethylated CpG regions and 183 hypomethylated regions identified in a microarray study in cDLBCL (non-breed specific in an Italian population of dogs) (107), 94.3% (953/1011) hypermethylated and 23.5% (43/183) hypomethylated DMRs found in the canine microarray study by Ferraresso *et al.* had a corresponding

DMR in this study. There was more correlation between studies when the genes versus the genomic coordinates were compared. When the genes that contained hypermethylated DMRs were compared between canine studies, 100% (451/451) of the genes in Ferraresso *et al.*'s study, 81.1% (146/180) of the genes in Epiphanio *et al.*'s restriction enzyme study and 40.8% (136/333) of the genes in Yamazaki *et al.*'s human microarray study were represented in the current study (Figure 6). None of the genes were common to all 4 studies. The design of our study was the most similar to the custom canine microarray study by Ferraresso *et al.* The Infinium study used a human microarray and used buffy coat cells as opposed to the primary tumor cells in lymph nodes used in the current study and in the canine microarray study. Our study had the least amount of overlap with the Infinium study. The DREAM study used a DLBCL cell line and used PBMCs from a normal dog as a comparator as opposed to lymph node cells in the canine microarray study and immunophenotypically sorted B cells in the current study. When the genes that contained hypomethylated DMRs were compared between canine studies, 22.4% (32/143) of the genes in Ferraresso *et al.*'s study, and 10% (18/180) of the genes in Epiphanio *et al.*'s study were represented in the current study, Figure 7. Less correlation in hypomethylated DMRs compared to hypermethylated DMRs was also seen in a previous study when the canine CLBCL-1 cell line was compared to other lymphoma cell lines (108).

Figure 6: Comparison of the hypermethylated genes identified with previous whole methylation canine studies. The Infinium study used the human HM850 microarray in buffy coats from dogs with high-grade B-cell lymphoma. The DREAM study used restriction endonuclease digestion in the CLBL-1 cell line. The microarray study used a canine microarray in whole lymph nodes of dogs with DLBCL.

Figure 7: Comparison of the hypomethylated genes identified with previous whole methylation canine studies. The Infinium study used the human HM850 microarray in buffy coats from dogs with high-grade B-cell lymphoma. The microarray study used a canine microarray in whole lymph nodes of dogs with DLBCL.

There was a low degree of conservation in the hypomethylome compared to the hypermethylome. The hypermethylome has been shown to be stable and distinct amongst different human tumor types (141). To the author's knowledge the hypomethylome has not been shown to differentiate cancer from normal or between cancer types. Hypomethylated DMRs may be less conserved than hypermethylated DMRs and may have a less significant functional consequence. Hypomethylated DMRs may be more dynamic across time, breed, region and lymphoma subtype. The mean fold change of the rpkm values of the hypermethylated and hypomethylated DMRs were 6.92 (range 2.00-324.11) and

2.52 (range 2069.43) respectively (unpaired two-tailed t-test, $p < 0.0001$). The lower fold change of hypomethylated DMRs compared to hypermethylated DMRs could contribute to the dynamic nature of the hypomethylome or alternatively would have required more power to find a statistical difference and would have been more subject to variability (e.g. technical, patient, environment, etc).

Genes involved in neurogenic and neuroendocrine tissues were recurrently overrepresented in cDLBCL methylation studies

Fourteen genes with a DMR in our study and 2 other canine studies were, *SHANK2*, *NEUROG3*, *KLF10*, *GSX2*, *TRIL*, *PTF1A*, *PAX2*, *KCNJ6*, *ATOH1*, *VAX1*, *RASGRF1*, *ONECUT1*, *CACNA1A*, and *EMX1*, Figure 6. None of these genes were in the COSMIC cancer database. All of these genes had a hypermethylated DMR in the promoter-TSS region which was not surprising since this was selected for in the Infinium and canine microarray study. Many of these genes are involved in neurologic pathways, including *ONECUT1*, *NEUROG3*, *RASGRF1*, *SHANK2*, *GSX2*, *CACNA1A*, *VAX1*, *PTF1A*, *ATOH1*, *PAX2*, *EMX1* and *KCNJ6*. Many of these genes are also involved with diabetes and other pancreatic or metabolic diseases, including *ONECUT1*, *RASGRF1*, *NEUROG3* and *PTF1A*. The overlap between the genes in the neurologic and pancreatic pathways are likely attributed to the neuroendocrine origin of beta-cells and that the methylome of DLBCL truly overlaps with just one of these pathways, likely neuronal, or less likely DLBCL overlaps with both of these pathways.

RASGRF1 encodes for a guanine nucleotide exchange factor, which activates the Ras GTPase by promoting the exchange of bound GDP for GTP. In the current study there was 1 hypermethylated DMR in the promoter-TSS and 1 hypomethylated DMR in intron 13/25. *RASGRF1* promoter hypermethylation has been associated with gastric cancer, normal mucosa adjacent to gastric tumors, pancreatic cancer and colorectal cancer (142-144). It is highly expressed in adult neurons in the CNS. The role of *RASGRF1* in tumorigenesis is unknown but it has been shown to be a TSG in gastric cancer perhaps by suppressing Cdc42-mediated cellular processes, such as tumor cell invasion and transformation (142). The promoter and first exon of *RASGRF2*, the other member of the family and also a TSG (145), was also hypermethylated and the last exon was hypomethylated.

ONECUT1 is a transcription factor that activates genes that are responsible for differentiation and development of ectodermal and endodermal tissues, such as *FOXA*, *PDX1*, *NEUROG3*, *HNF4*, *HNF1B* and *PROX1* (146). The *ONECUT* family of genes is highly conserved from *Drosophila* to humans and thus studies in model systems have contributed greatly to its understanding. *ONECUT1* has been shown to act like a TSG in the liver and pancreas. *ONECUT1* has been hypermethylated in many cancers including, chronic lymphocytic leukemia (147) and pancreatic (148), penile (149), cervical squamous cell (150) and hepatocellular carcinomas (151). There also was a hypermethylated DMR in each of the promoter-TSS-first exon regions of the other 2 members of the *ONECUT* family, *ONECUT2* and *ONECUT3*. In total there were 21 hypermethylated DMRs and no hypomethylated DMRs for the

ONECUT family. Hypermethylation of the *ONECUT* genes thus appears to be a deliberate versus a random event in cDLBCL. It is also interesting that the promoter-TSS of *NEUROG3*, a target of *ONECUT1*, and *PTRF1A*, which regulates *ONECUT1*, were also hypermethylated (152).

Canine whole exome study in Golden Retrievers

Genes that were recurrently mutated via DNA and DNA methylation changes in DLBCL in Golden Retrievers included SETD2, SOX17, WDR17 and DOK6

To identify genes that are recurrently altered in cDLBCL, the hypermethylated DLBCL regions were compared to genes that were found to be significantly mutated in an exome study of B-cell lymphomas in Golden Retrievers. Fourteen genes overlapped, *TP53, SOCS2, SETD2, MAP3K14, ENSCAFG00000030482, ERG, NTRK1, SOX17, ENSCAFG00000029913, PKD1, DOK6, ENSCAFG00000031200, ENSCAFG00000031827* and *WDR17*. This represents 19.7% of all of the mutated genes that were found in the Golden Retriever (95). There was a DMR in the promoter-TSS region for *SETD2, NTRK1, DOK6, WDR17, SOX17, ENSCAFG00000030482, ENSCAFG00000029913, ENSCAFG00000031200* and *ENSCAFG00000031827* (human ortholog, *ZNF519*). There were no human orthologs for *ENSCAFG00000030482, ENSCAFG00000029913, ENSCAFG00000031200* and *ENSCAFG00000031827* in UCSC. There was also a hypomethylated DMR for *DOK6, WDR17* and *NTRK1*.

SETD2 is the single gene that is responsible for adding the H3K36me3 mark in mammals (153). H3K36me3 is enriched in actively transcribed genes and has a role in genomic stability and RNA processing and initiation. *SETD2* functions as a TSG in cancers. *SETD2* is one of the most commonly mutated genes in human renal cell carcinoma and canine osteosarcoma (6, 153, 154). In lymphoma, *SETD2* mutations are the most common in enteropathy associated T cell lymphoma, with a prevalence of inactivating mutations in 32% of patients (155) but is uncommonly mutated in hDLBCL (COSMIC, accessed July 6 2019). There were also hypermethylated DMRs in the promoter-TSS of the following H3K26me3 readers, *DNMT3A* (7 hypermethylated DMRs) and *MORF4L1* which is important in RNA splicing. A ChIP study can be done to test the hypothesis that hypermethylation of *SETD2* leads to loss of H3K36me3 and RNA-seq can be done to evaluate changes in transcript processing. *WEE1* inhibitors can target *SETD2* negative tumors via synthetic lethality (156). An *in vitro* study of a cDLBCL cell line to evaluate the therapeutic potential of *WEE1* inhibitors is warranted.

SOX17 is a transcription factor that antagonizes Wnt signaling by promoting beta-catenin/TCF degradation (157). *SOX17* is a TSG in colonic adenocarcinoma, breast cancer (158) and endometrial cancer (159). *SOX17* was also hypermethylated in hDLCL (160) and in the current study.

WDR17 (WD repeat-containing protein 17) and *NTRK3* have been implicated as oncogenic drivers in human peripheral T cell lymphoma (161). *DOK6* is a member

of the *DOK* (downstream of kinase) family of intracellular adaptors that play a role in the *RET* signaling cascade. *DOK6* was lost via deletion in a significant number of patients with hDLBCL (162). *DOK6* has been shown to be a TSG in breast carcinoma (163) and gastric cancer (164).

Limitations and Future Directions

The methylation profiles could not be correlated with clinical outcome since the patients were not treated with the same drug protocol. It is recommended that future studies be done in patients enrolled in clinical trials since this allows for consistent staging, treatment and adequate follow up of patients. Approximately 16-32.7% of patients with aggressive LSA (mostly B cell) do not obtain a complete remission with a CHOP-based chemotherapy protocol (165-169). There is currently no test that can reliably identify these patients. Clinical correlations can answer the question of whether the methylome can identify these patients and if so is there an epigenetic mechanism that can explain chemoresistance.

Histology was not done on all of the patients and thus the diagnosis of cDLBCL could not be confirmed. In the current study, all dogs had frequent mitoses on cytology and medium to large-sized CD21+ B cells that were CD4-, CD5- and CD8- with variable MHC class II expression on flow cytometry. This flow cytometry profile has been consistent with histologically diagnosed cDLBCL in previous studies (170, 171) but this phenotype has not been directly compared to different types of B-cell

lymphoma. This profile also almost exclusively correlates with histologically diagnosed cDLBCL at the Clinical Immunology Lab at Colorado State University (personal communication). Since cDLBCL is the most common WHO histologic diagnosis in dogs with high-grade B-cell lymphomas, 93-96%, the dogs included in this study were most likely to have cDLBCL (172-174). Histopathologic evaluation by a diagnostic pathologist that specializes in lymphoproliferative diseases would be recommended for future studies.

RNA was not collected from most of the patients and thus gene expression cannot be correlated with methylation profiles. Absent or low methylation cannot be definitively attributed to hypomethylation since an alternative hypothesis includes structural variants such as deletions. Similarly high methylation cannot be definitively attributed to hypermethylation since amplifications cause a similar methylation profile and can also be a cause of copy number bias. Future experiments should also include whole genome, exome and RNA sequencing for a complete illustration of the genomic landscape.

MIRA-seq preferentially interrogates regions of DNA that are highly methylated and/or have high-CpG-density. Regions that are moderately or lowly methylated can only be inferred by lack of representative reads. Bias can also be introduced into enrichment-based experiments with variations in the efficiencies of fragmentation, bead purification and library construction (175). To minimize these biases input DNA for each MIRA reaction can be done.

Lastly, generalizations cannot be made to other breeds since only purebred Golden Retrievers were included in the cDLBCL cohort but the methylation profile in cDLBCL likely is conserved amongst different breeds due to the similarities of the results of this study to Ferraresso *et al.*'s study (107). Expansion of this study to include other breeds, T-cell lymphomas and other B-cell lymphomas (e.g. follicular, marginal zone and mantle cell lymphoma) are warranted. Results from these future studies could provide evidence that cDLBCL is similar across breeds, the methylome of aggressive T-cell lymphomas is distinct from cDLBCL and cDLBCL shows higher epigenetic variability and hypermethylation than less aggressive forms of lymphoma.

Although the hypermethylome of cDLBCL has the potential to be dynamic due to the reversibility of DNA methylation, it showed marked similarities among these patients. Future experiments analyzing the methylome over the course of chemotherapy, at remission, relapse and refractory statuses, could help answer the following questions: does the cDLBCL hypermethylome recede in remission or do certain hypermethylation marks persist? Does it change at relapse or in refractory disease? Can epigenetics explain acquired chemoresistance?

DNA methylation has been shown to be altered before DNA mutations develop in colonic adenocarcinoma and other cancers. Large scale profiling of a panel of robust cDLBCL methylation marks in lymph node aspirates of normal but at-risk dogs (e.g.

older Golden Retrievers) could help determine if DNA methylation changes also predate DNA mutations in cDLBCL. If normal individuals are found to carry hypermethylation marks, these dogs can be followed to determine if and when they develop lymphoma. Serial lymph node aspirates can also show the timeline of epimutations and mutations as normal lymphocytes evolve into lymphoma. A human study showed an increased risk of developing mature B-cell lymphomas (NHL, CLL, multiple myeloma) with global hypomethylation and promoter hypermethylation in *HOXA9* and *CDH1* in blood samples taken 0-10 years before diagnosis (176). Interestingly, there was no association between abnormal methylation marks and time to diagnosis suggesting that aberrant DNA methylation can be an early change in lymphomagenesis.

Conclusions

An epigenetic signature of cDLBCL was consistent in Golden Retrievers and appeared to be similar to that identified in different breeds in different countries. cDLBCL in Golden Retrievers was homogenous and subgroups were not seen. Genes that were recurrently mutated via DNA and DNA methylation changes in cDLBCL in Golden Retrievers included *SETD2*, *SOX17*, *WDR17* and *DOK6*. The methylome was characterized by hypermethylated DMRs in the promoter-TSS-first exon region and followed the distribution of CGIs. Genes with hypermethylated DMRs in the promoter-TSS-first exon often also had multiple hyper and hypomethylated DMRs distributed throughout the gene suggesting multiple layers of epimutations.

Epigenetically dysregulated pathways included neurogenic (e.g. *RASGRF1* and *ONECUT*), *HOX*, *WNT* and *TBX* pathways.

Acknowledgements

The authors would like to acknowledge Senthil Kumar, Hans Rindt and Maren Fleer at the University of Missouri for technical assistance.

CHAPTER 3:

Unraveling the chaotic genomic landscape of canine osteosarcoma with current sequencing technologies and bioinformatic approaches

Shirley Chu^a, Zachary L Skidmore^c, Jason Kunisaki^c, Jason R Walker^c, Malachi Griffith^c, Obi L Griffith^{bc}, and Jeffrey N Bryan^a

^aDepartment of Veterinary Medicine and Surgery, University of Missouri, 900 E. Campus Drive, Columbia, MO 65211, USA

^bDepartment of Medicine, Washington University, 660 S. Euclid Ave, St. Louis, MO, 63110, USA

^cMcDonnell Genome Institute, Washington University, 4444 Forest Park Ave, St. Louis, MO, 63108, USA

ABSTRACT

Background: Osteosarcoma (OSA) is an orphan disease in children but is one of the most common cancers in adult large and giant breed dogs. OSA tumors in both species characteristically have complex karyotypes. Other parallels that make dogs a good model for human OSA include the similarity of biologic behavior of the tumors, shared driver mutations and feasibility of performing clinical trials on dogs. Recurrent mutations most commonly are from structural variants (SV) versus single nucleotide variants (SNV). Whole genome sequencing (WGS) is thus crucial to identify structural variants. WGS, exome and RNA-seq were done on the primary and pulmonary metastatic lesion in two dogs with appendicular OSA. This study

represents the first WGS analysis of a metastatic canine OSA lesion and comprehensively illustrates the mutational landscape of somatic SNVs, SVs, copy number alterations, loss of heterozygosity (LOH), and gene fusions. The results from this study will illustrate the genetic similarities and differences between pediatric and canine OSA and thus the utility of dogs as a translational model and identify actionable driver pathways.

Results: Germline LOH was seen in both dogs but was more common in the Old English Sheepdog compared to the Labrador Retriever. A missense mutation in *TP53*, deletions in *CDKN2A/B*, *PTEN*, *DLG2*, *WRN* and *FAS* and amplifications in *MITF* and *FOXP1* were found in all samples. *NOTCH2*, *HSP90AB1*, *CDK4* and *CTNNB1* were highly expressed in all samples. Uniparental disomy in *TP53* was the most significant driver mutation. Chromosome 26 was the most affected by complex chromosomal rearrangements in both dogs.

Conclusions: Canine and human OSA are characterized by genomic instability and genomic heterogeneity, that are caused by loss of cell cycle checkpoint integrity and DNA damage response pathways. *TP53*, *RB1*, Wnt, NOTCH and PI3K/Akt pathways were commonly dysregulated and are candidate actionable driver pathways in canine and human OSA. Additionally *HSP90AB1* and *MITF* are candidate actionable driver oncogenes in canine OSA but are not commonly mutated in human OSA. Differences to human OSA included differences in the type of *TP53* mutations (exclusively SNV in dogs and SV are more common in humans), absence of dysregulation of a telomerase maintenance mechanism in canine OSA and

importance of *CDKN2A/B* deletion in canine OSA. Our data lend support that dogs are a good genetic model for human OSA.

Keywords

Osteosarcoma, canine, whole genome, exome, RNA-seq, metastasis

Introduction:

Osteosarcoma is a rare disease in humans, but it is the 8th most common cancer in children, and 30% of affected children succumb to their cancer within 5 years (177). On the other hand, it is estimated to occur up to 100 times more commonly in large/giant breed dogs, e.g. at a prevalence of 6.2% in Greyhounds (178), compared to an incidence of 5.4 cases/million per year in human males (179). Dogs are arguably the best animal model to study human OSA due to the many similarities (180-184). Both cancers are histologically similar and are treated with therapies that include surgery and platinum-based chemotherapy. Risk factors for both canine and human OSA are large body size (185) and radiation therapy exposure (186). OSA risk thus increases with increased cell divisions and DNA damage respectively. Chromosomal features such as aneuploidy, copy number and structural variations, and genomic instability are also common in both species. The most recurrently mutated gene in human OSA is *TP53*, 100% (187), and this is also true in canine OSA, 71-83% (6, 154). Common pathways determined to be OSA driver pathways are also dysregulated, such as the Wnt and PI3K/mTOR pathways (188).

At least 63 human OSA primary tumors and 16 metastatic tumors with paired normal samples have been sequenced with whole genome sequencing (WGS), whole exome sequencing (WES), and RNA-seq (St. Jude PeCan database). In comparison, 24 canine OSA primary and 0 canine OSA metastatic tumors have been sequenced with WGS, WES, and RNA-seq (6). Exome studies in canine OSA have been more

common, with 93 primary canine OSA and 10 metastatic canine OSA exomes sequenced to date (6, 154). Given the complex karyotype and heterogeneity of human and presumably canine OSA, sequencing of a large number of canine patients will be required to classify mutations into drivers and passengers.

WES is usually restricted to the protein coding regions, which is ~3% of the canine genome. WES would be sufficient for identifying CNVs and the small-scale mutations (e.g. indels, single nucleotide variants, SNVs) but would fail to capture the majority of the structural variants that make up the complex karyotype of OSA. This is the first study to perform combined WGS, WES, and RNA-seq on both the primary and metastatic lesion and thus provides the most in-depth genomic study of a canine OSA patient to date. This is also the first study to perform loss of heterozygosity (LOH) analysis in canine patients with OSA. The sequence data from the primary and metastatic lesions were used together to identify all of the mutations that were present in either or both lesions. Presence in both lesions served as internal validation that the mutation was less likely an artifact.

The purposes of this study were to (1) add to the growing genomic database in canine OSA; (2) adapt a comprehensive human cancer informatics pipeline for the canine species; (3) compare the primary and metastatic lesions to establish a sequence of events of tumor evolution; (4) evaluate the genetic similarities between pediatric and canine OSA and thus the utility of dogs as a translational model; and (5) identify actionable driver pathways.

Materials and methods

Samples

Tissues were obtained from the Animal Cancer Tissue Repository at Colorado State University from client-owned dogs that were diagnosed and treated for osteosarcoma with limb amputation, chemotherapy and pulmonary metastasectomy. Matched normal tissue for each case was obtained from normal skeletal muscle. Informed client consent was obtained for tissue banking. Tissues for DNA analysis were snap frozen and stored at -80°C. Tissues for RNA analysis were stored in RNAlater at 4°C.

Library preparation and sequencing

TruSeq stranded ribosome-depleted libraries were constructed from 500ng total RNA and single indexed. External RNA Controls Consortium (ERCC) spike-in was added to the RNA libraries (Invitrogen, 4456740). Exome capture utilized the Nimblegen Exome-Plus reagent (Nimblegen, 4000022470), covering ~152 Mb, as previously described (189). The whole genome and exome libraries from each sample were pooled and sequenced on one lane of the Illumina HiSeqX as 2 x 151 bp reads (eWGS). The RNA libraries were pooled and sequenced on 2 lanes of the HiSeq2500 1T platform as 2 x 125 bp reads.

Overview of bioinformatic analysis

The McDonnell Genome Institute's (MGI) cancer informatics pipelines, namely the Genome Modeling System, was adapted for the canine genome. Single nucleotide

variants (SNVs) were identified with SAMtools, VarScan2 and Strelka. Insertions and deletions (indels) were identified via GATK, Pindel, VarScan2 and Strelka. Variants were removed if they met the following criteria: tumor variant allele frequency (VAF) ≤ 2.5 , tumor variant reads ≤ 4 , tumor coverage ≤ 10 , normal VAF ≥ 10 and normal coverage ≤ 10 . Remaining variants were annotated with VEP using the canine reference canFam3.1 (Ensembl version 83) and retained if the observed consequence was predicted to result in a change in translation. Variants were also removed if they were found in the Dog Genome SNP Database. Differential expression analysis was performed via DESeq2 and p-value correction with the Benjamini-Hochberg procedure. Structural variants (SVs) were detected with Manta. Copy number alterations (CNAs) were detected with CopyCat. All coding variants were manually reviewed. All chromosomes were analyzed except the Y chromosome.

Variant Calling and filtering

Variants were identified via the somatic validation pipeline within the Genome Modeling System (190) for both samples in this cohort. Briefly, single nucleotide variants (snvs) were identified via SAMtools (r982), VarScan (2.3.6), and Strelka (1.0.11) with identical parameters for WGS and WES samples unless otherwise noted. Similarly, insertions and deletions were identified via GATK-Somatic-Indel (5336), Pindel (0.5), VarScan-Somatic, and StrelkaVariant read counts were obtained via bam-readcount (0.7; -b 20 -q 20) and variants were removed if they met the following criteria: tumor vaf ≤ 2.5 , tumor variant reads ≤ 4 , tumor coverage

≤ 10 , normal vaf ≥ 10 and normal coverage ≤ 10 . Remaining variants were annotated via VEP (2.7, --pick) using Ensembl (version 83) and retained if the observed consequence were any of the following: "coding_unknown", "essential_splice_site", "essential_splice_site, coding_unknown, intronic", "essential_splice_site, intronic", "essential_splice_site, within_non_coding_gene", "essential_splice_site, within_non_coding_gene", "frameshift_coding", "frameshift_coding, non_synonymous_coding", "frameshift_coding, splice_site", "frameshift_coding, splice_site, intronic", "non_synonymous_coding", "non_synonymous_coding, splice_site", "splice_site, intronic", "splice_site, intronic, within_non_coding_gene", "splice_site, synonymous_coding", "splice_site, within_non_coding_gene, within_non_coding_gene", "stop_gained", "stop_gained, frameshift_coding", "stop_gained, splice_site", "stop_lost", "within_mature_mirna". Variants were then removed if they were found in the Dog Genome SNP Database.

Additional ad-hoc filtering of SNVs included location on chr1-chr38 or chrX, ≥ 7 variant reads of support and $\geq 4\%$ VAF in the WGS, exome or RNA data for the tumor or metastatic lesion and ≤ 4 reads of support in the WGS or exome data for the normal sample.

RNA-seq analysis

Differential expression analysis was performed for tissue type (metastasis versus tumor) via the R package DEseq2 (1.14.1). Briefly Htseq counts were obtained on a

gene level and genes with a cumulative read count < 1 across all samples were removed. Differential expression of genes was calculated using a negative binomial distribution with a multi-factor design taking into account sample and tissue type. Adjusted p-values were calculated using the Benjamini-Hochberg procedure.

Structural variant and fusion detection

Structural variant (SV) detection was performed with Manta. Genes affected by SVs were annotated with biomaRt. Ensembl gene names were recorded for the entire region affected by deletion, duplication, or inversion and for the breakpoints (plus 10kb flanking base pairs) for translocations. Fusion prediction was performed with Pizzly. Gene fusions with ≥ 5 split or paired reads of support were considered for manual review using Svviz. Gene fusions failed manual review if either gene was in a scaffold, both genes belonged to the same gene family, the second gene was adjacent to the first gene (read through), there was no SV that could explain the gene fusion, or a BLAST of reads of one gene aligned to the other fusion partner.

Loss of heterozygosity (LOH)

Somatic LOH was defined by first identifying a set of germline SNPs (normal bam files) with a VAF between 0.4 and 0.6 (minimum 20x coverage). The VAF for these variants was then calculated from tumor samples and plotted. Segmentation analysis was performed to identify regions of somatic LOH. These regions (plus 10kb flanking base pairs) were annotated with biomaRt. Germline LOH was assessed and plotted alongside somatic LOH using common canine SNP sites as

regions for bam-readcount (normal bam files). The VAF was calculated from ref and variant allele counts for any SNPs with coverage $\geq 20x$. Regions of germline LOH represent blind spots for somatic LOH as they contain no heterozygosity to be lost.

Copy-number detection

Custom annotation scripts were used to estimate mapability and gc content for the canine (canFam3.1) reference genome as described in the copyCat documentation (<https://github.com/chrisamiller/copyCat>). Next, the GMS tool 'bam-windows' was run to estimate coverage across the genome (window size = 10,000). Gap and entry_point files were obtained from the UCSC genome browser. These inputs were used with the copyCat software (<https://github.com/chrisamiller/copyCat>) to estimate absolute CN values and identify copy-number segments and segment values. Genes were annotated to CN segments with biomaRt.

Results

Samples

Both patients in this study had the typical signalment expected of a dog with OSA, adult large breed dog, Table 1. The overall survival time of both dogs was ~1.5 years which is longer than the median survival time of ~10 months seen in dogs with OSA treated with standard of care therapy consisting of amputation and carboplatin, Table 2 (191). The longer survival in this study is likely due to the selection of dogs that had a pulmonary metastasectomy. Sequencing metrics are presented in Table 3. The tumor purity was too low in the metastatic lesion of the Lab to make conclusions about the mutations in this sample.

Table 1: Signalment and diagnostic information for the dogs that were sequenced in this study

Breed	Age (yrs)	Sex	Staging	Tumor Location	Histopathologic description
Labrador Retriever	8	MC	Thoracic and limb radiographs, Technetium scan, minimum database, RLN histopathology at surgery	Distal radius	Osteoblastic. MI=not given but many mitoses were noted
Old English Sheepdog	7	MI	Thoracic and limb radiographs, minimum database, RLN histopathology at surgery	Distal tibia	Osteoblastic. MI=20-30

TTP: time to progression; OST: overall survival time; RLN: regional lymph node; MI=mitotic index

Table 2: Treatment and outcome for the dogs that were sequenced in this study

Patient	Treatment	TTP (days)	Reason for failure	Further treatment	OST (days)
Labrador Retriever	Amputation, 6 total doses of chemotherapy with alternating doxorubicin with valproic acid and carboplatin	291	Lung metastasis	Pulmonary metastasectomy	548
Old English Sheepdog	Amputation, 6 total doses of chemotherapy with alternating doxorubicin and carboplatin	429	Lung and kidney metastasis	Pulmonary metastasectomy and nephrectomy	569

TTP: time to progression; OST: overall survival time

Table 3: Sequencing coverage and tumor purity for each sample

Patient	WGS	WES	RNA (x10 ⁶ reads)	Estimated Tumor Purity ^a
Labrador Retriever	Normal: 22x Primary: 34x Met: 32x	Normal: 100x Primary: 179x Met: 130x	Primary: 168 Met: 357	Primary: >80% Met: >11%
Old English Sheepdog	Normal: 26x Primary: 26x Met: 33x	Normal: 110x Primary: 147x Met: 131x	Primary: 208 Met: 260	Primary: >85% Met: >64%

WGS: whole genome sequencing; WES: whole exome sequencing

^aTumor purity was estimated by the variant allele frequency of *TP53* in each lesion.

SINGLE NUCLEOTIDE VARIANTS (SNVs)

Figure 1: Filtering algorithm for single nucleotide variants (SNVs)

Thirty-seven and 28 SNVs passed manual review in the Lab and Sheepdog sample respectively, Figure 1. The number of mutations in the primary tumor resulting in amino acid changes in this study was 34 (Lab) and 17 (Sheepdog). Only 1 gene was common in both dogs, *TP53*, Figure 2. The variant allele frequency (VAF) for *TP53* was 0%, 76.7-87.5% and 64.7-89.1% in the WGS/exome/RNA of the matched normal, primary and metastatic lesion respectively in the Sheepdog. The variant

allele frequency (VAF) for *TP53* was 0%, 79.1-94.4% and 11.6-33.3% in the WGS/exome/RNA of the matched normal, primary and metastatic lesion respectively in the Lab. The VAF was high in all tumor samples except the metastatic sample in the Lab, indicating that the tumor purity and the prevalence of the mutation in the tumor cell population was high in all but one sample. Both SNVs in *TP53* were predicted to have a moderate impact on the protein product.

Figure 2. 88.6% of the mutations in *TP53* (ENSCAFT00000026465.3) in canine OSA occurred in the DNA binding domain.

Genes with SNVs that were predicted to have a high impact on function included, *TNN*, *ARHGAP12*, *SNX9*, *SMARCA5* and *ATP8B4* in the Lab and *GZMK*, *ABCG4* and *DNAH14* in the Sheepdog. The corresponding SNV in the human ortholog was not found in the COSMIC database for any cancers (accessed 9-Jul-2019). *SMARCA5* was identified as a potential oncogenic driver in a mouse model of OSA (192). *SNX9*, a scaffold involved in actin assembling and clathrin-mediated endocytosis, has been implicated in mitosis, invasion and metastasis (193).

Genes in the COSMIC cancer database with a SNV included, *TP53* and *MYO5A* in the Lab and *TP53*, *NOTCH2*, *ZNF521* and *PICALM* in the Sheepdog. All of these SNVs were predicted to have a moderate impact on function. The *PICALM* SNV in the Sheepdog was a non-synonymous mutation in exon 9/21 or 10/20. A *PICALM-DLG2* fusion was also detected in the primary and metastatic lesion of the Sheepdog. The gene fusion resulted in deletion of *PICALM* with ~0.5-1.5 copies per cell (discussed in more detail as a gene fusion below). A non-synonymous SNV was found in exon 8 (of 35 or 34) of *NOTCH2*; this mutation has not been reported in the COSMIC database (accessed 9-Jul-2019) but has been reported in a metastatic canine OSA lesion (6). This mutation did not lead to decreased gene expression compared to the Lab. *NOTCH2* was also amplified in the Sheepdog with the *NOTCH2* SNV with ~3 copies in the primary and metastatic lesion. *NOTCH2* was also a highly expressed gene in all samples (lowest expression in the metastatic lesion from the Lab that also had low tumor purity). Interestingly, there were no mutations in *NOTCH1*. The *ZNF521* SNV was in exon 4/8 or 3/6 and there were no associated CNAs in this gene.

Low tumor purity in the metastatic lesion of the Lab precluded comparisons of the SNV landscape between the primary and metastatic lesion for this patient. In the Sheepdog there was strong correlation between the SNVs in the primary and metastatic lesions, Figure 3. Seven percent (2/28) of the SNVs were in the primary lesion only and 39.3% (11/28) were in the metastatic lesion only. Metastasis-specific SNVs were found in *ABCB9*, *ABCG4*, *CORO6*, *IRGD*, *GZMK*, *NOTCH2*, *OR2T35*, *TMEM212*, *ZNF529*, *LOC102152223* and *ENSCAFG00000029779*. The presence of

more SNVs in the metastatic versus primary lesion suggests either the development of further mutation, which could be random, at the metastatic site. Alternatively a higher burden of SNVs in the metastatic lesion may be due to selection and clonal expansion of rare mutants with increased metastatic potential at the primary site. Deeper sequencing of the primary tumor and functional studies of the metastasis-specific genes would be required to determine if the latter is true.

Figure 3: Strong correlation between SNVs seen in the primary compared to the metastatic lesion in the Sheepdog. VAF, variant allele frequency.

Kataegis was not seen

Kataegis is a phenomenon characterized by localized hypermutation with C→T and C→G substitutions at TpCpX sites. The mutations usually cluster over several hundred base pairs and alternate between stretches of C→T mutations and

stretches of G→A mutations. Kataegis likely occurs during repair of a breakpoint during one cycle of replication on one strand of DNA. Kataegis was not seen in the samples in this study but was seen in 50% of pediatric OSA, Figure 4 (187).

Figure 4. Absence of kataegis in canine OSA. Rainfall plot showing the intermutation distance versus genomic position for validated SNVs.

CHROMOSOMAL LESIONS

A comparison of the genes that were mutated via a SNV or SV in each dog is shown in Figure 5. Translocations, amplifications, deletion and inversions were seen in both dogs. A summary of the mutations identified in both dogs can be seen in Figures 6a and 6b. The Labrador had a more normal karyotype than the Sheepdog. Deletions were the most common SV. The pattern of SVs was complex and suggests

chromoanagenesis, or chromosomal regeneration. Chromoanagenesis can produce a vast number of mutations in a single event and can lead to rapid tumor evolution. In the current study more specifically, chromothripsis or chromosome shattering resulting from a single event was suggested by the genes that were reassembled in random order and orientation on chr26 with resulting loss of chromosomal segments. The 2 replication based processes, chromoanasythesis and chromoplexy were suggested by the clusters of gene amplifications and deletions that occurred and the chain of translocations with preserved orientation with frequent deletions at the breakpoints respectively.

Figure 5. Genes were more commonly mutated by SVs versus SNVs in canine OSA. Venn diagram depicting the proportion of (a) SNVs and (b) SVs that were shared between dogs and between the primary and metastatic lesion

Figures 6a and b. Plots depicting SNVs, CNAs, LOH and SVs in coding regions. Gene labels are shown for all SNVs and selected genes with other types of mutations. Genes in bold indicate genes that were common to both patients. Structural variants are depicted in the innermost circle.

a. Sheepdog

b. Lab

Copy number alternations (CNAs) and loss of heterozygosity (LOH) were common

In both dogs, segments of chromosomes were not heterozygous in the germline. This was expected on chrX since both dogs were male. On the autosomes this was attributed to loci that have limited variability within a breed or individual due to inbreeding. This was especially evident in the Sheepdog where LOH could thus not be interpreted in all sections of the chromosomes, Figure 7b.

Figures 7a and b: Chromosomal segments were not heterozygous in the germline. The selected chromosome, chr14, was also less affected by chromosomal lesions than other chromosomes (normal in the Sheepdog). Representative CNA (top panel), somatic LOH (middle panel) and germline LOH (bottom panel) plots from the primary lesions in the Lab and Sheepdog for chr14.

(b) Sheepdog

Low tumor purity in the metastatic lesion of the Lab precluded CNA and LOH analysis of this sample. CNA and LOH analysis was interpreted for the 3 other samples. Every single chromosome was affected by SVs except chr14 in the Sheepdog, Figure 7b. The Lab had a more normal karyotype than the Sheepdog (Figure 6). Chr1, 3, 6-10, 13, 15, 17, 24 and 31 were least affected by CNA and LOH in both patients (Figure 8). LOH patterns were shared in both patients on chr5, chr12, chr16, chr20, chr26, chr37 and chr38. LOH patterns were very similar in the primary and metastatic lesion of the Sheepdog with the exception of an additional large deletion and resulting LOH on chr2 and gain of LOH without a CNV, likely via whole chromosomal loss and gain at mitosis, in the metastatic lesion.

Figure 8: Summary of loss of heterozygosity (LOH) for both patients. The top 2 rows are from the Lab (metastatic and primary lesion) and the bottom 2 rows are from the Sheepdog (metastatic and primary lesion). Areas with LOH (high VAF or variant allele frequency difference) are represented in a lighter color.

Copy-number neutral LOH (acquired uniparental disomy, aUPD) was seen in *TP53* in all samples, Figure 9. A duplication event in the *TP53* loci occurred after the missense mutation in *TP53*. The segmental aUPD likely occurred due to mitotic recombination or duplication following unsuccessful repair of a dsDNA break (194). Along with the *TP53* missense mutation, deletion of the wild type copy and amplification of the mutant copy, the proposed second and third hits, functionally makes both patients *TP53*^{mut/mut}.

Figures 9a and b: Segmental copy number neutral loss of heterozygosity (CN-LOH) in the middle of chr5 in both dogs. Representative CNA (top panel), somatic LOH (middle panel) and germline LOH (bottom panel) plots from the primary lesions in the Lab and Sheepdog for chr5.

(a) Lab

(b) Sheepdog

PI3K/Akt pathway was commonly dysregulated

PTEN was located on the chromosome with the most chromosomal lesions, chr26.

Deletion was noted in all samples, with heterozygous in the primary tumor from the Lab and homozygous in the primary and metastatic lesion of the Sheepdog, Figure 10b. CNAs were not detected in *MTOR* or *AKT1* but an *AKT2* amplification was seen in the primary lesion in the Sheepdog.

Figures 10a and b. Chr26 was the most affected by structural variants (SVs). (a) Representative CNA (top panel), somatic LOH (middle panel) and germline LOH (bottom panel) plots from the primary lesion in the Lab for chr26. (b) Close up view of CNAs for *PTEN* in the Sheepdog primary lesion (top panel), Sheepdog metastatic lesion (middle panel) and Lab primary lesion (bottom panel).

RB1 pathway was commonly dysregulated

LOH and focal deletion of *CDKN2A/B* was seen in both dogs resulting in ~ 0.5 copy per cell, Figure 11. The locus in the Sheepdog that was deleted included *CDKN2A*, *CDKN2B* and a lincRNA, *ENSCAFG00000032828*. The gene-coding region in this deleted locus in the Lab was larger and also included *MTAP* and another lincRNA,

ENSCAFG00000034922. Amplification or deletion was not noted in *CDK4*, which is inhibited by *CDKN2A/B*, but was highly expressed in all samples, especially the metastatic lesion in the Sheepdog. *RB1* was deleted in the Labrador via a chromosomal deletion and consequently *RB1* expression was lower in the Labrador.

Figure 11. A focal deletion was noted in the *CDKN2A/B* locus, chr11:41,223,260-41,246,135 overlapped in both dogs. Representative CNA (top panel), somatic LOH (middle panel) and germline LOH (bottom panel) plots from the primary lesions in the Sheepdog and Lab for chr11. The plots for the Sheepdog metastatic lesion are almost identical to the primary lesion and are not shown.

Somatic deletions of genes that predispose patients to hereditary cancer risk syndromes were common

Germline losses of *TP53*, *WRN*, *RB1*, *RECQL4* and *BLM* are risk factors for hereditary OSA. Somatic loss of *TP53* (via SNV) and *WRN* (via deletion) were seen in all samples in this study. There were ~1-1.5 copies of *WRN* per cell with LOH. Somatic deletion of *RB1* was seen in the Lab only.

NF2 was deleted in both primary tumors with ~1-1.5 copies per cell but not in the metastatic lesion. Loss of the germline *NF2* leads to neurofibromatosis type 2, where multiple tumors form in the nervous system. The Sheepdog primary lesion also had an *NFAT5-NF2* translocation which would be predicted to functionally delete *NF2*.

Dysregulation of a telomerase maintenance mechanism (TMM) was not seen

Deletion of *TERT*, the enzymatic subunit of telomerase, was seen in the primary and metastatic sample of the Sheepdog. *ATRX* gain was seen in the Lab primary and Sheepdog metastatic sample and moderate expression was seen in all samples. *ATRX* and *DAXX* make up a chromatin remodeling protein that suppresses alternative lengthening of telomeres (ALT). The changes seen in this study are thus the opposite of what would be expected to give cancer cells unlimited growth potential.

Amplified and highly expressed genes

Upregulation of HSP90AB1 provided tumor cells a survival advantage

Heat shock protein, *HSP90AB1*, was one of the most highly expressed genes and was amplified in the Sheepdog primary and metastatic lesion with ~7-9 copies per cell. Expression was higher in the Sheepdog than the Lab and higher in the metastatic lesion than the primary lesion.

Wnt signaling was upregulated

Beta-catenin (*CTNNB1*) was another one of the most highly expressed genes and was amplified in the Sheepdog primary and metastatic lesion with ~3-4 copies per cell. *DKK3* a negative regulator of the Wnt pathway was deleted with ~1.3 copies per cell, in the Sheepdog metastatic lesion and Lab primary lesion. *DKK3* expression was also low in all samples.

Oncogenic pathways in human melanoma were active in canine OSA

MITF was moderately expressed in all samples and also was amplified in all samples with ~4 copies per cell. In this chromosomal segment, *FOXP1* and *EIF4E3* were also similarly amplified. In humans, *MITF*, *FOXP1* and *EIF4E3* are also on the same chromosomal segment.

Complex chromosomal rearrangements were seen in both dogs

Chr26 was the most severely affected and most consistently affected chromosome in both dogs. None of the amplified genes on chr26 were in the COSMIC database.

COSMIC TSG that were deleted and lowly to negligibly expressed on chr26 include *PTEN*, *FAS*, *SEPT5* and *LZTR1* in all lesions examinable and *CHEK2*, *ZNRF3*, *NF2* and *CLTCL1* in both primary lesions with lowly to negligible expression in all lesions. These TSG were clustered on 3 chromosomal segments. The first segment included *CHEK2*, *ZNRF3* and *NF2*. The second segment was broken up into 3 separate segments in the metastatic lesion in the Sheepdog but was in one segment in the primary lesions and included *SEPT5*, *CLTCL1* and *LZTR1*. The third segment included *PTEN* and *FAS*. The distances between the genes on each segment are conserved in humans. The major difference in dogs compared to humans is that the segment containing *PTEN* and *FAS* is not on the same chromosome as the first 2 segments in humans (chr10 and chr22).

In addition to chr26, chr12 and chr20 were the most common chromosomes to have SVs in both dogs. Differences in SV distribution between the dogs included, a predilection for SVs in chrX in the Lab and chr4, 12, 16, 17, 25, 30, 34 and 36 in the Sheepdog. Although similar genes were mutated via SVs in both dogs, the translocation partner for these genes were different in the two dogs. Translocation with chr4 was the most common in the Sheepdog but there were no translocations with chr4 in the Lab. This suggests that chr26 was shattered but was reassembled in a random way.

The complex pattern of chromosomal rearrangements can be explained by chromothripsis or sequential chromosomal lesions. Chromothripsis or chromosome

shattering, results from a single event in which numerous dsDNA breaks occur in one or a few chromosomes. The broken pieces are reassembled in random order and orientation via non-homologous end joining. Segments of chromosomes are lost. It is estimated to occur in 20-25% human OSAs (187, 195). Definitive evidence of chromothripsis can be elusive but would require molecular cytogenetics and analysis of the SVs in intergenic regions (196).

Expressed gene fusions were uncommon

490 and 377 potential expressed gene fusions were detected in the Lab and Sheepdog respectively. Gene fusions were further filtered with the requirement that each gene fusion have ≥ 5 reads of support (split and paired reads covering the breakpoint). Manual review of the resulting 21 and 17 potential gene fusions resulted in 1 and 3 validated gene fusions in the Lab and Sheepdog respectively. The gene fusions were as follows, *GRK3-HPS4* in the Lab and *ITGB6-CAMK2D*, *EIF4E3-MITF* and *PICALM-DLG2* in the Sheepdog. The *GRK3-HPS4* fusion was via an inversion on chr26 (Figure 12a). The *GRK3-HPS4* gene fusion was present in the primary and metastatic lesion of the Lab and the fusion protein was lowly expressed. This fusion leads to rearrangement of exons 7-20 of *GRK3* and likely results in loss of function.

The *PICALM-DLG2* fusion was seen in both the primary and metastatic lesion. The *PICALM-DLG2* fusion was via a deletion on chr21 (Figure 12b). The deletion resulted in loss of *CCDC83*, *SYTL2*, *CCDC89*, *CREBZF*, *TMEM126A* and *TMEM126B* and

~0.5 copy of these genes in the affected cells. These deleted genes were not expressed except for low expression of *TMEM126A* and *TMEM126B*. The fusion product was lowly expressed and thus the biological significance of this gene fusion is likely loss of function of the deleted genes and likely the fusion partners. This patient also had both a SNV in *PICALM*. Both dogs had a deletion in *DLG2*.

Figure 12. Few SVs resulted in expressed gene fusions. (a) *GRK3-HPS4* was caused by an inversion in the Lab primary and metastatic lesion. (b) *PICALM-DLG2* was caused by a deletion in the Sheepdog primary and metastatic lesion.

Discussion

Similar with human OSA, SNVs were not common

The number of mutations in the primary tumor resulting in amino acid changes in this study was 17 (Sheepdog) and 34 (Lab). This compares to a similar study in pediatric OSA, where 25.2 mutations were seen per case (range 5–103) (187).

SNVs were seen in the COSMIC cancer associated genes, *TP53*, *PICALM*, *NOTCH2* and *ZNF521*. *PICALM*, phosphatidylinositol binding clathrin assembly protein, is ubiquitously expressed and has a role in endocytosis or cycling of membrane proteins, autophagy and iron and cholesterol homeostasis (197-199). *PICALM* has been implicated in NSCLC, T-ALL and AML, where a translocation results in fusion with *MLLT10/AF10*, *ALK* or *KMT2A*. The oncogenic mechanism in leukemia may be impairment of differentiation of hematopoietic cells by upregulation of the *HOXA* cluster genes. In this study the deletion resulted in an expressed *PICALM-DLG2* fusion. The SNV in *PICALM* has been seen in an endometroid carcinoma (TCGA-AP-A0LM-01; COSS1783352). *NOTCH* receptor activation in osteoblasts inhibits differentiation but is also crucial for bone development and homeostasis. *NOTCH2* also specifically enhances osteoclastogenesis (200). Activation of *NOTCH1* and *NOTCH2* overexpression has been found in human OSA and *NOTCH1* can induce OSA in murine models

(201, Zanotti, 2016 #34). *NOTCH2* SNV mutations have been seen in many cancers, including DLBCL, bladder cancer and marginal zone lymphoma, where it can function as an oncogene or a tumor suppressor gene (COSMIC, accessed 9-Jul-2019). But only one *NOTCH2* mutation has been reported in human OSA and only one in canine OSA as far as the author is aware (6). The SNV in the human case was in an acceptor splice site (5' exon 34/34) in a metastatic OSA lesion, COSM6959094 (187). Interestingly, similar with this study the c.1366C>A SNV in the canine case was also in a metastatic lesion. Functional studies to determine the effect of this SNV in canine OSA cell lines are warranted. *ZNF521*, a transcription co-factor, is involved in the control of progenitor cells, including osteoid progenitors. Overexpression has been tumorigenic in medulloblastoma cell lines and a translocation results in a fusion with *PAX5* in ALL (202, 203). A missense mutation was found in a pediatric patient with OSA (187). The canine OSA SNV in *ZNF521* was not found in the COSMIC database.

In a previous canine OSA exome study, *SETD2* was the second most frequently altered gene with a frequency of SNVs of 21% and 0% CNAs (154). *SETD2* point mutations were not found in our study. Nine additional significantly mutated genes were identified in this previous study, *TANGO2*, *PRORS1P*, *MAGEA9*, *LOXHD1*, *HIST1H2AH*, *F6XVW9*, *OR4D5*, *MYTIL* and *RPL18*. SNVs were not seen in these genes in the current study or in a recently published study (6). Deletion in *TANGO2* and *F6XVW9* was seen in one dog in the current study. Amplification but with almost no expression in *MAGEA9* and *OR4D5* was seen in one dog in the current study.

Amplification with high expression in *RPL18* was seen in one dog in the current study. Mutations in these genes are likely passenger mutations or common in only a subset of canine OSA.

Similar with human OSA, TP53 was the most commonly mutated gene

Missense mutations in *TP53* are the most common genetic alteration in human cancers, accounting for >75% of mutations in *TP53* (204). The DNA-binding domain of *TP53* is the most common location for a SNV in canine and human OSA (205-208). The SNV in the Lab was described before (6) and the SNV in the Sheepdog has not been described before. Both SNVs were in exon 6, in the DNA-binding domain. Mutations in the DNA-binding domain commonly compromise the ability of this transcription factor to transactivate downstream genes that arrest the cell cycle, repair damaged DNA and induce apoptosis. Mutations thus lead to genomic and chromosomal instability. A strong association between loss of *TP53* and OSA has been shown in previous studies. Loss of *TP53* in humans is known to be a risk factor for OSA (209). OSA is the most common sarcoma in patients with germline mutations in *TP53* and Li-Fraumeni syndrome (204). *TP53* has a protective role in mesenchymal stem cells against developing osteosarcoma (210). Mutations in *TP53* have been seen in 100% of human OSA patients with WGS, most commonly via translocation, especially in intron 1 (187). Human sporadic OSA is the only tumor that has been found to have a translocation in intron 1 of *TP53* despite *TP53* being the most commonly mutated gene in all cancers (211). *TP53* mutations were found in 40.4% of appendicular and 45.5% of canine axial OSA (208). In the current study

SVs were not seen in *TP53* but 100% of patients had SNVs in the DNA-binding domain of *TP53* with LOH.

MUTATIONS OCCURRED MORE COMMONLY VIA CHROMOSOMAL LESIONS THAN BY SINGLE NUCLEOTIDE VARIANTS (SNVs)

PI3K/Akt and cell cycle checkpoint pathways were dysregulated as a result of chromosomal mutations in canine and human OSA

PTEN is an inhibitor of the PI3K/Akt pathway that is crucial for cell growth and survival especially in times of stress. Loss of *PTEN* results in activation of *AKT* and *MTOR* pathways. *PTEN* mutations are recurrently mutated in 56% human OSA (188, 212). *PTEN* deletions in both dogs likely led to permanent activation of the PI3K/Akt pathway (213). Tumor responses to an mTOR inhibitor and pan-AKT inhibitor have been seen in a PDX mouse model with either a *PTEN* loss or *AKT1* gain (212). Ridaforolimus maintenance was shown to minimally but significantly increase progression free survival (PFS) in adults with metastatic osteosarcomas (214). An NIH clinical trial with rapamycin therapy following standard of care therapy consisting of amputation and chemotherapy has just been completed in canine OSA and the results are pending.

Deletion in *CDKN2A/B* was seen in all samples and deletion in *RB1* was seen in the Lab. Homozygous deletion of *CDKN2A/B* in mesenchymal stem cells was an early

event in human OSA (215). In previous studies, deletions in *CDKN2A/B* were seen in 72% of dogs and in at least 20% of nuclei of all human OSA patients tested ((154, 216). Deletions in *CDKN2A/B* are also common in canine histiocytic sarcoma, 62.8%, T-cell lymphoma, 55.6%, and canine fibrosarcoma, 100% (217-219). *CDK4* along with *CDK6* are responsible for the G1 to S phase cell cycle transition. Loss of INK4 inhibitors such as *CDKN2B* can further lead to overactivity of *CDK4*. *CDK4* phosphorylates pRB and leads to its inactivation. In the dogs in this study, *CDK4* overexpression was likely a consequence of *CDKN2B* deletion. *CDK4/6* inhibitors have been approved for estrogen receptor positive, *HER2*- metastatic breast cancers with concurrent anti-estrogen therapy. Clinical trials for *KRAS*-mutant non-small-cell lung carcinoma (NSCLC), mantle cell lymphoma with overexpression of *CCND1* via translocation and liposarcoma with *CDK4* amplification amongst other cancers are underway. Selective *CDK4/6* inhibitors may be more effective in the Sheepdog compared to the Lab due to the presence of 2 copies of wild type *RB1*. Germline variants in the *CDKN2A/B* loci have been previously associated with risk of developing OSA in 3 high risk dog breeds (Irish Wolfhounds, Rottweilers and Greyhounds) (220) and also of developing histiocytic sarcoma in Bernese Mountain dogs (221). The germline SNV may predispose the locus to a chromosomal lesion or the germline SNV and *CDKN2A/B* deletion may be unrelated events. Regardless, *CDKN2A/B* is a major driver in canine OSA.

Somatic deletions were seen in genes that are commonly mutated in human patients with hereditary OSA, TP53, WRN and RB1

WRN was deleted in all samples. Loss of germline *WRN* results in Werner syndrome, which is characterized by premature aging. *WRN* encodes a DNA helicase in the RECQ family that activates the ATR CHEK1-induced S-phase checkpoint in response to ssDNA breaks and repair via homologous recombination versus NHEJ (222). Loss of *WRN* thus leads to NHEJ mediated erroneous joining of chromosomal segments resulting in chromosomal instability and deletions (223).

NF2 was deleted in both primary tumors. *NF2*, neurofibromatosis 2, produces the gene product Merlin, a moesin-ezrin-radixin-like protein. Merlin halts *EGFR* signaling in the RAS/MEK pathway when cells are in contact. *NF2* also inhibits the Wnt signaling pathway (224). *NF2* is known as a TSG in the COSMIC cancer database and has been best described as a TSG in hereditary cancers. Loss of somatic *NF2* occurs in spontaneous schwannomas, meningiomas, mesothelioma, glioblastoma, breast, colorectal, skin, clear renal cell, prostatic and hepatic carcinomas. Mice with heterozygous loss of *NF2* are prone to develop cancers, especially OSA, fibrosarcoma and hepatocellular carcinoma. The role of *NF2* as a driver specifically in spontaneous OSA has been controversial with studies supporting its role (192) and some arguing against its role (225). Nevertheless, loss of *NF2* was a common event in the primary lesion of canine OSA either as a driver or passenger.

Unlike human OSA, evidence of alternative lengthening of telomeres (ALT) was not seen in canine OSA

Unlike 85% and 95% of other human and canine cancers respectively that overexpress *TERT*, human OSAs maintain their telomeres via homologous recombination in the ALT pathway (226, 227). Like human OSA, *TERT* was not expressed in any of the samples in the current study. Further, a mutation in the *TERT* promoter also was not seen. This is in contrast to a previous study utilizing TRAP (Telomeric Repeat Amplification Protocol) analysis that showed 73% of canine OSAs were positive for telomerase activity. This could be due to poor correlation between telomerase activity and *TERT* expression. An alternative explanation could be due to differences in study population. Both dogs in this study survived longer than the average dog with OSA, 1.5 years versus 10 months. *TERT*-negative tumors and tumors that are also ALT negative, including OSA, have been reported in humans to have a favorable prognosis (228). A possible explanation for both dogs having *TERT*-negative tumors despite previous reports of *TERT*-positive canine OSA is that this study selected for dogs with a good prognosis since only dogs with pulmonary metastasectomy were included. A third type of telomere maintenance mechanism is the non-defined telomere maintenance mechanism (NDTMM). This has been described in human glioblastoma, OSA and metastases of cutaneous melanoma and was seen in 22% of human tumors (229). We propose that the dogs in this study utilized the NDTMM to support unlimited growth or their tumors did not contain a significant number of immortalized cells. Alternatively, since changes to *TERT*, *ATRX* and *DAXX* were not seen with eWGS and RNA-seq, the

maintenance of telomere length may be still via *TERT* and ALT but at the epigenetic or post-transcriptional level or via accessory genes in the *TERT* and ALT pathways. Further analysis of the data for *TERT* or ALT gene signatures and the application of IHC or promoter methylation to the samples can be done.

Upregulated genes and pathways that may be targetable included, HSP90AB1, EP300 and the Wnt pathway

HSP90AB1 is in the COSMIC cancer consensus database and is commonly overexpressed in cancers. *HSP90AB1* functions as a chaperone to stabilize proteins during translation, and stabilize mutant (oncoproteins), denatured or unstable proteins thus preventing their degradation. It has been considered crucial for tumor cell survival. Since a mutation was not seen *HSP90AB1* in the Lab, expression may be increased simply as a result of the cytosolic environment of the tumor (e.g. increased translation, mutant proteins and stress). In the Sheepdog the gene amplification and accompanying high expression may have offered a selective advantage for the tumor cells and thus may be a driver mutation for this dog. Inhibition of *HSP90AB1* was effective in a canine OSA cell line (230).

Overexpression of *HSP90AB1* can be targeted with an HSP90 inhibitor. STA-1474, an HSP90 inhibitor, has been shown to be safe and effective in dogs with mast cell tumors (231).

CTNNB1, beta-catenin, is the key protein that activates the Wnt signaling pathway. Activation of the Wnt pathway leads to cell proliferation and the maintenance of

stemness by turning on expression of *MYC*, cyclin D1, *COX2* and matrix metalloproteinases. Activation of the Wnt pathway has been implicated in many cancers including colon, breast, brain and liver cancers (232). The Wnt pathway on the other hand is crucial for osteoblast lineage determination by mesenchymal stem cells and thus is crucial for skeletal development (233). Overexpression of *CTNNB1* has been reported in 81% and 67% of canine primary and metastatic OSA lesions and 70% of human OSA via IHC (234, 235). Many drugs have been shown to decrease signaling in the Wnt pathway such as a vitamin D, tankyrase inhibitor, polyphyllin I and dihydrotanshinone I (236-238). Currently there are no FDA-approved drugs that specifically target the Wnt pathway but the results of early phase trials are pending (238). Sorafenib, a receptor tyrosine kinase (RTK) inhibitor, in combination with a *MEK* inhibitor inactivates *CTNNB1* signaling in human hepatocellular carcinoma (239). Toceranib, the only FDA-approved RTK inhibitor in dogs may have a clinical benefit in a small subset of dogs with metastatic OSA, 0-5% of dogs had a partial response and 5-43.5% of dogs had stable disease in previous studies (240-242). The effects of toceranib on *CTNNB1* signaling have not been investigated as far as the author is aware. Further study into inhibitors of *CTNNB1* in canine OSA and *CTNNB1* inhibition in dogs that respond to toceranib are warranted.

MITF and *FOXP1*, which are oncogenes in the COSMIC database, were amplified in all samples. *MITF* is a transcription factor that is crucial for the development and proliferation of melanocytes. Other members of the MiT (microphthalmia) family of

transcription factors include *TFEB*, *TFEC* and *TFE3*. *TFEB* was amplified in the Sheepdog and *TFE3* was amplified in the Lab with ~3-4 copies per cell. *MITF* has not been implicated in OSA but is amplified and is a driver gene in human melanoma via the RAS/MEK pathway (243). In melanoma cell lines *MITF* amplification is commonly accompanied by *CDKN2A* inactivation and *BRAF* mutation. In this study both dogs had *MITF* amplification and *CDKN2A* deletion and the Sheepdog had *BRAF* amplification with ~4 copies per cell. *MITF* inactivation leads to chemosensitivity and thus *MITF* directed therapies have potential anti-tumor potential in canine OSA (243). *MITF* is a downstream target of *EP300*, a histone acetyltransferase and an oncogene in a subset of melanomas. *EP300* was amplified in the metastatic lesion of the Sheepdog. *MITF* activity has been linked to acquired resistance to MAPK inhibitors (244) and thus it is predicted that MAPK inhibitors will be ineffective in canine OSA. There are no current *MITF* targeted therapies. Instead *EP300* targeted therapy have shown potential in preclinical models of *MITF* overexpressing melanomas (244). Canine osteosarcoma may be a good preclinical model for evaluating drugs in human melanoma.

HER2 was not overexpressed or mutated

Overexpression of *ERBB2/HER2/NEU* has been reported in some studies of canine (40%) and human OSA (43-80% of cases), although *HER2* amplification and mutation is rare (245). On the other hand overexpression of *HER2* has not been demonstrated in other human OSA studies (246). Interestingly in our study *HER2* had low expression and was not mutated either by a SNV or SV. *HER2* is an

oncogene. *HER2* is a RTK and its activation leads to MAPK, PI3K and Akt cell signaling. In the dogs in this study through loss of *PTEN*, PI3K and Akt pathways were activated and thus inhibition of the upstream *HER2* receptor may not be as effective as a therapeutic target than downstream targets. Trastuzumab (*HER2* monoclonal antibody) has yet to show a benefit in human OSA (247) but early results from *HER2* CAR-T cell therapy for recurrent/refractory *HER2*-positive human OSA showed clinical benefit (stable disease or partial remission) in a small subset, 28.6%, of patients (246). In a phase 1 study the *HER2* vaccine showed promise in canine OSA in the minimal residual disease setting compared to a historical control cohort treated with standard of care treatment (248). Perhaps activation of the immune system to remove *HER2* expressing cells is sufficient to overcome activation of downstream *HER2* pathways in a subset of patients. The results of the current phase 2 NIH clinical trial in dogs with OSA treated with the *HER2* vaccine are eagerly anticipated.

Chromosome 26 was shattered and reassembly resulted in numerous chromosomal lesions

CHEK2, *ZNRF3* and *NF2* were deleted in both primary lesions and have been implicated in human OSA (224, 225, 249-252). *CHEK2* is a crucial kinase in the DNA damage checkpoint cascade that stabilizes p53 when dsDNA breaks are detected. *ZNRF3* is part of the E3 ubiquitin ligase family and inhibits cell growth and suppresses invasion and induces apoptosis via regulation of the Wnt pathway. *SEPT5* and *LZTR1* were deleted in all lesions. *SEPT5* is a fusion partner with *MLL* in

human acute leukemias (253). *SEPT5* deletion is likely a passenger mutation in canine OSA. Germline loss of *LZTR1*, leucine-zipper-like transcriptional regulator 1, is a cause of a type of neurofibromatosis, like *NF2* (254). *CLTCL1* was deleted in both primary lesions. Decreased *CLTCL1*, clathrin, has been identified as an early change in breast carcinomas (255) and as a driver in oral squamous cell carcinoma (256). *PTEN* and *FAS* were deleted in all lesions. *FAS* is the extracellular receptor for the death receptor pathway that leads to apoptosis or immune destruction. *FAS* is frequently downregulated in human OSA (257-259). These genes are located on 3 loci on chr26 but are on 2 different loci in humans. Chr10 and chr22 are not common hotspots for fragmentation in human cancers but have been described in pancreatic carcinoma (195, 260). Localization of these TSGs onto one chromosome could contribute to the increased risk of OSA in dogs compared to humans. Fragility of canine chr26 was not seen in BrdU studies (261, 262). Analysis of WGS data from other canine cancers can determine if fragmentation of chr26 is specific to canine OSA.

Similar to human OSA, expressed gene fusions were uncommon and were not recurrent in canine OSA

The biological significance of the gene fusions likely was the inactivation of the fusion partners versus a gain of function of an oncogene. The following fusions were seen, *GRK3-HPS4* in the Lab and *ITGB6-CAMK2D*, *EIF4E3-MITF* and *PICALM-DLG2* in the Sheepdog. *PICALM* mutations have been implicated in other cancers but not in human or canine osteosarcoma. *PICALM-AF10* fusions in ALL most commonly occur

in introns 17-19 in *PICALM* and can induce acute leukemia in mice (263). Most of the protein, except the final 4, 20, 35, or 55 of the 652 amino acids, depending on the transcript, of *PICALM* are retained in the fusion transcripts in humans. In the Sheepdog 126 of the 636 amino acids were missing. *DLG2*, Disks Large Homolog 2, encodes a multi-domain scaffold protein in the Hippo signaling pathway that is important for defining epithelial polarity during cell division and loss of function of *DLG2* contributes to cell cycle progression and invasion in cancer cells (264). *DLG2* was one of the most highly recurrent mutations, 52.9%, in pediatric OSA (187) and was shown to be a TSG in a cross-species study in mice, humans and dogs (264). Both dogs had a deletion in *DLG2*, which was more prevalent than the 25-50% mutation rate in the previous studies (6, 187, 264). In the human study, mutations were exclusively via SVs on one chromosome and almost exclusively via deletions. Fusion to another gene was not detected. Interestingly, the *DLG2* fusion seen here had a similar result to the deletions seen in human OSA. *GRK3*, G-protein-coupled receptor kinase 3, overexpression has been shown to promote growth and metastasis via angiogenesis in human prostatic carcinoma (265). This gene fusion likely is not a driver for canine OSA and occurred as a result of the background of genetic instability

Conclusions

Canine and human OSA share many genetic similarities. First, both are characterized by chromosomal instability. Recurrent mutations were thus most likely from chromosomal lesions than SNVs with deletions as the most common

chromosomal lesion. Chromosomal instability could be explained by the loss of cell cycle checkpoint integrity, e.g. *CDKN2A/B* and *RB1*, and DNA damage response pathways, e.g. *TP53*, *CHEK2* and *WRN*. Similar to human OSA, biallelic loss of *TP53* was the most common mutation in canine OSA. Second, both are characterized by genetic heterogeneity, which likely is a result from chromosomal instability. There was almost no correlation between the SNV profile from the 2 dogs and only moderate correlation between the primary and metastatic lesion from the same patient. More correlation was seen between the SV profiles from the 2 dogs but was still low and again only moderate correlation between the primary and metastatic lesion from the same patient. The metastatic lesion had a higher mutational load than the primary lesion. This suggests that further mutations were gained during the metastatic process and in the new niche. Further sequencing studies with more paired tumor and metastatic lesions (including sequencing of multiple metastatic lesions) can help determine if the mutations seen in the metastatic lesions are random and due to genetic instability or specific to the metastatic process. Recurrent gene fusions or translocations were not found, providing further support of genetic heterogeneity. This suggests that the method of dsDNA repair led to random chromosomal fusions. Random fusions can explain the complex karyotypes inferred from the WGS data in this study. Lastly, potentially actionable driver pathways and genes were dysregulated including, *HSP90AB1*, cell cycle checkpoint, DNA damage response, NOTCH, mTOR, Wnt, and MAPK pathways.

Canine and human OSA shared more similarities than differences. A major difference was that *TP53* was mutated by a missense mutation and LOH in dogs compared to a translocation in the first intron in humans. As far as the author is aware a translocation in *TP53* has not been described in canine OSA. Compared to human OSA where it is theorized that genetic instability is the first lesion that leads to widespread SV (187), both dogs in our study had highly prevalent SNV mutations in *TP53*. It can be theorized in canine OSA that a SNV mutation in *TP53* is the first lesion and because of the crucial role of *TP53* in maintaining genetic stability, its loss then leads to widespread chromosomal lesions. Other differences included, *CDKN2A/B* deletion was more commonly observed in canine OSA (and other canine cancers) compared to human cancers and genomic evidence of a telomere maintenance mechanism, *HER2* upregulation and kataegis were not seen. Many of the key genetic drivers were shared between canine and human OSA but the pathogenesis of OSA in these 2 species likely differs. The most active part of the long bones that bear the most weight are affected in both species but pediatric OSA is most common during the adolescent growth spurt (puberty) and canine OSA is most common well beyond fusion of growth plates. This indicates that a longer amount of time is required in dogs than humans to gain sufficient mutations. Mechanisms of DNA repair or apoptosis may be more effective or critical genes in OSA development may be more protected from mutations in canine compared to human osteoblasts. The vast majority of human OSA cases in adults occur secondary to germline mutations. Canine OSA may occur secondary to a germline SNP or SV that has been inadvertently selected for in large/giant breed dogs.

Regardless of these differences the resulting driver pathways that are dysregulated coincide. Our data lend support that dogs are a good genetic model for human OSA. The high frequency of OSA in dogs relative to humans, similarity in body size of patients and general ease of performing medical treatments in dogs makes dogs a good clinical model. In addition to current clinical trials in canine OSA, dogs may be good models for clinical trials that target tumors with *TP53* mutations, such as *MDM2* or *MDM4* inhibitors, drugs that restore wild type *TP53* activity (e.g. PRIMA-1), deplete mutant *TP53* (e.g. statins and HSP90 inhibitors) or inhibit proteins that have a synthetically lethal interaction with *TP53* null cells (e.g. *CHK1*, *ATM/CHK2*, *Plk1*, *Wee1*, and *MK2*) (266).

The major limitation to the current study was the small sample size, which prevented generalized conclusions about differences between canine and human OSA.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Funding Information

This work was supported by a Clinician Scientist Grant from the American Kennel Club's Canine Health Foundation.

Acknowledgements

We would like to thank Nimblegen for the donation of the canine exome capture reagent.

Conclusions and Future Directions

In the first project in this dissertation, it was found that FOSCC is not a good model for PV-induced squamous cell carcinoma in humans. If PVs have a role in FOSCC it is at most a supplementary or uncommon role and elucidation of other factors, e.g. environmental factors, would make a bigger impact on the role of FOSCC as a naturally occurring HNSCC model and clinical outcome of FOSCC. To address the second objective of this study, the virome of the FOSCC and normal feline oral mucosa consisted of feline foamy virus (35% and 22%), feline torque teno virus (5% and 0%), alphaherpesvirus (5% and 0%), FIV (0% and 11%), Epstein-Barr virus (5% and 0%) and feline papillomavirus (5% and 0% respectively). It was highly interesting that the only FOSCC sample that had papillomavirus also had human Epstein Barr virus (EBV). Contamination from an infected person during sample handling is not likely since contamination would have needed to come from a substantial amount of human bodily fluids. To further explore a causative effect of EBV in FOSCC PCR for EBV and FPV can be done on a large number (>150) of FFPE FOSCC samples and a matched number of normal samples. A more common multifactorial cause of FOSCC is likely exposure to a carcinogen. Future studies may include epidemiologic studies to evaluate risk factors to exposure to potential carcinogens since only one such study has been done (59). This study evaluated risk factors associated with bathing, teeth brushing, indoor/outdoor exposure, home heating sources, second-hand smoke, flea control products and diet and found that canned diets, canned tuna in the diet, and flea collars were significantly associated with risk of FOSCC. ViroCap proved to be an effective tool to study the virome in

cats. Virally induced cancers exist in cats, ferrets, mice, woodchucks, cows and humans but aside from papillomavirus papillomas that have transformed into SCC, the author is not aware of any known virus associated cancer in dogs. The most likely cancers to have a viral cause in dogs are acute leukemias due to the anecdotal seasonality of the disease and prevalence in young or middle aged dogs. ViroCap can be easily applied to blood samples from these dogs.

The second project concluded that an epigenetic signature of cDLBCL was consistent in Golden Retrievers and appeared to be similar in different breeds in different countries. cDLBCL in Golden Retrievers was homogenous and subgroups were not seen. Genes that were recurrently mutated via DNA and DNA methylation changes in cDLBCL in Golden Retrievers included *SETD2*, *SOX17*, *WDR17* and *DOK6*. The methylome was characterized by hypermethylated DMRs in the promoter-TSS-first exon region and followed the distribution of CGIs. Genes with hypermethylated DMRs in the promoter-TSS-first exon often also had multiple hyper- and hypomethylated DMRs distributed throughout the gene suggesting multiple layers of epimutations. Epigenetically dysregulated pathways included neurogenic (e.g. *RASGRF1* and *ONECUT*), *HOX*, *WNT* and *TBX* pathways.

Current work on this study includes, testing a methylation biomarker panel on lymph node aspirates from normal but at-risk Golden Retrievers and Golden Retrievers with cDLBCL to further test the hypothesis that DNA methylation changes are more recurrent than DNA mutation changes in cDLBCL and to determine if there is a population of dogs that carry these at-risk marks. We are

also continuing to collect lymph node aspirates in normal but at risk Golden Retrievers. Offering an incentive such as a free exam, oncology consultation and lymph node aspirate if lymphadenopathy is noticed post baseline aspirate would help with accrual and follow up. The pre-LSA aspirates will be tested with MIRA-seq if they develop LSA. This targeted approach will hopefully determine if recurrent DNA methylation biomarkers occur before LSA develops and if this occurs early enough to allow intervention. Other current works include, correlation of methylation data with RNA data obtained in this study by RNA-sequencing (polyA selection) and Nanostring microarray and with microarray and RNA-seq in previous studies (96, 98, 267, 268). We will also perform in-depth genomic analysis, using the pipeline used in the OSA study, with matched normal WGS, WES and RNA-seq data obtained from this study. The methylation landscape can then be correlated with the DNA mutational landscape and expression for the first time in cDLBCL. To date, WGS also has not been done in cDLBCL. Most recently, the 3 hDLBCL subtypes, germinal center B-cell, activated B-cell and unclassified hDLBCL, have more recently been further divided into 4 subgroups based off of genetic abnormalities, MCD (co-occurrence of *MYD88* and *CD79B* mutations), BN2 (*BCL6* fusions and *NOTCH2* mutations), N1 (*NOTCH1* mutations) and EZB (*EZH2* mutations and *BCL2* translocations) (269). The 3 subtypes have not been seen thus far in cDLBCL with exome, methylation and gene expression data. If these classical SVs are not seen in the WGS and WES data, cDLBCL likely would not be a good model for any treatments and diagnostics, that are specific to a hDLBCL subtypes.

Many of the differentially methylated genes in cDLBCL have also been found in other cancer types. MIRA-seq was also done on a TCC cell line and cDLBCL cell line (17-71) with derived tumor initiating cells (TICs). Reduced representation bisulfite sequencing was also recently done on canine gliomas. Analysis of the methylome data from these datasets with the DLBCL datasets will be done to test the hypothesis that most of the identified DMRs are pan-cancer DMRs and fewer DMRs are cancer-type specific. To truly test the pan-cancer methylome hypothesis additional MIRA-seq studies will need to be done in canine cancers. Osteosarcoma (OSA) and hemangiosarcoma (HSA) are recommended for future study since both of these cancers have a >90% metastatic rate and metastasize via the hematogenous route. The potential of finding a pre-cancerous or early stage biomarker in blood samples that can lead to early treatment and improved prognosis is enticing. Whole genome and transcriptome studies have been done in OSA and thus methylation data could add to the understanding of the landscape of OSA and functional consequences of methylation can be determined. MIRA-seq can also be applied to DLBCL in other breeds of dogs and other types of NHL to test the hypothesis that the methylome is consistent across breeds but differs between types of NHL with increased epigenetic heterogeneity corresponding to aggressiveness of disease (270).

The third project concluded that canine and human OSA shared more similarities than differences. Genetic similarities, included chromosomal instability and heterogeneity. Recurrent mutations were more commonly via chromosomal

lesions than single nucleotide variants. Chromosomal instability was likely a result of biallelic deletions in *TP53*. Chromosome 26 appeared to be the most vulnerable chromosome as it was shattered into many small pieces but joined back together randomly in all samples. Potentially actionable driver pathways and genes were dysregulated including, *HSP90AB1*, cell cycle checkpoint, DNA damage response, NOTCH, mTOR, Wnt, and MAPK pathways. The major difference between canine and human OSA was that *TP53* was mutated by a missense mutation and LOH in dogs compared to a translocation in the first intron in humans. The only human cancer that this translocation has only ever been shown to occur in is OSA but has never been shown to occur in canine OSA. Other differences included, *CDKN2A/B* deletion was more commonly observed in canine OSA (and other canine cancers) compared to human cancers and genomic evidence of a telomere maintenance mechanism, *HER2* upregulation and kataegis were not seen. Future directions include, functional studies with *HSP90AB1*, *CDK4*, Wnt, and *NOTCH2* inhibitors in OSA cell lines to validate these targets for future clinical studies, demonstrating the unique vulnerability of chr26 in an expanded set of canine tumors, associating germline structural variants and predisposition to cancer, predicting neoepitopes, ultradeep sequencing of primary and metastatic tumors to gain insight in tumor evolution, and monitoring treatment efficacy and screening of at risk populations with circulating tumor DNA bearing *TP53* mutations.

Many more large-scale genomic studies are being pursued in veterinary oncology and the amounting data is becoming overwhelming. Funding in veterinary

medicine remains limited for sequencing studies and thus it is arguably more important in veterinary compared to human medicine to use previous sequencing data to answer new questions. Each bioinformatics project analyzes a minimal portion of the sequenced data and an even smaller portion of the data is presented. It is thus crucial to share raw genomic data, thoroughly describe patient samples including previous research done with the samples, library preparation methods and data analysis methods to facilitate collaborative research.

As the number of genomes of veterinary patients that are sequenced increases it will be imperative to organize this information into a user-friendly and easily accessible database. In human oncology numerous databases exist. For example, there are databases for mining cancer associated SNVs, e.g. COSMIC and CIViC, evaluating protein expression in cancers, e.g. the Human Protein Atlas, evaluating drugability of mutations, e.g. DGIdb, reporting of genomic variations and phenotypes, e.g. ClinVar, and more comprehensive tumor databases, e.g. Cancer Genetics Web and PeCan (Pediatric Cancers) to name a few. The author is not aware of similar databases in veterinary oncology. Before the amount of data becomes unmanageable, details of completed studies are forgotten, and interpretation files are lost, now is the ideal time to collaborate to build a strong foundation to support genomics in veterinary oncology.

BIBLIOGRAPHY

1. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-74.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;291(5507):1304-51.
4. Murchison EP, Wedge DC, Alexandrov LB, Fu B, Martincorena I, Ning Z, et al. Transmissible [corrected] dog cancer genome reveals the origin and history of an ancient cell lineage. *Science*. 2014;343(6169):437-40.
5. Liu D, Xiong H, Ellis AE, Northrup NC, Rodriguez CO, Jr., O'Regan RM, et al. Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer. *Cancer Res*. 2014;74(18):5045-56.
6. Gardner HL, Sivaprakasam K, Briones N, Zismann V, Perdigones N, Drenner K, et al. Canine osteosarcoma genome sequencing identifies recurrent mutations in DMD and the histone methyltransferase gene SETD2. *Commun Biol*. 2019;2:266.
7. Hendricks WPD, Zismann V, Sivaprakasam K, Legendre C, Poorman K, Tembe W, et al. Somatic inactivating PTPRJ mutations and dysregulated pathways identified in canine malignant melanoma by integrated comparative genomic analysis. *PLoS Genet*. 2018;14(9):e1007589.
8. Samirkumar B. Amin KJA, C. Elizabeth Boudreau, Emmanuel Martinez-Ledesma, Emre Kocakavuk, Kevin C. Johnson, Floris P. Barthel, Frederick S. Varn, Cynthia Kassab, Xiaoyang Ling, Hoon Kim, Mary Barter, Chew Yee Ngan, Margaret Chapman, Jennifer W. Koehler, Andrew D. Miller, C. Ryan Miller, Brian F. Porter,

Daniel R. Rissi, Christina Mazcko, Amy K. LeBlanc, Peter J. Dickinson, Rebecca Packer, Amanda R. Taylor, John H. Rossmeisl Jr, Amy Heimberger, Jonathan M. Levine, Roel G. W. Verhaak. Comparative molecular life history of spontaneous canine and human gliomas. *bioRxiv*.

9. Wylie TN, Wylie KM, Herter BN, Storchl GA. Enhanced virome sequencing using targeted sequence capture. *Genome Res*. 2015;25(12):1910-20.
10. Fidel J, Lyons J, Tripp C, Houston R, Wheeler B, Ruiz A. Treatment of oral squamous cell carcinoma with accelerated radiation therapy and concomitant carboplatin in cats. *J Vet Intern Med*. 2011;25(3):504-10.
11. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015 A Systematic Analysis for the Global Burden of Disease Study. *Jama Oncol*. 2017;3(4):524-48.
12. Wise-Draper TM, Draper DJ, Gutkind JS, Molinolo AA, Wikenheiser-Brokamp KA, Wells SI. Future directions and treatment strategies for head and neck squamous cell carcinomas. *Translational research : the journal of laboratory and clinical medicine*. 2012;160(3):167-77.
13. Herrero R, Castellsague X, Pawlita M, Lissowska J, Kee F, Balaram P, et al. Human papillomavirus and oral cancer: the International Agency for Research on Cancer multicenter study. *J Natl Cancer Inst*. 2003;95(23):1772-83.

14. Ajila V, Shetty H, Babu S, Shetty V, Hegde S. Human Papilloma Virus Associated Squamous Cell Carcinoma of the Head and Neck. *J Sex Transm Dis.* 2015;2015:791024.
15. Supsavhad W, Dirksen WP, Martin CK, Rosol TJ. Animal models of head and neck squamous cell carcinoma. *Vet J.* 2016;210:7-16.
16. Munday JS, Dittmer KE, Thomson NA, Hills SF, Laurie RE. Genomic characterisation of *Felis catus* papillomavirus type 5 with proposed classification within a new papillomavirus genus. *Vet Microbiol.* 2017;207:50-5.
17. Munday JS. Papillomaviruses in felids. *Veterinary journal.* 2014;199(3):340-7.
18. Munday JS, Kiupel M, French AF, Howe L, Squires RA. Detection of papillomaviral sequences in feline Bowenoid in situ carcinoma using consensus primers. *Veterinary Dermatology.* 2007;18(4):241-5.
19. Lange CE, Tobler K, Markau T, Alhaidari Z, Bornand V, Stockli R, et al. Sequence and classification of FdPV2, a papillomavirus isolated from feline Bowenoid in situ carcinomas. *Veterinary Microbiology.* 2009;137(1-2):60-5.
20. Munday JS, Gibson I, French AF. Papillomaviral DNA and increased p16(CDKN2A) protein are frequently present within feline cutaneous squamous cell carcinomas in ultraviolet-protected skin. *Veterinary Dermatology.* 2011;22(4):360-6.
21. Gil da Costa RM, Peleteiro MC, Pires MA, DiMaio D. An Update on Canine, Feline and Bovine Papillomaviruses. *Transbound Emerg Dis.* 2016.

22. Yamashita-Kawanishi N, Sawanobori R, Matsumiya K, Uema A, Chambers JK, Uchida K, et al. Detection of felis catus papillomavirus type 3 and 4 DNA from squamous cell carcinoma cases of cats in Japan. *J Vet Med Sci.* 2018;80(8):1236-40.
23. Skor O, editor Presence of Papillomavirus DNA in Feline Squamous Cell Carcinoma and Injection Site-Sarcoma. Veterinary Cancer Society Conference; 2015 October 15-17, 2015; Fairfax County, Virginia.
24. Anis EA, O'Neill SH, Newkirk KM, Brahmbhatt RA, Abd-Eldaim M, Frank LA, et al. Molecular characterization of the L1 gene of papillomaviruses in epithelial lesions of cats and comparative analysis with corresponding gene sequences of human and feline papillomaviruses. *Am J Vet Res.* 2010;71(12):1457-61.
25. O'Neill SH, Newkirk KM, Anis EA, Brahmbhatt R, Frank LA, Kania SA. Detection of human papillomavirus DNA in feline premalignant and invasive squamous cell carcinoma. *Vet Dermatol.* 2011;22(1):68-74.
26. Supsavhad W, Dirksen WP, Hildreth BE, Rosol TJ. p16, pRb, and p53 in Feline Oral Squamous Cell Carcinoma. *Vet Sci.* 2016;3(3).
27. Chen EC, Miller SA, DeRisi JL, Chiu CY. Using a pan-viral microarray assay (Virochip) to screen clinical samples for viral pathogens. *Journal of visualized experiments : JoVE.* 2011(50).
28. Baldwin DA, Feldman M, Alwine JC, Robertson ES. Metagenomic assay for identification of microbial pathogens in tumor tissues. *mBio.* 2014;5(5):e01714-14.
29. Wylie KM, Wylie TN, Cahill AG, Macones GA, Tuuli MG, Stout MJ. The vaginal eukaryotic DNA virome and preterm birth. *Am J Obstet Gynecol.* 2018;219(2):189 e1- e12.

30. Wylie KM, Wylie TN, Buller R, Herter B, Cannella MT, Storch GA. Detection of Viruses in Clinical Samples Using Metagenomic Sequencing and Targeted Sequence Capture. *J Clin Microbiol*. 2018.
31. Papineau A, Berhane Y, Wylie TN, Wylie KM, Sharpe S, Lung O. Genome Organization of Canada Goose Coronavirus, A Novel Species Identified in a Mass Die-off of Canada Geese. *Sci Rep*. 2019;9(1):5954.
32. Cancer Genome Atlas Research N, Analysis Working Group: Asan U, Agency BCC, Brigham, Women's H, Broad I, et al. Integrated genomic characterization of oesophageal carcinoma. *Nature*. 2017;541(7636):169-75.
33. Cancer Genome Atlas Research N, Albert Einstein College of M, Analytical Biological S, Barretos Cancer H, Baylor College of M, Beckman Research Institute of City of H, et al. Integrated genomic and molecular characterization of cervical cancer. *Nature*. 2017;543(7645):378-84.
34. Chambers G, Ellsmore VA, O'Brien PM, Reid SW, Love S, Campo MS, et al. Association of bovine papillomavirus with the equine sarcoid. *J Gen Virol*. 2003;84(Pt 5):1055-62.
35. Brandt S, Tober R, Corteggio A, Burger S, Sabitzer S, Walter I, et al. BPV-1 infection is not confined to the dermis but also involves the epidermis of equine sarcoids. *Vet Microbiol*. 2011;150(1-2):35-40.
36. Sykes JE. *Canine and Feline Infectious Diseases*. St. Louis, MO: Saunders; 2014.

37. Faraji F, Zaidi M, Fakhry C, Gaykalova DA. Molecular mechanisms of human papillomavirus-related carcinogenesis in head and neck cancer. *Microbes Infect.* 2017;19(9-10):464-75.
38. Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, Bristow CA, et al. Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci U S A.* 2014;111(43):15544-9.
39. Munday JS, French AF. *Felis catus* papillomavirus types 1 and 4 are rarely present in neoplastic and inflammatory oral lesions of cats. *Res Vet Sci.* 2015;100:220-2.
40. Munday JS, Howe L, French A, Squires RA, Sugiarto H. Detection of papillomaviral DNA sequences in a feline oral squamous cell carcinoma. *Res Vet Sci.* 2009;86(2):359-61.
41. Altamura G, Eleni C, Meoli R, Cardeti G, Friedrich KG, Borzacchiello G. Tongue Squamous Cell Carcinoma in a European Lynx (*Lynx Lynx*): Papillomavirus Infection and Histologic Analysis. *Vet Sci.* 2018;5(1).
42. Thaiwong T, Sledge DG, Collins-Webb A, Kiupel M. Immunohistochemical Characterization of Canine Oral Papillary Squamous Cell Carcinoma. *Vet Pathol.* 2018;55(2):224-32.
43. Sabbatini S, Savini F, Gallina L, Scagliarini A, Bassi P, Bettini G. p16 Immunostaining of Canine Squamous Cell Carcinomas Is Not Associated with Papillomaviral DNA. *PLoS One.* 2016;11(7):e0159687.

44. Munday JS, French A, Harvey CJ. Molecular and immunohistochemical studies do not support a role for papillomaviruses in canine oral squamous cell carcinoma development. *Vet J.* 2015;204(2):223-5.
45. Porcellato I, Brachelente C, Guelfi G, Reginato A, Sforza M, Bongiovanni L, et al. A retrospective investigation on canine papillomavirus 1 (CPV1) in oral oncogenesis reveals dogs are not a suitable animal model for high-risk HPV-induced oral cancer. *PLoS One.* 2014;9(11):e112833.
46. Zaugg N, Nespeca G, Hauser B, Ackermann M, Favrot C. Detection of novel papillomaviruses in canine mucosal, cutaneous and in situ squamous cell carcinomas. *Vet Dermatol.* 2005;16(5):290-8.
47. Teifke JP, Lohr CV, Shirasawa H. Detection of canine oral papillomavirus-DNA in canine oral squamous cell carcinomas and p53 overexpressing skin papillomas of the dog using the polymerase chain reaction and non-radioactive in situ hybridization. *Vet Microbiol.* 1998;60(2-4):119-30.
48. Meyers JM, Grace M, Uberoi A, Lambert PF, Munger K. Inhibition of TGF-beta and NOTCH Signaling by Cutaneous Papillomaviruses. *Front Microbiol.* 2018;9:389.
49. Viarisio D, Muller-Decker K, Accardi R, Robitaille A, Durst M, Beer K, et al. Beta HPV38 oncoproteins act with a hit-and-run mechanism in ultraviolet radiation-induced skin carcinogenesis in mice. *PLoS Pathog.* 2018;14(1):e1006783.
50. Smith KT, Campo MS. "Hit and run" transformation of mouse C127 cells by bovine papillomavirus type 4: the viral DNA is required for the initiation but not for maintenance of the transformed phenotype. *Virology.* 1988;164(1):39-47.

51. Gaukroger JM, Bradley A, Chandrachud L, Jarrett WF, Campo MS. Interaction between bovine papillomavirus type 4 and cocarcinogens in the production of malignant tumours. *J Gen Virol.* 1993;74 (Pt 10):2275-80.
52. Munday JS, Dunowska M, Hills SF, Laurie RE. Genomic characterization of *Felis catus* papillomavirus-3: a novel papillomavirus detected in a feline Bowenoid in situ carcinoma. *Vet Microbiol.* 2013;165(3-4):319-25.
53. Munday JS, Fairley R, Atkinson K. The detection of *Felis catus* papillomavirus 3 DNA in a feline bowenoid in situ carcinoma with novel histologic features and benign clinical behavior. *J Vet Diagn Invest.* 2016;28(5):612-5.
54. Munday JS, French AF, Gibson IR, Knight CG. The presence of p16 CDKN2A protein immunostaining within feline nasal planum squamous cell carcinomas is associated with an increased survival time and the presence of papillomaviral DNA. *Vet Pathol.* 2013;50(2):269-73.
55. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature.* 2015;517(7536):576-82.
56. Kang H, Kiess A, Chung CH. Emerging biomarkers in head and neck cancer in the era of genomics. *Nat Rev Clin Oncol.* 2015;12(1):11-26.
57. Gelwan E, Malm IJ, Khararjian A, Fakhry C, Bishop JA, Westra WH. Nonuniform Distribution of High-risk Human Papillomavirus in Squamous Cell Carcinomas of the Oropharynx: Rethinking the Anatomic Boundaries of Oral and Oropharyngeal Carcinoma From an Oncologic HPV Perspective. *Am J Surg Pathol.* 2017;41(12):1722-8.

58. Martin CK, Tannehill-Gregg SH, Wolfe TD, Rosol TJ. Bone-invasive oral squamous cell carcinoma in cats: pathology and expression of parathyroid hormone-related protein. *Vet Pathol.* 2011;48(1):302-12.
59. Bertone ER, Snyder LA, Moore AS. Environmental and lifestyle risk factors for oral squamous cell carcinoma in domestic cats. *J Vet Intern Med.* 2003;17(4):557-62.
60. Freedman ND, Abnet CC, Leitzmann MF, Hollenbeck AR, Schatzkin A. Prospective investigation of the cigarette smoking-head and neck cancer association by sex. *Cancer.* 2007;110(7):1593-601.
61. Guha N, Warnakulasuriya S, Vlaanderen J, Straif K. Betel quid chewing and the risk of oral and oropharyngeal cancers: A meta-analysis with implications for cancer control. *International Journal of Cancer.* 2014;135(6):1433-43.
62. Altieri A, Bosetti C, Gallus S, Franceschi S, Dal Maso L, Talamini R, et al. Wine, beer and spirits and risk of oral and pharyngeal cancer: a case-control study from Italy and Switzerland. *Oral Oncology.* 2004;40(9):904-9.
63. Alke A, Schwantes A, Zemba M, Flugel RM, Lochelt M. Characterization of the humoral immune response and virus replication in cats experimentally infected with feline foamy virus. *Virology.* 2000;275(1):170-6.
64. Winkler IG, Flugel RM, Lochelt M, Flower RLP. Detection and molecular characterisation of feline foamy virus serotypes in naturally infected cats. *Virology.* 1998;247(2):144-51.

65. Jarosova V, Hrazdilova K, Filipejova Z, Schanilec P, Celer V. Whole genome sequencing and phylogenetic analysis of feline anelloviruses. *Infect Genet Evol.* 2015;32:130-4.
66. Zhang W, Li L, Deng X, Kapusinszky B, Pesavento PA, Delwart E. Faecal virome of cats in an animal shelter. *J Gen Virol.* 2014;95(Pt 11):2553-64.
67. Shulman LM, Davidson I. Viruses with Circular Single-Stranded DNA Genomes Are Everywhere! *Annu Rev Virol.* 2017;4(1):159-80.
68. Focosi D, Antonelli G, Pistello M, Maggi F. Torquetenovirus: the human virome from bench to bedside. *Clin Microbiol Infect.* 2016;22(7):589-93.
69. Zhu CX, Shan TL, Cui L, Luo XN, Liu ZJ, Tang SD, et al. Molecular detection and sequence analysis of feline Torque teno virus (TTV) in China. *Virus Res.* 2011;156(1-2):13-6.
70. Okamoto H, Takahashi M, Nishizawa T, Tawara A, Fukai K, Muramatsu U, et al. Genomic characterization of TT viruses (TTVs) in pigs, cats and dogs and their relatedness with species-specific TTVs in primates and tupaia. *J Gen Virol.* 2002;83(Pt 6):1291-7.
71. Wohlfarth P, Leiner M, Schoergenhofer C, Hopfinger G, Goerzer I, Puchhammer-Stoeckl E, et al. Torquetenovirus Dynamics and Immune Marker Properties in Patients Following Allogeneic Hematopoietic Stem Cell Transplantation: A Prospective Longitudinal Study. *Biol Blood Marrow Transplant.* 2018;24(1):194-9.

72. McManus CM, Levy JK, Andersen LA, McGorray SP, Leutenegger CM, Gray LK, et al. Prevalence of upper respiratory pathogens in four management models for unowned cats in the Southeast United States. *Vet J.* 2014;201(2):196-201.
73. Sjodahl-Essen T, Tidholm A, Thoren P, Persson-Wadman A, Bolske G, Aspan A, et al. Evaluation of different sampling methods and results of real-time PCR for detection of feline herpes virus-1, *Chlamydomphila felis* and *Mycoplasma felis* in cats. *Vet Ophthalmol.* 2008;11(6):375-80.
74. Milman G, Smith KC, Erles K. Serological detection of Epstein-Barr virus infection in dogs and cats. *Vet Microbiol.* 2011;150(1-2):15-20.
75. Waugh EM, Gallagher A, McAulay KA, Henriques J, Alves M, Bell AJ, et al. Gammaherpesviruses and canine lymphoma: no evidence for direct involvement in commonly occurring lymphomas. *J Gen Virol.* 2015;96(Pt 7):1863-72.
76. Huang SH, Kozak PJ, Kim J, Habineza-Ndikuyeze G, Meade C, Gaurnier-Hausser A, et al. Evidence of an oncogenic gammaherpesvirus in domestic dogs. *Virology.* 2012;427(2):107-17.
77. Heller M, Gerber P, Kieff E. DNA of herpesvirus pan, a third member of the Epstein-Barr virus-Herpesvirus papio group. *J Virol.* 1982;41(3):931-9.
78. Beatty JA, Troyer RM, Carver S, Barrs VR, Espinasse F, Conradi O, et al. *Felis catus* gammaherpesvirus 1; a widely endemic potential pathogen of domestic cats. *Virology.* 2014;460-461:100-7.
79. Ertl R, Korb M, Langbein-Detsch I, Klein D. Prevalence and risk factors of gammaherpesvirus infection in domestic cats in Central Europe. *Virol J.* 2015;12:146.

80. Troyer RM, Lee JS, Vuyisich M, Chain P, Lo CC, Kronmiller B, et al. First Complete Genome Sequence of *Felis catus* Gammaherpesvirus 1. *Genome Announc.* 2015;3(6).
81. She Y, Nong X, Zhang M, Wang M. Epstein-Barr virus infection and oral squamous cell carcinoma risk: A meta-analysis. *PLoS One.* 2017;12(10):e0186860.
82. Levy JK, Crawford PC, Tucker SJ. Performance of 4 Point-of-Care Screening Tests for Feline Leukemia Virus and Feline Immunodeficiency Virus. *J Vet Intern Med.* 2017;31(2):521-6.
83. Westman ME, Malik R, Hall E, Sheehy PA, Norris JM. Comparison of three feline leukaemia virus (FeLV) point-of-care antigen test kits using blood and saliva. *Comp Immunol Microbiol Infect Dis.* 2017;50:88-96.
84. Lopez NA, Jacobson RH. False-positive reactions associated with anti-mouse activity in serotests for feline leukemia virus antigen. *J Am Vet Med Assoc.* 1989;195(6):741-6.
85. Lopez NA, Jacobson RH, Scarlett JM, Center SA, Randolph JF, Scott FW. Sensitivity and specificity of blood test kits for feline leukemia virus antigen. *J Am Vet Med Assoc.* 1989;195(6):747-51.
86. Gomes-Keller MA, Gonczi E, Grenacher B, Tandon R, Hofman-Lehmann R, Lutz H. Fecal shedding of infectious feline leukemia virus and its nucleic acids: a transmission potential. *Vet Microbiol.* 2009;134(3-4):208-17.
87. Tandon R, Cattori V, Gomes-Keller MA, Meli ML, Golder MC, Lutz H, et al. Quantitation of feline leukaemia virus viral and proviral loads by TaqMan real-time polymerase chain reaction. *J Virol Methods.* 2005;130(1-2):124-32.

88. Chiu ES, Hoover EA, VandeWoude S. A Retrospective Examination of Feline Leukemia Subgroup Characterization: Viral Interference Assays to Deep Sequencing. *Viruses*. 2018;10(1).
89. Polani S, Roca AL, Rosensteel BB, Kolokotronis SO, Bar-Gal GK. Evolutionary dynamics of endogenous feline leukemia virus proliferation among species of the domestic cat lineage. *Virology*. 2010;405(2):397-407.
90. Dunn JR, Gimeno IM. Current status of Marek's disease in the United States and worldwide based on a questionnaire survey. *Avian Dis*. 2013;57(2 Suppl):483-90.
91. Cubie HA. Diseases associated with human papillomavirus infection. *Virology*. 2013;445(1-2):21-34.
92. Kampmann M-L, Børsting C, Morling N. Decrease DNA contamination in the laboratories. *Forensic Science International: Genetics Supplement Series*. 2017;6:e577-e8.
93. Vandewoestyne M, Van Hoofstat D, De Groote S, Van Thuyne N, Haerinck S, Van Nieuwerburgh F, et al. Sources of DNA Contamination and Decontamination Procedures in the Forensic Laboratory. *Journal of Forensic Research*. 2012;2(2):3.
94. Seelig DM, Avery AC, Ehrhart EJ, Linden MA. The Comparative Diagnostic Features of Canine and Human Lymphoma. *Vet Sci*. 2016;3(2).
95. Elvers I, Turner-Maier J, Swofford R, Koltookian M, Johnson J, Stewart C, et al. Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res*. 2015;25(11):1634-45.

96. Bushell KR, Kim Y, Chan FC, Ben-Neriah S, Jenks A, Alcaide M, et al. Genetic inactivation of TRAF3 in canine and human B-cell lymphoma. *Blood*. 2015;125(6):999-1005.
97. Seelig DM, Ito D, Forster CL, Yoon UA, Breen M, Burns LJ, et al. Constitutive activation of alternative nuclear factor kappa B pathway in canine diffuse large B-cell lymphoma contributes to tumor cell survival and is a target of new adjuvant therapies. *Leuk Lymphoma*. 2017;58(7):1702-10.
98. Mudaliar MA, Haggart RD, Miele G, Sellar G, Tan KA, Goodlad JR, et al. Comparative gene expression profiling identifies common molecular signatures of NF-kappaB activation in canine and human diffuse large B cell lymphoma (DLBCL). *PLoS One*. 2013;8(9):e72591.
99. Burton JH, Mazcko C, LeBlanc A, Covey JM, Ji J, Kinders RJ, et al. NCI Comparative Oncology Program Testing of Non-Camptothecin Indenoisoquinoline Topoisomerase I Inhibitors in Naturally Occurring Canine Lymphoma. *Clin Cancer Res*. 2018;24(23):5830-40.
100. Marconato L, Frayssinet P, Rouquet N, Comazzi S, Leone VF, Laganga P, et al. Randomized, placebo-controlled, double-blinded chemoimmunotherapy clinical trial in a pet dog model of diffuse large B-cell lymphoma. *Clin Cancer Res*. 2014;20(3):668-77.
101. Gardner HL, Rippy SB, Bear MD, Cronin KL, Heeb H, Burr H, et al. Phase I/II evaluation of RV1001, a novel PI3Kdelta inhibitor, in spontaneous canine lymphoma. *PLoS One*. 2018;13(4):e0195357.

102. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503-11.
103. Richards KL, Motsinger-Reif AA, Chen HW, Fedoriw Y, Fan C, Nielsen DM, et al. Gene profiling of canine B-cell lymphoma reveals germinal center and postgerminal center subtypes with different survival times, modeling human DLBCL. *Cancer Res*. 2013;73(16):5029-39.
104. Baylin SB, Jones PA. Epigenetic Determinants of Cancer. *Cold Spring Harb Perspect Biol*. 2016;8(9).
105. Merlo A, Herman JG, Mao L, Lee DJ, Gabrielson E, Burger PC, et al. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. *Nat Med*. 1995;1(7):686-92.
106. Shaknovich R, Geng H, Johnson NA, Tsikitas L, Cerchietti L, Greally JM, et al. DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. *Blood*. 2010;116(20):e81-9.
107. Ferraresso S, Arico A, Sanavia T, Da Ros S, Milan M, Cascione L, et al. DNA methylation profiling reveals common signatures of tumorigenesis and defines epigenetic prognostic subtypes of canine Diffuse Large B-cell Lymphoma. *Sci Rep*. 2017;7(1):11591.
108. Yamazaki J, Jelinek J, Hisamoto S, Tsukamoto A, Inaba M. Dynamic changes in DNA methylation patterns in canine lymphoma cell lines demonstrated by genome-wide quantitative DNA methylation analysis. *Vet J*. 2018;231:48-54.

109. Epiphanio TMF, Fernandes N, de Oliveira TF, Lopes PA, Ressio RA, Goncalves S, et al. Global DNA methylation of peripheral blood leukocytes from dogs bearing multicentric non-Hodgkin lymphomas and healthy dogs: A comparative study. *PLoS One*. 2019;14(3):e0211898.
110. Glickman LG, Nita; Thorpe, Roland. The Golden Retriever Club of America National Health Survey 1998-199899. Golden Retriever Club of America, 1999.
111. Villamil JA, Henry CJ, Hahn AW, Bryan JN, Tyler JW, Caldwell CW. Hormonal and sex impact on the epidemiology of canine lymphoma. *J Cancer Epidemiol*. 2009;2009:591753.
112. Comazzi S, Marelli S, Cozzi M, Rizzi R, Finotello R, Henriques J, et al. Breed-associated risks for developing canine lymphoma differ among countries: an European canine lymphoma network study. *BMC Vet Res*. 2018;14(1):232.
113. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011;17(1):10-2.
114. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20.
115. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
116. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics*. 2014;30(2):284-6.

117. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 2015;16:22.
118. Lizio M, Mukarram AK, Ohno M, Watanabe S, Itoh M, Hasegawa A, et al. Monitoring transcription initiation activities in rat and dog. *Sci Data.* 2017;4:170173.
119. Dreos R, Ambrosini G, Perier RC, Bucher P. The Eukaryotic Promoter Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res.* 2015;43(Database issue):D92-6.
120. Jung M, Kadam S, Xiong W, Rauch TA, Jin SG, Pfeifer GP. MIRA-seq for DNA methylation analysis of CpG islands. *Epigenomics.* 2015;7(5):695-706.
121. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature biotechnology.* 2010;28(10):1106-14.
122. Liu QR, Zhang PW, Zhen Q, Walther D, Wang XB, Uhl GR. KEPI, a PKC-dependent protein phosphatase 1 inhibitor regulated by morphine. *J Biol Chem.* 2002;277(15):13312-20.
123. Dedinszki D, Kiss A, Markasz L, Marton A, Toth E, Szekely L, et al. Inhibition of protein phosphatase-1 and -2A decreases the chemosensitivity of leukemic cells to chemotherapeutic drugs. *Cell Signal.* 2015;27(2):363-72.
124. Compagno M, Lim WK, Grunn A, Nandula SV, Brahmachary M, Shen Q, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature.* 2009;459(7247):717-21.

125. Zhang Y, Li C, Xue W, Zhang M, Li Z. Frequent Mutations in Natural Killer/T Cell Lymphoma. *Cell Physiol Biochem*. 2018;49(1):1-16.
126. Zhang J, Jima D, Moffitt AB, Liu Q, Czader M, Hsi ED, et al. The genomic landscape of mantle cell lymphoma is related to the epigenetically determined chromatin state of normal B cells. *Blood*. 2014;123(19):2988-96.
127. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood*. 2013;122(7):1256-65.
128. Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, Forster J, et al. Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. *PLoS One*. 2013;8(12):e83244.
129. Coppe A, Andersson EI, Binatti A, Gasparini VR, Bortoluzzi S, Clemente M, et al. Genomic landscape characterization of large granular lymphocyte leukemia with a systems genetics approach. *Leukemia*. 2017;31(5):1243-6.
130. Pasqualucci L, Compagno M, Houldsworth J, Monti S, Grunn A, Nandula SV, et al. Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J Exp Med*. 2006;203(2):311-7.
131. Shah N, Sukumar S. The Hox genes and their roles in oncogenesis. *Nat Rev Cancer*. 2010;10(5):361-71.
132. Stratthdee G, Holyoake TL, Sim A, Parker A, Oscier DG, Melo JV, et al. Inactivation of HOXA genes by hypermethylation in myeloid and lymphoid malignancy is frequent and associated with poor prognosis. *Clin Cancer Res*. 2007;13(17):5048-55.

133. Chambwe N, Kormaksson M, Geng H, De S, Michor F, Johnson NA, et al. Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes. *Blood*. 2014;123(11):1699-708.
134. Choi JH, Li Y, Guo J, Pei L, Rauch TA, Kramer RS, et al. Genome-wide DNA methylation maps in follicular lymphoma cells determined by methylation-enriched bisulfite sequencing. *PLoS One*. 2010;5(9).
135. Kanduri M, Sander B, Ntoufa S, Papakonstantinou N, Sutton LA, Stamatopoulos K, et al. A key role for EZH2 in epigenetic silencing of HOX genes in mantle cell lymphoma. *Epigenetics*. 2013;8(12):1280-8.
136. Velichutina I, Shaknovich R, Geng H, Johnson NA, Gascoyne RD, Melnick AM, et al. EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood*. 2010;116(24):5247-55.
137. Choi HJ, Lee HB, Park HK, Cho SM, Han HJ, Lee SJ, et al. EZH2 Expression in Naturally Occurring Canine Tumors. *Comp Med*. 2018;68(2):148-55.
138. Wansleben S, Peres J, Hare S, Goding CR, Prince S. T-box transcription factors in cancer biology. *Biochim Biophys Acta*. 2014;1846(2):380-91.
139. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene*. 2017;36(11):1461-73.
140. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*. 2011;12(2):R18.
141. Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylation. *Hum Mol Genet*. 2007;16 Spec No 1:R50-9.

142. Takamaru H, Yamamoto E, Suzuki H, Nojima M, Maruyama R, Yamano HO, et al. Aberrant methylation of RASGRF1 is associated with an epigenetic field defect and increased risk of gastric cancer. *Cancer Prev Res (Phila)*. 2012;5(10):1203-12.
143. Tan AC, Jimeno A, Lin SH, Wheelhouse J, Chan F, Solomon A, et al. Characterizing DNA methylation patterns in pancreatic cancer genome. *Mol Oncol*. 2009;3(5-6):425-38.
144. Chen H, Xu Z, Yang B, Zhou X, Kong H. RASGRF1 Hypermethylation, a Putative Biomarker of Colorectal Cancer. *Ann Clin Lab Sci*. 2018;48(1):3-10.
145. Chen H, Suzuki M, Nakamura Y, Ohira M, Ando S, Iida T, et al. Aberrant methylation of RASGRF2 and RASSF1A in human non-small cell lung cancer. *Oncol Rep*. 2006;15(5):1281-5.
146. Kropp PA, Gannon M. Onecut transcription factors in development and disease. *Trends Dev Biol*. 2016;9:43-57.
147. Yassi M, Shams Davodly E, Mojtabanezhad Shariatpanahi A, Heidari M, Dayyani M, Heravi-Moussavi A, et al. DMR Fusion: A differentially methylated region detection tool based on the ranked fusion method. *Genomics*. 2018;110(6):366-74.
148. Mishra NK, Guda C. Genome-wide DNA methylation analysis reveals molecular subtypes of pancreatic cancer. *Oncotarget*. 2017;8(17):28990-9012.
149. Kuasne H, Colus IM, Busso AF, Hernandez-Vargas H, Barros-Filho MC, Marchi FA, et al. Genome-wide methylation and transcriptome analysis in penile carcinoma: uncovering new molecular markers. *Clin Epigenetics*. 2015;7:46.
150. Lai HC, Lin YW, Huang TH, Yan P, Huang RL, Wang HC, et al. Identification of novel DNA methylation markers in cervical cancer. *Int J Cancer*. 2008;123(1):161-7.

151. Song MA, Tiirikainen M, Kwee S, Okimoto G, Yu H, Wong LL. Elucidating the landscape of aberrant DNA methylation in hepatocellular carcinoma. *PLoS One*. 2013;8(2):e55761.
152. Thompson N, Gesina E, Scheinert P, Bucher P, Grapin-Botton A. RNA profiling and chromatin immunoprecipitation-sequencing reveal that PTF1a stabilizes pancreas progenitor identity via the control of MNX1/HLXB9 and a network of other transcription factors. *Mol Cell Biol*. 2012;32(6):1189-99.
153. Li J, Duns G, Westers H, Sijmons R, van den Berg A, Kok K. SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget*. 2016;7(31):50719-34.
154. Sakthikumar S, Elvers I, Kim J, Arendt ML, Thomas R, Turner-Maier J, et al. SETD2 is recurrently mutated in whole-exome sequenced canine osteosarcoma. *Cancer Res*. 2018.
155. Moffitt AB, Ondrejka SL, McKinney M, Rempel RE, Goodlad JR, Teh CH, et al. Enteropathy-associated T cell lymphoma subtypes are characterized by loss of function of SETD2. *J Exp Med*. 2017;214(5):1371-86.
156. Pfister SX, Markkanen E, Jiang Y, Sarkar S, Woodcock M, Orlando G, et al. Inhibiting WEE1 Selectively Kills Histone H3K36me3-Deficient Cancers by dNTP Starvation. *Cancer Cell*. 2015;28(5):557-68.
157. Sinner D, Kordich JJ, Spence JR, Opoka R, Rankin S, Lin SC, et al. Sox17 and Sox4 differentially regulate beta-catenin/T-cell factor activity and proliferation of colon carcinoma cells. *Mol Cell Biol*. 2007;27(22):7802-15.

158. Fu DY, Tan HS, Wei JL, Zhu CR, Jiang JX, Zhu YX, et al. Decreased expression of SOX17 is associated with tumor progression and poor prognosis in breast cancer. *Tumour Biol.* 2015;36(10):8025-34.
159. Zhang Y, Bao W, Wang K, Lu W, Wang H, Tong H, et al. SOX17 is a tumor suppressor in endometrial cancer. *Oncotarget.* 2016;7(46):76036-46.
160. Martin-Subero JI, Kreuz M, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E, et al. New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood.* 2009;113(11):2488-97.
161. Simpson HM, Khan RZ, Song C, Sharma D, Sadashivaiah K, Furusawa A, et al. Concurrent Mutations in ATM and Genes Associated with Common gamma Chain Signaling in Peripheral T Cell Lymphoma. *PLoS One.* 2015;10(11):e0141906.
162. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet.* 2011;43(9):830-7.
163. Ghanem T, Bracken J, Kasem A, Jiang WG, Mokbel K. mRNA expression of DOK1-6 in human breast cancer. *World J Clin Oncol.* 2014;5(2):156-63.
164. Leong SH, Lwin KM, Lee SS, Ng WH, Ng KM, Tan SY, et al. Chromosomal breaks at FRA18C: association with reduced DOK6 expression, altered oncogenic signaling and increased gastric cancer survival. *NPJ Precis Oncol.* 2017;1(1):9.
165. Lautscham EM, Kessler M, Ernst T, Willimzig L, Neiger R. Comparison of a CHOP-LAsp-based protocol with and without maintenance for canine multicentric lymphoma. *Vet Rec.* 2017;180(12):303.

166. Curran K, Thamm DH. Retrospective analysis for treatment of naive canine multicentric lymphoma with a 15-week, maintenance-free CHOP protocol. *Vet Comp Oncol.* 2016;14 Suppl 1:147-55.
167. Burton JH, Garrett-Mayer E, Thamm DH. Evaluation of a 15-week CHOP protocol for the treatment of canine multicentric lymphoma. *Vet Comp Oncol.* 2013;11(4):306-15.
168. Childress MO, Ramos-Vara JA, Ruple A. Retrospective analysis of factors affecting clinical outcome following CHOP-based chemotherapy in dogs with primary nodal diffuse large B-cell lymphoma. *Vet Comp Oncol.* 2018;16(1):E159-E68.
169. Sorenmo K, Overley B, Krick E, Ferrara T, LaBlanc A, Shofer F. Outcome and toxicity associated with a dose-intensified, maintenance-free CHOP-based chemotherapy protocol in canine lymphoma: 130 cases. *Vet Comp Oncol.* 2010;8(3):196-208.
170. Curran KM, Schaffer PA, Frank CB, Lana SE, Hamil LE, Burton JH, et al. BCL2 and MYC are expressed at high levels in canine diffuse large B-cell lymphoma but are not predictive for outcome in dogs treated with CHOP chemotherapy. *Vet Comp Oncol.* 2017;15(4):1269-79.
171. Rout ED, Avery PR. Lymphoid Neoplasia: Correlations Between Morphology and Flow Cytometry. *Vet Clin North Am Small Anim Pract.* 2017;47(1):53-70.
172. Ponce F, Marchal T, Magnol JP, Turinelli V, Ledieu D, Bonnefont C, et al. A morphological study of 608 cases of canine malignant lymphoma in France with a

focus on comparative similarities between canine and human lymphoma morphology. *Vet Pathol.* 2010;47(3):414-33.

173. Vezzali E, Parodi AL, Marcato PS, Bettini G. Histopathologic classification of 171 cases of canine and feline non-Hodgkin lymphoma according to the WHO. *Vet Comp Oncol.* 2010;8(1):38-49.

174. Valli VE, San Myint M, Barthel A, Bienzle D, Caswell J, Colbatzky F, et al. Classification of canine malignant lymphomas according to the World Health Organization criteria. *Vet Pathol.* 2011;48(1):198-211.

175. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet.* 2010;11(3):191-203.

176. Wong Doo N, Makalic E, Joo JE, Vajdic CM, Schmidt DF, Wong EM, et al. Global measures of peripheral blood-derived DNA methylation as a risk factor in the development of mature B-cell neoplasms. *Epigenomics.* 2016;8(1):55-66.

177. Anninga JK, Gelderblom H, Fiocco M, Kroep JR, Taminiau AH, Hogendoorn PC, et al. Chemotherapeutic adjuvant treatment for osteosarcoma: where do we stand? *Eur J Cancer.* 2011;47(16):2431-45.

178. Rosenberger JA, Pablo NV, Crawford PC. Prevalence of and intrinsic risk factors for appendicular osteosarcoma in dogs: 179 cases (1996-2005). *J Am Vet Med Assoc.* 2007;231(7):1076-80.

179. Ottaviani G, Jaffe N. The epidemiology of osteosarcoma. *Cancer Treat Res.* 2009;152:3-13.

180. Makielski KM, Mills LJ, Sarver AL, Henson MS, Spector LG, Naik S, et al. Risk Factors for Development of Canine and Human Osteosarcoma: A Comparative Review. *Vet Sci*. 2019;6(2).
181. Simpson S, Dunning MD, de Brot S, Grau-Roma L, Mongan NP, Rutland CS. Comparative review of human and canine osteosarcoma: morphology, epidemiology, prognosis, treatment and genetics. *Acta Vet Scand*. 2017;59(1):71.
182. Fenger JM, London CA, Kisseberth WC. Canine osteosarcoma: a naturally occurring disease to inform pediatric oncology. *ILAR J*. 2014;55(1):69-85.
183. Morello E, Martano M, Buracco P. Biology, diagnosis and treatment of canine appendicular osteosarcoma: similarities and differences with human osteosarcoma. *Vet J*. 2011;189(3):268-77.
184. Mueller F, Fuchs B, Kaser-Hotz B. Comparative biology of human and canine osteosarcoma. *Anticancer Res*. 2007;27(1A):155-64.
185. Mirabello L, Pfeiffer R, Murphy G, Daw NC, Patino-Garcia A, Troisi RJ, et al. Height at diagnosis and birth-weight as risk factors for osteosarcoma. *Cancer Causes Control*. 2011;22(6):899-908.
186. Gillette SM, Gillette EL, Powers BE, Withrow SJ. Radiation-induced osteosarcoma in dogs after external beam or intraoperative radiation therapy. *Cancer Res*. 1990;50(1):54-7.
187. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E, et al. Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Rep*. 2014;7(1):104-12.

188. Perry JA, Kiezun A, Tonzi P, Van Allen EM, Carter SL, Baca SC, et al. Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc Natl Acad Sci U S A*. 2014;111(51):E5564-73.
189. Broeckx BJ, Hitte C, Coopman F, Verhoeven GE, De Keulenaer S, De Meester E, et al. Improved canine exome designs, featuring ncRNAs and increased coverage of protein coding genes. *Sci Rep*. 2015;5:12810.
190. Griffith M, Griffith OL, Smith SM, Ramu A, Callaway MB, Brummett AM, et al. Genome Modeling System: A Knowledge Management Platform for Genomics. *PLoS Comput Biol*. 2015;11(7):e1004274.
191. Selmic LE, Burton JH, Thamm DH, Withrow SJ, Lana SE. Comparison of carboplatin and doxorubicin-based chemotherapy protocols in 470 dogs after amputation for treatment of appendicular osteosarcoma. *J Vet Intern Med*. 2014;28(2):554-63.
192. Moriarity BS, Otto GM, Rahrman EP, Rathe SK, Wolf NK, Weg MT, et al. A Sleeping Beauty forward genetic screen identifies new genes and pathways driving osteosarcoma development and metastasis. *Nat Genet*. 2015;47(6):615-24.
193. Bendris N, Schmid SL. Endocytosis, Metastasis and Beyond: Multiple Facets of SNX9. *Trends Cell Biol*. 2017;27(3):189-200.
194. Tuna M, Knuutila S, Mills GB. Uniparental disomy in cancer. *Trends Mol Med*. 2009;15(3):120-8.

195. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011;144(1):27-40.
196. Kinsella M, Patel A, Bafna V. The elusive evidence for chromothripsis. *Nucleic Acids Res*. 2014;42(13):8231-42.
197. Scotland PB, Heath JL, Conway AE, Porter NB, Armstrong MB, Walker JA, et al. The PICALM protein plays a key role in iron homeostasis and cell proliferation. *PLoS One*. 2012;7(8):e44252.
198. Mercer JL, Argus JP, Crabtree DM, Keenan MM, Wilks MQ, Chi JT, et al. Modulation of PICALM Levels Perturbs Cellular Cholesterol Homeostasis. *PLoS One*. 2015;10(6):e0129776.
199. Moreau K, Fleming A, Imarisio S, Lopez Ramirez A, Mercer JL, Jimenez-Sanchez M, et al. PICALM modulates autophagy activity and tau accumulation. *Nat Commun*. 2014;5:4998.
200. Canalis E, Schilling L, Yee SP, Lee SK, Zanotti S. Hajdu Cheney Mouse Mutants Exhibit Osteopenia, Increased Osteoclastogenesis, and Bone Resorption. *J Biol Chem*. 2016;291(4):1538-51.
201. Tanaka M, Setoguchi T, Hirotsu M, Gao H, Sasaki H, Matsunoshita Y, et al. Inhibition of Notch pathway prevents osteosarcoma growth by cell cycle regulation. *Br J Cancer*. 2009;100(12):1957-65.
202. Spina R, Filocamo G, Iaccino E, Scicchitano S, Lupia M, Chiarella E, et al. Critical role of zinc finger protein 521 in the control of growth, clonogenicity and tumorigenic potential of medulloblastoma cells. *Oncotarget*. 2013;4(8):1280-92.

203. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007;446(7137):758-64.
204. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, et al. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci U S A*. 2003;100(14):8424-9.
205. Mendoza S, Konishi T, Dernell WS, Withrow SJ, Miller CW. Status of the p53, Rb and MDM2 genes in canine osteosarcoma. *Anticancer Res*. 1998;18(6A):4449-53.
206. van Leeuwen IS, Cornelisse CJ, Misdorp W, Goedegebuure SA, Kirpensteijn J, Rutteman GR. P53 gene mutations in osteosarcomas in the dog. *Cancer Lett*. 1997;111(1-2):173-8.
207. Johnson AS, Couto CG, Weghorst CM. Mutation of the p53 tumor suppressor gene in spontaneously occurring osteosarcomas of the dog. *Carcinogenesis*. 1998;19(1):213-7.
208. Kirpensteijn J, Kik M, Teske E, Rutteman GR. TP53 gene mutations in canine osteosarcoma. *Vet Surg*. 2008;37(5):454-60.
209. Mirabello L, Yeager M, Mai PL, Gastier-Foster JM, Gorlick R, Khanna C, et al. Germline TP53 variants and susceptibility to osteosarcoma. *J Natl Cancer Inst*. 2015;107(7).
210. He Y, de Castro LF, Shin MH, Dubois W, Yang HH, Jiang S, et al. p53 loss increases the osteogenic differentiation of bone marrow stromal cells. *Stem Cells*. 2015;33(4):1304-19.

211. Ribi S, Baumhoer D, Lee K, Edison, Teo AS, Madan B, et al. TP53 intron 1 hotspot rearrangements are specific to sporadic osteosarcoma and can cause Li-Fraumeni syndrome. *Oncotarget*. 2015;6(10):7727-40.
212. Sayles LC, Breese MR, Koehne AL, Leung SG, Lee AG, Liu HY, et al. Genome-Informed Targeted Therapy for Osteosarcoma. *Cancer Discov*. 2019;9(1):46-63.
213. Porta C, Paglino C, Mosca A. Targeting PI3K/Akt/mTOR Signaling in Cancer. *Front Oncol*. 2014;4:64.
214. Demetri GD, Chawla SP, Ray-Coquard I, Le Cesne A, Staddon AP, Milhem MM, et al. Results of an international randomized phase III trial of the mammalian target of rapamycin inhibitor ridaforolimus versus placebo to control metastatic sarcomas in patients after benefit from prior chemotherapy. *J Clin Oncol*. 2013;31(19):2485-92.
215. Mohseny AB, Szuhai K, Romeo S, Buddingh EP, Briaire-de Bruijn I, de Jong D, et al. Osteosarcoma originates from mesenchymal stem cells in consequence of aneuploidization and genomic loss of Cdkn2. *J Pathol*. 2009;219(3):294-305.
216. Mohseny AB, Tieken C, van der Velden PA, Szuhai K, de Andrea C, Hogendoorn PC, et al. Small deletions but not methylation underlie CDKN2A/p16 loss of expression in conventional osteosarcoma. *Genes Chromosomes Cancer*. 2010;49(12):1095-103.
217. Hedan B, Thomas R, Motsinger-Reif A, Abadie J, Andre C, Cullen J, et al. Molecular cytogenetic characterization of canine histiocytic sarcoma: A spontaneous model for human histiocytic cancer identifies deletion of tumor suppressor genes

and highlights influence of genetic background on tumor behavior. BMC Cancer. 2011;11:201.

218. Aguirre-Hernandez J, Milne BS, Queen C, O'Brien PC, Hoather T, Haugland S, et al. Disruption of chromosome 11 in canine fibrosarcomas highlights an unusual variability of CDKN2B in dogs. BMC Vet Res. 2009;5:27.

219. Thomas R, Seiser EL, Motsinger-Reif A, Borst L, Valli VE, Kelley K, et al. Refining tumor-associated aneuploidy through 'genomic recoding' of recurrent DNA copy number aberrations in 150 canine non-Hodgkin lymphomas. Leuk Lymphoma. 2011;52(7):1321-35.

220. Karlsson EK, Sigurdsson S, Ivansson E, Thomas R, Elvers I, Wright J, et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. Genome Biol. 2013;14(12):R132.

221. Shearin AL, Hedan B, Cadieu E, Erich SA, Schmidt EV, Faden DL, et al. The MTAP-CDKN2A locus confers susceptibility to a naturally occurring canine cancer. Cancer Epidemiol Biomarkers Prev. 2012;21(7):1019-27.

222. Palermo V, Rinalducci S, Sanchez M, Grillini F, Sommers JA, Brosh RM, Jr., et al. CDK1 phosphorylates WRN at collapsed replication forks. Nat Commun. 2016;7:12880.

223. Oshima J, Huang S, Pae C, Campisi J, Schiestl RH. Lack of WRN results in extensive deletion at nonhomologous joining ends. Cancer Res. 2002;62(2):547-51.

224. Kim M, Kim S, Lee SH, Kim W, Sohn MJ, Kim HS, et al. Merlin inhibits Wnt/beta-catenin signaling by blocking LRP6 phosphorylation. Cell Death Differ. 2016;23(10):1638-47.

225. Stemmer-Rachamimov AO, Nielsen GP, Rosenberg AE, Louis DN, Jones D, Ramesh V, et al. The NF2 gene and merlin protein in human osteosarcomas. *Neurogenetics*. 1998;2(1):73-4.
226. Yazawa M, Okuda M, Setoguchi A, Nishimura R, Sasaki N, Hasegawa A, et al. Measurement of telomerase activity in dog tumors. *J Vet Med Sci*. 1999;61(10):1125-9.
227. Shay JW, Bacchetti S. A survey of telomerase activity in human cancer. *Eur J Cancer*. 1997;33(5):787-91.
228. Ulaner GA, Huang HY, Otero J, Zhao Z, Ben-Porat L, Satagopan JM, et al. Absence of a telomere maintenance mechanism as a favorable prognostic factor in patients with osteosarcoma. *Cancer Res*. 2003;63(8):1759-63.
229. Barthel FP, Wei W, Tang M, Martinez-Ledesma E, Hu X, Amin SB, et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat Genet*. 2017;49(3):349-57.
230. McCleese JK, Bear MD, Fossey SL, Mihalek RM, Foley KP, Ying W, et al. The novel HSP90 inhibitor STA-1474 exhibits biologic activity against osteosarcoma cell lines. *Int J Cancer*. 2009;125(12):2792-801.
231. London CA, Acquaviva J, Smith DL, Sequeira M, Ogawa LS, Gardner HL, et al. Consecutive Day HSP90 Inhibitor Administration Improves Efficacy in Murine Models of KIT-Driven Malignancies and Canine Mast Cell Tumors. *Clin Cancer Res*. 2018;24(24):6396-407.
232. Hajra KM, Fearon ER. Cadherin and catenin alterations in human cancer. *Genes Chromosomes Cancer*. 2002;34(3):255-68.

233. Hill TP, Spater D, Taketo MM, Birchmeier W, Hartmann C. Canonical Wnt/beta-catenin signaling prevents osteoblasts from differentiating into chondrocytes. *Dev Cell*. 2005;8(5):727-38.
234. Stein TJ, Holmes KE, Muthuswamy A, Thompson V, Huelsmeyer MK. Characterization of beta-catenin expression in canine osteosarcoma. *Vet Comp Oncol*. 2011;9(1):65-73.
235. Haydon RC, Deyrup A, Ishikawa A, Heck R, Jiang W, Zhou L, et al. Cytoplasmic and/or nuclear accumulation of the beta-catenin protein is a frequent event in human osteosarcoma. *Int J Cancer*. 2002;102(4):338-42.
236. Stratford EW, Daffinrud J, Munthe E, Castro R, Waaler J, Krauss S, et al. The tankyrase-specific inhibitor JW74 affects cell cycle progression and induces apoptosis and differentiation in osteosarcoma cell lines. *Cancer Med*. 2014;3(1):36-46.
237. Chang J, Li Y, Wang X, Hu S, Wang H, Shi Q, et al. Polyphyllin I suppresses human osteosarcoma growth by inactivation of Wnt/beta-catenin pathway in vitro and in vivo. *Sci Rep*. 2017;7(1):7605.
238. Krishnamurthy N, Kurzrock R. Targeting the Wnt/beta-catenin pathway in cancer: Update on effectors and inhibitors. *Cancer Treat Rev*. 2018;62:50-60.
239. Huynh H, Ong R, Goh KY, Lee LY, Puehler F, Scholz A, et al. Sorafenib/MEK inhibitor combination inhibits tumor growth and the Wnt/betacatenin pathway in xenograft models of hepatocellular carcinoma. *Int J Oncol*. 2019;54(3):1123-33.

240. London C, Mathie T, Stingle N, Clifford C, Haney S, Klein MK, et al. Preliminary evidence for biologic activity of toceranib phosphate (Palladia((R))) in solid tumours. *Vet Comp Oncol.* 2012;10(3):194-205.
241. Laver T, London CA, Vail DM, Biller BJ, Coy J, Thamm DH. Prospective evaluation of toceranib phosphate in metastatic canine osteosarcoma. *Vet Comp Oncol.* 2018;16(1):E23-E9.
242. Kim C, Matsuyama A, Mutsaers AJ, Woods JP. Retrospective evaluation of toceranib (Palladia) treatment for canine metastatic appendicular osteosarcoma. *Can Vet J.* 2017;58(10):1059-64.
243. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature.* 2005;436(7047):117-22.
244. Kim E, Zucconi BE, Wu M, Nocco SE, Meyers DJ, McGee JS, et al. MITF Expression Predicts Therapeutic Vulnerability to p300 Inhibition in Human Melanoma. *Cancer Res.* 2019;79(10):2649-61.
245. Flint AF, U'Ren L, Legare ME, Withrow SJ, Dernell W, Hanneman WH. Overexpression of the erbB-2 proto-oncogene in canine osteosarcoma cell lines and tumors. *Vet Pathol.* 2004;41(3):291-6.
246. Wang W, Zhao HF, Yao TF, Gong H. Advanced development of ErbB family-targeted therapies in osteosarcoma treatment. *Invest New Drugs.* 2019;37(1):175-83.
247. Ebb D, Meyers P, Grier H, Bernstein M, Gorlick R, Lipshultz SE, et al. Phase II trial of trastuzumab in combination with cytotoxic chemotherapy for treatment of

- metastatic osteosarcoma with human epidermal growth factor receptor 2 overexpression: a report from the children's oncology group. *J Clin Oncol*. 2012;30(20):2545-51.
248. Mason NJ, Gnanandarajah JS, Engiles JB, Gray F, Laughlin D, Gaurnier-Hausser A, et al. Immunotherapy with a HER2-Targeting *Listeria* Induces HER2-Specific Immunity and Demonstrates Potential Therapeutic Effects in a Phase I Trial in Canine Osteosarcoma. *Clin Cancer Res*. 2016;22(17):4380-90.
249. Miller CW, Ikezoe T, Krug U, Hofmann WK, Tavor S, Vegesna V, et al. Mutations of the *CHK2* gene are found in some osteosarcomas, but are rare in breast, lung, and ovarian tumors. *Genes Chromosomes Cancer*. 2002;33(1):17-21.
250. Kovac M, Blattmann C, Ribi S, Smida J, Mueller NS, Engert F, et al. Exome sequencing of osteosarcoma reveals mutation signatures reminiscent of *BRCA* deficiency. *Nat Commun*. 2015;6:8940.
251. Zhou C, Jiang CQ, Zong Z, Lin JC, Lao LF. miR-146a promotes growth of osteosarcoma cells by targeting *ZNRF3*/*GSK-3beta*/*beta*-catenin signaling pathway. *Oncotarget*. 2017;8(43):74276-86.
252. Petrilli AM, Fernandez-Valle C. Role of Merlin/*NF2* inactivation in tumor biology. *Oncogene*. 2016;35(5):537-48.
253. Cerveira N, Bizarro S, Teixeira MR. *MLL-SEPTIN* gene fusions in hematological malignancies. *Biol Chem*. 2011;392(8-9):713-24.
254. Evans DGR, Salvador H, Chang VY, Erez A, Voss SD, Druker H, et al. Cancer and Central Nervous System Tumor Surveillance in Pediatric Neurofibromatosis 2 and Related Disorders. *Clin Cancer Res*. 2017;23(12):e54-e61.

255. Sens-Abuazar C, Napolitano EFE, Osorio CA, Krepischi AC, Ricca TI, Castro NP, et al. Down-regulation of ANAPC13 and CLTCL1: Early Events in the Progression of Preinvasive Ductal Carcinoma of the Breast. *Transl Oncol*. 2012;5(2):113-23.
256. Al-Hebshi NN, Li S, Nasher AT, El-Setouhy M, Alsanosi R, Blancato J, et al. Exome sequencing of oral squamous cell carcinoma in users of Arabian snuff reveals novel candidates for driver genes. *Int J Cancer*. 2016;139(2):363-72.
257. Lafleur EA, Jia SF, Worth LL, Zhou Z, Owen-Schaub LB, Kleinerman ES. Interleukin (IL)-12 and IL-12 gene transfer up-regulate Fas expression in human osteosarcoma and breast cancer cells. *Cancer Res*. 2001;61(10):4066-71.
258. Lafleur EA, Koshkina NV, Stewart J, Jia SF, Worth LL, Duan X, et al. Increased Fas expression reduces the metastatic potential of human osteosarcoma cells. *Clin Cancer Res*. 2004;10(23):8114-9.
259. Worth LL, Lafleur EA, Jia SF, Kleinerman ES. Fas expression inversely correlates with metastatic potential in osteosarcoma cells. *Oncol Rep*. 2002;9(4):823-7.
260. Cai H, Kumar N, Bagheri HC, von Mering C, Robinson MD, Baudis M. Chromothripsis-like patterns are recurring but heterogeneously distributed features in a survey of 22,347 cancer genome screens. *BMC Genomics*. 2014;15:82.
261. Stone DM, Jacky PB, Hancock DD, Prieur DJ. Chromosomal fragile site expression in dogs: I. Breed specific differences. *Am J Med Genet*. 1991;40(2):214-22.
262. Stone DM, Stephens KE. Bromodeoxyuridine induces chromosomal fragile sites in the canine genome. *Am J Med Genet*. 1993;46(2):198-202.

263. Caudell D, Aplan PD. The role of CALM-AF10 gene fusion in acute leukemia. *Leukemia*. 2008;22(4):678-85.
264. Shao YW, Wood GA, Lu J, Tang QL, Liu J, Molyneux S, et al. Cross-species genomics identifies DLG2 as a tumor suppressor in osteosarcoma. *Oncogene*. 2019;38(2):291-8.
265. Li W, Ai N, Wang S, Bhattacharya N, Vrbanac V, Collins M, et al. GRK3 is essential for metastatic cells and promotes prostate tumor progression. *Proc Natl Acad Sci U S A*. 2014;111(4):1521-6.
266. Thoenen E, Curl A, Iwakuma T. TP53 in bone and soft tissue sarcomas. *Pharmacol Ther*. 2019.
267. Mooney M, Bond J, Monks N, Eugster E, Cherba D, Berlinski P, et al. Comparative RNA-Seq and microarray analysis of gene expression changes in B-cell lymphomas of *Canis familiaris*. *PLoS One*. 2013;8(4):e61088.
268. Giannuzzi D, Marconato L, Cascione L, Comazzi S, Elgendy R, Pegolo S, et al. Mutational landscape of canine B-cell lymphoma profiled at single nucleotide resolution by RNA-seq. *PLoS One*. 2019;14(4):e0215154.
269. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N Engl J Med*. 2018;378(15):1396-407.
270. De S, Shaknovich R, Riester M, Elemento O, Geng H, Kormaksson M, et al. Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet*. 2013;9(1):e1003137.

VITA

Shirley Chu is from Vancouver, British Columbia. She graduated from the University of British Columbia with a Bachelor of Science with a major in Microbiology and Immunology. She then earned a Doctorate of Veterinary Medicine degree from the Western College of Veterinary Medicine at the University of Saskatchewan. She completed a small animal rotating internship and medical oncology residency at the University of Missouri. During her residency she enrolled in the Doctorate of Philosophy program in Veterinary Pathobiology.