Novel Approaches to the Molecular Characterization of Primary Canine Osteosarcoma

by

Rebecca L. Nance

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 5, 2023

Keywords: cancer, osteosarcoma, dog, precision medicine, transcriptomic sequencing

Approved by

Bruce F. Smith, Chair, Professor, Department of Pathobiology
Richard C. Bird, Professor, Department of Pathobiology
Maninder Sandey, Associate Professor, Department of Pathobiology
Xu Wang, Associate Professor, Department of Pathobiology

Abstract

Osteosarcoma (OSA) is an aggressive malignant bone tumor that often affects pediatric humans as well as dogs. Canine OSA shares many similarities with the human condition, including clinical presentation and molecular profiles, and therefore serves as an excellent model to study the disease. OSA is a difficult tumor to dissociate and sequence due to its bony and brittle composition. Furthermore, it is characterized by extreme genetic complexity and significant intra- and inter-tumoral heterogeneity. While great strides have been made in identifying key mutations driving carcinogenesis, the treatment and prognosis for OSA have remained largely unchanged for 40 years. New approaches to studying OSA are critical for improving patient outcomes for both humans and dogs.

Differential gene expression analysis represents one approach to analyzing important differences in the molecular profiles between normal and tumor tissue. A prerequisite to successful sequencing and differential analysis is the isolation of high-quality RNA from both normal and neoplastic tissue. In the case of OSA, normal bone corresponds to the non-neoplastic and pre-cancerous tissue representative of osteosarcoma. Bone is a dynamic tissue consisting of many different cell types embedded within a rigid matrix. Removal of the bone marrow is an important consideration for isolating RNA that is unique to OSA progenitors for an appropriate differential gene expression analysis. Therefore, a method to isolate RNA from normal canine bone was first established for subsequent transcriptomic sequencing of both unaffected tissue and tumorous lesions.

Transcriptomic bulk sequencing of primary OSA provides a global view of the changes involved in OSA tumorigenesis. Differential gene expression analysis identifies dysregulated genes in the tumors compared to the normal bone. Patient-matched OSA and normal bone were

collected and sequenced to identify genes that were commonly dysregulated among the group. However, this approach requires a large sample size to accommodate the statistical analysis and fails to account for inter-individual differences. To circumvent this, we supplemented the group differential profile with a novel individual-level analysis by deriving individual fold-change differences from the group's significant genes. The results confirm the hyper-variability in OSA and the need to study OSA at the individual level.

Single-cell sequencing (SCseq) has emerged in the past five years as a valuable tool for interrogating the molecular changes occurring in diseased tissue at the level of individual cells. However, an essential requirement for this approach includes highly viable cells (>60% live cells) dissociated from the primary tissue. OSA has proved difficult to manipulate and dissociate due to the matrix composition, resulting in poor cell viability unsuitable for SCseq. Single-nuclei isolation and sequencing have recently evolved as an alternative using 10x Genomics technology. This approach requires high-quality nuclei but overcomes the need for viable whole cells as input. With protocol adjustments, we have successfully isolated high-quality nuclei from a primary canine OSA used for single-nuclei multiome sequencing. Multiome sequencing includes transcriptomic profiling using RNA sequencing as well as epigenomic profiling using ATAC (Assay for Transposase-Accessible Chromatin) sequencing. This dual sequencing approach was used to characterize the tumor microenvironment and elucidate changes occurring at single-nuclei resolution in an individual patient.

This dissertation provides a thorough background of canine OSA, describes a novel method to isolate RNA from canine bone for proper comparative sequencing analysis, presents bulk RNA sequencing of 7 canine OSA tumors along with an individualized approach to differential gene analysis, and concludes with single-nuclei multiome sequencing of a primary

OSA tumor to characterize the tumor microenvironment. Importantly, these approaches can be translated to study human OSA for comparison to canine OSA molecular pathways and for the development of more targeted and effective therapies.

Acknowledgments

First and foremost, I am forever grateful for the guidance, patience, encouragement, and teachings of my mentor, Dr. Bruce Smith. Our regular scientific discussions inspired my intellectual curiosity and drove my passion for research. He always welcomed and encouraged my endless questions and without him, I would not be the scientist I am today. Dr. Smith provided an environment that supported my mental, physical, and emotional well-being and importantly, he gave me the freedom to pursue my own research interests. I could not have asked for a better advisor and friend.

Secondly, I am thankful for Dr. Payal Agarwal for her wisdom, experience, and training in the laboratory. She taught me the foundations of many experimental procedures and techniques that I will carry throughout my career. Earning her confidence and trust in the lab was, and remains, one of my proudest achievements.

I am also grateful for the many teachers and professors I've had along the way who inspired and supported my endless pursuit for knowledge and my relentless questions. Special thanks to my committee members, Drs. Richard C. Bird, Xu Wang, and Maninder Sandey, for their support, guidance, and insightful discussion. Additionally, I am appreciative of the mentorship and friendship gained from Drs. Emily Graff, Amanda Gross, and Payal Agarwal, who have continually inspired me by showing what a woman, wife, and mother in science looks like.

Finally, I could not have accomplished this pursuit without my dog, Stella. She was my unwavering source of love, support, and companionship. When I was stressed or discouraged, she never failed to put a smile on my face and peace in my heart.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AO/PI	Acridine orange/propidium iodide
ATAC	Assay for Transposase-Accessible Chromatin
BCP	Bromochloropropane
DEG	Differentially expressed gene(s)
FDA	U.S. Food and Drug Administration
FDR	False discovery rate
GEM	Gel bead-in-emulsion
GEX	Gene expression
GO	Gene ontology
GWAS	Genome-wide association study
H&E	Hematoxylin and eosin
lncRNA	Long non-coding RNA
LSI	Latent semantic indexing
OSA	Osteosarcoma
PBS	Phosphate buffered saline
PCA	Principal component analysis
RIN	RNA integrity number
RT-PCR	Reverse-transcriptase polymerase chain reaction
SCseq	Single-cell sequencing
SNP	Single-nucleotide polymorphism
SVD	Singular value decomposition
TF-IDF	Term frequency inverse document frequency

TME	Tumor microenvironment
UMI	Unique molecular identifiers
WGS	Whole-genome sequencing

CHAPTER 1

Review of Literature

Reprinted (in part) from Nance RL, Sajib AM, Smith BF. Canine models of human cancer: Bridging the gap to improve precision medicine. Prog Mol Biol Transl Sci. 2022;189(1):67-99.

1.1 Cancer and the precision medicine approach to treatment

Cancer is a highly complex and heterogeneous disease characterized by extreme genomic instability, sustained cell proliferation, suppressed cell death, induced angiogenesis, stimulated cell invasion, and evaded immune destruction[1]. Mutations in oncogenes and/or tumor suppressor genes contribute to initial aberrant cellular properties and can result in the accumulation of additional genetic changes that drive tumor formation and progression. Our understanding of tumorigenesis has grown exponentially in the past decade as the molecular processes continue to be elucidated, largely due to the efforts of translational research made possible by the reduced cost of next-generation sequencing and the development of new technologies. In recent years, the study and treatment of cancer has evolved from an organ-centric approach to one that seeks to characterize tumors based on their molecular profile. Advances in the field of "omics" technologies have made it possible to integrate genomic, transcriptomic, and proteomic information to create a more comprehensive strategy to improve patient outcomes and lower treatment costs[2]. Extensive sequencing has exposed the overwhelming complexity of cancer and the overall genetic uniqueness of each individual tumor, regardless of type or stage, which often share common mutations in critical genes and dysregulated pathways. This observation gave rise to the concept of precision or personalized medicine, which seeks to overcome both intra- and inter-tumoral heterogeneity as well as inter-individual variation in drug response[3–5]. By providing insight into druggable targets such as pathways, receptors, or genes, critical

therapeutic decisions and the development of novel therapies are guided by the molecular landscape of an individual tumor.

However, many obstacles compromise the adequate implementation of such treatment approaches, including the lack of appropriate infrastructure and communication among medical doctors and research scientists[6]. Additionally, the clinical validity of precision medicine faces some scrutiny due to the lack of appropriate or efficient standardization[7]. As such, the clinical utility is criticized, and additional evidence is needed to support the claim that "omics"-based personalized medicine generates an overall cost reduction[8,9]. There is a critical need for an infrastructure that supports the integration of a patient's genomic data with relevant clinical information such as age, diagnosis, family history, symptom severity, and other confounding variables such as lifestyle, diet, and drug interactions[2].

In spite of these obstacles, many molecular-based treatments have been successfully implemented in humans. In fact, more than 1 out of 4 drugs approved by the U.S. Food and Drug Administration (FDA) since 2014 are precision medicine therapies[10]. According to an FDA news release, the first approved drug for cancer treatment based solely on molecular features rather than tumor type was released in May 2017. The drug, known as pembrolizumab, or more commonly as Keytruda, was granted accelerated approval for the treatment of unresectable or metastatic solid tumors with biomarkers for high microsatellite instability or deficient mismatch repair. Amid a global pandemic in the first half of 2020, the FDA approved a record total of 21 precision drugs for cancer treatment; of these, 7 were characterized by novel molecular entities and 14 were previously approved drugs with expanded molecular indications[11]. These include the tyrosine kinase inhibitor imatinib, which was approved for use in chronic myeloid leukemia tumors bearing the t(9,22)(q34;q11) translocation which generates the fusion gene BCR-ABL1[12].

Breast cancer treatment has improved substantially with the identification of molecular subtypes, including the application of the anti-HER2 kinase inhibitor neratinib, monoclonal antibody trastuzumab that targets the HER2 receptor, and PI3Kα inhibitor alpelisib in patients bearing HER2 amplification and/or overexpression and PIK3CA mutations, respectively[13–16] . These are just a few of the many successful applications of precision medicine in the treatment of cancer that continue to improve patient outcomes.

1.2 Limitations of rodent models

Despite significant progress in precision oncology research and treatment, development is hindered by the extremely low percentage of therapeutics that show similar outcomes in humans and pre-clinical rodent studies[17,18]. Indeed, the rate of effective translation from animal models to clinical trials is less than 8%[19]. The abundant failures of murine models are disappointing, time-consuming, and expensive. While rodent models remain fundamentally important in the study of cancer, they frequently do not accurately recapitulate the full complexity of human cancer, including tumor heterogeneity, microenvironment, accumulated genomic instability, immune influence, and metastatic disease, and therefore suffer significant translational limitations[19–23].

The controlled environment of laboratory rodents is thought to contribute to the inaccurate representation of human disease. These inbred animals are typically homogeneous in terms of genetics, size, age, sex, and diet, and are housed in pathogen-free environments that are unrepresentative of natural environments[24]. The use of young mice is controversial given the age-related changes that occur to the immune system and the propensity of cancer occurrence in older humans[25]. Even patient-derived xenografts with implantation of humanized immune

system components lack critical functions and responses that are representative of human malignancies[24].

In the past decade, the discovery of the dynamic interactions between the immune system and gut microbiome has uncovered a new field of research[26]. The gut microbiome has emerged as an important and influential factor affecting immune function, and consequently, tumorigenesis and response to treatment[27]. In a 2015 study, Sivan et al uncovered significant differences in the growth rate of subcutaneous B16.SIY melanoma in genetically similar mice housed in two different facilities as a result of immune-mediated differences in tumor-specific T cells and intra-tumoral CD8+ cell accumulation. After cohousing the two groups of mice prior to tumor implantation, the team found these differences in tumor growth and immune response disappeared, suggesting an environmental effect. They evaluated the gut microbiome in both groups, performed fecal transplantations, and discovered that *Bifidobacterium* conferred an anti-tumor immune response that was augmented with anti-PDL1 monoclonal antibody immunotherapy[28]. This data suggests that commensal gut microbiota can influence the immune response to tumor and immunotherapy effectiveness and, therefore, should be considered in the selection of disease models.

In addition to the molecular and physiological dissimilarities in rodent models, differences in drug metabolism are also likely to blame for the poor translational outcomes. Tumor response to drugs, particularly chemotherapy, depends on a plethora of elements, including tumor-specific factors such as histology, growth rate, and drug-resistance mechanisms, as well as individual-specific features such as physiology, absorption efficiency, and immunologic response[29]. A 1976 study found differences in the drug sensitivities of human and murine bone marrow upon exposure to chemotherapeutic agents, calling into question the

applicability of mouse models in the evaluation of chemotherapy and combination therapy response[30]. Transgenic mice bearing alterations in metabolic enzymes, such as humanized cytochrome P450, and transplantation of human liver hepatocytes have been developed to better mimic *in vivo* drug response[31,32]. Despite these improvements, the implementation of these models does not generate improved translational benefits to human responses.

Simply put, rodent models do not faithfully mimic the inherent tumor evolution and drug response observed in humans; a natural disease model exhibiting spontaneous tumor development is likely to generate more promising results[33]. The success of precision medicine is critically dependent upon a reliable and appropriate model for monitoring tumor development, progression to metastases, immune influence, and therapeutic safety and effectiveness.

1.3 Emergence of the canine model

In 2003, the National Cancer Institute's Center for Cancer Research (CCR) recognized the value of spontaneously occurring tumors in pet animals as models of human disease and launched the Comparative Oncology Program (NCI-COP). The establishment of this program inspired many researchers to explore the genetics of companion dogs. Two years later, whole-genome shotgun sequencing (WGS) of a boxer named Tasha resulted in the first high-quality draft sequence of the dog, covering ~99% of the genome (CanFam2.0)[34]. After the completion of the canine genome project in 2005, researchers had the genomic information necessary to launch a gene expression database of normal canine tissue[34,35]. The Canine Normal Tissue Database established the foundation for a more robust assessment of the biological functions of specific genes and cross-species comparisons of gene expression. In fact, during the development of this

database, Briggs et al found remarkable similarities in tissue-specific expression profiles between dogs and humans, thus providing additional support for the dog's role in translational research.

1.3.1 Evolution of the canine genome

Man's best friend, *Canis lupus familiaris*, has evolved alongside humans for centuries. Diverging from the gray wolf, the earliest record of a confirmed domestic dog was discovered in a maxillary fragment discovered in the Kesserloch Cave in Switzerland and dated to 12.225+/-45 bp or c. 14,100-14,600 BP from a 2012 study[36]. First discovered in 1928 and later confirmed in 1964, karyotype analysis reveals 38 pairs of autosomal acrocentric chromosomes and a pair of metacentric sex chromosomes in the dog[37]. Dogs were the first species to be domesticated, occurring more than 15,000 years ago when humans were hunter-gatherers[38]. Since then, selection for certain phenotypic traits such as temperament and function has resulted in over 400 breeds that can be differentiated by distinct genetic components and morphological variation[39–41]. Linkage disequilibrium in the dog is roughly 20-50 times more extensive within breed groups than in the human population, even for isolated populations, which allows genome-wide association studies (GWAS) to be accomplished with significantly fewer SNPs. For example, only about 10,000 canine SNPs are necessary as opposed to the ~500,000 SNPs needed in humans, providing an economical advantage to disease mapping[42].

Conserved synteny analyses show that human and dog genomes are more closely related to each other in terms of nucleotide divergence and chromosomal aberrations than either one is to rodent genomes, including mice[20,43]. The average rate of nucleotide divergence in both humans and dogs is approximately 0.35 substitutions per site, a rate that is much more similar than that observed in the mouse[42]. Inter-chromosomal rearrangements occur less frequently in

the canine genome than in rodents but significantly more than in humans. However, the rate of intra-chromosomal reshuffling is more similar between humans and dogs than rodents[42]. Despite evolutionary divergence, the canine and human genomes share remarkable similarities that highlight the importance of the dog as an essential animal model.

1.3.2 Similarities in canine and human immune system, drug metabolism, and microbiome

The relatedness between dogs and humans extends beyond genetics to include immune system development and function. Unlike rodents, both human and canine immune systems are fully developed before birth and maturation is believed to continue postnatally[44]. Studies of X-linked severe combined immunodeficiency caused by mutations in the common gamma chain, a component of at least six different cytokine receptors, have compared the outcomes on lymphocyte development and function in the human, dog, and mouse. These results have shown the effects on B cells, T cells, natural killer cells, and postnatal T cell development are almost identical in dogs and humans, while mice show distinctly different responses[44]. Similar findings of age-related changes in immune cell distribution have been found in humans and dogs. Age-related decline in the absolute numbers of B-, T-, and CD4+ cells, but not in CD8+ cells, in dogs resemble previously published data in humans[45,46]. The major immune effector cells have been characterized in dogs, including CD4+, CD8+, CD90+, and dendritic cells, and show similar homology to humans[47,48]

Differences in pharmacokinetics and pharmacodynamics has been observed in various dog breeds. Many breed-specific features affecting drug absorption and metabolism have been elucidated, including polymorphisms in cytochrome P450 enzymes and cyclooxygenase 2

inhibitors. Due to these differences which can lead to inconsistent results, caution should be used in studies employing multiple different breeds. All domestic dog breeds and other canids lack the n-acetyltransferase genes that detoxify many drugs and metabolites, including arylamine and hydrazine compounds[49]. While canine pharmacodynamics and metabolism are not perfect models for humans, the similarities can be useful for drawing comparative conclusions.

The gut microbiome can influence drug pharmacokinetics and the immune system and therefore should be considered in experimental models. In a 2018 study cataloging the canine gut microbiome, researchers found more similarities to the human microbiome than to mouse or pig[50]. Furthermore, gut microbial changes associated with colorectal cancer, irritable bowel syndrome, and diabetes mellitus have shown comparable results between human and dog[51–53].

1.4 Cancer in the dog

Dogs develop spontaneous tumors that share similar biological and histopathological properties as human neoplasms at comparable rates[54]. The predicted cancer incidence is similar in dogs and humans, with an estimated 300 per 100,000 people or dogs per year at risk of developing a tumor[43]. In addition to the relative ease of sample acquisition, dogs also share similar environments and are exposed to comparable risk factors as humans[8]. Age, nutrition, sex, and environment influence tumorigenesis in similar ways in both species[55]. While unfortunate for man's best friend, the condensed lifespan of dogs allows the accelerated study of tumors that would normally take decades to develop in humans to be performed in 2-3 years[56]. Veterinary clinical trials are less expensive and patients are also typically more homogeneous in terms of pre-treatment and age than human clinical trials[57]. Owners can evaluate their dog's quality of life in response to treatment within their natural environment, a task that is not easily accomplished

and often not considered in laboratory rodent studies. Additionally, with less stringent control on patient records and confidentiality, access to patient data is more easily available in dogs.

Some tumors have also shown a significant breed disposition, providing the opportunity to study the genetic components, such as germline mutations, that influence the risk of tumor development[58]. Genome mapping studies have revealed breeds with significantly increased risk for specific tumor types, including osteosarcoma in the Greyhound and Rottweiler, hemangiosarcoma in the Golden Retriever, lymphoma in the Golden Retriever and Boxer, melanoma in the Scottish Terrier, mast cell tumors in the Pug and Golden Retriever, transitional cell carcinoma in the Scottish Terrier, brain tumors in Boxers, and stomach cancer in Chows[59,60]. Breed-specific cancers show patterns that reflect the inherited mutations that predispose human children and families to an increased risk of specific cancers[61]. Genome-wide association studies (GWAS) in breed groups have successfully identified inherited risk loci for several tumor types, which will be discussed in the sections below. Due to the genetic homogeneity within a breed, GWAS analysis can be accomplished using a smaller number of canine patients compared with human studies, providing a more efficient opportunity to identify risk loci that may be translational to humans[61].

Comparative gene expression studies in human and canine tumors have exposed the striking similarities in terms of genotypic and phenotypic properties, including specific germline and somatic mutations, chromosomal aberrations, physiological processes, enriched gene sets or pathways, and molecular markers in various cancers that are not naturally reflected in rodent models. In addition to the genetic and environmental aspects driving tumor formation, the immune system, gut microbiome, and drug metabolism play a critical role, especially for studying the effects of immunotherapy. Dogs and humans share more commonalities in their

immune system development, function, microbiome composition, and metabolism than rodents. In addition to genomic and pathophysiological similarities, dogs share similar environmental influences that may impact cancer. For these reasons, canine models of cancer are superior to rodents in oncological research. In the following sections, we will discuss these similarities in depth for each tumor type. While dogs also serve as valuable models for numerous other human disorders, including inherited retinal degeneration, dilated cardiomyopathy, cystinuria, Duchenne muscular dystrophy, chronic obstructive pulmonary disorder, and rare human disorders such as Caffey, van den Dende-Gupta, and Raine syndromes, the focus of this chapter will be on cancer specifically[43,62–64]. Ultimately, our canine companions serve as promising models for uncovering the molecular dynamics of tumorigenesis, improving clinical outcomes, lowering treatment costs, evaluating drug safety and efficacy, and advancing the era of precision medicine in the treatment of cancer – for both man and man's best friend[65].

1.5 Canine osteosarcoma

Osteosarcoma (OSA) is a highly aggressive bone tumor that typically originates in the appendicular skeleton of pediatric patients and large breed dogs. The canine model has been extensively studied and is perhaps the most well-known cancer model due to its striking similarity to human OSA. In both humans and dogs, primary tumors and metastases occur in similar sites, show comparable clinical presentations, and share treatment approaches[66]. Metastases primarily occur in the lungs and drug-resistance is exceedingly common. The standard of care for dogs involves amputation followed by chemotherapy, though the one-year survival rate is less than 50%[67].

OSA is characterized by complex karyotypes with significant chromosomal aberrations[68]. In fact, in terms of gene expression, OSA is one of the most heterogeneous tumor types[69]. Flow cytometry analysis of multiple sections from 25 untreated human OSA tumors indicated DNA aneuploidies present in 86% and multiple aneuploid DNA stemlines in 48% of highly malignant cases, rates that far exceed those from other tumors[70]. Contrary to human disease, canine OSA shows a disposition for large breed dogs, suggesting a heritable component[71]. A genome-wide association study in Greyhounds, Rottweilers, and Irish Wolf Hounds identified 33 loci associated with OSA, accounting for 50-80% of the disease risk[71]. Notably, none of the genomic regions overlapped between the three breeds. A non-coding regulatory element upstream of the *CDKNA2A/B* locus was identified as a main causal variant, causing subsequent dysregulation of *CDKN2A/ARF*. *CDKN2A/ARF* encodes multiple cyclin-dependent kinase inhibitors, including INK4a/b and ARF, which inactivate D-cyclins to promote G1 progression and control senescence via the RB (Retinoblastoma) and p53 pathways[71]. Germline variants in the regulatory regions of *CDKN2A* may explain the observed heritability and predisposition of OSA development. Investigation of the copy number aberrations via comparative genomic hybridization arrays in 12 Greyhounds and 10 Rottweilers revealed substantial conservation in regional DNA copy numbers between the two breeds. Furthermore, the results showed striking similarities in both the frequency and distribution of copy number aberrations (CNAs) between dogs and humans, including *MYC* gain (dog=60%, human=67%), *RB1* loss (dog=36%, human=33%), and *CDKN2A/B* loss (dog=73%, human=67%)[71].

Perhaps the most well documented genetic alteration in both human and canine OSA involves *p53*. Mutations in *p53* have been reported in 41% of primary canine OSA tumors with the majority consisting of point mutations (74%) and fewer being deletions (26%)[72]. In contrast,

approximately 20% of primary human OSA tumors have been reported to harbor *p53* mutations, most of which include genomic deletions[73].

At the gene expression level, canine and human OSA are nearly indistinguishable. A cross-species gene signature study using oligonucleotide arrays to examine a set of orthologous genes could not distinguish the human and canine diseases by hierarchical clustering[74]. The researchers identified two genes, interleukin-8 (IL-8) and solute carrier family I member 3 (SLCIA3), which were expressed in all dogs, but only a subset of pediatric patients. These genes were validated by protein expression using tissue microarray immunohistochemistry and were found to be associated with poor outcomes in an independent population of pediatric OSA patients[74]. The identification of these progression-associated genes would likely not have been uncovered without the cross-species approach. In another study, Scott et al evaluated the tumor transcriptional profiles using RNA sequencing of OSA tumors and cell lines derived from human, mouse, and dog. Despite the complex genomic instability characterizing OSA, their results showed conserved inter-tumoral expression and patterns of transcriptional variation that were distinct from other tumor types[68].

In a 2011 study, genome-wide expression profiling revealed two molecular subtypes in 27 OSA tumors obtained from dogs prior to any clinical intervention. The cell lines derived from the tumors clustered into two distinct branches; Branch A, consisting of 16 samples, and Branch B, consisting of 10 samples, with two reciprocally expressed groups of genes, defined as “gene cluster 1” and “gene cluster 2”. The pathways evident in gene cluster 1, which were overexpressed in the Branch A samples, consisted of genes responsible for mitosis, chromosome segregation, and mitotic spindle formation. The cluster 2 genes, overexpressed in Branch B dogs, were predicted to be associated with cancer cell-microenvironment interactions. Upon analysis of

patient outcomes, the researchers found that dogs from Branch A exhibited a worse survival than those from Branch B[75]. Scott et al proposed that this may be due to the greater metastatic potential of more aggressive tumors, resulting in their ability to overcome limitations established by the tumor niche and rapidly divide and survive regardless of their environment.

Canine studies have demonstrated the practicality and effectiveness of precision medicine in the treatment of OSA. A multi-site feasibility study evaluated the turn-around time for the molecular analysis of 20 dogs with OSA[76]. The team established a processing pipeline that generated a report of tumor expression within 5 days of sample receipt, demonstrating that the practical application of precision medicine in the clinical setting is reasonable[76]. Another canine case study was used to show the efficacy of personalized therapy for OSA. Davis et al used a primary OSA tumor sample from a 7-year-old Golden Retriever to establish a cell line, which was then screened for sensitivity to a panel of kinase inhibitors. The molecule with the highest activity was the Src and Abl inhibitor dasatinib, which had not been previously administered in dogs. After immunohistochemical staining to confirm overexpression of Src, the dog began treatment with dasatinib and a dose of 0.75 mg/kg/day was established and well-tolerated. Twenty-four months after initial diagnosis and eight months after completion of the precision medicine adjuvant therapy, the canine patient remained disease-free with no evidence of metastases. For comparison, the average survival rate after initial diagnosis is 33-65% for 12 months and 16-28% for 24 months[77]. While these results are based only on a single case, this 2013 study confirms the critical role dogs play in demonstrating the efficacy and feasibility of precision medicine therapy.

OSA is perhaps the most well-documented of tumors in canine comparative oncology. The shared characteristics with human OSA provide support for the canine model. In OSA,

where treatment has remained essentially unchanged for 40 years, there is an urgent need to develop and improve therapies, and it is likely that dogs hold the key to benefitting patient survival.

CHAPTER 2

A Method for Isolating RNA from Canine Bone

Reprinted from Nance R, Agarwal P, Sandey M, Starenki D, Koehler J, Sajib AM, Smith BF. A method for isolating RNA from canine bone. Biotechniques. 2020 Jun;68(6):311-317.

2.1 Introduction

Isolation of high-quality RNA from tissue is necessary to evaluate the molecular basis of gene expression in that tissue. Messenger RNA, as well as a variety of non-coding RNA, provides insight into gene activity and therefore cellular processes and pathways that are active in a given tissue. Common approaches to analysis of RNA include quantitative and end-point reverse-transcriptase PCR (RT-PCR) and transcriptomic sequencing. The quality of RNA obtained is a key factor in validating the significance of RNA sequencing and analysis[78]. While established methods for RNA isolation often provide sufficient yield, especially for soft tissues, they are not applicable to some tissues, such as bone.

The single-step method of isolating total RNA using an acid guanidinium thiocyanate-phenol-chloroform mixture has been a widely used and well-established technique for over thirty years[79]. This approach results in cell lysis as well as inhibition of RNases and DNases. With the addition of chloroform, phase separation allows for the isolation of RNA, DNA, and/or protein. RNA is precipitated using propanol and a high salt solution. This technique is flexible and variations in the protocol have accommodated many diverse tissues. For example, in tissues with high proteoglycan content such as cartilage, undesirable precipitates are often formed during the RNA precipitation step. In a modification described by Lee et al, an additional phase separation and high salt precipitation avoided the formation of unwanted precipitates[80]. While the traditional single-step “Chomczynski” method typically provides high yield, high quality RNA,

it requires meticulous technique to avoid carry-over of unwanted material when pipetting the aqueous layer. As a result, column-based methods have been growing in popularity as a quick and easy alternative, but are effectively limited in certain tissues with inadequate yield[80]. Higher yield of RNA is often obtained from the Chomczynski method in comparison to column-based methods[80]. A technique defined as the "TRIspin method" has been previously described that combines both approaches by following the initial steps of the Chomczynski method with a column-based method[81]. This method was originally developed to provide maximal reproducible amounts of high-quality RNA from dense, hypocellular connective tissue of rabbits[81].

Tissue homogenization is an essential step to releasing RNA from cells. Several methods have been described such as bead dissociation, liquid nitrogen grinding, cryogenic mill pulverization, and cryosectioning[80,82–84]. Some tissues present unique challenges for RNA isolation and require altered homogenization protocols to obtain sufficient RNA. Bone in particular is difficult to manipulate due to the rigid nature of the proteoglycan-rich matrix, which makes homogenizing the tissue challenging, and the low cell-to-matrix ratio, which provides little RNA to work with[80].

Bone is a dynamic, living tissue which is constructed and maintained by three types of cells: osteoblasts, osteoclasts, and osteocytes. Osteoblasts, which originate from mesenchymal stem cells, are responsible for the creation of new bone and represent 4-6% of the total cell population. Osteoclasts, which are of macrophage origin, remove old bone matrix, helping to remodel bone in response to external forces. Osteocytes, the most abundant cells in bone, are differentiated osteoblasts that have become embedded in the bone matrix and coordinate the bone remodeling activities of osteoblasts and osteoclasts. The extracellular bone matrix consists of inorganic salts, largely phosphate and calcium ions, and an organic matrix composed

predominantly of collagenous proteins and some noncollagenous proteins[85]. The majority of bone studies have utilized formalin fixation, decalcification, and paraffin embedding which results in a highly altered RNA profile and loss of RNA integrity[78]. Cryosectioning is widely considered the best source of intact bone RNA[78]. A method of isolating RNA from osteocytes in human femurs has been established by Eisenberger et al. Their technique employs the use of tape-assisted cryosectioning followed by hematoxylin-eosin staining to guide microdissection using an UVa-nitrogen laser. RNA was then isolated from the microdissected osteocytes via the TRIspin method of Reno et al[78]. While this method is appropriate for characterizing the profile of osteocytes specifically, it requires equipment that is not always readily available and excludes osteoblasts and osteoclasts, which are additional key components of bone composition. While osteocytes are the most abundant cell type in bone, characterization of all three cell types is important for evaluating the molecular basis of disease, particularly osteosarcoma.

Contaminating tissue such as bone marrow and periosteum are problematic for isolating RNA that is unique to osteoblasts, osteoclasts, and osteocytes. The downstream transcriptomic analysis of cancerous versus noncancerous tissue requires the isolation of RNA from tissue of the same origin as that of the tumor. With the exception of a single murine study, current methods for isolating RNA from all three bone cell types do not account for the contaminating marrow RNA[82]. This is especially important as hematopoietic precursor cells express many stem cell markers that could be shared with neoplastic cells, especially the small subpopulation of radio- and chemoresistant neoplastic cells with a more stem-like phenotype[86].

Preventing RNA degradation prior to isolation is an important consideration. Snap freezing samples in liquid nitrogen as early as possible after harvesting has been utilized as a suitable means of reducing RNA loss[80,83].

This study describes a technique to isolate sufficient quality RNA from canine bone cells for downstream transcriptomic sequencing. Samples were obtained from dogs undergoing limb amputation due to osteosarcoma. The phalanges were selected as the bone of choice due to their location distal and distant from the tumor as well as their relative ease of accessibility. Structurally, the phalanges are long bones, with an extended, bone marrow-filled diaphysis composed of dense cortical bone and epiphyseal caps of trabecular bone on either end. The periosteum is an external layer of connective tissue surrounding the bone[87]. A protocol to isolate RNA from bone should include adaptations to address these contaminating tissues and isolate only cellular bone matrix RNA. The resulting protocol yielded sufficient quality RNA suitable for downstream applications.

2.2 Materials and Methods

2.2.1 Sample preparation and bone marrow removal

Phalanx bone specimens were obtained from seven dogs undergoing limb amputation for osteosarcoma at the Auburn University College of Veterinary Medicine. Since amputations were performed for clinical treatment of disease, the procurement of tissues from amputated legs were exempt from Auburn University Institutional Animal Care and Use Committee review. In all cases, the tumor was at least one joint space proximal to the phalanx. Upon amputation of the leg, the second phalanx was dissected free of the surrounding tissue. Typically, two to three phalanges were obtained from each amputated limb. Phalanx sizes varied based on the breed and size of the dog but were approximately 12 mm in length and 8 mm in width after trimming. After obtaining the phalanx, the exterior soft tissue and periosteum were removed using a scalpel, scissors, and clean laboratory wipes (Kimwipe, Kimberly Clark, USA). After sufficient soft

tissue removal, the epiphyses were removed and discarded using large shearing cutters. The diaphysis was placed into a 1.5 mL microcentrifuge tube and centrifuged at 10,000 x g for 10 minutes at room temperature to remove the bone marrow, which was collected and stored at -80°C. The bone was then transferred to a clean 1.5 mL microcentrifuge tube, covered with 1X phosphate buffered saline (PBS) (Corning), and centrifuged at 10,000 x g for 5 minutes. After transferring the bone to a clean 1.5 mL microcentrifuge tube, this PBS wash was repeated to ensure the least amount of contaminating bone marrow remained within the medullary cavity. The prepared bone was then snap frozen in liquid nitrogen and stored at -80°C for up to 8 months prior to RNA extraction. The total time for bone preparation was approximately 30 minutes.

2.2.2 Histological staining

To verify that the bone samples were free from contaminating tissue, histological staining was performed for both processed and unprocessed samples. The processed bones underwent the bone marrow and periosteum removal protocols, including the outer tissue removal, centrifugation, and PBS washing. For the unprocessed samples, the epiphyses were simply removed. The specimens were placed in 10% formalin, decalcified using 7% hydrochloric acid, and embedded in paraffin. They were cut in 5 μm transverse sections and stained using hematoxylin and eosin according to established standard methods.

2.2.3 Bone tissue homogenization and RNA extraction

A steel mortar and pestle and metal spatula were made RNAse-free by wiping with RNase Away (Thermo Fisher Scientific), wrapping individually in aluminum foil, baking at

350°F for 6 hours or overnight, and then storing in -80°C. Bone samples were removed from storage in -80°C, weighed to compare total yield across various sized dogs, and subsequently ground into a fine powder using the mortar and pestle on top of a bed of dry ice. The crushed bone powder was divided evenly into two 1.5 mL microcentrifuge tubes, each containing 1 mL of pre-chilled Tri-Reagent (Molecular Research Center) and approximately 0.5 g of zirconium oxide beads (Next Advance, ceria stabilized zirconium oxide, 0.5 mm diameter) and kept on ice. The samples then underwent 4 rounds of a 30 second spin in a Bullet Blender (Next Advance) at a speed of "5.5" with a 1-minute incubation on ice between rounds. After a 10-minute incubation at room temperature, 100 uL of bromochloropropane (BCP) (Molecular Research Center) was added to each microcentrifuge tube, vortexed thoroughly, incubated for 5 minutes at room temperature, and then centrifuged at 20,000 x g for 15 minutes at 4°C.

Initially, after removal of the aqueous layer, a second extraction was performed by adding 100 uL of RNase-free water to the tubes containing the organic layer. After thoroughly mixing the contents, they were centrifuged again at 20,000 x g for 15 min at 4°C, and the aqueous layer was removed and combined with the aqueous layer from the first extraction. Subsequent experiments determined that this additional extraction step was unnecessary, and RNA yield was not largely affected by omitting this step. The genomic DNA column included in the Qiagen RNeasy kit was excluded because initial experiments showed a significant reduction in RNA yield when this column was included in the protocol. As a result, we incorporated DNAse to address genomic DNA contamination by treating the aqueous extract with 10 uL DNAse I with 1/10 volume 10x DNAse I Reaction Buffer (Thermo Fisher Scientific). The reaction was incubated for 10 minutes at room temperature. Following the manufacturer's protocol, DNase was inactivated by adding 10 uL of 25 mM EDTA and heated for 10 minutes at

65°C. Following this step, 20 ug of RNA grade glycogen (Thermo Fisher Scientific) was added to each tube in an attempt to increase RNA yield in the subsequent steps. All data shown in this paper were obtained using glycogen; however, subsequent experiments have demonstrated that glycogen does not improve RNA yields (data not shown).

An equal volume of 70% ethanol was added and mixed well by pipetting. The sample was then applied to an RNA column (RNeasy Micro Kit, Qiagen), placed in a 2 mL collection tube and centrifuged for 1 minute at 21,000 x g at room temperature (all subsequent centrifugations were performed at room temperature). In many cases, this step was repeated to include all of the sample due to the maximum loading capacity of 700 uL for the columns. After discarding the eluate, 700 uL of Buffer RW1 was added to each column, incubated for 2 minutes at room temperature, and centrifuged for 1 minute at 21,000 x g. The flow through was discarded and 500 uL of Buffer RPE was added to each column, incubated for 2 minutes at room temperature, and centrifuged for 1 minute at 21,000 x g. Then, the flow through was discarded and 500 uL of 80% ethanol (prepared using RNase-free water) was applied to each column, incubated for 2 minutes at room temperature, and centrifuged for 3 minutes at 21,000 x g. After the flow through was discarded and the collection tube was replaced, the columns were spun at 21,000 x g for 5 minutes with the lids open. The collection tube was replaced and 17 uL of RNase-free water (preheated to 65°C) was applied to the center of the column membrane. After a 10-minute incubation at room temperature, RNA was obtained via centrifugation for 5 minutes at 21,000 x g. Due to the column retaining volume of 2 uL according to the manufacturer, the total eluted volume was 15 uL.

2.2.4 RNA quantity and quality

The RNA was quantified using a Nanodrop 2000 instrument and assessed for purity using the absorbance values at 260, 280, and 230 nm. Pure RNA is considered to have absorbance ratios A260/280 of 2.0 and A260/230 ratios of 2.0-2.2. The samples were then analyzed using the Agilent Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany) and an RNA Integrity Number (RIN) derived.

2.3 Results and Discussion

2.3.1 Bone Samples

Bone specimens were obtained from the phalanges of seven dogs undergoing limb amputation for osteosarcoma at the Auburn University College of Veterinary Medicine. Immediately after receiving the sample, care was taken to remove all possible external contaminating non-bone tissue, such as periosteum and connective tissue. After cutting off the epiphyses, centrifugation of the diaphysis yielded a pellet of material assumed to be bone marrow and/or fat from the marrow cavity. Washes of the bone with PBS yielded additional visible material in the buffer.

2.3.2 Histological staining

Histological staining of processed and unprocessed bone indicated that the diaphysis was free from contaminating tissue. Processed samples were subjected to the contaminating tissue removal protocol, including the mechanical removal of exterior soft tissue and the centrifugation and PBS washing to remove contaminants within the diaphysis. The epiphyses were simply removed from the unprocessed samples, with no further manipulations to the bone. In the unprocessed sections, yellow bone marrow adipocytes can be seen within the diaphysis and

external connective tissue envelopes the bone (Figure 2.1 B,C). The processed section shows a clear diaphysis, free from bone marrow or fat, as well as an uncontaminated external surface (Figure 2.1A).

Figure 2.1. Histological staining of processed vs. unprocessed samples to evaluate removal of contaminating non-bone tissue.

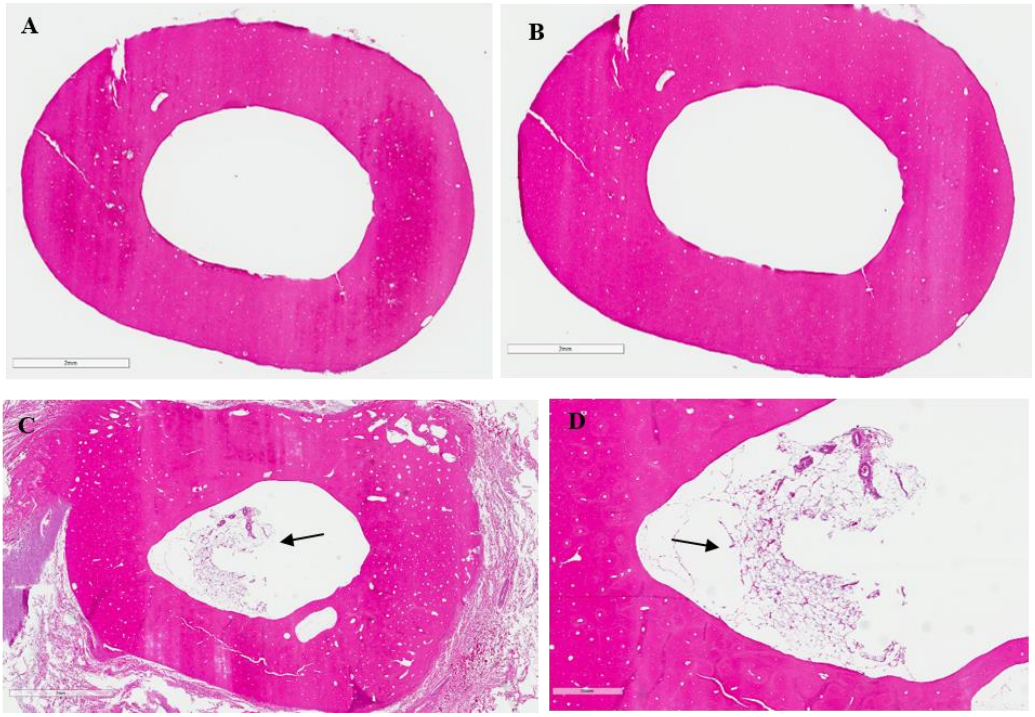


Figure 2.1. H&E-stained cross section of phalanx bone after processing to remove contaminating tissue (A, B). Processing included physical manipulation to remove exterior connective tissue and the epiphyses, centrifugation to remove tissue within the diaphysis, and washes with PBS to further remove any non-bone tissue. The diaphysis and bone exterior appear free from contaminating tissue (C, D). H&E-stained cross section of phalanx bone with no manipulations aside from removal of the epiphyses. In addition to connective tissue surrounding the bone, the diaphysis contains what is presumed to be bone marrow or fat (indicated by arrows).

2.3.3 RNA preparation

A steel mortar and pestle was used to hammer the frozen bone into a fine powder and zirconium oxide beads in combination with a Bullet Blender further aided in homogenizing the tissue. Keeping the samples on dry ice served to minimize the RNA lost to RNases. Alternating cycles of bead dissociation in the BulletBlender with incubations on ice was suitable to preventing sample overheating and thus RNA degradation. Tri-Reagent was used to liberate RNA from the bone cells and bromochloropropane (BCP) allowed phase separation of RNA, DNA, and protein so that RNA could be isolated. Our initial experiments utilizing the one-step acid guanidinium thiocyanate-phenol-chloroform extraction method yielded minimal RNA that was highly contaminated, presumably with guanidinium isothiocyanate, based on the absorbance ratios at 260/230. Thus, we implemented the "TRIspin" method developed by Reno and colleagues that employs the use of an RNA column following the single-step guanidinium isothiocyanate method. This adaptation produced significantly higher yield and quality RNA that proved suitable for downstream transcriptomic sequencing. The genomic DNA column included in Qiagen's RNeasy column kit was omitted due to a lack of RNA yield when this column was used. Alternatively, we included a DNase treatment to the extracted RNA prior to applying to the RNA column to eliminate genomic DNA contamination. We attempted to increase RNA yield with a second extraction of the aqueous layer from the organic phenol-chloroform mixture, but further experiments showed that this did not significantly impact the yield. Glycogen was added in an attempt to further increase RNA yield, but this did not prove to increase the RNA obtained, and in some cases, resulted in a slight decrease in yield (data not shown). RNA was eluted from the column in RNase free water and evaluated for quantity and quality using a Nanodrop

instrument and the Agilent Bioanalyzer 2100. In all seven bones, RNA of sufficient quantity and quality, based on Nanodrop data and RIN values, was obtained for transcriptomic sequencing.

2.3.4 Evaluation of RNA quantity and quality

Sufficient RNA was obtained in each case to proceed with analysis. RNA concentrations ranged from a low of 3.7 ug per gram of tissue to a high of 38.1 ug per gram of tissue (Table 2.1). The mean RNA yield obtained from the bone samples was 14.7 ug/gram and the median RNA yield was 12.7 ug/gram. RNA purity was initially assessed by the ratio of absorbance at 260 nm to the absorbance at 280 nm. These values ranged from 1.70 to 2.04, with only one value (1.70) below 2, indicating high-quality RNA. The ratio of absorbance at 260 nm to absorbance at 230 nm ranged from 0.95 to 2.14, indicating possible carry-over of guanidinium isothiocyanate, which absorbs at about 260 nm, in the samples with 260/230 ratios below 2.

Table 2.1. Bone RNA quantity and quality.

Dog	Sample Weight (g)	RNA yield (ug)	A260/280	A260/230	RIN
1	0.9778	5.5	2.01	1.52	6.90
2	1.1300	14.3	2.04	1.93	6.60
3	1.1828	16.8	2.01	1.41	6.50
4	0.9788	10.2	2.03	1.82	4.50
5	0.5952	2.2	1.70	0.95	7.10
6	0.3933	15.0	2.00	2.14	7.20
7	1.1372	20.8	*	*	6.70

*data not available

As RNA degradation increases, the 18S and 28S ribosomal subunit peaks decrease while additional smaller peaks appear. In samples with significant degradation, 18S and 28S peaks will be difficult to distinguish. In the electrophoretograms pictured in Figure 2.2, 18S and 28S ribosomal subunits are visible as distinct peaks denoted with an asterisk (Figure 2.2). The initial peak present on the left in each electrophoretogram is the marker used as an internal standard. While there is variability between the height of rRNA peaks and the presence of additional, sometimes well-defined peaks (as seen in dog 7), these electrophoretograms demonstrate moderately intact RNA suitable for sequencing.

Figure 2.2. Electrophoretograms of bone RNA.

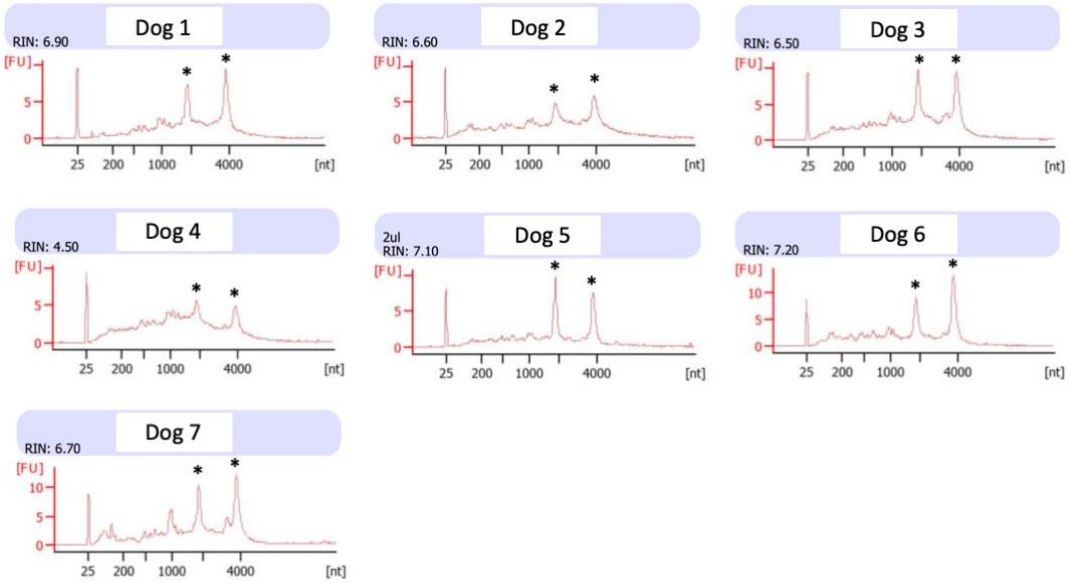


Figure 2.2. Electrophoretograms showing RNA quality and RIN values obtained for bone RNA using the Agilent Bioanalyzer 2100. Distinct peaks representing the 18S and 28S ribosomal subunits are indicated with asterisks (*). RIN values provide a more accurate level of RNA integrity and range from 1 to 10, with 10 indicating fully intact RNA and 1

being completely degraded RNA. The absence of additional significant spikes and anomalies indicates the RNA is of moderate integrity.

RNA integrity number (RIN) is based on an algorithm that evaluates the entire electrophoretic signature to determine a more accurate level of RNA integrity than previous methods that rely on subjective interpretation of electrophoretogram data. RIN is reported as a number ranging from 1 to 10, with 1 corresponding to highly degraded RNA and 10 equating to mostly intact RNA. RIN values obtained from the seven bone samples ranged from 4.50 to 7.20, with a mean of 6.54 and a median of 6.70, indicating moderately intact RNA (Table 2.1). Dog 4, with the lowest RIN of 4.5, shows the poorest RNA quality as evidenced by the increased background and decreased rRNA peak distinction present in the electrophoretogram (Figure 2.2). Nonetheless, this RNA was judged to be of sufficient quality for sequencing.

2.4 Conclusions

Bone composition presents unique challenges to RNA isolation due to the rigid, hypocellular bone matrix and the presence of contaminating tissue, such as bone marrow, fat, and connective tissue. Homogenization of the brittle, mineralized matrix was accomplished using a combination of tissue pulverization using a mortar and pestle on dry ice and bead dissociation, which effectively disrupted the bone matrix prior to RNA liberation using acid guanidinium isothiocyanate-phenol. As expected in a hypocellular tissue such as bone, the total RNA quantity obtained from these samples was relatively low and therefore limits the downstream applications. In this case, sufficient amounts for transcriptomic sequencing were achieved. A

critical prerequisite for valid transcriptomic sequencing and analysis is high quality RNA. RIN values confirmed the RNA we obtained was of high integrity.

Removal of unwanted tissue is a critical component of isolating RNA that is unique to bone cells. While a previous study by Kelly et al has shown that centrifugation is sufficient to remove bone marrow in the murine femur, a combination of PBS wash and centrifugation ensured the least amount of contaminating marrow possible[82]. While removal of the highly cellular bone marrow results in significantly lower RNA yield due to the low cell population present in bone, our technique yielded sufficient quantity for sequencing purposes. External soft tissue was removed using physical manipulation. Although histological staining indicated the bone samples were free from external and internal contaminating tissue, it is not possible to conclude that the samples consist solely of RNA from osteoblasts, osteoclasts, and osteocytes.

As an alternative to the one-step method of RNA isolation using guanidinium isothiocyanate which yielded negligible RNA, we adapted the TRIspin method developed by Reno et al that combines the one-step method followed by an RNA column. This technique provided maximal amounts of high-quality RNA from all seven bone samples.

RNA isolation is a key component to analyzing gene expression and disease characteristics. The difficulties associated with manipulating bone present unique challenges to RNA isolation. Extracting RNA from bone cells independent of contaminating cell types is essential for downstream comparison of gene expression specific to cells of the bone matrix. In this study, we have established a technique for isolating RNA from canine phalanges, independent on contaminating tissue, which successfully yields RNA of sufficient quantity and quality for downstream transcriptomic sequencing.

CHAPTER 3

Transcriptomic Analysis of Canine Osteosarcoma from a Precision Medicine Perspective Reveals Limitations of Differential Gene Expression Studies

Reprinted from Nance RL, Cooper SJ, Starenki D, Wang X, Matz B, Lindley S, Smith AN, Smith AA, Bergman N, Sandey M, Koehler J, Agarwal P, Smith BF. Transcriptomic Analysis of Canine Osteosarcoma from a Precision Medicine Perspective Reveals Limitations of Differential Gene Expression Studies. Genes (Basel). 2022 Apr 13;13(4):680.

3.1 Introduction

Osteosarcoma (OSA) is a highly aggressive and spontaneous tumor of the bone seen primarily in the appendicular skeleton of pediatric patients. Approximately 10–20% of patients exhibit macro-metastatic lesions at the time of diagnosis, while 80–90% of patients are presumed to harbor micro-metastases[88]. Metastases occur almost exclusively in the lungs and, once present, make management difficult. The 5-year survival rate for cases with detectable metastasis at the time of diagnosis is approximately 20–25%[88,89]. The current standard-of-care treatment includes surgical resection and combination chemotherapy[88]. However, 30–40% of patients will relapse within 3 years of starting treatment[90]. Metastectomy is considered the second-line treatment and improves 5-year survival rates to 40%[91]. Moreover, osteosarcoma is particularly proficient in acquiring multiple drug-resistant pathways, a major limiting factor for patient survival. Subsequent reoccurrence is exceedingly common and repeated systemic chemotherapy is frequently required; unfortunately, the eventual resistance of the metastases to treatment is inevitable[91]. Despite significant advances in the cancer therapy domain, the standard-of-care treatment and survival rates for OSA have remained essentially unchanged for 40 years[88,92,93].

For these reasons, it is imperative to explore and implement improved models to develop more effective treatment approaches.

Canine OSA has the potential to serve as an excellent model for the human disease due to similar clinical, histological, and molecular characteristics[74,94,95]. Primary tumors occur at similar sites with comparable histological presentation, response to treatment, and occurrence and distribution of metastases[66]. Further support for dogs as a valuable translational model is evidenced by a cross-species gene signature study which used oligonucleotide arrays to examine expression from a limited set of orthologous genes. The researchers found indistinguishable gene expression patterns between canine and pediatric OSA[74]. Studies utilizing whole genome and/or whole exome sequencing have reached similar conclusions[94,95]. The primary difference between canine and human OSA is the age of onset. Canine OSA has a bimodal age distribution, with peaks at 18–24 months and 7 years, though older, larger breed dogs are typically affected more often. On the other hand, human OSA primarily occurs in the second decade of life[91]. OSA prevalence is also greater in dogs, occurring 27 times more frequently than in humans, and progression occurs rapidly[67]. With amputation of the affected limb in combination with chemotherapy, the 1-year survival rate for dogs is approximately 45%[67].

New molecular tools, such as next-generation sequencing, have allowed significant improvements to be made in the identification of common genetic changes that are associated with specific types of cancer[96]. This technology has inaugurated the era of personalized or precision medicine, which utilizes an individual's specific genetic changes to guide treatment. This approach seeks to classify individual patients into groups based on the presence of key gene mutations that directly impact the tumor's sensitivity to specific chemotherapeutic agents. In this

manner, treatments can be selected that have a higher likelihood of efficacy due to a better understanding of the relevant functioning pathways in that particular tumor.

At the molecular level, OSA, whether human or canine, is characterized by substantial genetic complexity and instability[97]. Relapsed OSA is considerably more complex, involving multiple drug-resistant pathways. The unique genetic complexity of OSA poses limitations for therapy, and precision medicine is not immune to these limitations[98]. The personalized treatment of two high risk human OSA patients based on comprehensive molecular profiling via next-generation exome sequencing showed no significant benefit to overall health or disease progression[99]. However, the targeted therapy was implemented after chemotherapy and refractory disease appeared, when drug resistance mechanisms are highly convoluted and make interpretation difficult. Fortunately, precision medicine is now entering a new phase where the transcriptomic analysis of individual patients may provide unique perspectives for treatment by addressing the intrinsic heterogeneity of gene expression in tumors, both within and between patients. A common approach to identifying targetable components of a tumor involves the analysis of differentially expressed genes (DEGs) in tumors compared to normal tissue.

Patient samples and clinical information are important factors to consider in differential gene studies and are unfortunately often overlooked. The source of normal tissue should be derived from the tumor's cell-of-origin, and, in the case of OSA, that includes osteoblasts and osteocytes. Prior studies of differential gene expression in OSA have used RNA isolated from a canine osteoblast cell line[95]; tissues from unrelated organs such as liver, lymph node, and kidney[74]; tissue harvested adjacent to the tumor, potentially jeopardizing the normality of the sample[100]; bone tissue from unrelated patients[101]; or have not clearly described the origin of the matched normal samples[102,103]. While these data are undoubtedly useful to advancing our

understanding of the disease, evaluation using primary normal bone cells harvested distally from the tumor within the same patient may yield more appropriate results. Furthermore, consistency in chemotherapy status among patients at the point of sample collection is important for drawing conclusions related to tumor status. The administration of chemotherapeutic agents provides a selective pressure that alters the tumor's cellular population and phenotype[98,104]. Analysis of tumor gene expression prior to any chemotherapy provides a snapshot of the tumor phenotype, independent of acquired drug resistance or variation in individual drug response. Many studies in human OSA have utilized tissues harvested from individuals after the onset of chemotherapy, making interpretation of the results difficult[99,105,106].

DEG analysis combines the tumor and normal samples into two distinct groups to derive statistically meaningful DEGs that are generalizable across individuals. However, combining patients into a group implies that the samples are similar, if not identical, thereby contradicting and disregarding the intra- and inter-tumoral heterogeneity that forms the basis of precision medicine. While this type of analysis is critical for determining statistical differences in the group, individual-level analysis can provide additional insight into the differences among patients.

Given these limitations of traditional group differential gene expression studies, we supplemented this analysis with a more novel approach of evaluating individual tumors based on genes identified in the group approach. This study uses transcriptomic sequencing of RNA derived from seven primary canine osteosarcoma tumors with patient-matched normal bone to identify DEGs and pathways. Importantly, the normal bone samples, which serve as the comparator for determining baseline gene expression levels, were harvested from the phalanges of each patient. Furthermore, all samples were obtained prior to chemotherapy or evidence of

macrometastatic lung disease. We explore the resulting DEGs in terms of group analysis for tumor vs. bone, as well as the discrete analysis of individual patients to identify the extent of heterogeneity in gene expression among individuals. These results highlight the phenotypic diversity of primary OSA among individuals and provide a supplemental approach to traditional methods of analyzing DEGs, particularly when the goal is an application in precision medicine.

3.2 Materials and Methods

3.2.1 Description of data

Animals: Samples were obtained from seven dogs undergoing routine limb amputation for the clinical treatment of osteosarcoma (OSA) at the Auburn University College of Veterinary Medicine. Patient characteristics are briefly summarized in Table 3.1. The histopathology of tissue immediately adjacent to that used for RNA extraction was performed to confirm OSA and ensure that intact neoplastic tissue was entered into the experimental pipeline. Thoracic radiographs performed prior to amputation indicated no evidence of pulmonary macro-metastatic disease. In all cases, samples were obtained prior to chemo-therapy treatment.

Table 3.1. Summary of canine osteosarcoma patients.

Patient	Sex	Neuter Status	Age	Tumor Site	Breed	Tumor RIN	Bone RIN
A	M	Castrated	10 yrs, 6 mos	Left distal radius	Golden Retriever	5.0	6.6
B	M	Castrated	7 yrs	Left distal femur	Rottweiler	7.4	4.5

C	M	Castrated	7 yrs	Left distal radius	Doberman Pinscher	6.9	6.5
D	M	Castrated	11 yrs, 9 mos	Left proximal humerus	Greyhound	6.4	6.7
E	F	Spayed	7 yrs, 6 mos	Left distal radius	Great Pyrenees	6.9	6.9
F	M	Castrated	7 yrs	Right distal radius	Golden Retriever	8.3	7.1
G	F	Spayed	9 yrs, 9 mos	Right distal tibia	Greyhound	6.7	7.2

Table 3.1. Tumor and patient-matched normal bone samples were obtained from seven dogs undergoing limb amputation for osteosarcoma. Age is based on the date of limb amputation and breed is reported per the owner. (RIN = RNA integrity number, yrs = years, mos = months).

3.2.2 RNA isolation and sequencing

Normal bone RNA: Patient-matched samples were collected to obtain normal bone RNA to allow transcript levels to be compared with the tumors. The second phalanx was removed from the amputated leg within one hour of amputation. In all cases, the tumor was at least one joint space proximal to the phalanx. Due to the location, which was distal and distant from the primary tumor, it is unlikely that the normal bone sample had undergone any neoplastic events related to the tumor. For each dog, bone preparation and RNA extraction for the phalanx samples was performed according to Nance et al, as described in chapter two[107].

Tumor RNA: For each dog, the tumor was dissected into approximately 3 mm by 3 mm samples, flash frozen in liquid nitrogen, and stored at −80 °C until RNA extraction. RNA isolation was accomplished using a homogenizer in combination with acid guanidinium

thiocyanate-phenol-chloroform extraction. After weighing, approximately 100–200 mg of the frozen tumor samples was added to a 14 mL snap cap tube (Falcon) containing 2 mL of chilled Tri-Reagent (Molecular Research Center, Cincinnati, OH, USA). The sample was subjected to mechanical homogenization in short but frequent bursts until sufficiently homogenized, as determined visually. The samples were divided into two microcentrifuge tubes and after 10 min incubation at room temperature, 100 μL of bromochloropropane (BCP) (Molecular Research Center) was added to each, vortexed thoroughly, and incubated for 5 min at room temperature. After centrifugation at 20,000× g for 15 min at 4 °C, the RNA-containing aqueous layer was carefully transferred to a new tube. Contaminating DNA was removed by the addition of DNase I Reaction Buffer (10% of the total volume) and 10 μL DNase I (Thermo Fisher Scientific), followed by a 10 min incubation at room temperature. DNase was inactivated by the addition of 10 μL of EDTA and heated at 65 °C for 10 min. Following this, 10% volume sodium acetate (3 M, pH 5.2) (VWR International) was added to each tube, followed by 70% volume isopropyl alcohol, and the samples were briefly vortexed. Following a 15-min incubation at −20 °C, the samples were centrifuged at 20,000× g for 20 min at 4 °C. The supernatant was decanted, the pellet dislodged in 1 mL 70% ethanol, and centrifuged at 20,000× g for 30 min at 4 °C. After the supernatant was removed, the pellet was air dried for approximately 5–10 min and subsequently resuspended in 15 μL of RNase-free water. The RNA was quantified using a Nanodrop 2000 instrument (Thermo Fisher Scientific, Waltham, MA, USA).

RNA Sequencing: The samples were commercially prepared and sequenced (HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA). Initial quality control analysis was performed using the Agilent Bioanalyzer 2100 (Agilent Technologies,Santa Clara, CA, USA) and an RNA integrity number (RIN) was generated. RNA sequencing libraries were produced

with 500 ng of total RNA using the TruSeq PolyA library kit (Illumina) to deplete samples of any RNA aside from polyadenylated mRNA. They were pooled and sequenced on two lanes of the Illumina HiSeq v4 (PE, 50 bp, 25 M reads) yielding an average of 34 M reads per sample.

3.2.3 Bioinformatic pipeline and data analysis

The trimming of adapters and the first leading base was performed using Trimmomatic (v0.40)[108] with a minimum length of 36 bp and raw quality was assessed using FASTQC (v0.10.1)[109]. After trimming, approximately 32–45 million sequence reads remained for each sample and FASTQC was used to ensure all bases had a Phred quality score above 28. Reads were mapped to the indexed canine reference genome (CanFam3.1) obtained from ENSEMBL (release 103) using HiSat2 (v2.2.1)[110] and a table of mapped read counts was generated with Stringtie (v1.3.3)[111]. On average, 93% of the reads mapped to the canine reference genome. The bioinformatic pipeline is summarized in Figure 3.1.

Figure 3.1. Overview of bioinformatic pipeline used to process and analyze data.

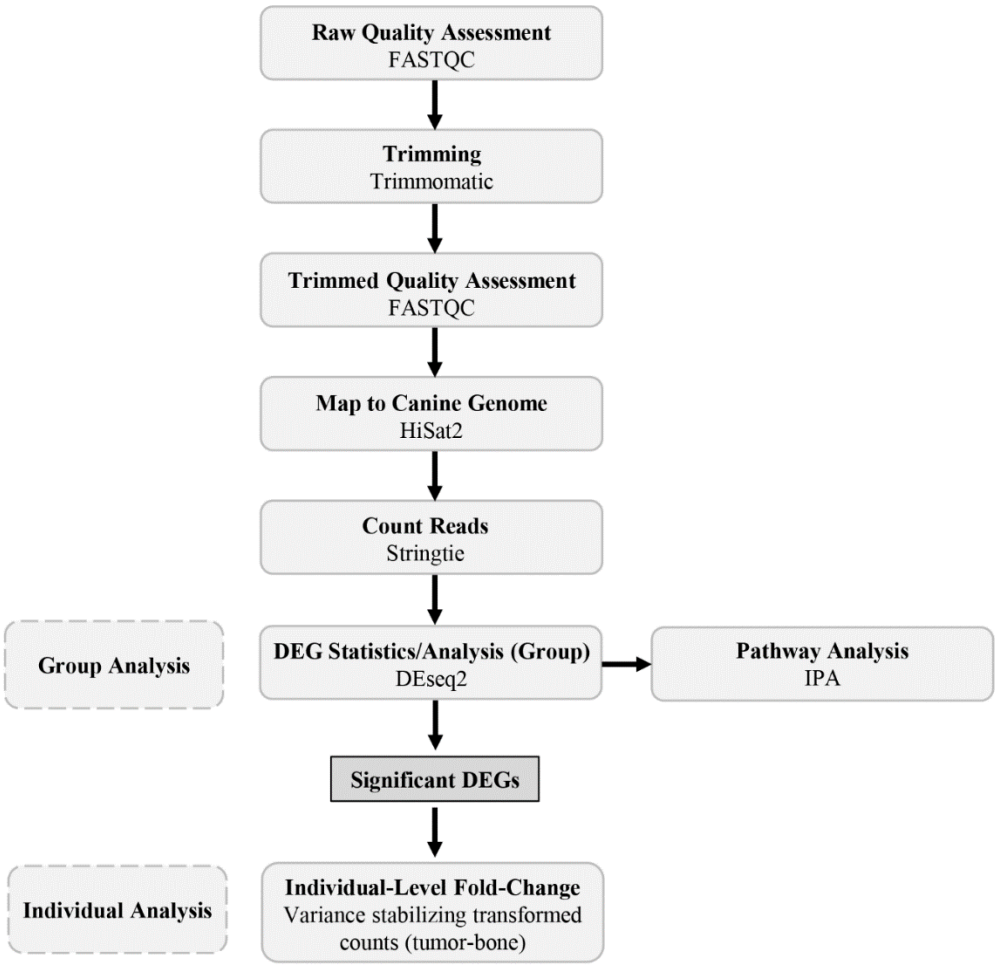


Figure 3.1. RNA sequence data were subjected to modest trimming with Trimmomatic and quality analysis with FASTQC prior to mapping to the reference canine genome with HiSat2, counting reads with Stringtie, and DEG statistics with DEseq2 to generate significant DEGs (FDR < 0.05, FC > 2, ≤2) which were carried forward in pathway and individual analyses. Using the significant DEGs from the group analysis, fold-change values were generated for each patient to produce the individual-level data. Bioinformatic tools and packages utilized are indicated.

3.2.4 Group analysis

The statistical analysis and identification of differentially expressed genes was performed using DEseq2 (v3.14)[112] to generate the traditional group analysis results. These results provide a

list of genes that are broadly differentially expressed among samples. The DEseq2 package applies a general linear model with a negative binomial distribution and applies the Benjamini–Hochberg procedure to control for the false discovery rate (FDR). The pre-filtering of genes with less than 1 read was per-formed prior to statistical analysis. A multi-factor design was used for statistical analysis to include patient ID as a term in the design formula (design = ~dog + tissue source). This design has been recommended in the DEseq2 vignette for analyzing paired samples because it accounts for differences between individuals. Log2 fold-changes were calculated relative to bone. To extract significant DEGs while minimizing noise, the data were filtered using a false discovery rate (FDR) less than 0.05, base mean greater than 10, and log2 fold-change greater than 1 and less than −1 (corresponding to a fold-change of 2 and −2, respectively). The variance stabilizing transformation function in DEseq2 was used to transform the data to fit an approximately homoscedastic distribution and to remove the dependence of the variance on the mean[112]. The results of this transformation were used for the visualization and clustering of the results to generate a heatmap and a principal component analysis plot. Hallmark pathway analysis and gene ontology (GO) enrichment analysis of the upregulated genes was performed using Metascape with a p-value cut-off of 0.01, minimum overlap of 3, and minimum enrichment of 1.5[113]. Due to the constraints of gene nomenclature in Metascape, only the up-regulated genes with identified human orthologs (total 670) were included in the analysis. Raw counts are available in the supplementary information.

3.2.5 Individual analysis

To explore how the DEGs varied among individual patients, we first filtered the data to include only the significant genes as defined by FDR < 0.05, base mean > 10, and an FC cut-off

of 2 from the classical group analysis. We then calculated log2 fold-change values for each gene by subtracting the variance-stabilized transformed counts (on the log2 scale) of bone from tumors for each dog. We used these results to determine the number one upregulated and downregulated gene in each patient.

3.3 Results

3.3.1 RNA quality

RIN numbers were used to determine sequencing suitability based on RNA integrity. Samples with an RIN below 5 cannot guarantee reliable sequencing results; for all except two (patient B bone RIN = 4.5, patient A tumor RIN = 5.0), the RIN was above 5, indicating moderately intact RNA. The average RIN for tumor and bone was 6.8 and 6.5, respectively.

3.3.2 Group analysis

For the traditional group analysis, we compared the gene expression from all tumor tissues to all normal tissues and found a total of 3742 differentially expressed genes with a false discovery rate (FDR) of less than 0.05. After further filtering using a base mean great-er than 10 and fold-change values greater than 2 and less than −2, there were 2031 significant DEGs. Of these, 803 genes were upregulated with a fold-change greater than 2, and 1228 genes were downregulated with a fold-change less than −2 in tumor compared to normal bone (Figure 3.2A). Hierarchical clustering of the 2031 significant DEGs shows that the normal and tumor samples cluster together in terms of over- and under-expression (Figure 3.2B). This is as expected based on studies in other organisms where as many as a third of genes show evidence for altered expression in tumors compared to normal tissue.

Figure 3.2: Differential gene expression analysis reveals over 3000 significant genes.

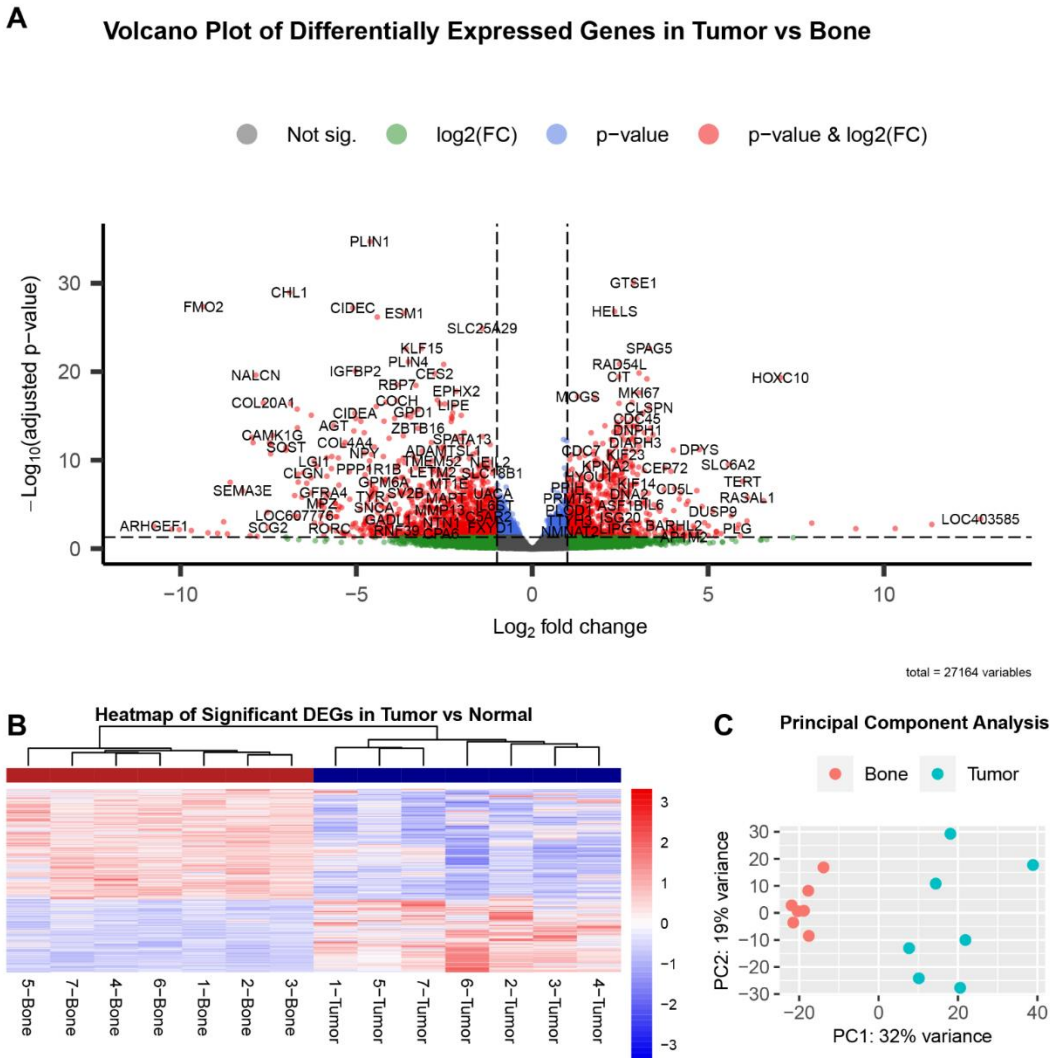


Figure 3.2. Analysis of differential gene expression in tumor vs normal group data reveals over 3000 significant DEGs. Volcano plot shows many genes are highly differentially expressed between tumor vs. normal (A). Genes indicated in red are significant in terms of both adjusted p-value (<0.05) and fold-change (>2 and ≤2). The sign of fold-change (positive or negative) was retained and is reported in terms of tumor compared to bone. Heatmap of the significant genes (FDR < 0.05) shows that these DEGs easily differentiate tumor from normal tissue (B). Each row represents a gene and upregulation is indicated in red, while downregulation is shown in blue. Patient and sample ID is indicated underneath the corresponding column. PCA plot shows

grouping of normal bone samples (red circles) distinct from tumor samples (blue circles) as expected (C). The dispersion and variability of the tumor samples is thought to be related to intra- and inter-tumoral heterogeneity.

Principal component analysis was performed to further evaluate sample clustering (Figure 3.2C). As expected, all bone samples clustered together, distinct from the tumor samples. To evaluate the biologically significant DEGs, we subset the significant results based on a fold-change cut-off of 2 and sorted them based on the log2 fold-change or adjusted *p*-value (FDR). Ordering based on log2 fold-change allows us to evaluate the most highly over- or under-expressed genes that still meet the adjusted p-value below 0.05. On the other hand, ordering the gene list based on an adjusted p-value provides DEGs with higher confidence in shared trends among individuals. The top 10 downregulated and upregulated genes are listed in Table 3.2. Many of the canine ENSEMBL gene identifications did not correspond to known gene symbols or human orthologs for the top 10 genes when ordered based on smallest and largest log2 fold-change (Table 3.2A), but when ordered based on smallest adjusted *p*-value, all 10 genes corresponded to known gene symbols (Table 3.2B). Based on sorting by padj, the top upregulated genes included *GTSE1*, *HELLS*, *SPAG5*, *RAD54L*, and *IQGAP3*. The top downregulated genes included *PLIN1*, *CL1*, *FMO2*, *CIDEC*, and *ESM1*.

Table 3.2. Top 10 up- and down-regulated genes from group analysis.

A

Top 10 Downregulated Genes in Tumor, Ordered by Log2 Fold-Change				
ENSEMBL	**SYMBOL**	**log2FC**	**padj**	**GENENAME**
ENSCAFG00000028799	ARHGEF1	-10.75	2.93E-03	Rho guanine nucleotide exchange factor 1
ENSCAFG00000041995	NA	-10.25	6.33E-03	NA
ENSCAFG00000007622	NA	-10.04	6.95E-03	NA
ENSCAFG00000005350	NA	-9.69	9.07E-03	NA
ENSCAFG00000014986	FMO2	-9.35	4.40E-28	flavin containing dimethylaniline monoxygenase 2
ENSCAFG00000042006	NA	-9.22	1.63E-02	NA
ENSCAFG00000049609	NA	-8.94	1.84E-02	NA
ENSCAFG00000008109	NA	-8.77	9.28E-04	NA
ENSCAFG00000029213	LOC607979	-8.67	2.22E-02	eukaryotic translation initiation factor 3, subunit L pseudogene
ENSCAFG00000011465	NA	-8.59	3.20E-08	NA
Top 10 Upregulated Genes in Tumor, Ordered by Log2 Fold-Change				
ENSEMBL	**SYMBOL**	**log2FC**	**padj**	**GENENAME**
ENSCAFG00000044295	NA	13.49	1.31E-04	NA
ENSCAFG00000009135	LOC403585	12.75	4.34E-04	serum amyloid A1
ENSCAFG00000032019	NLRP12	11.36	1.85E-03	NLR family pyrin domain containing 12
ENSCAFG00000013213	NA	10.32	5.31E-03	NA
ENSCAFG00000043115	NA	9.20	5.53E-03	NA
ENSCAFG00000007173	NA	7.95	1.31E-03	NA
ENSCAFG00000006648	HOXC10	7.06	4.67E-20	homeobox C10
ENSCAFG00000015211	APOBEC3Z1	6.58	3.81E-06	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ENSCAFG00000008986	RASAL1	6.12	1.68E-06	RAS protein activator like 1
ENSCAFG00000035513	LOC111093651	6.11	7.53E-04	uncharacterized LOC111093651

B

Top 10 Downregulated Genes in Tumor, Ordered by Adjusted p-value				
ENSEMBL	**SYMBOL**	**log2FC**	**padj**	**GENENAME**
ENSCAFG00000011986	PLIN1	-4.61	1.91E-35	perilipin 1
ENSCAFG00000006248	CHL1	-6.90	1.11E-29	cell adhesion molecule L1 like
ENSCAFG00000014986	FMO2	-9.35	4.40E-28	flavin containing dimethylaniline monoxygenase 2
ENSCAFG00000005266	CIDEC	-5.10	6.25E-28	cell death inducing DFFA like effector c
ENSCAFG00000018381	ESM1	-3.63	2.38E-27	endothelial cell specific molecule 1
ENSCAFG00000006392	ACKR4	-4.41	6.62E-27	atypical chemokine receptor 4
ENSCAFG00000030764	SLC25A29	-1.42	1.31E-25	solute carrier family 25 member 29
ENSCAFG00000003807	KLF15	-3.14	2.28E-23	Kruppel like factor 15
ENSCAFG00000001854	AQP7	-3.58	2.66E-23	aquaporin 7
ENSCAFG00000015323	PLIN4	-3.52	7.60E-22	perilipin 4
Top 10 Upregulated Genes in Tumor, Ordered by Adjusted p-value				
ENSEMBL	**SYMBOL**	**log2FC**	**padj**	**GENENAME**
ENSCAFG00000000782	GTSE1	2.88	8.51E-31	G2 and S-phase expressed 1
ENSCAFG00000008090	HELLS	2.34	1.54E-27	helicase, lymphoid specific
ENSCAFG00000018724	SPAG5	3.33	2.28E-23	sperm associated antigen 5
ENSCAFG00000004272	RAD54L	2.48	1.35E-21	RAD54 like
ENSCAFG00000016616	IQGAP3	3.04	1.41E-20	IQ motif containing GTPase activating protein 3
ENSCAFG00000010114	CIT	2.46	3.84E-20	citron rho-interacting serine/threonine kinase
ENSCAFG00000006648	HOXC10	7.06	4.67E-20	homeobox C10
ENSCAFG00000016090	TOP2A	3.26	6.47E-20	DNA topoisomerase II alpha
ENSCAFG00000013255	MKI67	3.04	2.35E-18	marker of proliferation Ki-67
ENSCAFG00000008478	MOGS	1.30	6.47E-18	mannosyl-oligosaccharide glucosidase

Table 3.2. When ordered by lowest and highest log2 fold-change, many of the ENSEMBL gene identifications do not correspond to known gene symbols. When ordered by smallest adjusted p-value (FDR), all ENSEMBL IDs correspond to known symbols

Hallmark pathway analysis of the upregulated DEGs revealed G2M checkpoint (M5901), E2F targets (M5925), MTORC1 signaling (M5924), and epithelial mesenchymal transition (M5930) as the most enriched pathways in tumors (Figure 3.3A). The top pathways from gene ontology (GO) enrichment analysis included the mitotic cell cycle (GO:0000278), regulation of cell cycle process (GO:0010564), DNA metabolic process (GO:0006259), and spindle

organization (GO:0007051) (Figure 3.3B). These results are consistent with previously published data.

Figure 3.3. Pathway analysis of upregulated DEGs.

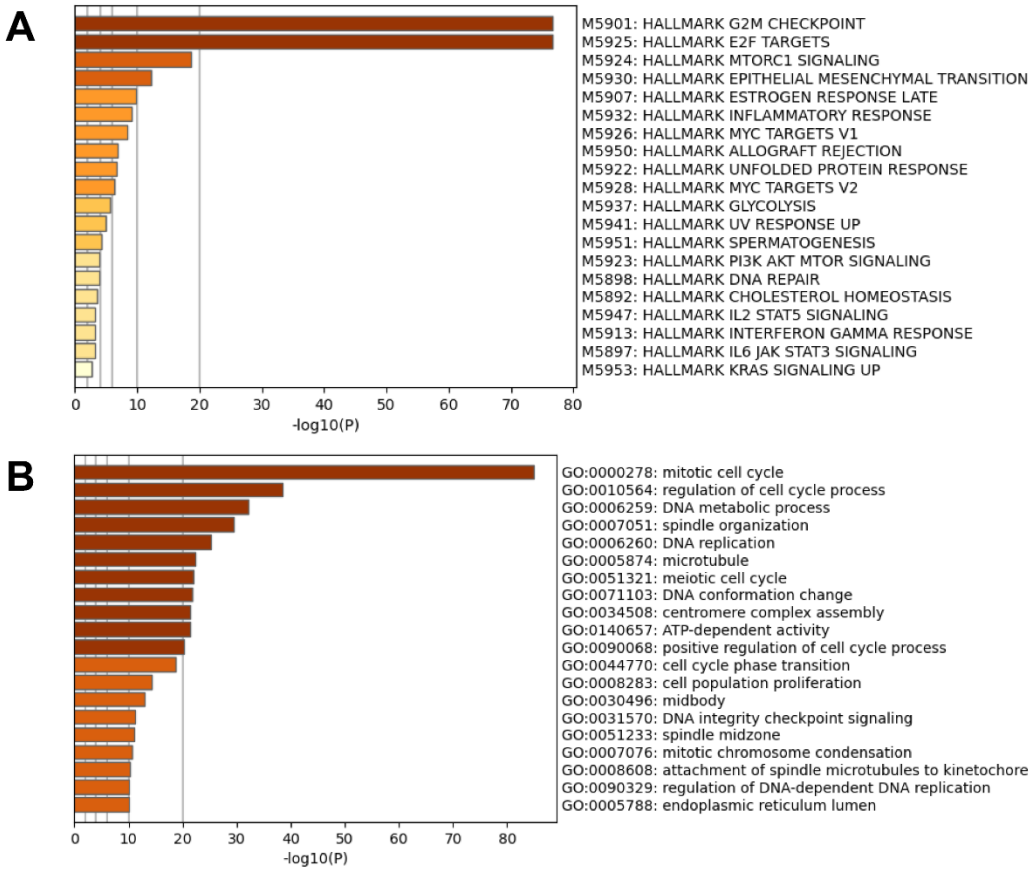


Figure 3.3. Pathway analysis of the upregulated DEGs using hallmark pathways (A) and gene ontology (GO) enrichment terms (B).

The classical group analysis identifies DEGs that are broadly dysregulated among patients. As a whole, each of the tumors exhibited similar patterns in terms of DEGs and these are distinct from normal bone. However, expression was not necessarily consistent across patients for all DEGs. To observe how the top genes for the group analysis varied between

individuals, we plotted the normalized counts of the number one upregulated and downregulated gene according to the smallest adjusted *p*-values (Figure 3.4B,D) and the largest log2 fold-change difference (Figure 3.4A,C). The top upregulated gene according to the largest fold-change difference was ENSCAFG00000044295 (L2FC = 13.5, FDR = 1.3 × 10−4), which, according to ENSEMBL, encodes an uncharacterized protein of 120 amino acids in length. The top upregulated gene with the smallest FDR was *GTSE1* (L2FC = 2.9, FDR = 8.6 × 10−31), or G2 and S phase-expressed protein 1, which encodes a protein that binds the tumor suppressor protein P53 to repress its ability to induce apoptosis in response to DNA damage. The top downregulated gene according to fold-change was *ARHGEF1* (L2FC = −10.7, FDR = 2.9 × 10−3), or rho guanine nucleotide exchange factor 1, which encodes a protein that may be involved in forming a complex with G proteins and the stimulation of rho-dependent signals. The top downregulated gene with the smallest FDR was *PLIN1* (L2FC = −4.6, FDR = 1.9 × 10−35), or perilipin 1, which encodes a protein involved in adipocyte lipid metabolism. Visualization of these normalized gene counts shows some variation among patients, some even showing an inverse relationship despite having an FDR < 0.05 (Figure 3.4C).

Figure 3.4. Top genes from group analysis.

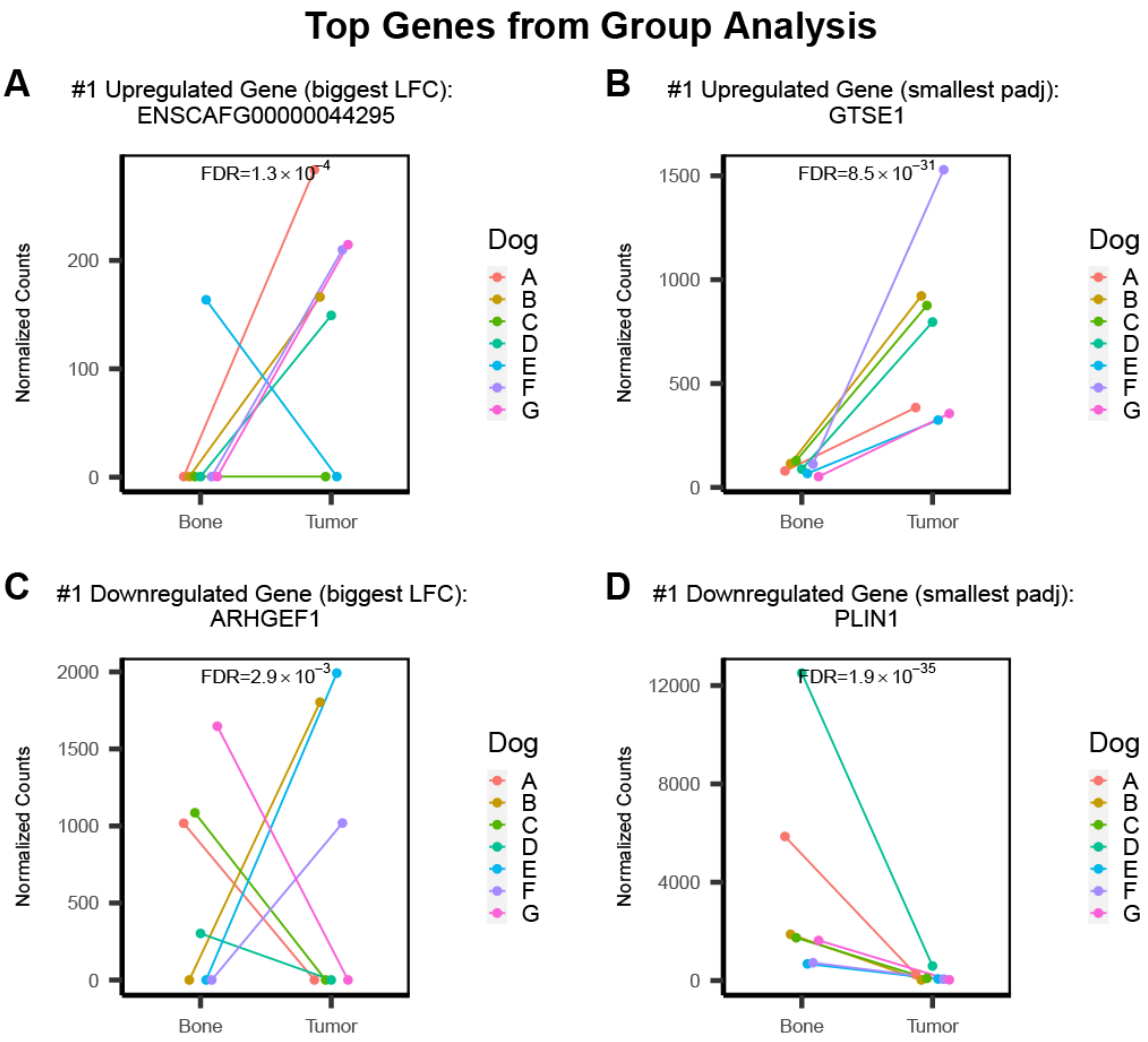


Figure 3.4. Plots of the normalized counts of the top up- and down-regulated genes in the classical group analysis show some variation among individual patients. The number one gene when ordered by log2 fold-change shows variation among individual patients (A,C). The top gene when ordered by smallest adjusted p-value shows less variation in terms of direction of fold-change, but some variation between individuals is still evident (B,D).

3.3.3 Individual analysis

To supplement the classical group analysis, we sought to identify the top DEGs in each individual patient by calculating log2 fold-change values for each pair of tumor and normal

samples using the significant genes from the group analysis. Figure 3.5 depicts the normalized counts of the top upregulated gene in each patient. While the results show a general trend shared among individuals (Figure 3.5C–F), the top genes for patient A and B are heavily skewed by substantial counts in the tumor of that specific transcript, and minimal to no counts in the other patients (Figure 3.5A,B). The top upregulated gene for patient A, ENSCAFG00000041995, encodes a predicted long non-coding RNA of 1359 bp with unknown functions (Figure 3.5A). Interestingly, this gene was found to be the second top downregulated gene when ordered by fold-change for the group analysis, with a log2 fold-change of −10.25 and FDR of 0.006 (Table 3.2). For patient B, the top upregulated gene was LOC403585, also known as serum amyloid A1, which encodes a protein that is ex-pressed in response to inflammation (Figure 3.5B). In the group analysis, this gene was found to be the second top upregulated gene when ordered by fold-change, with a log2 fold-change of 12.75 and an FDR of 0.0004. The top upregulated gene for both patients C and G was *TFPI2* (tissue factor pathway inhibitor 2), which encodes a serine proteinase inhibitor that has been identified as a tumor suppressor in a variety of cancers (Figure 3.5C). The top upregulated gene in patient D, *COL11A1* (collagen type XI α 1 chain), encodes a protein component of pro-collagen type XI, a major component of bone tissue (Figure 3.5D). For patient E, the top upregulated gene was *SFRP2* (secreted frizzled related protein 2), which encodes a protein involved in Wnt signaling (Figure 3.5E). The top upregulated gene in Patient F was ENSCAFG00000028460, which encodes two long non-coding RNA transcripts with unknown functions (Figure 3.5F). These three genes, *COL11A1*, *SFRP2*, and ENSCAFG00000028460, were not observed in the top 20 down- or up-regulated genes from the group analysis.

Figure 3.5. Top upregulated gene in each patient.

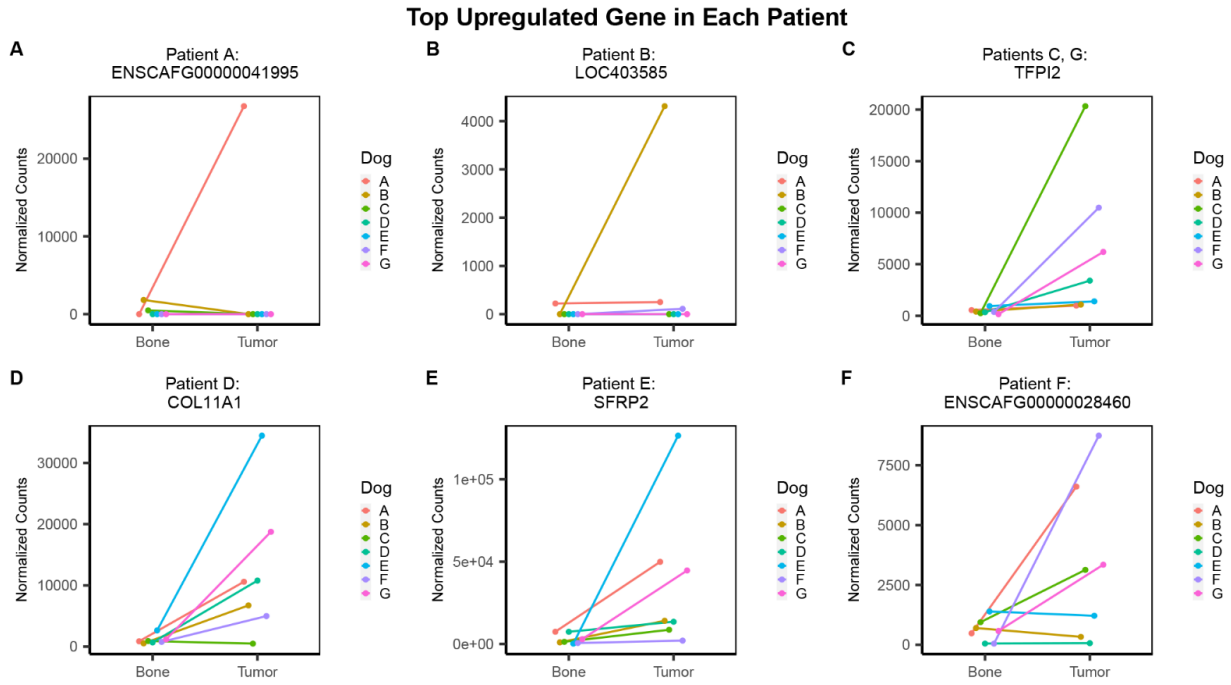


Figure 3.5. Plots of the top upregulated gene for each dog based on individual-level analysis reveal further variation among patients. The normalized counts of the top upregulated gene in each patient is shown. The top upregulated gene for patient A was ENSCAFG00000041995 (A). The top upregulated gene for patient B was LOC403585 (B). Patients C and G shared the same top upregulated gene, *TFPI2* (C). The top upregulated gene for patient D was *COL11A1* (D). The top up-regulated gene for patient E was *SFRP2* (E). The top upregulated gene for patient F was ENSCAFG00000028460 (F).

Similarly, the top downregulated gene for each patient was determined and more overlap among individuals was observed (Figure 3.6). The top downregulated gene for patient A was *CYTL1* (cytokine-like 1), which encodes a protein that is expressed in bone marrow and cord blood mononuclear cells with the CD34 surface receptor[114] (Figure 3.6A). The top downregulated gene for patients B, E, and G was ENSCAFG00000034058, which encodes a

1501 bp long non-coding RNA with uncharacterized function (Figure 3.6B). The top downregulated gene for patient C was *MYOC* (myocilin) which encodes a protein involved in cytoskeletal function (Figure 3.6C). The top downregulated gene in patients D and F was *MEPE* (matrix extracellular phosphoglycoprotein) which encodes a protein component of the extracellular matrix of bone (Figure 3.6D).

Figure 3.6. Top downregulated gene in each patient.

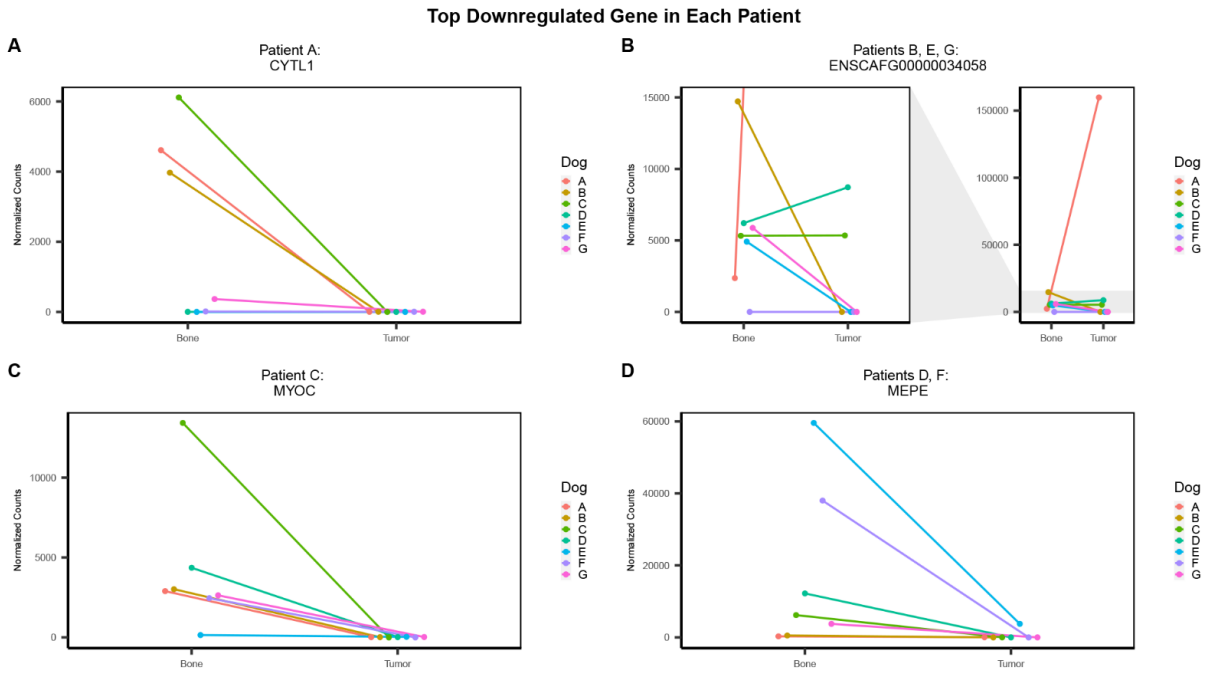


Figure 3.6. The top downregulated gene for each individual dog is shared among some patients. The normalized counts of the top downregulated gene in each patient are shown. The top downreg-ulated gene in patient A was *CYTL1* (A). Patients B, E, and G share the same top downregulated gene, ENSCAFG00000034058 (B). The top downregulated gene in patient C was *MYOC* (C). Patients D and F share the same top downregulated gene, *MEPE* (D).

To compare the individual and group analysis results, we extracted the log2 fold-change values and adjusted *p*-values (FDR) from the group analysis for each top gene from the individual analysis (Table 3.3). While the results show similarities, the top upregulated gene in patient A (ENSCAFG00000041995) shows distinct opposition in terms of the direction of fold-change from the group analysis (L2FC = −10.25). In fact, while this gene is significantly over-expressed in patient A's tumor, it is modestly downregulated in two tumors (patients B and C) and shows no evidence of expression in tumor or normal bone for the remaining patients (Figure 3.5A). Furthermore, this gene was identified as the second most downregulated gene from the group results when ordered by log2 fold-change (Table 3.2).

Table 3.3. Comparison of group and individual-level analyses.

		Patient	Gene		FDR	Log2FC
Individual Analysis Results	**Top Upregulated Gene**	A	ENSCAFG00000041995	**Group Analysis Results**	6.39E-03	-10.25
		B	LOC403585		4.37E-04	12.75
		C	TFPI2		3.60E-06	3.23
		D	COL11A1		1.49E-09	3.18
		E	SFRP2		7.36E-06	3.19
		F	ENSCAFG00000028460		4.15E-02	1.79
		G	TFPI2		3.60E-06	3.23
	Top Downregulated Gene	A	CYTL1		2.34E-04	-6.65
		B	ENSCAFG00000034058		3.31E-02	-5.92
		C	MYOC		2.58E-13	-7.97
		D	MEPE		1.06E-12	-7.47
		E	ENSCAFG00000034058		3.31E-02	-5.92
		F	MEPE		1.06E-12	-7.47
		G	ENSCAFG00000034058		3.31E-02	-5.92

Table 3.3. The top up- and downregulated gene for each patient according to log2 fold-change values from individual analysis shows some similarities to group analysis results. The top upregulated gene in patient A (ENSCAFG00000041995) shows distinct opposition in terms of direction of fold-change in the group results.

3.4 Discussion

Osteosarcoma is a highly complex and volatile bone tumor primarily seen in pediatric patients for which treatment has remained stagnant for almost 40 years. It is well-established that canine OSA is an ideal model for studying the human disease, as well as developing and testing therapeutic alternatives, such as precision medicine concepts. Using dogs, we aim to fill gaps in the current knowledge of OSA tumorigenesis while identifying potential therapeutic targets that can be tested and evaluated in a clinical setting.

While the technology of precision medicine has rapidly advanced, the overall clinical efficacy of such therapies leaves much to be desired, due primarily to the inherent heterogeneity and rapid evolution of drug-resistant mechanisms that define cancer[115]. Though cancer is traditionally considered a disease of genotypic origin, the phenotype reflects the accumulated genetic complexity and actionable targets that may be exploited. Transcriptomic sequencing provides a snapshot of gene expression and a link between genotypic and phenotypic landscape. In the context of cancer, where aberrant transcriptional patterns are pervasive, transcriptome profiling can identify and quantify changes in gene activity that are distinctive of tumors. Using this information, we can identify differentially expressed genes associated with various cellular processes and pathways that may be broadly expressed among patients to guide current treatment options and develop new therapies.

Traditional differential gene analysis groups individuals together to derive statistically meaningful DEGs; however, this approach fails to account for intra- and inter-tumoral heterogeneity and therefore is limited in its application for precision medicine. Deriving individual fold-change values provides additional insight into the differences among patients and

potential targets that may be beneficial for a subset of patients. Individual-level analyses are difficult due to the lack of statistical inference that can be derived from a single sample. To address this limitation, we selected only the significant DEGs from the group analysis to be carried forward to the individual-level analysis, thereby providing some statistical basis for the individual results. However, it must be recognized that some distinctive gene expression alterations may be so specific to one individual that they will be lost from this analysis.

Using this concept, we sequenced the transcriptomes of seven primary canine OSA tumors and patient-matched normal bone samples to derive differentially expressed genes (DEGs). In contrast to previously published studies, all samples were collected before chemotherapy and evidence of macro-metastatic lung disease to limit confounding interpretation of the results. Furthermore, normal bone samples were collected from each patient to generate suitable baseline gene expression levels. To our knowledge, this is the first study to derive normal expression for each patient using a biologically appropriate sample (bone) which is representative of the tumor's cell-of-origin (osteoblasts and osteocytes, depleted of bone marrow) in primary osteosarcoma prior to chemotherapy. This pairwise comparison reduces biological variability, increases statistical power, and provides a more thorough perspective of differential gene expression.

The clustering of the DEGs derived from the traditional group analysis showed dis-tinct relationships between tumor and normal samples as a group. The wide dispersion and variability in the tumor samples, as evidenced in the PCA plot (Figure 3.2C), is believed to be related to the intra- and inter-tumoral heterogeneity. The bone tissues were depleted of bone marrow and any cells other than those embedded within the bone matrix, whereas the tumor samples may contain many different types of cells at various stages of differentiation (for example, infiltrating

immune cells and de-differentiated cells with a more stem-cell-like phenotype). Furthermore, even tumors of the same type or source exhibit heterogeneity in their molecular profile[116].

We used traditional group analysis to statistically identify the top genes that are broadly dysregulated among the patients. Significant DEGs were selected based on padj < 0.05, FC > 2 or < −2, and a base mean count > 10, and these results were ranked based on log2 fold-change as well as an adjusted *p*-value. The top upregulated genes from the group analysis included *GTSE1* (when sorted by padj) and ENSCAFG00000044295 (when sorted by log2 fold-change). GTSE1 regulates the G1/S cell cycle transition and has been reported to be overexpressed in other human cancers[117–120]. *GTSE1* has been implicated in conferring cisplatin resistance in human osteosarcoma, though more studies are necessary to confirm its role in OSA tumorigenesis[121]. ENSCAFG00000044295 encodes a 120 amino acid protein with uncharacterized functions that does not correlate to any human ENSEMBL IDs. The top downregulated genes included *PLIN1* (when sorted by padj) and *ARHGEF1* (when sorted by log2 fold-change). The downregulation of *PLIN1* mRNA has been reported in breast cancer and is considered a tumor suppressor in breast cancer progression[122,123]. The dysregulation of rho GTPases, including *ARHGEF1*, are commonly reported in a variety of cancers[124].

We showed that the group results can sometimes be misleading due to the heterogeneity among individuals. In some cases, a bi-directional change in expression was observed among patients. To circumvent this, we used individual-level analysis to derive log2 fold-change values for each patient using only the significant genes identified in the group analysis. We used this information to identify the top upregulated and downregulated genes in each patient and compare their expression across individuals. The top downregulated gene was shared between patients B, E, and G (ENSCAFG00000034058), as well as for patients D and F (*MEPE*).

ENSCAFG00000034058 encodes a long non-coding RNA with unknown functions. Co-expression network analysis has revealed MEPE as a dysregulated gene in human osteosarcoma[125]. The top upregulated gene was different for all individuals (A: ENSCAFG00000041995, B: *LOC403585*, D: *COL11A1*, E: *SFRP2*, F: ENSCAFG00000028460), except for patients C and G (*TFPI2*). The upregulation of *COL11A1* and *SFRP2* has been reported in human osteosarcoma[106,126]. *TFPI2*, a tumor suppressor gene, has been shown to be upregulated in other cancers, including colorectal, gastric, and prostate[127–129]. Both ENSCAFG00000041995 and ENSCAFG00000028460 encode uncharacterized long non-coding RNAs that do not correlate to known human ENSEMBL gene IDs. More studies are needed to evaluate their role in OSA tumorigenesis.

The top upregulated gene in patient A (ENSCAFG00000041995) showed an inverse relationship in terms of the direction of fold-change in the group analysis. In fact, this gene was the second-most downregulated gene according to the group results sorted by log2 fold-change, despite having an adjusted p-value less than 0.05. The observed differences in gene expression among individuals may be due to (1) the effects of different mutations in combination with genetic and/or environmental effects, (2) the stage of the tumor, or (3) initial RNA quality. This example represents the hallmark conclusion of this study; despite relatively stringent filtering conditions to minimize noise and confidently identify the top DEGs, traditional group analyses can be misleading and lead to the identification of therapeutic targets that may be ineffective for most patients.

Consistent with the conventional theme of research, this study is not without limitations. With a small sample size of seven, the statistical power is limited. As with all next-generation sequencing studies, computational challenges may limit the downstream interpretation. For

example, the extent of gene annotation in the canine reference genome (CanFam3.1) limited our ability to identify the appropriate human orthologs. Additionally, it is becoming increasingly evident that non-coding RNAs such as micro-RNA and long non-coding RNA (lncRNA) play a dynamic role in tumor progression[102,103]. This study utilized a sequencing approach that excluded RNAs without poly-A tails. However, some lncRNAs are produced and processed with poly-A tails and as such were included in the sequencing. Furthermore, the heterogeneity of tumors makes it difficult to under-stand and characterize the phenotypic landscape, its components, and how those components affect tumor progression. It is also difficult to identify the distribution of these changes; for example, a highly upregulated gene may be vastly over-expressed, but only prevalent in only a small subset of cells. The clonal hypothesis implies a spatial distribution of different tumor cell populations, and using a single small section for sequencing may impair our ability to detect other cellular subsets. With the advancement of single-cell RNA sequencing and the development of accompanying bioinformatics tools, we may be able to elucidate the details of a tumor's phenotype more clearly. While more studies are needed to evaluate the role these dysregulated genes play in tumorigenesis, these results support the notion that traditional group DEG analysis should be supplemented by individual-level analysis in the context of precision medicine for the identification of potential therapeutic targets.

3.5 Conclusions

Canine OSA serves as an excellent model for determining molecular targets and developing and evaluating personalized precision treatment. The relatively ample availability of dogs with spontaneously occurring OSA provides a powerful and underused translational model. Limb amputation offers the opportunity to easily collect tumor samples for sequencing as well as

appropriate normal tissue, for example, from the distal phalanges, to allow comparisons within the same patient prior to chemotherapy[107].

The goal of precision medicine is to use an individual tumor's molecular profile, rather than tumor category or stage, to inform therapeutic decisions and design novel treatments. Group DEG analysis has been traditionally used to identify relevant genes, but these results do not account for intra- or inter-tumoral differences. Despite using a multi-factor paired design to account for differences between individuals, as well as strict filtering parameters to minimize noise, our results showed some conflicting elements and variation in DEG expression. These results suggest that identifying significantly up- or downregulated genes as potential therapeutic targets using traditional group analysis may not always be appropriate, even when stringent filtering conditions are used. While these conclusions may be well established in the bioinformatics community, they may be lesser known in the precision oncology field, where DEG analysis is pervasive. Supplementing DEG analyses with individual-level analyses provides additional insight into the intra- and inter-tumoral heterogeneity.

The novelty of this study lies in the sample set as well as the analytical workflow. In contrast to other published studies, (1) all samples were obtained prior to chemotherapy, which can alter the phenotype; (2) normal bone samples (depleted of contaminating tissue) from each patient were used as the comparator tissue for baseline gene expression; and (3) DEG analysis was supplemented by individual-level analysis and compared to group DEG analysis. This study demonstrates that the variation in DEG expression be-tween individuals, obtained using traditional group DEG analysis, is sufficient to warrant further individual-level analysis to identify more effective targets for precision therapy.

CHAPTER 4

Single-Nuclei Multiome (ATAC + Gene Expression) Sequencing of a Primary Canine Osteosarcoma Elucidates Intra-Tumoral Heterogeneity and Characterizes the Tumor Microenvironment

4.1 Introduction

Osteosarcoma (OSA) is a highly malignant bone tumor occurring most often in pediatric and adolescent humans as well as large breed dogs. It is highly heterogeneous and aggressive, with poor survival rates for both species. The median survival time for dogs undergoing amputation of the affected limb in combination with chemotherapy is approximately one year after diagnosis, with most dogs succumbing to metastases[130]. Due to its similarities, canine OSA represents a powerful translational model for understanding the human disease as well as designing and testing clinical therapeutics.

Tumor heterogeneity makes treatment difficult due to the evolution of cell subsets that impact tumor growth, metastasis, and drug resistance[131]. Both intrinsic and extrinsic factors contribute to tumor heterogeneity, including the accumulation of genetic mutations, epigenetic factors affecting cellular activity and identity, and microenvironmental influences such as cell-cell interactions[131]. Single-cell sequencing is a powerful approach to evaluating tumor heterogeneity by identifying various cell types and states within a tumor. Identifying the relative proportion of cells with aberrant transcription patterns provides critical information regarding the potential effectiveness of therapies.

However, obtaining viable cells after tissue dissociation is a prerequisite for single-cell sequencing and represents a major limitation to this technology, particularly for difficult tissues such as osteosarcoma. Since OSA is derived from bone, it often contains bone matrix which is

very rigid and difficult to homogenize. The dissociation procedure selects for cells that survive this process, potentially excluding rare cell types, and may also result in RNA degradation. Furthermore, enzymatic and/or mechanical tissue dissociation may alter the cell's phenotype by inducing a transcriptional stress response, resulting in artifacts upon sequencing[132,133]. In our laboratory, efforts to isolate single-cell suspensions from primary canine OSA, including various mechanical and enzymatic protocols, have resulted in poor cell viability unsuitable for single-cell sequencing. Additionally, removal of dead cells to increase viability may not accurately reflect the true biology of the tumor due to the selection of surviving cells.

In contrast, single-nuclei sequencing circumvents the cell viability challenge by lysing the cells to obtain nuclei. However, nuclei quality is a critical factor to consider, and nuclear membranes should appear intact and with minimal blebbing under high power microscopy after isolation. Importantly and in contradistinction to single-cell, single-nuclei sequencing can be performed on frozen archived samples.

Gene expression sequencing of single-*nuclei* differs from single-*cell* in the information it provides. Single-nuclei sequencing captures polyadenylated RNA transcripts that are actively being transcribed in the nucleus, whereas single-cell sequencing captures all polyadenylated RNA within the cell's cytoplasm. Despite these differences, single-nuclei sequencing has been shown to provide equivalent gene detection signatures and accurate cell identification while minimizing bias compared to single-cell sequencing[132–135].

10x Genomics uses a microfluidic-based approach to partition single nuclei into gel beads containing barcoded primers and enzymes. This technology allows single nuclei to be captured and barcoded so that after sequencing, reads can be traced back to the corresponding

cell/nucleus. Identifying individual cell phenotypes and genotypes can reveal the inherent intra-tumoral heterogeneity at high resolution and provide insight into tumorigenesis.

To date, this is the first study to use single-nuclei multiome sequencing, including ATAC (Assay for Transposase-Accessible Chromatin) and GEX (Gene Expression) sequencing, of a treatment-naïve primary canine osteosarcoma, to simultaneously capture the transcriptomic and epigenomic profiles in the same nucleus.

4.2 Materials and Methods

4.2.1 Patient/Sample Description

Osteosarcoma tissue was obtained from a 7-year-old male Doberman Pinscher presenting to the Wilford and Kate Bailey Small Animal Teaching Hospital at Auburn University for limb amputation. Importantly, the samples were obtained prior to chemotherapy, radiation, or evidence of pulmonary macro-metastatic disease. A sample was subjected to histopathology and confirmed to be osteoblastic osteosarcoma. The tumor specimen was diced into approximately 50 mg pieces, immediately flash frozen in liquid nitrogen, and stored at -80C. This sample was stored at -80C for approximately 5 years prior to nuclei isolation.

4.2.2 Nuclei Isolation

Nuclei were isolated from 42 mg flash frozen OSA tissue by following the Nuclei Isolation kit protocol from 10x Genomics (Pleasanton CA, USA, CG000505 Rev A) with minor adjustments to enhance homogenization while retaining nuclear morphology. The Lysis Buffer provided with the kit was diluted with phosphate buffered saline (PBS) to 0.5 strength and the sample was briefly homogenized (1-2 sec) using a bladed homogenizer on ice, followed by a 5-

minute incubation on ice. The nuclei isolation protocol from 10x Genomics was then followed according to the manufacturer's directions. Nuclei were visualized and counted using trypan blue (ThermoFisher Scientific, Waltham, MA, USA) and ViaStain acridine orange/propidium iodide (AO/PI) (PerkinElmer, Waltham, MA, USA) to determine quality and quantity using the Keyence BZ-X810 microscope with 100x oil-immersion objective. Nuclei were counted by hand using a hemocytometer in addition to using the Biorad TC20 automated cell counter (Biorad, Hercules, CA, USA) to determine the concentration for library preparation.

4.2.3 Library Preparation

Approximately 10,000 nuclei were used to generate ATAC and gene expression libraries. Libraries were prepared according to the Chromium Next GEM single-cell multiome ATAC + gene expression user guide (10x Genomics, Pleasanton, CA, USA, CG000338 Rev F). Briefly, the single-nuclei along with a master mix, 10x barcoded gel beads, and partitioning oil are loaded onto the Chromium Next GEM Chip J to generate single-nuclei gel bead-in-emulsions (GEMs). Pre-amplification PCR was performed and the GEX and ATAC libraries were split for further processing separately. The prepared libraries were shipped to Novogene (Sacramento, CA, USA) for 150 bp paired-end sequencing on the Illumina NovaSeq 6000 platform using the sequencing parameters recommended by 10x Genomics.

4.2.4 Bioinformatic Processing, Dimensional Reduction, and Weighted Nearest Neighbor Analysis

The canine reference genome "canFam6", also known as Dog_10K_Boxer_Tasha (GCF_000002285.5), was used to align and count reads using Cell Ranger (v7.1.0). Analysis was

accomplished using the Cell Ranger ARC Count (v2.0.2) pipeline and the output was loaded into Seurat (v4.3.0) for further processing. Seurat was used to filter the data using a feature threshold ($200 < n < 30000$) and counts threshold ($50 < n < 50000$). The SCTransform function in Seurat was used to normalize and transform the GEX data using a regularized negative binomial regression model, as described previously[136]. Dimensionality reduction was accomplished using PCA and UMAP embedding was used to visualize clusters using Seurat. ATAC data was processed using latent semantic indexing (LSI), which combines term frequency inverse document frequency (TF-IDF) normalization followed by singular value decomposition (SVD) of the top identified features. A weighted combination of the GEX and ATAC data was used to construct a Weighted nearest neighbor (WNN) graph and clusters were identified using the SLM algorithm in Seurat.

4.2.5 Cell Cluster Annotation

ScType was used to annotate cell clusters based on a given reference set of upregulated/downregulated markers and designated cell types[137]. A custom annotation set was created using single-cell markers accessed from CellMarker2.0 and annotated using ScType[138]. Annotation of "tumor" vs. "normal" clusters was based on differential expression analysis of the bulk RNA sequencing results produced in Nance et al[139]. This dataset included bulk RNA sequencing of 7 primary canine osteosarcoma tumors, including the tumor in the current study, along with patient-matched normal bone. Using this data, a custom annotation set was created to designate "tumor cells" from "normal cells" based on upregulated genes (log2 fold-change>2 and padj<0.05) in tumor and bone, respectively. ScType was then used to annotate the cell

clusters and overlay the results on the weighted nearest neighbor UMAP plot. All markers used for cluster annotation are listed in Supplemental Table 4.2.

4.2.6 Copy Number Variation (CNV) Analysis

Genomic copy number was inferred for the osteoblast clusters using the Bioconductor package inferCNV with cluster 2 (fibroblasts), cluster 3 (endothelial cells), cluster 4 (myeloid cells), cluster 5 (osteoclasts), cluster 6 (osteocytes), and cluster 8 (memory CD4+ T cells) as the normal reference[140].

4.2.7 Differential Gene Expression for Identification of Cluster Markers

Using the variance-stabilized GEX data, the 'FindAllMarkers' function in Seurat was used to identify positive markers for clusters compared to all remaining cells using the roc/standard AUC classifier test (min.pct=0.25 and logfc.threshold=0.25). The positive markers for each cluster were subjected to subsequent pathway analysis using all genes in the canine database as the reference.

4.2.8 Enriched Pathway Analyses

Gene set enrichment analysis using Hallmark and Canonical pathways in the Canis lupus familiaris genome was accomplished using the R package singleseqgset (v0.1.2.9000). To identify enriched GO Biological Processes among clusters, PANTHER (v17.0) was used to perform a statistical overrepresentation test (Fisher's exact test with FDR correction) using the GO Ontology database (DOI: 10.5281/zenodo.6799722 Released 2022-07-01)[141,142].

4.3 Results

To explore the cellular heterogeneity and microenvironment of canine OSA, single-nuclei multiome (ATAC + Gene Expression, GEX) sequencing was conducted on a primary canine OSA tumor lesion obtained prior to chemotherapy or macro-metastatic lung disease.

4.3.1 Quality and Dimensionality of the Single-Nuclei and Sequencing Data

After isolation and prior to library preparation, nuclei were assessed for quality and quantity using high power microscopy in combination with AO/PI fluorescent staining. The majority of nuclear membranes appeared intact with minimal blebbing (Figure 4.1). The nuclei were in sufficient quantity for library preparation, sequencing, and downstream analysis.

Figure 4.1. High-power microscopy to evaluate single-nuclei quality.

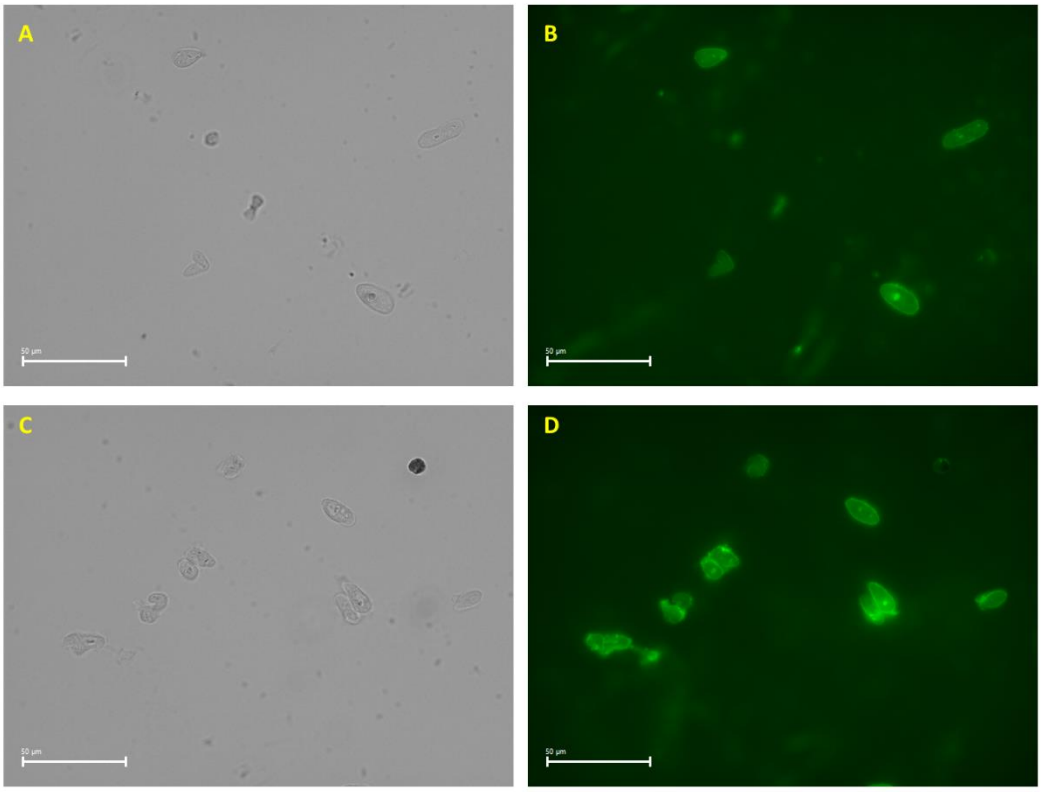


Figure 4.1. Phase contrast (A, C) and AO/PI-stained nuclei under fluorescent microscopy (B, D) shows nuclear membranes are intact with minimal blebbing.

Approximately 10,000 nuclei were used for multiome (ATAC + GEX) library preparation using 10x Genomics technology and sequenced (150 bp PE) on the Illumina NovaSeq 6000 platform. After classification of each barcode into cell and non-cell groups, there were an estimated 5969 total nuclei sequenced with 8462 median ATAC high-quality fragments per cell and 2603 median GEX genes per cell (Fig 4.2A). To examine true nuclei, filtering parameters (feature threshold 200-30000 and count threshold 50-50000) were applied to both ATAC and GEX data to eliminate empty droplets and doublets. After filtering, 5849 total nuclei and 23784 genes were included in the downstream analysis (Fig 4.2B).

Figure 4.2. Quality filtering to eliminate doublets and empty droplets.

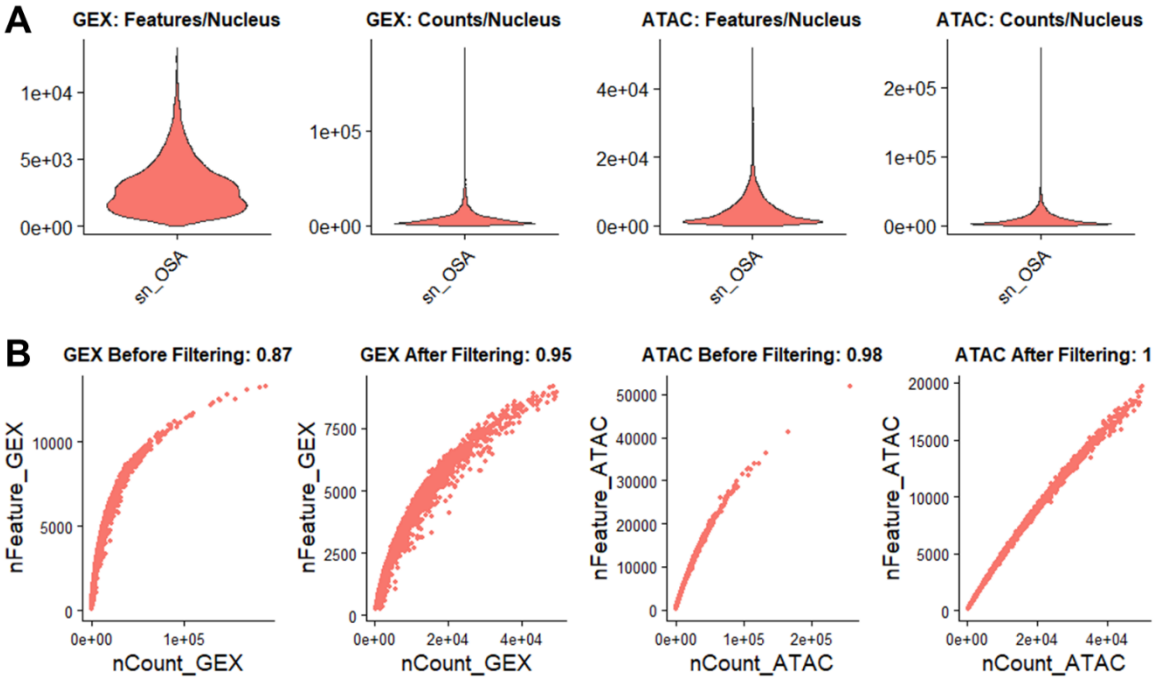


Figure 4.2. Violin plots of the total features and counts detected per nucleus for both GEX and ATAC data prior to filtering (A). Scatterplots of the number of features by counts per nucleus before and after filtering for both GEX and ATAC data (B). Each dot on the figure represents a single nucleus. “Features” refers to the number of genes detected in each nucleus, whereas “counts” refers to the number of molecules (UMIs, Unique Molecular Identifiers) detected in each nucleus.

4.3.2 Unsupervised Clustering to Evaluate Cellular Heterogeneity of Primary Canine OSA Reveals 9 Distinct Clusters

Based on unsupervised clustering using principal component analysis (PCA) and graph-based dimensional reduction, we identified 9 total cell clusters (c0-8) in the GEX (Fig 4.3A) and ATAC data (Fig 4.3B). The weighted nearest neighbor (WNN) procedure in Seurat v4 integrates multimodal data from the same cell to generate a unified representation of the dataset[143]. Using a weighted combination of the GEX and ATAC data, a WNN UMAP plot was generated to elucidate additional structure in the cellular clustering of canine osteosarcoma (Fig 4.3C).

Clusters 0, 1, and 2 were more closely grouped together and less discrete in the ATAC data, whereas the GEX data shows a more distinct relationship among these clusters (Fig 4.3C). These results suggest a relationship in the epigenetic programming of clusters 0, 1, and 2, despite differences in gene expression patterns. Conversely, the ATAC UMAP plot shows more separation between clusters 1 and 7, while the GEX UMAP plot shows a closer relationship and less clear distinction. This suggests that clusters 1 and 7 are closely related based on RNA expression, but display different accessible motifs.

Figure 4.3. Cellular heterogeneity in primary canine OSA reflected by 9 cell clusters.

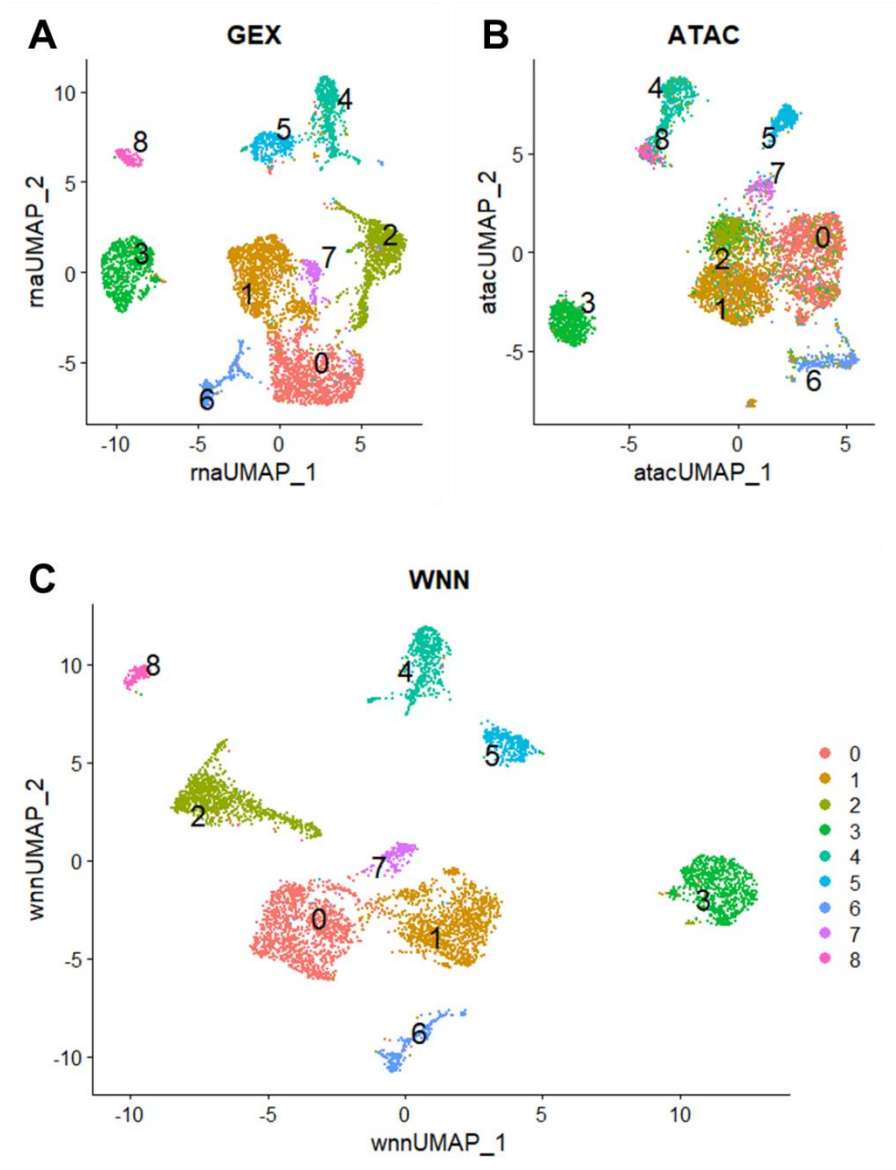


Figure 4.3. UMAP plot shows 9 total clusters (c0-8) for GEX (A), ATAC (B), and weighted nearest neighbor (WNN) graph which combines both modalities (C). Each dot represents a single nucleus, and the color corresponds to the cluster.

Clusters were numbered 0-8 and contained increasing numbers of cells. For example, cluster 0 contained the most cells (1284/5849 cells, 21.9%), while cluster 8 contained the least number of cells (162/5849 cells, 2.8%) (Table 4.1).

4.3.3 Cluster Annotation in Primary Canine OSA

To identify clusters based on cell type, a reference set containing single-cell markers for bone, osteosarcoma, and immune cells was used to annotate the clusters using ScType (Fig 4.4A)[137]. Osteoblasts were associated with clusters 0, 1, and 7 (2713/5849 cells, 46.4%); Fibroblasts were associated with cluster 2 (1023/5849 cells, 17.5%); Endothelial cells were associated with cluster 3 (798/5849 cells, 13.6%); Myeloid cells were associated with cluster 4 (548/5849 cells, 9.4%); Osteoclasts were associated with cluster 5 (333/5849 cells, 5.7%); Osteocytes were associated with cluster 6 (272/5859 cells, 4.6%); and Memory CD4+ T cells were associated with cluster 8 (162/5849 cells, 2.8%) (Fig 4.4A, Table 4.1).

Figure 4.4. Cluster annotation with known cell markers and bulk OSA tumor/normal bone RNAseq markers.

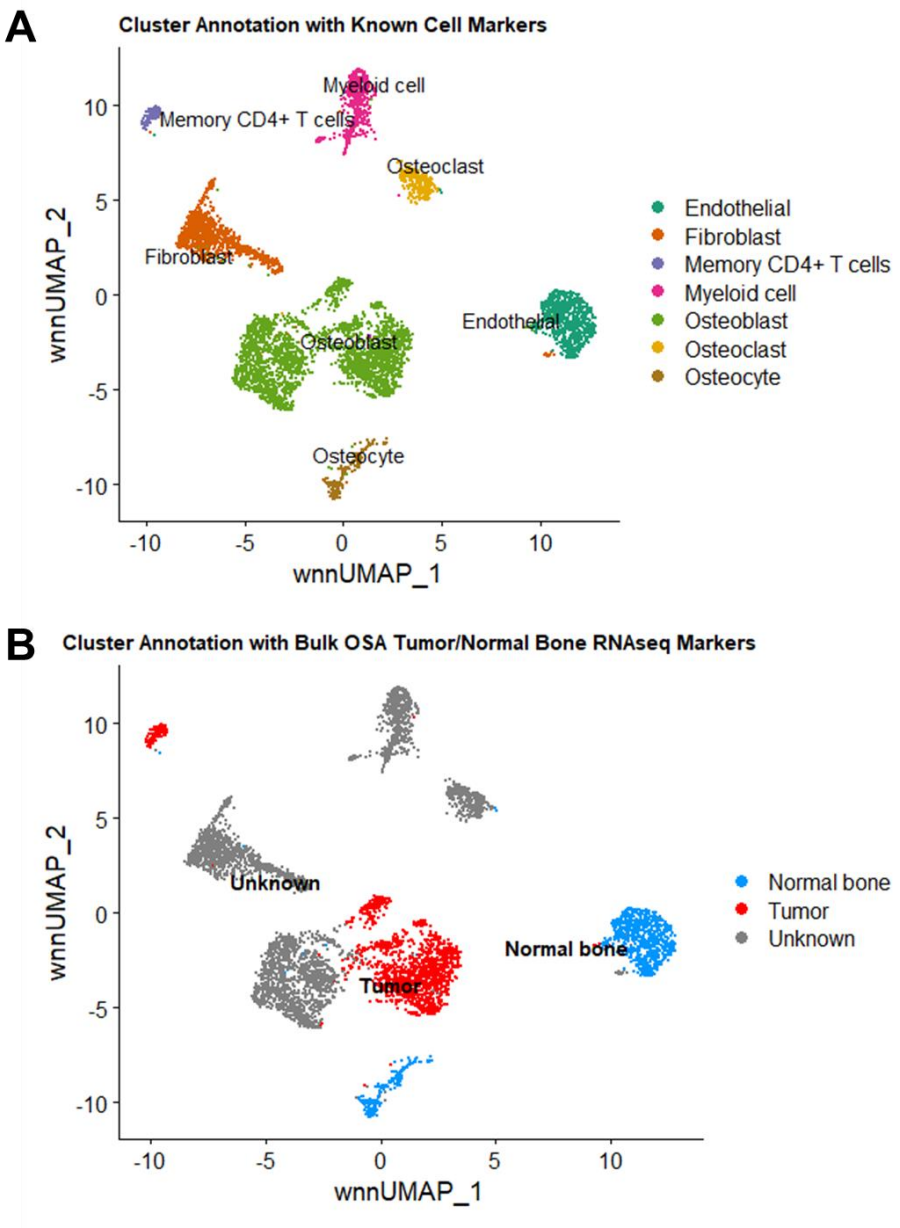


Figure 4.4. Cell cluster annotation based on known single-cell marker genes (A) and upregulated genes in the corresponding bulk OSA tumor/normal bone dataset (B). Each dot represents a single nucleus, and the color corresponds to the annotated cell group name.

To annotate the clusters based on “tumor” vs. “normal bone”, we generated an in-house annotation set using significantly upregulated genes and downregulated genes (padj<0.05 and FC

< -2 and >2) in the bulk canine OSA tumor vs. patient-matched normal bone dataset generated in Nance et al (Fig 4.4B)[139]. Based on these results, tumor cells were related to clusters 1 (osteoblasts), 7 (osteoblasts), and 8 (memory CD4+ T cells) (1591/5849 cells, 27.2%), which is consistent with the diagnosis of osteoblastic OSA. Normal bone was related to clusters 3 (endothelial cells) and 6 (osteocytes) (1070/5849 cells, 18.3%). Clusters 0 (osteoblasts), 2 (fibroblasts), 4 (myeloid cells), and 5 (osteoclasts) had unknown relation to the bulk RNA seq data (3188/5849 cells, 54.5%) (Fig 4.4B, Table 4.1).

Since bulk RNAseq is derived from a mixed population of cells, it is likely the tumor dataset included tumor-initiating cells that drive tumor formation, tumor-associated cells such as fibroblasts and immune infiltrates, and a small proportion of normal bone cells such as osteoclasts and endothelial cells. Therefore, we cannot distinguish between tumor cells that drive tumorigenesis vs. those that are passengers in the process based solely on these results. It is apparent, however, that bulk RNA sequencing captures only about 27% of the tumor's total cell population compared to single-nuclei RNA sequencing.

The sample processing technique to obtain normal bone for bulk RNA sequencing involved removal of bone marrow, which would include myeloid cells. Furthermore, osteoclasts also typically reside in the bone marrow, and while they can be recruited to the bone matrix, they occur in a much smaller proportion in comparison to other cell types[144]. Thus, we suspect that the inability to distinguish cluster 4 (myeloid cells) and cluster 5 (osteoclasts) as normal bone using markers derived from bulk RNA sequencing is due to sample processing which removed the majority of these cell types. Regardless, as cells of the tumor microenvironment, they are likely to play an important role in tumorigenesis.

Table 4.1. Cluster annotation using marker genes.

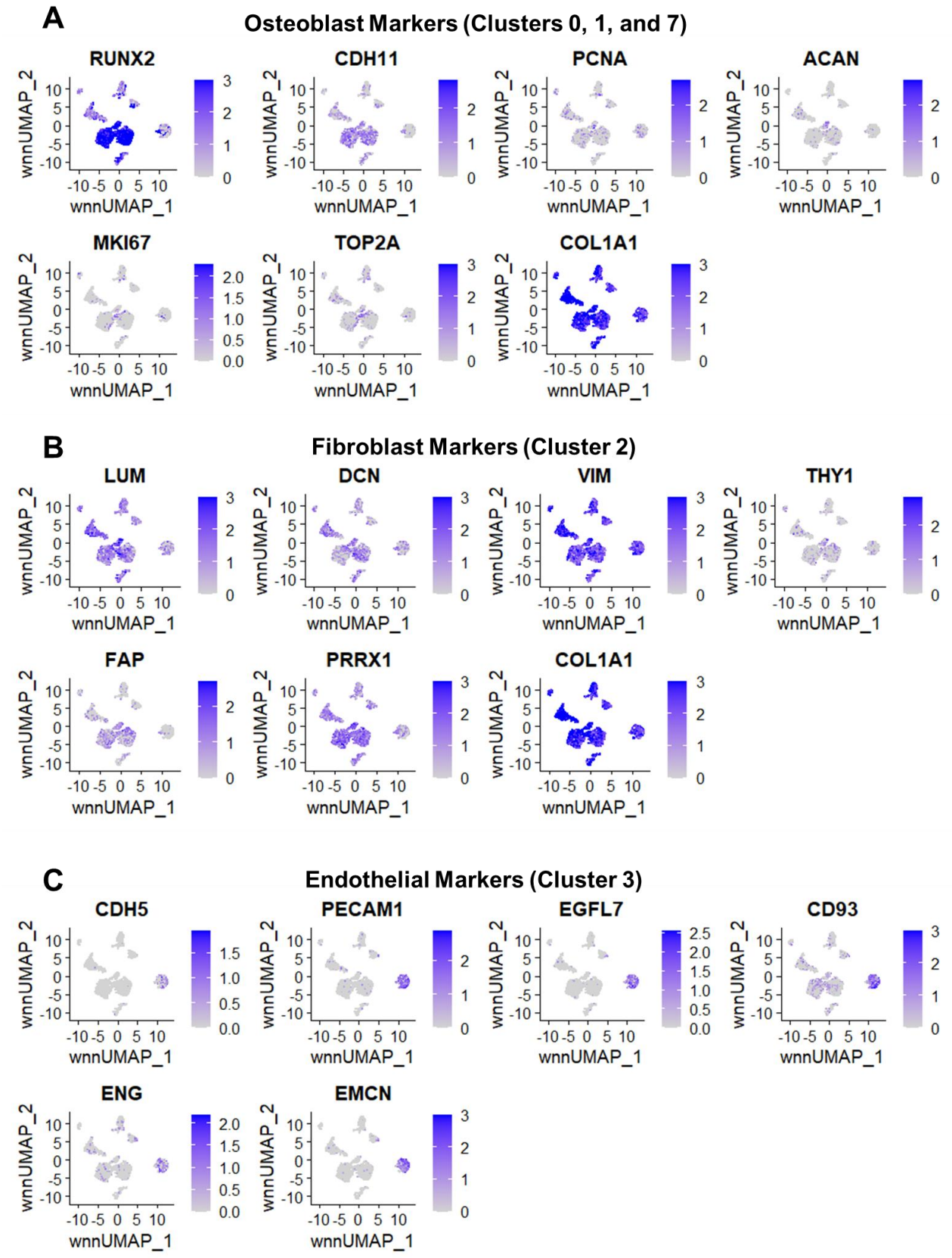
Cluster	# of Cells	% of Total Number of Cells	Bone/OSA/Immune Markers (ScType score)	Bulk Tumor/Normal Markers (ScType score)
0	1284	21.9%	Osteoblast (992)	Unknown (172)
1	1249	21.3%	Osteoblast (1452)	Tumor cells (1287)
2	1023	17.5%	Fibroblast (1113)	Unknown (136)
3	798	13.6%	Endothelial cell (6914)	Normal bone cells (419)
4	548	9.4%	Myeloid cell (3622)	Unknown (-4.84)
5	333	5.7%	Osteoclast (5236)	Unknown (-3.74)
6	272	4.6%	Osteocyte (700)	Normal bone cells (1276)
7	180	3.1%	Osteoblast (325)	Tumor cells (318)
8	162	2.8%	Memory CD4+ T cell (1364)	Tumor cells (77.7)

Table 4.1. Nine clusters (c0-8) were identified in primary canine osteosarcoma based on ATAC and GEX sequencing of 5849 individual nuclei. Clusters were annotated using known cell marker genes as well as marker genes derived from the bulk OSA tumor/normal bone dataset, which also contained the same patient's tumor and patient-matched normal bone.

Cluster annotation for each cell type was further inspected by plotting the expression of several marker genes derived from the CellMarker2.0 database that were used for annotation with ScType (Fig 4.5). Although this is not an exhaustive list of markers used for annotation, osteoblast markers included *RUNX2*, *CDH11*, *PCNA*, *ACAN*, *MKI67*, *TOP2A*, and *COL1A1* (Fig

4.5A); Fibroblast markers included *LUM*, *DCN*, *VIM*, *THY1*, *FAP*, *PRRX1*, and *COL1A1* (Fig 4.5B); Endothelial markers included *CDH5*, *PECAM1*, *EGFL7*, *CD93*, *ENG*, and *EMCN* (Fig 4.5C); Myeloid markers included *CD14* and *CD74* (Fig 4.5D); Osteoclast markers included *ATP6V0D2*, *DCSTAMP*, *CTSK*, *OCSTAMP*, *MMP9*, and *ACP5* (Fig 4.5E); Osteocyte markers included *GBLAP* (osteocalcin), *SPP1* (osteopontin), *CD86*, and *IBSP* (bone sialoprotein) (Fig 4.5F); Memory CD4+ T cell markers included *CD3E*, *CD3D*, *CTLA4*, *LCK*, *LTB*, and *CD2* (Fig 4.5G).

Figure 4.5. Markers for cell type based on cluster annotation.


A
Osteoblast Markers (Clusters 0, 1, and 7)
RUNX2
CDH11
PCNA
ACAN
MKI67
TOP2A
COL1A1
B
Fibroblast Markers (Cluster 2)
LUM
DCN
VIM
THY1
FAP
PRRX1
COL1A1
C
Endothelial Markers (Cluster 3)
CDH5
PECAM1
EGFL7
CD93
ENG
EMCN
wnnUMAP_1
wnnUMAP_2

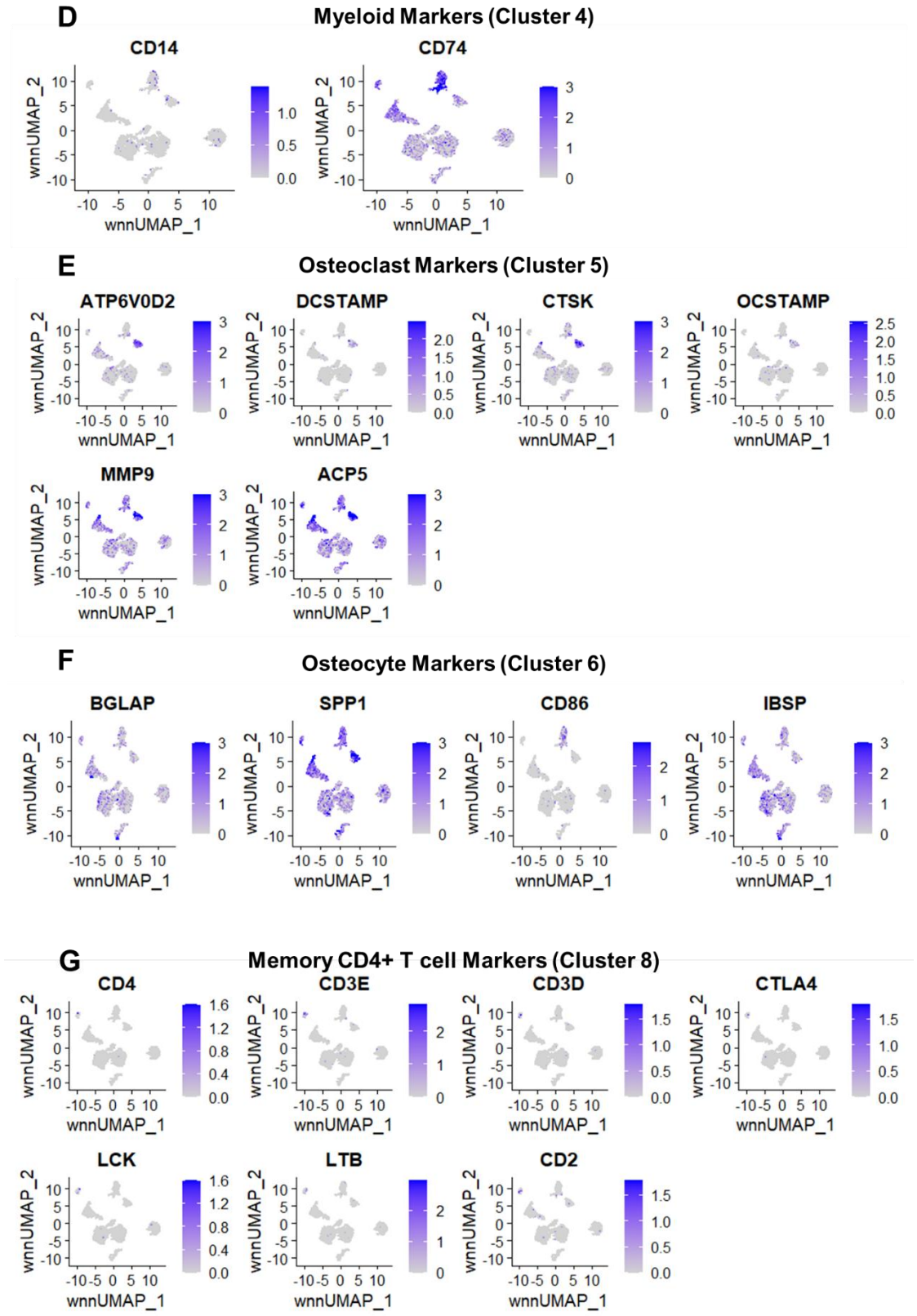

D
Myeloid Markers (Cluster 4)
CD14
CD74
E
Osteoclast Markers (Cluster 5)
ATP6V0D2
DCSTAMP
CTSK
OCSTAMP
MMP9
ACP5
F
Osteocyte Markers (Cluster 6)
BGLAP
SPP1
CD86
IBSP
G
Memory CD4+ T cell Markers (Cluster 8)
CD4
CD3E
CD3D
CTLA4
LCK
LTB
CD2
wnnUMAP_1
wnnUMAP_2

Figure 4.5. Marker gene expression based on cluster annotation for each cell type identified. Only some of the annotation marker genes are shown. Darker shades of purple indicate upregulated expression, gray indicates zero change.

Cluster annotation using the upregulated genes in bulk OSA tumor vs. corresponding normal bone RNAseq dataset included tumor markers HOXC10, SPAG5, TOP2A, IQGAP3, HELLS, MKI67, CLSPN, and RAD54L (Fig 4.6).

Figure 4.6. Markers for OSA tumor based on cluster annotation with bulk RNA sequencing results.

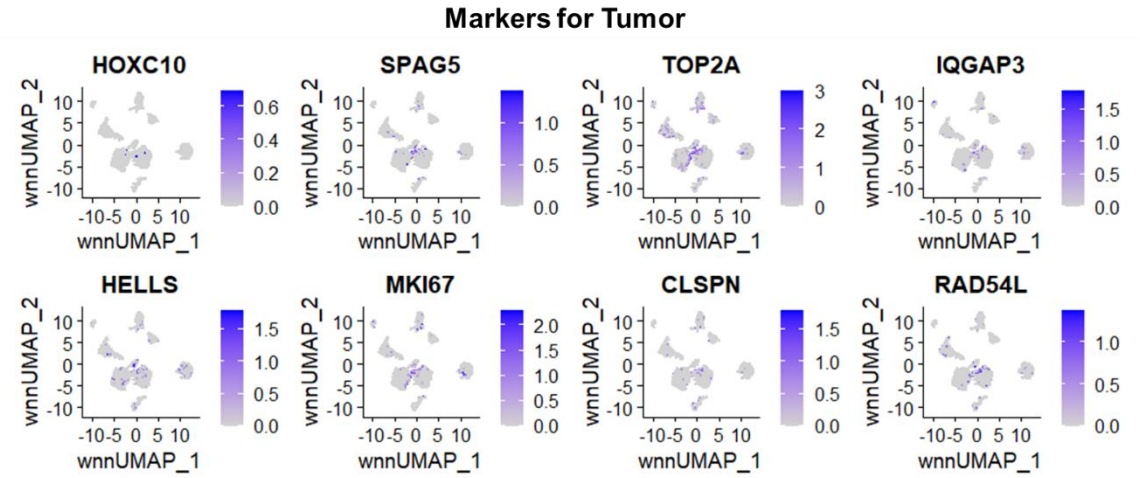


Figure 4.6. Expression of markers used for annotation based on upregulated genes in primary canine OSA/normal bone bulk RNA sequencing dataset. Darker shades of purple indicate upregulated expression, gray indicates zero change.

4.3.4 Copy Number Variation

OSA is characterized by significant genomic instability resulting in large-scale chromosomal copy number variations. To evaluate chromosomal structure, CNV analysis was performed for the osteoblasts (clusters 0, 1 and 7) using the remaining clusters as the normal reference. Significant amplifications were observed in chromosomes 12-14 and deletions were present in chromosomes 5, 18, and 20 in the osteoblast clusters (Fig 4.7A). Cluster 7 cells contain more CNVs compared to clusters 0 and 1 and their hierarchical relationship is reflected by the dendrogram. Compared to clusters 0 and 1, cluster 7 shows a distinct amplification of chromosome 24 and deletion of chromosome 26.

Figure 4.7. Heatmap of CNVs in osteoblastic clusters.

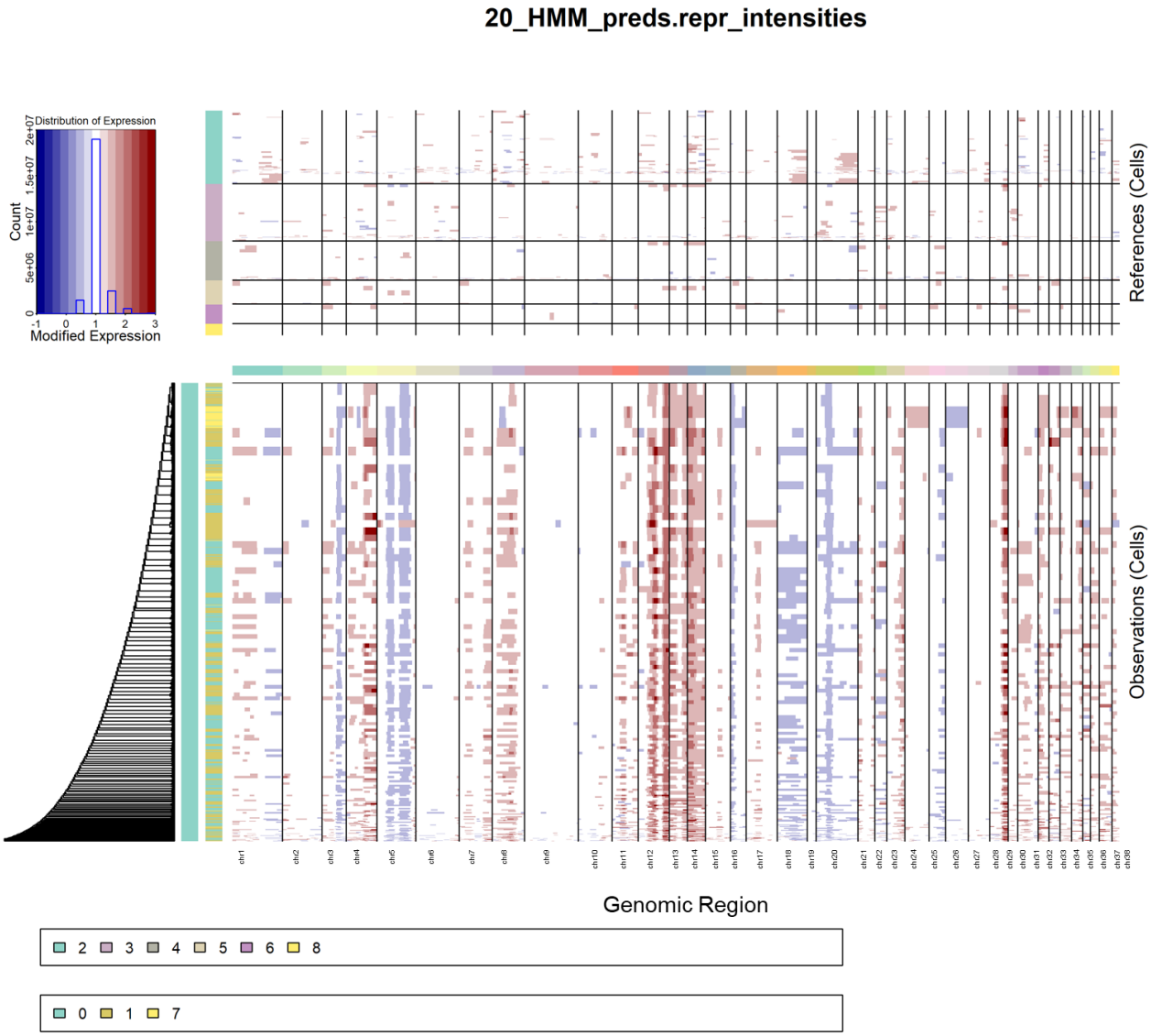


Figure 4.7. Heatmap of the CNVs identified in the osteoblast clusters 0, 1, and 7 using remaining clusters as the reference (A). Expression values for the normal cell clusters (depicted in the top heatmap) are subtracted from tumor cluster expression data (depicted in the bottom heatmap) to visualize differences. Rows are individual nuclei, columns are genes (ordered from left to right across the chromosomes); amplifications are colored red and deletions are colored blue.

4.3.5 Differentially Expressed Genes Define Clusters

After normalizing the UMI counts using a regularized negative binomial regression, highly variable features (genes) were identified to be used in downstream principal component analysis. Using the FindAllMarkers function in Seurat, markers were identified for every cluster compared to all remaining cells. The top 10 most highly variable genes according to the GEX data were *LDB2*, *PTPRG*, *ACP5*, *MMP9*, *CHRM3*, *CHAD*, *F13A1*, *SLC9B2*, *GPC5*, and *SLIT2* (Fig 4.8A). The top 5 differentially expressed genes defining each cluster were plotted on a heatmap (Fig 4.8B). Cluster 4 (myeloid), cluster 5 (osteoclasts), and cluster 8 (memory CD4+ T cells) share similar patterns of differential gene expression on the heatmap in Figure 4.8B, which is likely due to their related immunological functions and origins.

Figure 4.8. Differentially expressed genes define clusters in primary canine OSA.

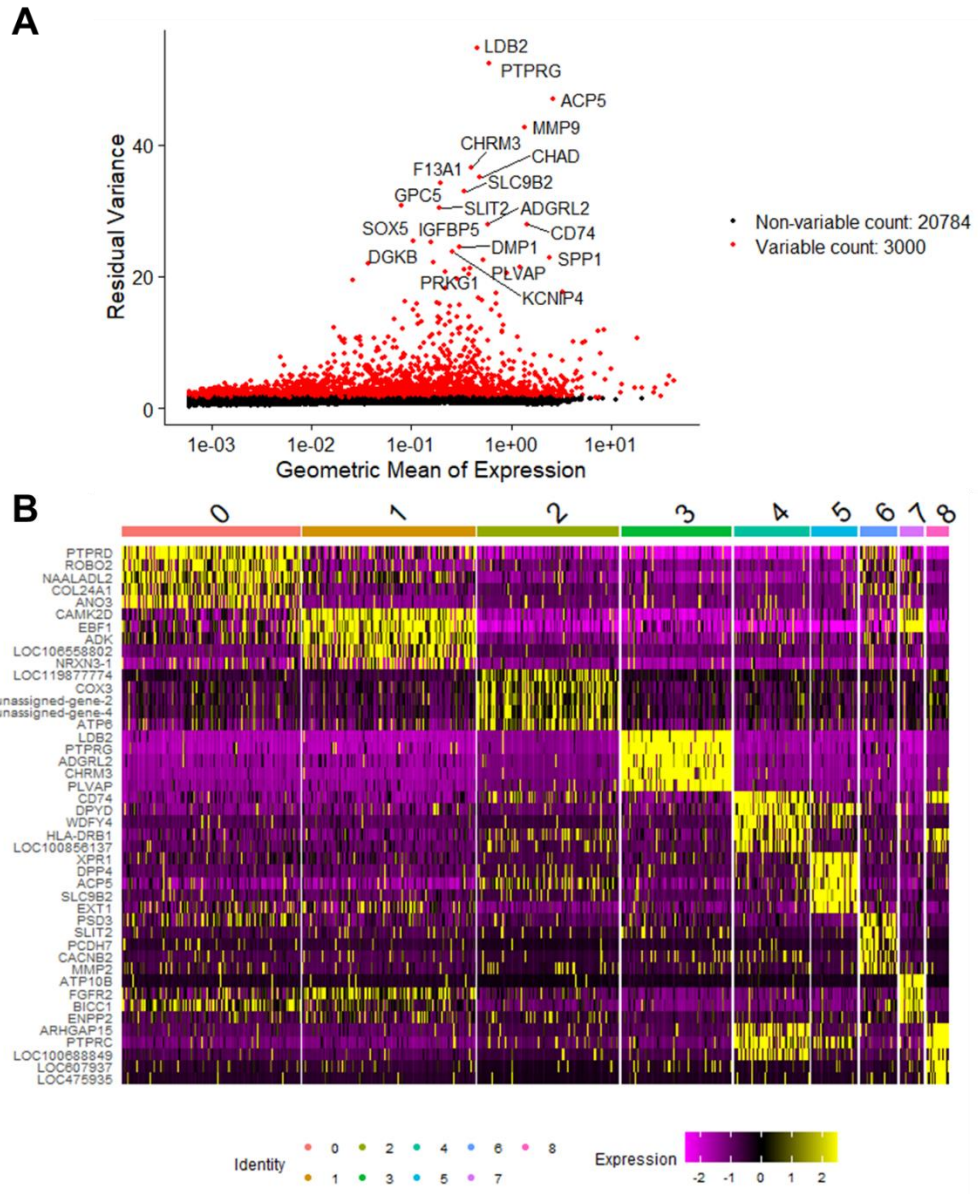


Figure 4.8. Volcano plot of the top 3000 variable genes in the dataset with the top 20 most highly variable genes labeled. Red dots indicate the differentially expressed genes (n=3000), black dots represent the non-variable genes (n=20784) (A). Heatmap of the top 5 differentially expressed genes in each cluster. Clusters are identified by color and number on the top *x*-axis, gene symbols are listed on the *y*-axis; yellow indicates upregulation and pink/purple indicates downregulation (B).

4.3.6 Gene Set Enrichment Analysis using Hallmark and Canonical Pathways

Using the markers identified for each cluster, gene set enrichment analysis was performed using Hallmark and Canonical pathways to identify variation among clusters (Fig 4.9A, B).

Interestingly, clusters 1 and 7 (tumorous osteoblasts) showed upregulation of G2M transition, E2F targets, MYC targets v2, and glycolysis, but cluster 0 (osteoblasts with unknown relation to tumor/normal) showed downregulation of these pathways. Cluster 0 also showed upregulation of the hypoxia response. Collectively, these results suggest that clusters 1 and 7 consist of actively dividing osteoblasts driving tumor expansion while cluster 0 may consist of necrotic and hypoxic osteoblasts.

Cluster 8 (memory CD4+ T cells) and cluster 4 (myeloid cells) share similar patterns of Hallmark and Canonical pathway expression, likely due to shared immunological functions (Fig 4.9A). Similarly, cluster 5 (osteoclasts) and cluster 4 (myeloid cells) displays similarities in Hallmark and Canonical pathways, which is explained by their shared macrophage functions.

Cluster 2 (fibroblasts) shared patterns of enriched Canonical pathways with cluster 6 (normal osteocytes), with the exception of Regulation of the Actin Cytoskeleton by Rho GTPases, G1 and S Phases, and Regulation Cascade of Cyclin Expression, which were all downregulated in osteocytes relative to fibroblasts (Fig 4.9B).

Figure 4.9. Gene set enrichment analysis among clusters using hallmark and canonical pathways.

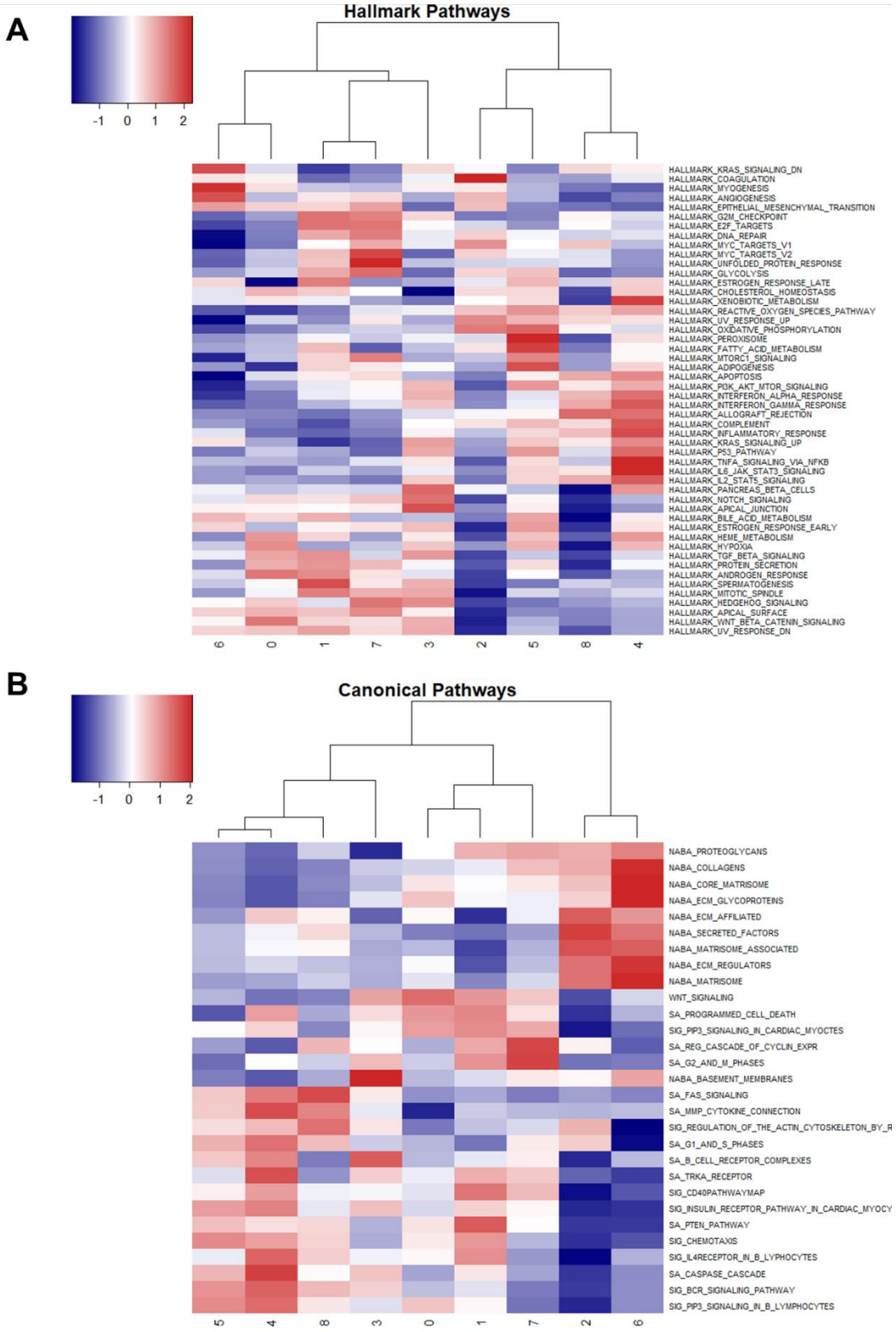

A
Hallmark Pathways
-1 0 1 2
HALLMARK_KRAS_SIGNALING_DN
HALLMARK_COAGULATION
HALLMARK_MYOGENESIS
HALLMARK_ANGIOGENESIS
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_G2M_CHECKPOINT
HALLMARK_E2F_TARGETS
HALLMARK_DNA_REPAIR
HALLMARK_MYC_TARGETS_V1
HALLMARK_MYC_TARGETS_V2
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_GLYCOLYSIS
HALLMARK_ESTROGEN_RESPONSE_LATE
HALLMARK_CHOLESTEROL_HOMEOSTASIS
HALLMARK_XENOBIOTIC_METABOLISM
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY
HALLMARK_UV_RESPONSE_UP
HALLMARK_OXIDATIVE_PHOSPHORYLATION
HALLMARK_PEROXISOME
HALLMARK_FATTY_ACID_METABOLISM
HALLMARK_MTORC1_SIGNALING
HALLMARK_ADIPOGENESIS
HALLMARK_APOPTOSIS
HALLMARK_PI3K_AKT_MTOR_SIGNALING
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_ALLOGRAFT_REJECTION
HALLMARK_COMPLEMENT
HALLMARK_INFLAMMATORY_RESPONSE
HALLMARK_KRAS_SIGNALING_UP
HALLMARK_P53_PATHWAY
HALLMARK_TNFA_SIGNALING_VIA_NFKB
HALLMARK_IL6_JAK_STAT3_SIGNALING
HALLMARK_IL2_STAT5_SIGNALING
HALLMARK_PANCREAS_BETA_CELLS
HALLMARK_NOTCH_SIGNALING
HALLMARK_APICAL_JUNCTION
HALLMARK_BILE_ACID_METABOLISM
HALLMARK_ESTROGEN_RESPONSE_EARLY
HALLMARK_HEME_METABOLISM
HALLMARK_HYPOXIA
HALLMARK_TGF_BETA_SIGNALING
HALLMARK_PROTEIN_SECRETION
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_SPERMATOGENESIS
HALLMARK_MITOTIC_SPINDLE
HALLMARK_HEDGEHOG_SIGNALING
HALLMARK_APICAL_SURFACE
HALLMARK_WNT_BETA_CATENIN_SIGNALING
HALLMARK_UV_RESPONSE_DN
6 0 1 7 3 2 5 8 4
B
Canonical Pathways
-1 0 1 2
NABA_PROTEOGLYCANS
NABA_COLLAGENS
NABA_CORE_MATRISOME
NABA_ECM_GLYCOPROTEINS
NABA_ECM_AFFILIATED
NABA_SECRETED_FACTORS
NABA_MATRISOME_ASSOCIATED
NABA_ECM_REGULATORS
NABA_MATRISOME
WNT_SIGNALING
SA_PROGRAMMED_CELL_DEATH
SIG_PIP3_SIGNALING_IN_CARDIAC_MYOCTES
SA_REG_CASCADE_OF_CYCLIN_EXPR
SA_G2_AND_M_PHASES
NABA_BASEMENT_MEMBRANES
SA_FAS_SIGNALING
SA_MMP_CYTOKINE_CONNECTION
SIG_REGULATION_OF_THE_ACTIN_CYTOSKELETON_BY_R
SA_G1_AND_S_PHASES
SA_B_CELL_RECEPTOR_COMPLEXES
SA_TRKA_RECEPTOR
SIG_CD40PATHWAYMAP
SIG_INSULIN_RECEPTOR_PATHWAY_IN_CARDIAC_MYOCY
SA_PTEN_PATHWAY
SIG_CHEMOTAXIS
SIG_IL4RECEPTOR_IN_B_LYPHOCYTES
SA_CASPASE_CASCADE
SIG_BCR_SIGNALING_PATHWAY
SIG_PIP3_SIGNALING_IN_B_LYMPHOCYTES
5 4 8 3 0 1 7 2 6

Figure 4.9. Heatmap of the Z-scores produced from gene set enrichment analysis using Hallmark Pathways (A) and Canonical Pathways (B). Red indicates upregulation, blue indicates downregulation. Clusters are numbered on the x-axis.

Collectively, the enriched Hallmark and Canonical pathway results validate the clustering and annotation results by confirming shared relationships and functions among common cell types and aid in the elucidation of osteoblast heterogeneity. Cluster 0 osteoblasts show distinct downregulation of cell cycle and upregulation of hypoxia pathways in comparison to osteoblasts in clusters 1 and 7, which suggests that the largest cell cluster identified is responding to the body's natural anti-tumor response. Furthermore, targeting this cluster of cells would likely not produce an effective response because these cells are not contributing to the active expansion of the tumor.

4.3.7 Enriched Pathway Analysis using GO Biological Processes

Using all genes in the canine genome database as a reference, a statistical overrepresentation test was performed on the significantly upregulated genes from each cluster to identify enriched pathways based on GO biological processes. Using a false discovery rate (FDR) cut-off of 0.05, the top 5 pathways based on fold enrichment for each cluster are depicted in Figure 4.10.

Enriched GO Biological Processes in cluster 0 osteoblasts included several pathways involved in regulation of cell adhesion. Cluster 1 osteoblasts were enriched for regulation of PI3K signaling, skeletal system development, and transmembrane receptor protein tyrosine kinase signaling. Upregulated pathways in cluster 2 (fibroblasts) were related to increased

cellular activity and protein production, including ribosomal assembly, mitochondrial electron transport, and translation. Cluster 3 (normal endothelial cells) was enriched for negative regulation of Rho-dependent protein serine/threonine kinase activity, regulation of macrophage colony-stimulating factor production, and cell migration involved in endocardial cushion formation (a specialized region of mesenchymal cells that give rise to heart structures). The upregulated GO Biological processes in cluster 4 (myeloid cells) included membrane raft localization/distribution, synapse pruning, negative regulation of granulocyte differentiation, and cell junction disassembly. Enriched pathways for osteoclasts in cluster 5 included macrophage fusion, dendritic cell homeostasis, positive regulation of CD8+ T cells, and glucuronoside metabolic/catabolic processes. Cluster 6 (normal osteocytes) upregulated processes included regulation of negative chemotaxis and cell-cell interactions and migration. Cluster 7 (tumor osteoblasts) was enriched for pathways related to anatomical structure and system/organism development. Enriched pathways for cluster 8 (memory CD4+ T cells) were related to antigen processing and presentation via MHC class I and positive regulation of T cell-mediated cytotoxicity (Fig 4.10).

Figure 4.10. Enriched GO biological processes among clusters.

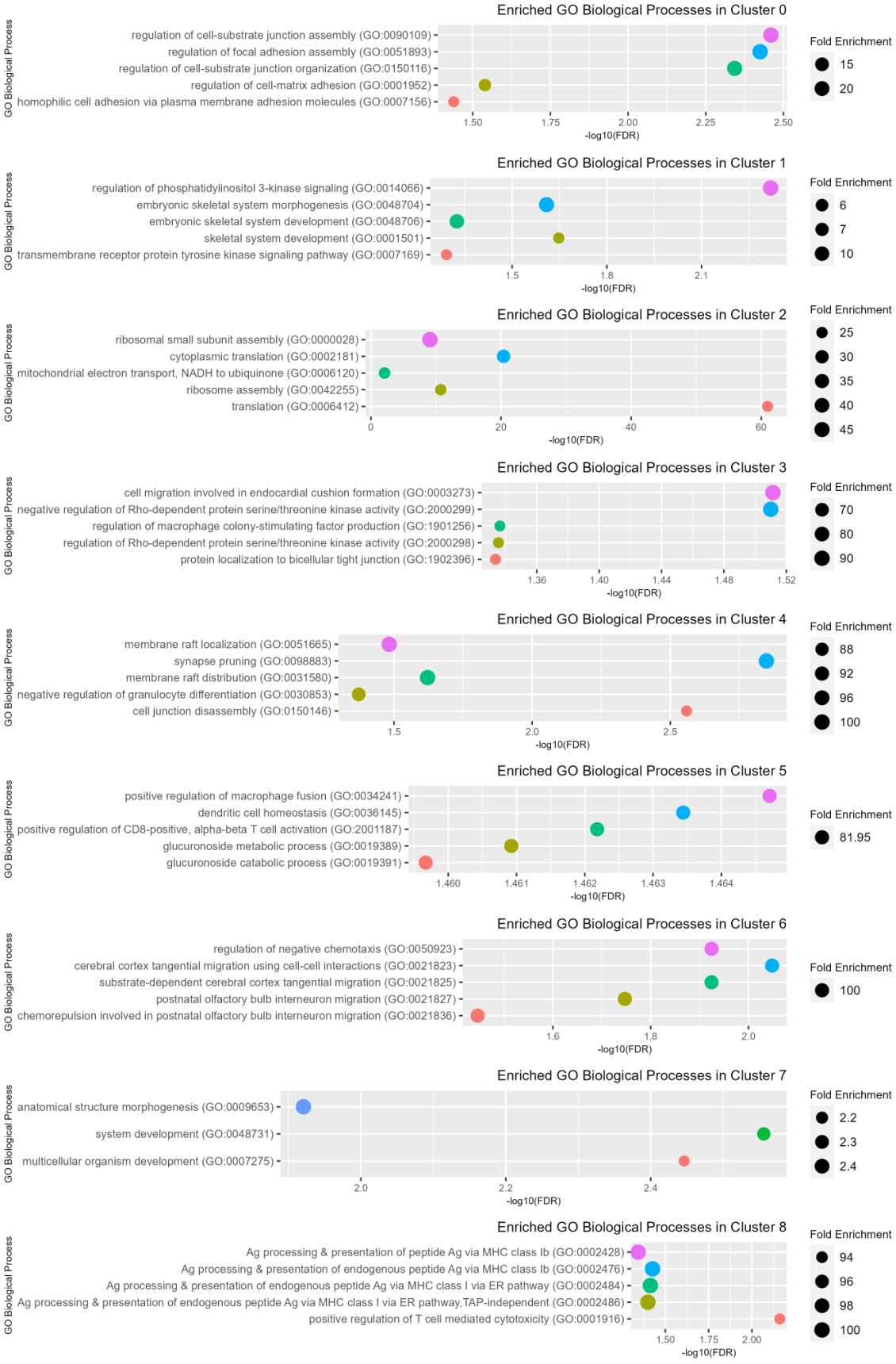

Enriched GO Biological Processes in Cluster 0
GO Biological Process
regulation of cell-substrate junction assembly (GO:0090109)
regulation of focal adhesion assembly (GO:0051893)
regulation of cell-substrate junction organization (GO:0150116)
regulation of cell-matrix adhesion (GO:0001952)
homophilic cell adhesion via plasma membrane adhesion molecules (GO:0007156)
1.50
1.75
2.00
2.25
2.50
-log10(FDR)
Fold Enrichment
15
20
Enriched GO Biological Processes in Cluster 1
GO Biological Process
regulation of phosphatidylinositol 3-kinase signaling (GO:0014066)
embryonic skeletal system morphogenesis (GO:0048704)
embryonic skeletal system development (GO:0048706)
skeletal system development (GO:0001501)
transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169)
1.5
1.8
2.1
-log10(FDR)
Fold Enrichment
6
7
10
Enriched GO Biological Processes in Cluster 2
GO Biological Process
ribosomal small subunit assembly (GO:0000028)
cytoplasmic translation (GO:0002181)
mitochondrial electron transport, NADH to ubiquinone (GO:0006120)
ribosome assembly (GO:0042255)
translation (GO:0006412)
0
20
40
60
-log10(FDR)
Fold Enrichment
25
30
35
40
45
Enriched GO Biological Processes in Cluster 3
GO Biological Process
cell migration involved in endocardial cushion formation (GO:0003273)
negative regulation of Rho-dependent protein serine/threonine kinase activity (GO:2000299)
regulation of macrophage colony-stimulating factor production (GO:1901256)
regulation of Rho-dependent protein serine/threonine kinase activity (GO:2000298)
protein localization to bicellular tight junction (GO:1902396)
1.36
1.40
1.44
1.48
1.52
-log10(FDR)
Fold Enrichment
70
80
90
Enriched GO Biological Processes in Cluster 4
GO Biological Process
membrane raft localization (GO:0051665)
synapse pruning (GO:0098883)
membrane raft distribution (GO:0031580)
negative regulation of granulocyte differentiation (GO:0030853)
cell junction disassembly (GO:0150146)
1.5
2.0
2.5
-log10(FDR)
Fold Enrichment
88
92
96
100
Enriched GO Biological Processes in Cluster 5
GO Biological Process
positive regulation of macrophage fusion (GO:0034241)
dendritic cell homeostasis (GO:0036145)
positive regulation of CD8-positive, alpha-beta T cell activation (GO:2001187)
glucuronoside metabolic process (GO:0019389)
glucuronoside catabolic process (GO:0019391)
1.460
1.461
1.462
1.463
1.464
-log10(FDR)
Fold Enrichment
81.95
Enriched GO Biological Processes in Cluster 6
GO Biological Process
regulation of negative chemotaxis (GO:0050923)
cerebral cortex tangential migration using cell-cell interactions (GO:0021823)
substrate-dependent cerebral cortex tangential migration (GO:0021825)
postnatal olfactory bulb interneuron migration (GO:0021827)
chemorepulsion involved in postnatal olfactory bulb interneuron migration (GO:0021836)
1.6
1.8
2.0
-log10(FDR)
Fold Enrichment
100
Enriched GO Biological Processes in Cluster 7
GO Biological Process
anatomical structure morphogenesis (GO:0009653)
system development (GO:0048731)
multicellular organism development (GO:0007275)
2.0
2.2
2.4
-log10(FDR)
Fold Enrichment
2.2
2.3
2.4
Enriched GO Biological Processes in Cluster 8
GO Biological Process
Ag processing & presentation of peptide Ag via MHC class Ib (GO:0002428)
Ag processing & presentation of endogenous peptide Ag via MHC class Ib (GO:0002476)
Ag processing & presentation of endogenous peptide Ag via MHC class I via ER pathway (GO:0002484)
Ag processing & presentation of endogenous peptide Ag via MHC class I via ER pathway,TAP-independent (GO:0002486)
positive regulation of T cell mediated cytotoxicity (GO:0001916)
1.50
1.75
2.00
-log10(FDR)
Fold Enrichment
94
96
98
100

Figure 4.10. The top 5 enriched GO Biological Processes and their associated fold enrichment and false discovery rate (FDR) for each cluster. Dot size correlates to the corresponding fold enrichment.

4.3.8 Comparison to Bulk Transcriptomic Sequencing Including the Same Tumor

In addition to characterizing the single-nuclei sequencing results of this tumor, we also sought to compare these results to bulk RNA sequencing of the same patient's tumor. Our lab has previously published bulk RNA sequencing results of the same patient's tumor along with six additional canine OSA tumors and patient-matched normal bone. The results, published in Nance et al, 2022, provide individual log2 fold-change values for each dog, in addition to bulk differential gene expression analysis. The patient belonging to the current study corresponds to patient C in the aforementioned manuscript[139].

The top upregulated genes in OSA tumor compared to normal bone based on bulk RNA sequencing included *GTSE1, HELLS, SPAG5, RAD54L, IQGAP3, CIT, HOXC10, TOP2A*, and *MKI67* (Fig 4.11A). The top upregulated genes in this patient based on bulk RNA sequencing of tumor and patient-matched normal bone included *TFPI2, DDX60, OAS1, CD5L, TERT, OAS2, RFGRIP1L, OAS3*, and *ANLN* (Fig 4.11B). Based on these results, the marker genes based on individual-level analysis capture more of the tumor's heterogeneity than marker genes derived from bulk RNA sequencing. These results validate our previously published approach to individual-level analysis using bulk RNA sequencing of tumor and patient-matched normal.

Figure 4.11. Expression of the top upregulated genes from the bulk OSA tumor/normal bone RNAseq results.

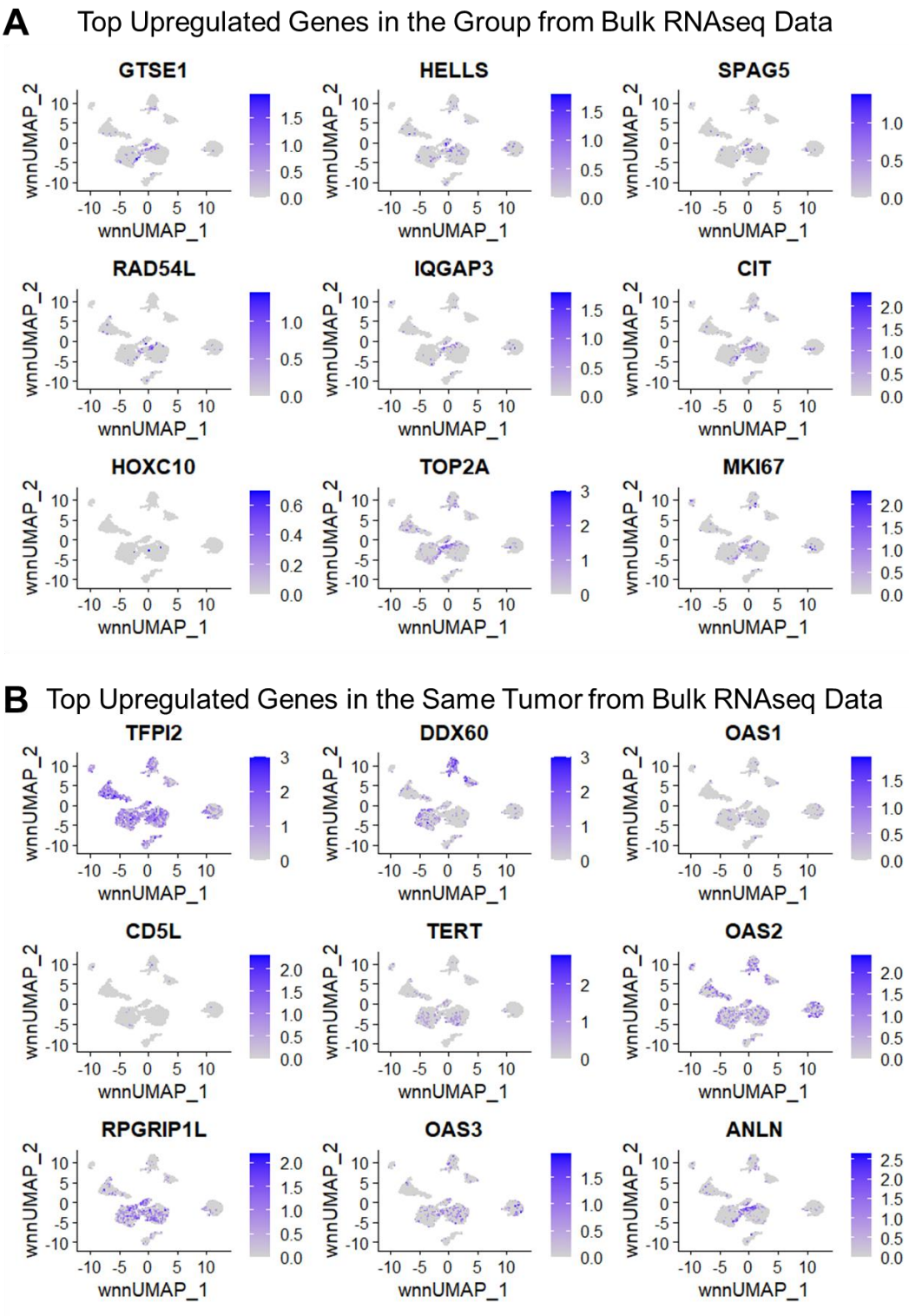


Figure 4.11. Expression of marker genes derived from the bulk OSA tumor/normal RNAseq dataset for the top upregulated genes in the group (A). Marker gene expression for the top

upregulated genes in the same patient from the bulk OSA tumor/normal RNAseq dataset (B). Darker shades of purple indicate upregulated expression, gray indicates zero change.

4.4 Discussion

To our knowledge, this is the first study to utilize single-nuclei multiome (ATAC + GEX) sequencing to characterize the molecular landscape of a treatment-naïve primary canine OSA tumor. Additionally, the results were compared to bulk RNA sequencing, including the same tumor and patient-matched normal bone to further evaluate intra-tumoral heterogeneity.

Since osteosarcoma is a difficult tumor to homogenize, a modification was made to the 10x Genomics nuclei isolation protocol which resulted in high-quality nuclei suitable for sequencing. This protocol variation included the use of a bladed homogenizer in 0.5 strength Lysis buffer (included with the 10x Genomics kit). Upon observation with fluorescent dye, the nuclear membranes appeared mostly intact with minimal blebbing. Despite the recommendations from 10x Genomics to store the samples long term in liquid nitrogen, storing them at -85°C did not appear to affect the quality of our results. While approximately 10,000 nuclei were subjected to library preparation, only 5969 nuclei were actually sequenced. This is expected due to the microfluidic partitioning process which relies on dilution to prevent multiple nuclei in one droplet[145].

Unsupervised clustering and weighted nearest neighbor analysis identified 9 cell clusters in primary canine OSA. As expected, the most abundant cell type present was osteoblasts, though these cells formed three distinct subclusters (clusters 0, 1 and 7). The second most abundant cell cluster contained fibroblasts (cluster 2) followed by endothelial cells (cluster 3).

The tumor immune microenvironment included myeloid cells (cluster 4), osteoclasts (cluster 5), and memory CD4+ T cells (cluster 8).

Cluster annotation using upregulated genes from bulk RNA sequencing of primary canine OSA and patient-matched normal bone resulted in the identification of cluster 1 (osteoblasts), cluster 7 (osteoblasts), and cluster 8 (memory CD4+ T cells) as OSA tumor while cluster 3 (endothelial) and cluster 6 (osteocytes) were associated with normal bone. Clusters 0 (osteoblasts), 2 (fibroblasts), 4 (myeloid), and 5 (osteoclasts) had unknown relation to the tumor/normal bone annotation. The inability of the annotation package to distinguish osteoclasts (cluster 5) and myeloid cells (cluster 4) as normal bone derivatives could be due to the relatively low proportion of these cells in comparison to osteocytes present in normal bone matrix. Markers for endothelial cells and myeloid cells are predicted to be expressed at low levels in this RNA dataset due to the bone processing technique to remove bone marrow, which contains the majority of myeloid cells and osteoclasts. It is also plausible that normal, non-transformed cells have altered gene expression patterns in response to the tumor microenvironment and signals from surrounding cells. The GEX profile of a normal cell in a normal environment would likely differ from that of a normal, non-neoplastic cell within the TME.

Secondly, while cluster 0 did not reflect the pattern of previously bulk sequenced OSA tumor, it is unlikely that cluster 0, which contains the most cells, contains normal osteoblasts due to the small percentage of osteoblasts in normal bone. Osteoblasts also undergo age-related decline, and these bone samples were obtained from geriatric dogs[146]. Furthermore, CNV analysis showed this cluster contains many large-scale chromosomal rearrangements. Hallmark and Canonical pathway analysis suggest these cells are hypoxic, necrotic, and not actively dividing. Therefore, we predict this cluster consists of dying osteoblastic tumor cells in response

to the body's apoptotic immune signals and intrinsic anti-tumor response. Alternatively, or perhaps in conjunction, the tumor could simply be outgrowing its blood and nutrient supply. On the other hand, the activities of clusters 1 and 7 indicate these cells are highly active and thus more likely to be contributing to tumor growth and expansion.

Osteosarcoma has been predicted to be a poorly immunogenic tumor and therefore immunotherapies have been ineffective at managing OSA[147]. Based on upregulated pathway analysis, immune cells in the TME (cluster 8) are actively participating in tumor antigen presentation via MHC class Ib and stimulating T-cell mediated cytotoxicity (Fig 4.9). The consequence of these actions is perhaps the dying osteoblasts in cluster 0. However, more studies are needed to understand how and why some cells escape the immune response.

Despite their prevalence in the tumor microenvironment, the inability to distinguish fibroblasts as tumor using the bulk data could be due to the low sample size (n=7) of the bulk RNA sequencing data. Fibroblasts are also present in a smaller proportion in comparison to osteoblastic OSA cells. This reflects a limitation in differential gene expression analysis of bulk RNA sequencing data, where differences are obtained from averages across the entire population. The top marker gene for cluster 2 fibroblasts was *TMSB10*, which encodes a protein involved in cytoskeleton organization and cell migration. According to The Human Protein Atlas (https://www.proteinatlas.org/ENSG00000034510-TMSB10/single+cell+type), *TMSB10* is typically expressed at low frequency in fibroblasts. Many of the other top marker genes for cluster 2 were involved in ribosomal assembly and function, including *RPS14*, *RPLP1*, *RPL36*, *RPS11*, and *RPS28*. These results suggest this cluster consists of highly active cells that are generating and secreting a large number of products that modulate the surrounding tumor microenvironment, consistent with the activities of cancer-associated fibroblasts (CAFs). As a

major component of the tumor microenvironment, CAFs secrete a variety of factors that play a key role in tumorigenesis, and their activation is predicted to be controlled via epigenetic regulation[148]. Factors secreted by fibroblasts have been shown to modulate osteoblasts and their extracellular matrix remodeling functions[149]. This dynamic relationship is reflected in the clustering of the ATAC data, where the osteoblast and fibroblast clusters are less defined, closely interconnected, and show significant overlap (Fig 4.3B). However, additional analyses are needed to elucidate this relationship.

Compared to single-cell sequencing, bulk RNA sequencing of the same sample captures only about 28% of the tumor's heterogeneity, highlighting one of the major limitations of bulk sequencing. On the other hand, single-cell/nuclei sequencing is quite expensive and remains a limiting factor in applying this technology. Nonetheless, single-nuclei multiome sequencing provides an unparalleled view into the TME and intra-tumoral landscape. This approach is critical to improving targeted therapies and patient outcomes. Although this study consisted of a sample size of one, the computational framework can be applied to additional tumors and as novel computational tools are developed, this data can be conveniently reanalyzed.

In summary, we have successfully applied single-nuclei multiome sequencing to characterize the intra-tumoral heterogeneity and immune landscape of a treatment-naïve primary canine osteosarcoma.

4.5 Acknowledgements

This work was made possible in part by a grant of high-performance computing resources and technical support from the Alabama Supercomputer Authority.

4.6 Data and Code Availability

All coding scripts used in this manuscript can be found at https://github.com/rln0005/OSA_snMultiomeSeq. The data will be deposited into the NCBI GEO repository.

4.7 Supplemental Data

Supplemental Table 4.2. All markers used for annotation with ScType.

Cell Type	Markers
ScType: Bone/OSA/Immune	
Antigen-presenting cell	CD83, CCR7, CD83
Articular chondrocyte	COL22A1, COL22A1
B cell	CD19, Siglech, SIGLECH, CD19
Cancer stem cell	PROM1
Cartilage cell	COL2A1
Cartilage progenitor cell	CENPF, UBE2C, PCLAF, CDC20, CDKN3, BIRC5, CKAP2
Chondrocyte	SCRG1, ACAN, SOX6, SOX5, SOX9, COL2A1
Dendritic cell	SIGLECH, Siglech
Distal proliferative progenitor cell	H1F3, JAG, MSX1, Hoxd13, TBX3, TBX2
Early mesenchymal cell	LY6A, CD34
Embryonic stem cell	LHX3, LSL1, ISL1, ISL2, ABCAM, Mnx1, FOXP1
Endothelial cell	CDH5, PECAM1, EGFL7, CD93, VWF, MYL1, ENG, EMCN
Epithelial cell	KRT42, KRT14, PERP
Fibroblast	LUM, DCN, COL1A1, THY1, FAP, PDPN, VIM, PRRX1, MSX1

Hemtopoietic stem cell	CD34, HOXB5
Lymphocyte	SET
Mesenchymal progenitor cell	PDGFRB, CSPG4, ITGB1
Mesenchymal stromal cell	CDH2, PDGFRB, THY1, ENG, ITGAV, NES, NT5E, SPP1, HEY1, ITGB1, LY6A, VIM
Natural killer cell	IL7R, NKG7, CD3D, KLRD1
Osteoblast	MKI67, ACAN, PCNA, CDH11, COL1A1, TOP2A, RUNX2, COL2A1, SOX9
Osteoclast	ACP5, ATP6V0D2, DCSTAMP, MMP9, ATP6V1B1, CTSK, OCSTAMP, PPARGC1B
Progenitor cell	CD34, LGR5
T cell	CD69, CD3D, CD3G, TRRDC, ICOS, CD3G, IRF8
Myeloid cell	CD14, FCGR3A, CD74
Memory CD4+ T cells	CD4,CD2,CD3D,CD3E,CD3G,CD3Z,CD25,CD45RA,CD62L,CD27,CD127,FOXP3,CCR7,CD45,CCR6,CD11b,CD30,CD45RO,CD6,CTLA4,IL2RA,GZMB,SELL,CCR7,S100A4,TRAC,LTB,CD52,TRBC2,SHISA5,LCK,THY1,DAPL1
Myeloid dendritic cells	ITGAX,CD83,CD1C,NRP1,CLEC4C,CD86,IL3RA,CD80,CD1A,ITGAX,CD40,HLA-DQA1,CD11c,HLA-DR,HLA-DPB1,HLA-DPA1,CLEC10A,CST3,GPR31,ODF3L1,PRB2,CD207,ARSE,CLEC141,MRC,EBLN1,CRIP3
Osteocyte	BGLAP, SPP1, CD86, CD14, IBSP
ScType: Bulk OSA Tumor/Normal Bone	
OSA tumor cells	HOXC10,GTSE1,SPAG5,TOP2A,IQGAP3,HELLS,MKI67,CLSPN,RAD54L,CDCA3,DPYS,RECQL4,CIT,SHCBP1,SLC6A2,CDC45,DEPDC1,HOXC6,BUB1,UBE2C,CCNB3,CDC6,DNPH1,NEK2,KIF18B,FOXM1,MMP12,TPX2,ORC1,PRC1,TERT,DIAPH3,MYBL2,KIF20A,KIF4A,UCHL1,CEP72,TACC3,CCR7,HTRA3,CENPT,ORC6,CCNB1,PKMYT1,CENPF,KIFC1,WDR62,TROAP,AURKA,COL11A1,DLGAP5,KIF23,SLC15A1,RASAL1,NCAPG,PLAUR,E2F8,KIF15,APOBEC3Z1,RRM2,CD5L,CDC20,CDCA2,SGO1,TONSL,KNL1,CCNB2,SERPINB2,POLE,FANCM,CDCA8,HOXC4,RAD51AP1,RACGAP1,TK1,UBE2T,PLK1,CCL2,PITX1,KIF14,BIRC5,ERCC6L,MCM10,OIP5,PAX1,SPON1,VEGFD,ANLN,HJURP,DOK5,NUF2,HROB,KIF22,ZWINT,KIF11,SAA1,PBK,LOC100855995,PIMREG,CDK1,DUSP9,BUB1B,CCNA2,CDC25C,CDCA5,MIS18A,NDC80,E2F2,TFPI2,CKAP2L,CCNO,PANX2,IL6,KPNA2,ARSJ,SPEF1,DNA2,ECT2,SPC24,MASTL,IGF2BP3,EXO1,SFRP2,LOC608024,ASPM,TNFRSF4,CENPU,PTTG1,INPP5J,LOC612762,ACOT7,ARHGEF39,NLRP12,ESCO2,LOC100688904,E2F7,ESPL1,LOC111093651,CENPK,KNSTRN

Normal bone cells	FMO2,CHL1,PLIN1,CIDEC,ACKR4,NALCN,ESM1,IGFBP2,COL20A1,LYPD1,AQP7,KRT23,PLIN4,ANGPTL7,KLF15,RBP7,CIDEA,GALNT15,MYOC,CAMK1G,LOC100686507,GPR39,AGT,COCH,PLP1,CALCR,MEPE,APOD,TRARG1,SBSN,SOST,CNKSR2,EPHA7,CLDN11,PLK5,CES2,COL4A4,SCN7A,GPD1,ESR1,NRG2,SCD5,CDH19,LEP,COL4A3,LGI1,CNTN1,ASPA,NPY,NGFR,AATK,RNF207,FABP4,ZBTB16,RHPN2,PLD5,CLGN,PROM2,GPR1,PCK1,TOX3,SLITRK3,PPP1R1B,CALB2,VIT,BMP3,DGKB,NRXN1,ALDH1A3,SEMA3E,OLFM2,EPHX2,HIF3A,LIPE,GPC5,NPY1R,MYRIP,FN3K,GRIA4,LOC100686484,PNPLA2,EPHA1,NAP1L2,KIAA1755,S100B,GFRA4,GPM6A,LOC486151,WSCD2,SYN2,ECRG4,LOC111091389,CHAD,RARRES1,EPCAM,TMEM52,GPIHBP1,ALK,FMO3,ADAMTSL1,LOC607776,LOC475605,TIMP4,LOC491723,NDRG2,CD36,CCDC3,MYOZ2,AIF1L,AMIGO2,LYZF2,TYR,LETM2,MPZ,LOC100686073,DMP1,GRIP1,MYORG,SHISA9,GDF10,CADM3,GOLT1A,PATJ,CCN5,ACSBG1,WDR88,AK5,SV2B,LEPR,BGLAP,CLEC3A,NMNAT3,ACSM5,CP,DPP10,LOC480425,KCNMB2,SPHKAP,BTNL9,SGK2,CCDC85A,RGN,PTGDS,HLF,NPY2R,NPR3,RDH16,LOC607314,FRZB,LIMCH1,MLXIPL,EVX2,PLLP,GYG2,TXNIP,ITIH1,PTX3,SNCA,ACSM3,DNAH3,ARMC4,LOC608319,EFHD1,WASF3,RERGL,CYTL1,PCDH20,BSPRY,LOC607487,ABCC6,FGL1,SPP1,KCNU1,LOC480825,LOC607460,ARMC3,FAM163A,AGBL3,EPHA5,MT1E,APOE,WIPF3,LOC477699,SH3RF2,GNA14,SCUBE1,PER2,UCN3,CNTFR,SPOCK3,LOC111089969,NKX6-1,LHX9,RSPO1,WIF1,LOC100856005,CLSTN2,RETN,PLPPR3,NPFFR2,NTF3,IFITM10,ASS1,TMC2,TRIM67,ITGBL1,PIP,FGF12,FMN2,PFKFB1,ADRA1D,LOC486353,MAPT,WNT6,HPSE2,PRKAR1B

CHAPTER 5

Conclusions and Future Directions

Osteosarcoma (OSA) is a complex tumor derived from bone that typically occurs spontaneously in young adults and large breed dogs. In both species, significant intra- and inter-tumoral heterogeneity make treatment difficult. Precision medicine seeks to optimize therapeutic efficiency by classifying tumors based on genetic or molecular profiling. Canine and human OSA share many similarities in terms of tumorigenesis, molecular characteristics, risk factors, treatment, and clinical and histological presentation. Genetic structure of the canine breed system has enabled the elucidation of complex diseases. Thus, dogs serve as a powerful translational model of naturally occurring cancer, particularly osteosarcoma. In dogs, OSA occurs more frequently than humans and standard treatment involves amputation and chemotherapy. Amputation enables researchers to collect tumor as well as corresponding patient-matched normal bone to identify differentially expressed genes which can be useful for guiding and designing therapies.

An often over-looked but critical component of differential expression analysis is the source of normal comparator tissue. With OSA arising from bone, the initial goals of this project were to develop and optimize a method to isolate RNA from normal canine phalanges, as described in chapter two. This process involved removing and washing the exterior periosteum and interior bone marrow cavity contents to generate RNA specific to the bone matrix. Due to the rigid, crystallized, and hypocellular nature of bone, several modifications were made to enable tissue homogenization and RNA extraction while minimizing degradation. These adjustments included use of a steel mortar and pestle on dry ice along with beaded homogenization. After tissue dissociation, the ‘TRIspin method’ was used to extract RNA, which

combined acid guanidinium thiocyanate-phenol-chloroform extraction with the use a column-based kit. This technique resulted in the isolation of RNA from normal canine bone in sufficient quantity and quality for sequencing.

In chapter three, I described differential gene expression analysis using RNA extracted from seven primary canine OSA tumors and, using the method developed in chapter two, from patient-matched normal bone samples. Using a traditional bulk approach in addition to a novel individual approach, I showed the limitations of bulk sequencing in capturing inter-tumoral heterogeneity between patients. While increasing the sample size would surely identify additional variation, bulk RNA sequencing fails to identify signals from rare yet important cell types, such as those within the tumor microenvironment (TME). Many cancer therapeutics are now targeting cells within the TME, such as immune-associated cells, to combat tumors and metastases. Furthermore, the significant intra-tumoral heterogeneity inherent in OSA suggests an important role in the TME in modulating tumor progression.

Single-cell sequencing provides extraordinary resolution of intra-tumoral heterogeneity by capturing the genome and/or transcriptome of individual cells. However, high cell viability (>60%) after tissue dissociation is critical to obtaining valuable and informative sequencing data. Numerous attempts were made to isolate viable single cells from primary canine OSA using a variety of mechanical and enzymatic dissociation techniques. However, these approaches consistently yielded poor cell viability (<30%), which suggests the loss of potentially important cell subtypes.

To overcome viability obstacles associated with single-cell sequencing, individual nuclei can be isolated rather than cells, though nuclei quality is still an important factor to consider. In the chapter 4 methods, I describe a modified approach to isolate nuclei from a primary canine

OSA involving brief bladed and chilled homogenization in 0.5 strength lysis buffer. Using high-power fluorescent microscopy, the nuclei were determined to be in sufficient quantity and quality for multiome (ATAC + GEX) library preparation and sequencing.

Single-nuclei multiome sequencing of a treatment-naïve primary canine OSA revealed nine cell clusters and their proportions based on unsupervised clustering of the ATAC and GEX data. Among the cell types identified included osteoblasts, fibroblasts, endothelial cells, osteoclasts, osteocytes, and immune-associated cells. Interestingly, approximately half of the osteoblastic cell clusters showed hypoxic and necrotic trends, while the other half showed high replication and cell cycle activities. Cancer-associated fibroblasts and macrophages composed about 17% and 9% of the tumor, respectively. The smallest cluster was identified as memory CD4+ T cells (3%), and while expression of CD4 was limited, these cells expressed several genes involved in antigen recognition by T cells.

Comparison with single-nuclei multiome sequencing of the same primary tumor shows limitations of bulk sequencing in capturing intra-tumoral heterogeneity, particularly from a precision medicine approach where the genome and/or transcriptome informs therapy. However, in the absence of single-cell sequencing data, individual-level differential expression data using bulk tumor and patient-matched normal are informative of the intra-tumoral heterogeneity.

Future directions of this project involve the identification of enhanced motifs using the ATACseq data to elucidate the shared regulatory regions and relationship between osteoblasts and cancer-associated fibroblasts in canine OSA. Further elucidation of the role that immune-associated cells, including the clusters defined as myeloid and memory CD4+ T cells, play in tumorigenesis is also necessary.

Additional tumors can be collected and sequenced using this pipeline to further define intra- and inter-tumoral heterogeneity in primary canine OSA. Moreover, sequencing and analysis of canine OSA metastases would be critically informative of disease progression and identification of druggable targets. Since amputation removes the entire primary tumor, subsequent treatment is targeted at the metastases. While metastases are derived from the original tumor, their migration process and altered microenvironment suggest there may be clonal differences in the tumor's cell population. Therefore, I believe obtaining canine OSA metastases for single-nuclei sequencing is the most critical component for identifying precision-based druggable targets moving forward.

To summarize, I have characterized the molecular landscape of primary canine osteosarcoma using single-nuclei multiome sequencing and compared the results to bulk RNA sequencing of the same tumor and patient-matched normal bone. These results provide an unprecedented perspective of the intra-tumoral heterogeneity in primary canine osteosarcoma and can be useful for designing precision medicine therapeutics.

REFERENCES

1. Hanahan D, Weinberg RA. The Hallmarks of Cancer. *Cell*. 2000;100(1):57-70. doi:10.1016/S0092-8674(00)81683-9

2. Morse BL, Kim RB. Is personalized medicine a dream or a reality? *Crit Rev Clin Lab Sci*. 2015;52(1):1-11. doi:10.3109/10408363.2014.950407

3. LeBlanc VG, Marra MA. Next-Generation Sequencing Approaches in Cancer: Where Have They Brought Us and Where Will They Take Us? *Cancers*. 2015;7(3):1925-1958. doi:10.3390/cancers7030869

4. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010;11(10):685-696. doi:10.1038/nrg2841

5. Rao SR, Somarelli JA, Altunel E, et al. From the Clinic to the Bench and Back Again in One Dog Year: How a Cross-Species Pipeline to Identify New Treatments for Sarcoma Illuminates the Path Forward in Precision Medicine. *Front Oncol*. 2020;10. doi:10.3389/fonc.2020.00117

6. Misra SC, Bisui S. Feasibility of Large Scale Implementation of Personalized Medicine in the Current Scenario. *Int J E-Health Med Commun IJEHMC*. 2016;7(2):30-49. doi:10.4018/IJEHMC.2016040103

7. Brunak S, Collin CB, Cathaoir KEÓ, et al. Towards standardization guidelines for in silico approaches in personalized medicine. *J Integr Bioinforma*. 2020;17(2-3). doi:10.1515/jib-2020-0006

8. Król M, Motyl T. Exploiting cancer genomics in pet animals to gain advantage for personalized medicine decisions. *J Appl Genet*. 2014;55(3):337-341. doi:10.1007/s13353-014-0206-0

9. Vizirianakis IS, Fatouros DG. Personalized nanomedicine: paving the way to the practical clinical utility of genomics and nanotechnology advancements. *Adv Drug Deliv Rev*. 2012;64(13):1359-1362. doi:10.1016/j.addr.2012.09.034

10. Kent J. FDA Approvals Advance Precision Medicine, Genomics Treatments. *Health IT Anal*. Published online February 25, 2020. https://healthitanalytics.com/news/fda-approvals-advance-precision-medicine-genomics-treatments

11. Ray T. FDA Approves Record Number of Precision Oncology Drugs in H1 2020. *Precis Oncol News*. Published online July 21, 2020. https://www.precisiononcologynews.com/cancer/fda-approves-record-number-precision-oncology-drugs-h1-2020#.YQnWcI43m71

12. Soverini S, Mancini M, Bavaro L, Cavo M, Martinelli G. Chronic myeloid leukemia: the paradigm of targeting oncogenic tyrosine kinase signaling and counteracting resistance for successful cancer therapy. *Mol Cancer*. 2018;17(1):49. doi:10.1186/s12943-018-0780-6

13. André F, Ciruelos E, Rubovszky G, et al. Alpelisib for PIK3CA-Mutated, Hormone Receptor–Positive Advanced Breast Cancer. *N Engl J Med*. 2019;380(20):1929-1940. doi:10.1056/NEJMoa1813904

14. Hyman DM, Piha-Paul SA, Won H, et al. HER kinase inhibition in patients with HER2- and HER3-mutant cancers. *Nature*. 2018;554(7691):189-194. doi:10.1038/nature25475

15. Smyth LM, Piha-Paul SA, Won HH, et al. Efficacy and Determinants of Response to HER Kinase Inhibition in HER2-Mutant Metastatic Breast Cancer. *Cancer Discov*. 2020;10(2):198-213. doi:10.1158/2159-8290.CD-19-0966

16. Gambardella V, Tarazona N, Cejalvo JM, et al. Personalized Medicine: Recent Progress in Cancer Therapy. *Cancers*. 2020;12(4):1009. doi:10.3390/cancers12041009

17. Jong M de, Maina T. Of Mice and Humans: Are They the Same?—Implications in Cancer Translational Research. *J Nucl Med*. 2010;51(4):501-504. doi:10.2967/jnumed.109.065706

18. Day CP, Merlino G, Van Dyke T. Preclinical Mouse Cancer Models: A Maze of Opportunities and Challenges. *Cell*. 2015;163(1):39-53. doi:10.1016/j.cell.2015.08.068

19. Mak IW, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. *Am J Transl Res*. 2014;6(2):114-118.

20. Ranieri G, Gadaleta CD, Patruno R, et al. A model of study for human cancer: Spontaneous occurring tumors in dogs. Biological features and translation for new anticancer therapies. *Crit Rev Oncol Hematol*. 2013;88(1):187-197. doi:10.1016/j.critrevonc.2013.03.005

21. Barutello G, Rolih V, Arigoni M, et al. Strengths and Weaknesses of Pre-Clinical Models for Human Melanoma Treatment: Dawn of Dogs' Revolution for Immunotherapy. *Int J Mol Sci*. 2018;19(3):799. doi:10.3390/ijms19030799

22. Riccardo F, Aurisicchio L, Impellizeri JA, Cavallo F. The importance of comparative oncology in translational medicine. *Cancer Immunol Immunother*. 2015;64(2):137-148. doi:10.1007/s00262-014-1645-5

23. Talmadge JE, Singh RK, Fidler IJ, Raz A. Murine Models to Evaluate Novel and Conventional Therapeutic Strategies for Cancer. *Am J Pathol*. 2007;170(3):793-804. doi:10.2353/ajpath.2007.060929

24. Park JS, Withers SS, Modiano JF, et al. Canine cancer immunotherapy studies: linking mouse and human. *J Immunother Cancer*. 2016;4(1):97. doi:10.1186/s40425-016-0200-7

25. Weyand CM, Goronzy JJ. Aging of the Immune System. Mechanisms and Therapeutic Targets. *Ann Am Thorac Soc*. 2016;13(Suppl 5):S422-S428. doi:10.1513/AnnalsATS.201602-095AW

26. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science*. 2012;336(6086):1268-1273. doi:10.1126/science.1223490

27. Ivanov II, Honda K. Intestinal commensal microbes as immune modulators. *Cell Host Microbe*. 2012;12(4):496-508. doi:10.1016/j.chom.2012.09.009

28. Sivan A, Corrales L, Hubert N, et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy. *Science*. 2015;350(6264):1084-1089. doi:10.1126/science.aac4255

29. Lawrence J, Cameron D, Argyle D. Species differences in tumour responses to cancer chemotherapy. *Philos Trans R Soc B Biol Sci*. 2015;370(1673):20140233. doi:10.1098/rstb.2014.0233

30. Gordon MY, Blackett NM. The Sensitivities of Human and Murine Hemopoietic Cells Exposed to Cytotoxic Drugs in an in Vivo Culture System. *Cancer Res*. 1976;36(8):2822-2826.

31. Sanoh S, Ohta S. Chimeric mice transplanted with human hepatocytes as a model for prediction of human drug metabolism and pharmacokinetics. *Biopharm Drug Dispos*. 2014;35(2):71-86. doi:10.1002/bdd.1864

32. Cheung C, Gonzalez FJ. Humanized Mouse Lines and Their Application for Prediction of Human Drug Metabolism and Toxicological Risk Assessment. *J Pharmacol Exp Ther*. 2008;327(2):288-299. doi:10.1124/jpet.108.141242

33. Tarone L, Barutello G, Iussich S, et al. Naturally occurring cancers in pet dogs as pre-clinical models for cancer immunotherapy. *Cancer Immunol Immunother*. 2019;68(11):1839-1853. doi:10.1007/s00262-019-02360-6

34. Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 2005;438(7069):803-819. doi:10.1038/nature04338

35. Briggs J, Paoloni M, Chen QR, Wen X, Khan J, Khanna C. A Compendium of Canine Normal Tissue Gene Expression. *PLoS ONE*. 2011;6(5). doi:10.1371/journal.pone.0017107

36. Napierala H, Uerpmann HP. A 'new' palaeolithic dog from central Europe. *Int J Osteoarchaeol*. 2012;22(2):127-137. doi:10.1002/oa.1182

37. Breen M, Thomas R. Karyotype and Chromosomal Organization. In: Ostrander EA, Giger U, Lindblad-Toh K, eds. *The Dog and Its Genome*. Cold Spring Harbor Laboratory Press.

38. Price TD, Gebauer AB. *Last Hunters - First Farmers : New Perspectives on the Prehistory Transition to Agriculture*. School of American Research Press; 1996.

39. Parker HG, Kim LV, Sutter NB, et al. Genetic Structure of the Purebred Domestic Dog. *Science*. 2004;304(5674):1160-1164. doi:10.1126/science.1097406

40. Ostrander E, Dreger D, Evans J. Canine Cancer Genomics: Lessons for Canine and Human Health. *Annu Rev Anim Biosci*. 2019;7(1):449-472.

41. vonHoldt BM, Pollinger JP, Lohmueller KE, et al. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature*. 2010;464(7290):898-902. doi:10.1038/nature08837

42. Wade CM, Karlsson EK, Mikkelsen TS, Zody MC, Lindblad-Toh K. The Dog Genome: Sequence, Evolution, and Haplotype Structure. In: *The Dog and Its Genome*. Cold Spring Harbor Laboratory Press; 2006.

43. Ostrander E, Urs G, Lindblad-Toh K, eds. *The Dog and Its Genome*. Cold Spring Harbor Laboratory Press; 2006.

44. Felsburg PJ. Overview of immune system development in the dog: comparison with humans. *Hum Exp Toxicol*. 2002;21(9-10):487-492. doi:10.1191/0960327102ht286oa

45. Greeley EH, Kealy RD, Ballam JM, Lawler DF, Segre M. The influence of age on the canine immune system. *Vet Immunol Immunopathol*. 1996;55(1):1-10. doi:10.1016/S0165-2427(96)05563-8

46. Sansoni P, Cossarizza A, Brianti V, et al. Lymphocyte Subsets and Natural Killer Cell Activity in Healthy Old People and Centenarians. *Blood*. 1993;82(9):2767-2773. doi:10.1182/blood.V82.9.2767.2767

47. Cobbold S, Metcalfe S. Monoclonal antibodies that define canine homologues of human CD antigens: summary of the First International Canine Leukocyte Antigen Workshop (CLAW). *Tissue Antigens*. 1994;43(3):137-154. doi:10.1111/j.1399-0039.1994.tb02315.x

48. Isotani M, Katsuma K, Tamura K, et al. Efficient generation of canine bone marrow-derived dendritic cells. *J Vet Med Sci*. 2006;68(8):809-814. doi:10.1292/jvms.68.809

49. Trepanier LA, Ray K, Winand NJ, Spielberg SP, Cribb AE. Cytosolic arylamine n-acetyltransferase (NAT) deficiency in the dog and other canids due to an absence of NAT genes. *Biochem Pharmacol*. 1997;54(1):73-80. doi:10.1016/S0006-2952(97)00140-8

50. Coelho LP, Kultima JR, Costea PI, et al. Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome*. 2018;6(1):72. doi:10.1186/s40168-018-0450-3

51. Jergens AE, Guard BC, Redfern A, et al. Microbiota-Related Changes in Unconjugated Fecal Bile Acids Are Associated With Naturally Occurring, Insulin-Dependent Diabetes Mellitus in Dogs. *Front Vet Sci*. 2019;0. doi:10.3389/fvets.2019.00199

52. Vázquez-Baeza Y, Hyde ER, Suchodolski JS, Knight R. Dog and human inflammatory bowel disease rely on overlapping yet distinct dysbiosis networks. *Nat Microbiol*. 2016;1(12):1-5. doi:10.1038/nmicrobiol.2016.177

53. Herstad KMV, Moen AEF, Gaby JC, Moe L, Skancke E. Characterization of the fecal and mucosa-associated microbiota in dogs with colorectal epithelial tumors. *PLOS ONE*. 2018;13(5):e0198342. doi:10.1371/journal.pone.0198342

54. MacEwen EG. Spontaneous tumors in dogs and cats: Models for the study of cancer biology and treatment. *Cancer Metastasis Rev*. 1990;9(2):125-136. doi:10.1007/BF00046339

55. Paoloni M, Khanna C. Translation of new cancer treatments from pet dogs to humans. *Nat Rev Cancer*. 2008;8(2):147-156. doi:10.1038/nrc2273

56. Parker HG, Shearin AL, Ostrander EA. Man's Best Friend Becomes Biology's Best in Show: Genome Analyses in the Domestic Dog. *Annu Rev Genet*. 2010;44:309-336. doi:10.1146/annurev-genet-102808-115200

57. Fowles JS, Brown KC, Hess AM, Duval DL, Gustafson DL. Intra- and interspecies gene expression models for predicting drug response in canine osteosarcoma. *BMC Bioinformatics*. 2016;17(1):93. doi:10.1186/s12859-016-0942-8

58. Gardner HL, Fenger JM, London CA. Dogs as a Model for Cancer. *Annu Rev Anim Biosci*. 2016;4:199-222. doi:10.1146/annurev-animal-022114-110911

59. Olson PN. Using the canine genome to cure cancer and other diseases. *Theriogenology*. 2007;68(3):378-381. doi:10.1016/j.theriogenology.2007.04.016

60. Waters DJ, Wildasin K. CANCER Clues from PET DOGS. *Sci Am*. 2006;295(6):94-101.

61. Schiffman JD, Breen M. Comparative oncology: what dogs and other species can teach us about humans with cancer. *Philos Trans R Soc B Biol Sci*. 2015;370(1673):20140231. doi:10.1098/rstb.2014.0231

62. Hytönen MK, Lohi H. Canine models of human rare disorders. *Rare Dis*. 2016;4(1):e1241362. doi:10.1080/21675511.2016.1241362

63. Kornegay JN, Bogan JR, Bogan DJ, et al. Canine models of Duchenne muscular dystrophy and their use in therapeutic strategies. *Mamm Genome*. 2012;23(1):85-108. doi:10.1007/s00335-011-9382-y

64. Chapman RW. Canine models of asthma and COPD. *Pulm Pharmacol Ther*. 2008;21(5):731-742. doi:10.1016/j.pupt.2008.01.003

65. Rowell JL, McCarthy DO, Alvarez CE. Dog Models of Naturally Occurring Cancer. *Trends Mol Med*. 2011;17(7):380-388. doi:10.1016/j.molmed.2011.02.004

66. Withrow SJ. *Withrow and MacEwen's Small Animal Clinical Oncology*. Elsevier Health Sciences; 2007.

67. 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. doi:10.1186/s13028-017-0341-9

68. Scott MC, Temiz NA, Sarver AE, et al. Comparative Transcriptome Analysis Quantifies Immune Cell Transcript Levels, Metastatic Progression, and Survival in Osteosarcoma. *Cancer Res*. 2018;78(2):326-337. doi:10.1158/0008-5472.CAN-17-0576

69. Heymann MF, Lézot F, Heymann D. The contribution of immune infiltrates and the local microenvironment in the pathogenesis of osteosarcoma. *Cell Immunol*. 2019;343:103711. doi:10.1016/j.cellimm.2017.10.011

70. Hiddemann W, Roessner A, Wörmann B, et al. Tumor heterogeneity in osteosarcoma as identified by flow cytometry. *Cancer*. 1987;59(2):324-328. doi:10.1002/1097-0142(19870115)59:2<324::aid-cncr2820590226>3.0.co;2-9

71. Karlsson EK, Sigurdsson S, Ivansson E, et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. *Genome Biol*. 2013;14(12):R132. doi:10.1186/gb-2013-14-12-r132

72. Kirpensteijn J, Kik M, Teske E, Rutteman GR. TP53 Gene Mutations in Canine Osteosarcoma. *Vet Surg*. 2008;37(5):454-460. doi:10.1111/j.1532-950X.2008.00407.x

73. Wunder JS, Gokgoz N, Parkes R, et al. TP53 Mutations and Outcome in Osteosarcoma: A Prospective, Multicenter Study. *J Clin Oncol*. Published online September 21, 2016. doi:10.1200/JCO.2005.04.074

74. Paoloni M, Davis S, Lana S, et al. Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. *BMC Genomics*. 2009;10:625. doi:10.1186/1471-2164-10-625

75. Scott MC, Sarver AL, Gavin KJ, et al. Molecular subtypes of osteosarcoma identified by reducing tumor heterogeneity through an interspecies comparative approach. *Bone*. 2011;49(3):356-367. doi:10.1016/j.bone.2011.05.008

76. Monks NR, Cherba DM, Kamerling SG, et al. A multi-site feasibility study for personalized medicine in canines with Osteosarcoma. *J Transl Med*. 2013;11(1):158. doi:10.1186/1479-5876-11-158

77. Davis LE, Hofmann NE, Li G, et al. A case study of personalized therapy for osteosarcoma. *Pediatr Blood Cancer*. 2013;60(8):1313-1319. doi:https://doi.org/10.1002/pbc.24512

78. Eisenberger S, Hoppe G, Pyerin W, Ackermann K. High-quality RNA preparation for transcript profiling of osteocytes from native human bone microdissections. *Anal Biochem*. 2004;335(2):260-266. doi:10.1016/j.ab.2004.09.037

79. Chomczynski P, Sacchi N. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nat Protoc*. 2006;1(2):581-585. doi:10.1038/nprot.2006.83

80. Lee JTY, Cheung KMC, Leung VYL. Extraction of RNA from tough tissues with high proteoglycan content by cryosection, second phase separation and high salt precipitation. *J Biol Methods*. 2015;2(2). doi:10.14440/jbm.2015.40

81. Reno C, Marchuk L, Sciore P, Frank CB, Hart DA. Rapid isolation of total RNA from small samples of hypocellular, dense connective tissues. *Biotechniques*. 1997;22(6):1082-. doi:10.2144/97226bm16

82. Kelly NH, Schimenti JC, Ross FP, van der Meulen MCH. A method for isolating high quality RNA from mouse cortical and cancellous bone. *Bone*. 2014;68:1-5. doi:10.1016/j.bone.2014.07.022

83. Carter LE, Kilroy G, Gimble JM, Floyd ZE. An improved method for isolation of RNA from bone. *Bmc Biotechnol*. 2012;12. doi:10.1186/1472-6750-12-5

84. Cepollaro S, Della Bella E, de Biase D, Visani M, Fini M. Evaluation of RNA from human trabecular bone and identification of stable reference genes. *J Cell Physiol*. 2018;233(6):4401-4407. doi:10.1002/jcp.26319

85. Florencio-Silva R, Sasso GRD, Sasso-Cerri E, Simoes MJ, Cerri PS. Biology of Bone Tissue: Structure, Function, and Factors That Influence Bone Cells. *Biomed Res Int*. Published online 2015. doi:10.1155/2015/421746

86. Kim WT, Ryu CJ. Cancer stem cell surface markers on normal stem cells. *Bmb Rep*. 2017;50(6):285-298. doi:10.5483/BMBRep.2017.50.6.039

87. Clarke B. Normal Bone Anatomy and Physiology. *Clin J Am Soc Nephrol*. 2008;3:S131-S139. doi:10.2215/Cjn.04151206

88. Geller DS, Gorlick R. Osteosarcoma: A Review of Diagnosis, Management, and Treatment Strategies. *O T E O R Co M A*.:14.

89. Luetke A, Meyers PA, Lewis I, Juergens H. Osteosarcoma treatment – Where do we stand? A state of the art review. *Cancer Treat Rev*. 2014;40(4):523-532. doi:10.1016/j.ctrv.2013.11.006

90. Longhi A, Errani C, De Paolis M, Mercuri M, Bacci G. Primary bone osteosarcoma in the pediatric age: State of the art. *Cancer Treat Rev*. 2006;32(6):423-436. doi:10.1016/j.ctrv.2006.05.005

91. Gorlick R, Khanna C. Osteosarcoma. *J Bone Miner Res Off J Am Soc Bone Miner Res*. 2010;25(4):683-691. doi:10.1002/jbmr.77

92. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: Data from the Surveillance, Epidemiology, and End Results Program. *Cancer*. 2009;115(7):1531-1543. doi:10.1002/cncr.24121

93. Sayles LC, Breese MR, Koehne AL, et al. Genome-Informed Targeted Therapy for Osteosarcoma. *Cancer Discov*. 2019;9(1):46-63. doi:10.1158/2159-8290.CD-17-1152

94. Sakthikumar S, Elvers I, Kim J, et al. SETD2 Is Recurrently Mutated in Whole-Exome Sequenced Canine Osteosarcoma. *Cancer Res*. 2018;78(13):3421-3431. doi:10.1158/0008-5472.CAN-17-3558

95. Gardner HL, Sivaprakasam K, Briones N, et al. Canine osteosarcoma genome sequencing identifies recurrent mutations in DMD and the histone methyltransferase gene SETD2. *Commun Biol*. 2019;2(1):1-13. doi:10.1038/s42003-019-0487-2

96. Zhao J, Dean DC, Hornicek FJ, Yu X, Duan Z. Emerging next-generation sequencing-based discoveries for targeted osteosarcoma therapy. *Cancer Lett*. 2020;474:158-167. doi:10.1016/j.canlet.2020.01.020

97. Martin JW, Squire JA, Zielenska M. The Genetics of Osteosarcoma. *Sarcoma*. 2012;2012:e627254. doi:10.1155/2012/627254

98. Lipinski KA, Barber LJ, Davies MN, Ashenden M, Sottoriva A, Gerlinger M. Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends Cancer*. 2016;2(1):49-63. doi:10.1016/j.trecan.2015.11.003

99. Subbiah V, Wagner MJ, McGuire MF, et al. Personalized comprehensive molecular profiling of high risk osteosarcoma: Implications and limitations for precision medicine. *Oncotarget*. 2015;6(38):40642-40654.

100. Simpson S, Dunning M, de Brot S, et al. Molecular Characterisation of Canine Osteosarcoma in High Risk Breeds. *Cancers*. 2020;12(9):2405. doi:10.3390/cancers12092405

101. Endo-Munoz L, Cumming A, Sommerville S, Dickinson I, Saunders NA. Osteosarcoma is characterised by reduced expression of markers of osteoclastogenesis and antigen presentation compared with normal bone. *Br J Cancer*. 2010;103(1):73-81. doi:10.1038/sj.bjc.6605723

102. Xie L, Yao Z, Zhang Y, et al. Deep RNA sequencing reveals the dynamic regulation of miRNA, lncRNAs, and mRNAs in osteosarcoma tumorigenesis and pulmonary metastasis. *Cell Death Dis*. 2018;9(7):1-16. doi:10.1038/s41419-018-0813-5

103. Zhao X, Wang Q, Lin F, et al. RNA Sequencing of Osteosarcoma Gene Expression Profile Revealed that miR-214-3p Facilitates Osteosarcoma Cell Proliferation via Targeting Ubiquinol-Cytochrome c Reductase Core Protein 1 (UQCRC1). *Med Sci Monit Int Med J Exp Clin Res*. 2019;25:4982-4991. doi:10.12659/MSM.917375

104. Glickman MS, Sawyers CL. Converting Cancer Therapies into Cures: Lessons from Infectious Diseases. *Cell*. 2012;148(6):1089-1098. doi:10.1016/j.cell.2012.02.015

105. Martson A, Kõks S, Reimann E, Prans E, Erm T, Maasalu K. Transcriptome analysis of osteosarcoma identifies suppression of wnt pathway and up-regulation of adiponectin as potential biomarker. *Genomics Discov*. 2013;1:3. doi:10.7243/2052-7993-1-3

106. Ho XD, Phung P, Q Le V, et al. Whole transcriptome analysis identifies differentially regulated networks between osteosarcoma and normal bone samples. *Exp Biol Med*. 2017;242(18):1802-1811. doi:10.1177/1535370217736512

107. Nance R, Agarwal P, Sandey M, et al. A method for isolating RNA from canine bone. *BioTechniques*. 2020;68(6):311-317. doi:10.2144/btn-2019-0153

108. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl*. 2014;30(15):2114-2120. doi:10.1093/bioinformatics/btu170

109. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Accessed October 26, 2021. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

110. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357-360. doi:10.1038/nmeth.3317

111. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290-295. doi:10.1038/nbt.3122

112. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

113. Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10(1):1523. doi:10.1038/s41467-019-09234-6

114. Liu X, Rapp N, Deans R, Cheng L. Molecular Cloning and Chromosomal Mapping of a Candidate Cytokine Gene Selectively Expressed in Human CD34+ Cells. *Genomics*. 2000;65(3):283-292. doi:10.1006/geno.2000.6170

115. Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. *Nat Rev Drug Discov*. 2018;17(3). doi:10.1038/nrd.2017.226

116. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta*. 2010;1805(1):105. doi:10.1016/j.bbcan.2009.11.002

117. Lin F, Xie YJ, Zhang XK, et al. GTSE1 is involved in breast cancer progression in p53 mutation-dependent manner. *J Exp Clin Cancer Res*. 2019;38(1):152. doi:10.1186/s13046-019-1157-4

118. Lai W, Zhu W, Li X, et al. GTSE1 promotes prostate cancer cell proliferation via the SP1/FOXM1 signaling pathway. *Lab Invest*. 2021;101(5):554-563. doi:10.1038/s41374-020-00510-4

119. Wu X, Wang H, Lian Y, et al. GTSE1 promotes cell migration and invasion by regulating EMT in hepatocellular carcinoma and is associated with poor prognosis. *Sci Rep*. 2017;7(1):5129. doi:10.1038/s41598-017-05311-2

120. Liu A, Zeng S, Lu X, et al. Overexpression of G2 and S phase-expressed-1 contributes to cell proliferation, migration, and invasion via regulating p53/FoxM1/CCNB1 pathway and predicts poor prognosis in bladder cancer. *Int J Biol Macromol*. 2019;123:322-334. doi:10.1016/j.ijbiomac.2018.11.032

121. Xie C, Xiang W, Shen H, Shen J. GTSE1 is possibly involved in the DNA damage repair and cisplatin resistance in osteosarcoma. *J Orthop Surg*. 2021;16(1):713. doi:10.1186/s13018-021-02859-8

122. Zhou C, Wang M, Zhou L, et al. Prognostic significance of PLIN1 expression in human breast cancer. *Oncotarget*. 2016;7(34):54488-54502. doi:10.18632/oncotarget.10239

123. Lu C, Wang X, Zhao X, Xin Y, Liu C. Long non-coding RNA ARAP1-AS1 accelerates cell proliferation and migration in breast cancer through miR-2110/HDAC2/PLIN1 axis. *Biosci Rep*. 2020;40(4):BSR20191764. doi:10.1042/BSR20191764

124. Jung H, Yoon SR, Lim J, Cho HJ, Lee HG. Dysregulation of Rho GTPases in Human Cancers. *Cancers*. 2020;12(5):1179. doi:10.3390/cancers12051179

125. Wang J sheng, Wang Y guo, Zhong Y sheng, et al. Identification of co-expression modules and pathways correlated with osteosarcoma and its metastasis. *World J Surg Oncol*. 2019;17(1):46. doi:10.1186/s12957-019-1587-7

126. Kim H, Yoo S, Zhou R, et al. Oncogenic role of SFRP2 in p53-mutant osteosarcoma development via autocrine and paracrine mechanism. *Proc Natl Acad Sci*. 2018;115(47):E11128-E11137. doi:10.1073/pnas.1814044115

127. Sierko E, Wojtukiewicz MZ, Kisiel W. The Role of Tissue Factor Pathway Inhibitor-2 in Cancer Biology. *Semin Thromb Hemost*. 2007;33(7):653-659. doi:10.1055/s-2007-991532

128. Zhang W, Yang C, Wang S, et al. SDC2 and TFPI2 Methylation in Stool Samples as an Integrated Biomarker for Early Detection of Colorectal Cancer. *Cancer Manag Res*. 2021;13:3601-3617. doi:10.2147/CMAR.S300861

129. Wang S, Xiao X, Zhou X, et al. TFPI-2 is a putative tumor suppressor gene frequently inactivated by promoter hypermethylation in nasopharyngeal carcinoma. *BMC Cancer*. 2010;10(1):617. doi:10.1186/1471-2407-10-617

130. Culp WTN, Olea-Popelka F, Sefton J, et al. Evaluation of outcome and prognostic factors for dogs living greater than one year after diagnosis of osteosarcoma: 90 cases (1997–2008). *J Am Vet Med Assoc*. 2014;245(10):1141-1146. doi:10.2460/javma.245.10.1141

131. Suvà ML, Tirosh I. Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges. *Mol Cell*. 2019;75(1):7-12. doi:10.1016/j.molcel.2019.05.003

132. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J Am Soc Nephrol*. 2019;30(1):23-32. doi:10.1681/ASN.2018090912

133. Denisenko E, Guo BB, Jones M, et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol*. 2020;21(1):130. doi:10.1186/s13059-020-02048-6

134. Bakken TE, Hodge RD, Miller JA, et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PloS One*. 2018;13(12):e0209648. doi:10.1371/journal.pone.0209648

135. Lake BB, Codeluppi S, Yung YC, et al. A comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type expression from nuclear RNA. *Sci Rep*. 2017;7(1):6031. doi:10.1038/s41598-017-04426-w

136. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20(1):296. doi:10.1186/s13059-019-1874-1

137. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat Commun*. 2022;13(1):1246. doi:10.1038/s41467-022-28803-w

138. Hu C, Li T, Xu Y, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res*. 2023;51(D1):D870-D876. doi:10.1093/nar/gkac947

139. Nance RL, Cooper SJ, Starenki D, et al. Transcriptomic Analysis of Canine Osteosarcoma from a Precision Medicine Perspective Reveals Limitations of Differential Gene Expression Studies. *Genes*. 2022;13(4):680. doi:10.3390/genes13040680

140. Tickle T, Tirosh I, Georgescu C, Brown M, Haas B. inferCNV of the Trinity CTAT Project. inferCNV of the Trinity CTAT Project. Published 2019. https://github.com/broadinstitute/inferCNV

141. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci*. 2022;31(1):8-22. doi:10.1002/pro.4218

142. Mi H, Thomas P. PANTHER Pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol Clifton NJ*. 2009;563:123-140. doi:10.1007/978-1-60761-175-2_7

143. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29. doi:10.1016/j.cell.2021.04.048

144. Khan IA, Bordoni B. Histology, Osteoclasts. In: *StatPearls*. StatPearls Publishing; 2023. Accessed May 27, 2023. http://www.ncbi.nlm.nih.gov/books/NBK554489/

145. Matuła K, Rivello F, Huck WTS. Single-Cell Analysis Using Droplet Microfluidics. *Adv Biosyst*. 2020;4(1):1900188. doi:10.1002/adbi.201900188

146. Almeida M. Aging mechanisms in bone. *BoneKEy Rep*. 2012;1:102. doi:10.1038/bonekey.2012.102

147. Yahiro K, Matsumoto Y. Immunotherapy for osteosarcoma. *Hum Vaccines Immunother*. 17(5):1294-1295. doi:10.1080/21645515.2020.1824499

148. Ping Q, Yan R, Cheng X, et al. Cancer-associated fibroblasts: overview, progress, challenges, and directions. *Cancer Gene Ther*. 2021;28(9):984-999. doi:10.1038/s41417-021-00318-4

149. Costa Fernandes CJ da, Zambuzzi WF. Fibroblast-secreted trophic factors contribute with ECM remodeling stimulus and upmodulate osteocyte gene markers in osteoblasts. *Biochimie*. 2020;168:92-99. doi:10.1016/j.biochi.2019.10.013